

**“ИХ ӨГӨГДЛИЙН САН ҮҮСГЭН БҮРДҮҮЛЭХ,
ХЭРЭГЛЭЭНД НЭВТРҮҮЛЭХЭД ШААРДЛАГАТАЙ
СИСТЕМ, ТЕХНИК ХЭРЭГСЭЛ, ХҮНИЙ НӨӨЦИЙН
ТООЦОО, СУДАЛГАА”**

ЗАХИАЛАГЧ: ЦАХИМ ХӨГЖИЛ, ИННОВАЦ, ХАРИЛЦАА ХОЛБООНЫ ЯАМ

ГҮЙЦЭТГЭГЧ: БИДСОЛЮШН ХХК

УЛААНБААТАР ХОТ
2025 ОН

АГУУЛГА

ОРШИЛ	3
Хууль эрх зүйн орчин	4
Их өгөгдлийн сангийн дэд бүтэц	4
1.1.1 ӨГӨГДЛИЙН ХЭМЖЭЭ БА ӨСӨЛТ /DATA VOLUME AND GROWTH/	10
1.1.2 ТОНОГ ТӨХӨӨРӨМЖ /HARDWARE/	12
ДҮГНЭЛТ	20
ИХ ӨГӨГДЛИЙН ПРОГРАММ ХАНГАМЖ, БАГАЖ ХЭРЭГСЛИЙН СУДАЛГАА /SOFTWARE AND TOOLS/	22
1.2.1 ӨГӨГДЛИЙН САНГИЙН ТЕХНОЛОГИ	22
1.2.2 МЭДЭЭЛЛИЙН ИНТЕГРАЦЛАЛ БА МЕНЕЖМЕНТ:	27
САЙН ТУРШЛАГА	29
1.2.3 МЭДЭЭЛЭЛ ЗАГВАРЧЛАХ ХЭРЭГСЭЛ /DATA MODELING TOOLS/	36
ДҮГНЭЛТ	43
АЮУЛГҮЙ БАЙДЛЫН ТӨЛӨВЛӨЛТ /SECURITY PIANNING/	46
ДҮГНЭЛТ	50
ХЯНАЛТ, ЗАСВАР ҮЙЛЧИЛГЭЭ /MONITORING AND MAINTENANCE/	51
ДҮГНЭЛТ	54
УР ЧАДВАРЫГ БИЙ БОЛГОХ, СУРГАЛТ	56
ДҮГНЭЛТ	61
ТӨСӨВ: ШААРДЛАГАТАЙ ЗАРДЛЫН ТӨСВИЙН ТООЦОО	62
ДҮГНЭЛТ	74
ИХ ӨГӨГДЛИЙН САНГИЙН АЖЛЫН ДААЛГАВАР	76
ИХ ӨГӨГДЛИЙН САНГ ҮҮСГЭН БҮРДҮҮЛЭХ, ТӨРӨЛЖСӨН МЭДЭЭЛЛИЙН САНГИЙН ӨГӨГДЛИЙГ ИХ ӨГӨГДЛИЙН САНД НЭГТГЭХ, МЭДЭЭЛЭЛ ХАРИУЦАГЧ БАЙГУУЛЛАГУУД БОЛОН ОЛОН НИЙТЭД АШИГЛУУЛАХ НӨХЦӨЛИЙГ БҮРДҮҮЛСЭН ПЛАТФОРМЫН АЖИЛ	82
4.1 ХЯЗГААРЛАЛТТАЙ МЭДЭЭЛЛИЙН ХАМГААЛАЛТ	88
4.2 ӨГӨГДӨЛ ТАТАХ БА СОЛИЛЦООНЫ СИСТЕМ	88
4.3 ХАДГАЛАХ ШИЙДЭЛ	88
4.4 УРЬДЧИЛАН БОЛОВСРУУЛАХ, ДҮРСЛЭХ ҮЕ ШАТ	89
4.5 МЭДЭЭЛЛИЙН САНГИЙН УЯЛДАА, ИНТЕГРАЦ	89
4.6 OPENDATA.GOV.MN СИСТЕМИЙН ШИНЭЧЛЭЛТ	89
4.7 ИХ ӨГӨГДЛИЙН САНГИЙН АШИГЛАЛТ	90
НЭГДСЭН ДҮГНЭЛТ	91
I. Монгол Улсын хууль, эрх зүйн баримт бичиг	131

II. Олон улсын стандарт, удирдамж.....	131
III. Олон улсын туршлага, бодлогын баримт бичиг	131
Эстони	131
Канад	131
Сингапур.....	131
IV. Технологийн баримт бичиг, бүтээгдэхүүний reference	131
V. Судалгаа, онолын материалууд	132

ОРШИЛ

Аливаа улс орны хувьд төрийн гол үүрэг нь хүн амын амжиргааны түвшинг дээшлүүлэх, эрүүл мэндийн үйлчилгээг хүртээмжтэй болгох, боловсролын чанарыг дээшлүүлэх, аюулгүй амьдрах баталгааг хангах, хөдөлмөрийн зах зээлийг дэмжих, эдийн засгийн өсөлтийг сайжруулах зэрэг байдаг. Эдгээр үйл ажиллагааг амжилттай хэрэгжүүлэхийн тулд төрийн бодлогыг боловсруулах, хэрэгжүүлсэн бодлогын үр нөлөөг үнэлэх, салбар тус бүрийн нөхцөл байдлыг тодорхойлох гол хүчин зүйл нь тоон мэдээлэл буюу дата юм. Төрийн бодлого боловсруулахад салбар хоорондын хамаарал, түүнийг илэрхийлсэн их өгөгдлийн сангийн хэрэгцээ улам нэмэгдэж байна. Өнөөдрийн нийгмийн хурдацтай хувьсан өөрчлөгдөж буй нөхцөл байдал, техник технологийн үсрэнгүй хөгжил ч нөгөө талаас их өгөгдлийн сангийн хэрэглээг нэмэгдүүлж байна.

Манай улсын хувьд ч их өгөгдлийн сан байгуулах үйл ажиллагааг хэрэгжүүлж буй нь цаг үетэйгээ нийцсэн үйл явц юм.

Бид их өгөгдлийн санг бий болгож үйл ажиллагаанд нэвтрүүлснээр төрийн бодлогыг нотолгоонд суурилсан оночтой байдлаар хэрэгжүүлэхэд, иргэддээ хүртээмжтэй, чирэгдэл багатай үйлчилгээг хэрэгжүүлэх бодит суурь болох юм.

Их өгөгдлийн сан байгуулагдсанаар:

- Төрийн бодлого боловсруулахад шаардагдах зардал, цаг хугацааг хэмнэх боломжтой.

Өнөөгийн нөхцөлд тулгардаг хамгийн түгээмэл асуудал нь тухайн нэг салбар, зорилтот бүлгийн нөхцөл байдлыг тодорхойлохын тулд зорилтот судалгааг зохион байгуулах замаар нөхцөл байдлын талаарх мэдээлэл цуглуулж түүн дээр шаардлагатай шийдвэр гаргалтуудыг хийж байгаа нь зардал ихтэй, хугацааны хоцрогдолтой, хүний хүчин зүйлээс хамаарсан, зөвхөн тухайн түүврийн хүрээг хамарсан, тогтмол бус тоон мэдээлэлтэй байх нөхцөл болж байна.

Иймд их өгөгдлийн санг байгуулж, тоон мэдээллийн нэгдсэн өгөгдөл солилцоог бий болгосноор улсын хэмжээнд өрнөж буй эдийн засаг, нийгмийн идэвхжил, хүндрэлийг бодит цагийн тоон өгөгдөлд суурилсан, зардал бага, нийт хүн амын хувьд нэгдсэн шугамаар үнэлэх боломж бүрдэнэ.

- Нийгэм, эдийн засгийн өнөөдрийн нөхцөл байдал төдийгүй ирээдүйд бий болох хэрэгцээ, шаардлагыг зорилтот бүлэг тус бүрээр оновчтой тодорхойлох боломжтой.

Их өгөгдлийн санг бий болгож түүхэн өгөгдлүүдийг бий болгосноор нийгэм, эдийн засаг, боловсрол, эрүүл мэнд зэрэг төрийн бодлогыг түүхэн өгөгдөл, шинээр үүссэн нөхцөл байдалд тулгуурлан дүн шинжилгээ хийх, ирээдүйд үүсэх хэрэгцээ, шаардлагыг оновчтойгоор үнэлэх бололцоо бий болно.

- Нотолгоонд суурилсан бодлогын шийдвэр гаргалт сайжирна.

Төрийн бодлого шийдвэр гарахаас өмнөх тооцоолол, судалгааны оновчтой байдал ихээр нэмэгдэнэ. Ингэснээр төрийн гаргасан шийдвэр бүр үндэслэл, үнэлгээ, үр дүнгийн хэмжилтүүдээр баталгаажиж явах нөхцөл байдал бүрэлдэнэ.

- Төрийн байгууллагуудын тоо мэдээллийн чанар сайжирч, зөрүүтэй байдал арилна.

Төрийн байгууллагуудын хувьд нэгдсэн өгөгдлийн шинжилгээ хийх боломжгүйгээс уялдаа холбоогүй, зөрүүтэй тоо, мэдээллийг чиг үүрэг бүрийн хувьд боловсруулж байна. Харин их өгөгдлийн санг бий болгосноор төрийн нэгдсэн шинжилгээний тогтолцоо бий болж, тоо мэдээллийн зөрүүтэй байдал арилахын зэрэгцээгээр мэдээллийн давхардал, ангилал кодын зөрүүтэй байдал арилна.

Хууль эрх зүйн орчин

Монгол Улсын Үндсэн хууль, Монгол Улсын Үндэсний аюулгүй байдлын үзэл баримтлал, Алсын хараа- 2050, Монгол Улсын Тогтвортой хөгжлийн зорилт -2030 зэрэг урт, дунд, богино хугацааны хөгжлийн бодлого, хөтөлбөрт “Төрийн мэдээллийн бүрэн бүтэн байдлыг хангахын тулд үнэн, бодитой, цаг үеэ олсон, шаардлагатай мэдээллийг төрийн мэдээллийн санд төвлөрүүлэх”, “Төрийн мэдээллийн нээлттэй нөөцийг бүрдүүлэх ажиллагааг эрчимжүүлж, тэдгээрийг цахим засгийн үйлчилгээнд үр нөлөөтэй ашиглах боломжийг бүрдүүлэх”, “Төрийн байгууллагуудын цахим мэдээллийн сан, түүний дэд бүтцийн аюулгүй байдлыг хангах, чадавхыг нэмэгдүүлж, мэдээллийн аюулгүй байдлын шийдэл боловсруулах тогтолцоог бэхжүүлэх” зэргийг тусгасан байдаг.

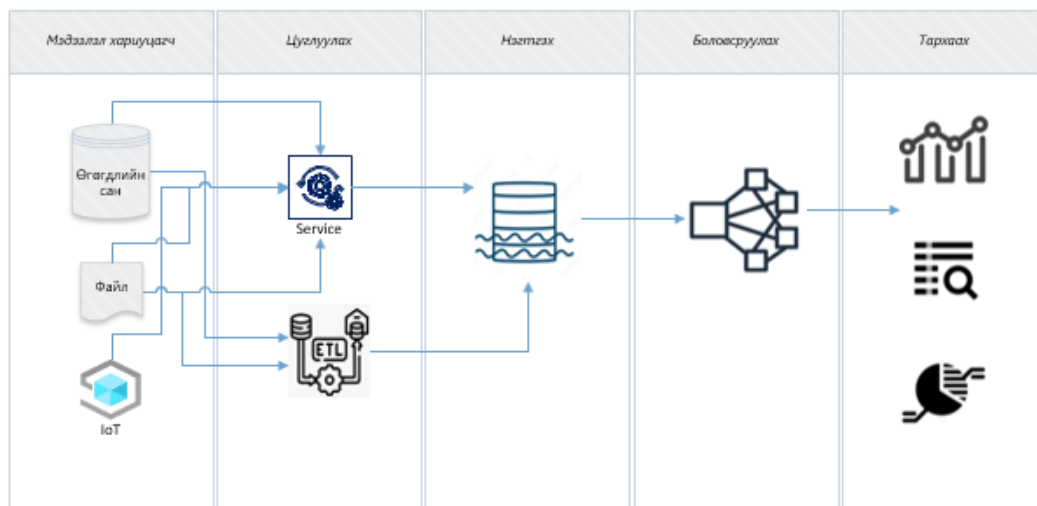
УИХ-с шинэчлэн найруулж батласан “Хүний хувийн мэдээлэл хамгаалах тухай”, “Нийтийн мэдээллийн ил тод байдлын тухай”, “Кибер аюулгүй байдлын тухай” хуулиуд болон Монгол Улсын Засгийн Газрын “Их өгөгдлийн сан” үүсгэх тухай 2023 оны 403 дугаар тогтоол батлагдсанаар Их өгөгдлийн сан байгуулахад манлайлан оролцох байгууллагуудыг тодорхойлсны зэрэгцээгээр мэдээлэл солилцох дэд бүтэц, мэдээллийн хязгаарлалтуудыг тодорхойлж мэдээлэл нийлүүлэгч байгууллагуудыг үүрэгжүүлж өгснөөр их өгөгдлийн сан бий болгох урьдчилсан нөхцөл бүрдсэн.

Их өгөгдлийн сангийн дэд бүтэц

Их өгөгдлийн сангийн нэгдсэн дэд бүтцийг дараах үндсэн дэд хэсгүүдээс бүрдэнэ гэж тооцлоо:

1. Өгөгдлийг мэдээлэл хариуцагч эх сурвалжаас цуглуулах (ACQUIRE)
2. Цуглуулсан өгөгдлийг хадгалах, зохион байгуулах (ORGANIZE)
3. Хадгалж зохион байгуулсан өгөгдлийг ашиглан дүн шинжилгээ хийх (ANALYZE)
4. Үр дүнг тархаах, ашиглах (DISSEMINATION)
5. Их өгөгдлийн дэд бүтцийг удирдан зохион байгуулах (MANAGE AND GOVERN)

Өгөгдлийн чанар, Метадата менежмент, Нууцлал аюулгүй байдал, Өгөгдлийн каталоги



Өгөгдлийг мэдээлэл хариуцагч эх сурвалжаас цуглуулах (ACQUIRE)

Мэдээлэл хариуцагч эх сурвалжаас татан цуглуулах буюу "Acquire" хэсэгт өгөгдлийг төрөл бүрийн эх сурвалжуудтай холбогдон нэгтгэх үйлдэл хийгдэнэ. Дараах аргуудаар өгөгдлийг цуглуулна.

- Тасралтгүй горимоор үүсч буй өгөгдөл (Streaming)
IoT буюу юмсын интернэт хөгжиж буйтай холбоотойгоор ухаалаг хот, ухаалаг үйлдвэр, ухаалаг дэд бүтэц (ухаалаг тоолуур, ухаалаг цахилгаан түгээх сүлжээ, ухаалаг тээврийн систем, ухаалаг гаалийн хяналт гэх мэтээр) зэрэг олон системүүдийн мэдрэгч төхөөрөмжүүдээс өгөгдлийн урсгалыг хүлээн авах.

Сүүлийн жилүүдэд мэдрэгч төхөөрөмжүүдээс өгөгдөл хүлээн авч нэгдсэн байдлаар боловсруулж бодит цагийн горимд тулгуурласан шийдвэр гаргалт хийх, үр дүнг нь иргэдийн амьдралын чанарыг сайжруулахад ашиглах чиглэлээр сонгодог жишээ болсон "Ухаалаг хот" төслүүд олон улс оронд хэрэгжээд байна.

Жишээ нь:

Сингапур, Барселона

Мэдрэгч болон CCTV ашиглан замын хөдөлгөөний нягтыг бодит цагийн горимд хяналт тавьдаг бөгөөд AI алгоритм нь гэрлэн дохиог хамгийн оновчтойгоор өөрчилдөг. Аль гудамж хамгийн их ачаалалтай байгааг тодорхойлж, замын чиглэлийг автоматаар өөрчилж чаддаг.

Сөүл, Амстердам

Цахилгаан хэрэглэгчийн цахилгааны хэрэглээг бодит цагаар хэмжиж, төв серверт дамжуулдаг. Иргэд өөрсдийн хэрэглээг шууд хянаж, хэтрэхээс сэргийлэх боломжийг бий болгосон. Энэ нь цаашлаад эрчим хүчний

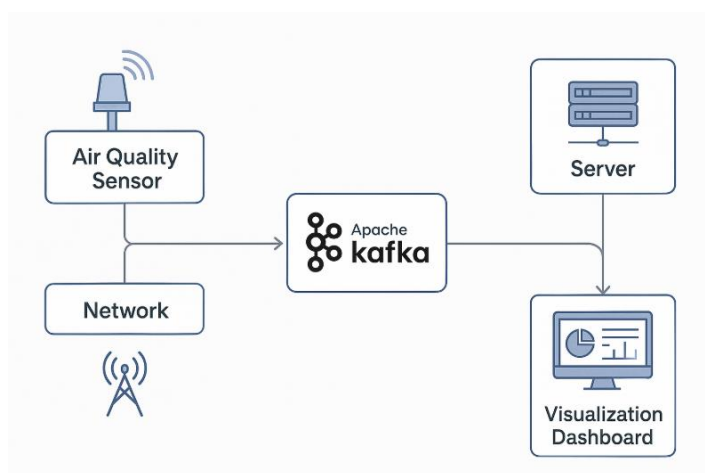
нийлүүлэлтийг эрчим хүчний хэрэглээтэй нь уялдуулж автомат тохируулга хийдэг байна.

Осло (Норвеги), Чэнду (Хятад)

Гудамж замын хөдөлгөөн мэдрэгчтэй гэрэлтүүлэг нь хүн, тээврийн хэрэгслийг мэдэрч түүний шаардлагатай орон зайд асдаг байна. Энэхүү шийдэл нь эрчим хүчний хэмнэлтийг 40-70% болгосон байна.

Улаанбаатар

Манай орны тухайд энэ чиглэлийн хөгжүүлэлт, шийдлүүд хийгдэж эхлээд байгааг дурдах хэрэгтэй. Тухайлбал цахилгааны ухаалаг тоолуур, агаарын чанарын автомат харуулуудыг дурдаж болно.



IoT төхөөрөмжөөс өгөгдөл цуглуулж боловсруулалт хийх дамжлага.

ҮЕ ШАТ	ҮҮРЭГ
1. IoT төхөөрөмжүүд	Хөдөлгөөн мэдрэгч, ухаалаг тоолуур, камер, GPS гэх мэт. Эдгээр нь өгөгдлийг тасралтгүй үүсгэдэг.
2. Edge Gateway / IoT Hub	Сүлжээний хязгаарлалттай орчинд өгөгдлийг урьдчилан боловсруулж, багасгаж, нэгдсэн хэлбэрт оруулна.
3. Streaming Platform	Өгөгдлийг дараалалтайгаар, саатал багатайгаар дамжуулна.
4. Stream Processing Engine	Бодит цагийн горимд дүн шинжилгээ хийж, тодорхой нөхцөлд (rule/AI model) үндэслэн шуурхай хариу үйлдэл үзүүлнэ.
5. Data Lake / Warehouse	Том өгөгдөл хадгалах, AI сургалт хийх, тайлан гаргах зорилгоор ашиглана.
6. Хэрэглэгчийн интерфэйс	Дашбоард, вэб, гар утасны апп зэргээр бодит цагийн мэдээ харуулна.

- Тодорхой шатлалтайгаар авч болох өгөгдөл (Staging/At Rest)
Их өгөгдлийн дэд бүтцийг байгуулахад төрийн байгууллагууд буюу суурь мэдээллийн болон төрөлжсөн мэдээллийн сангуудаас өгөгдлийн сангийн суурийг бүрдүүлнэ. Цаашид дан ганц төрийн байгууллагууд гэлтгүй хувийн хэвшил,

төрийн бус байгууллагууд зэрэг байгууллагуудаас өгөгдөл хүлээн авах, өгөгдлөөр үйлчлэх байдлаар их өгөгдлийн сан өргөжинө.

Өгөгдлийн эх сурвалжуудад дээрээс Их өгөгдлийн санд нийлүүлэх өгөгдлийг олон төрлөөр буюу бүтэцтэй болон хагас бүтэцтэй өгөгдлийн төрлөөр болон олон төрлийн файл хэлбэрээр тодорхой хугацааны давтамжтайгаар өгөгдлийг багцлан цуглуулна.

Их өгөгдлийн санд нийлүүлэх өгөгдлийн төрөл нь:

- Хүснэгтэн бүтэцтэй – Одоо өргөн хэрэглэгдэж буй RDBMs буюу өгөгдлийн удирдлагын сангийн системээс татан авсан олон улсын SQL стандартын дагуу бүрдсэн хүснэгтэн мэдээлэл (Төрийн ихэнх систем нь өгөгдлөө энэ хэлбэрт хадгалж байгаа юм),
- Хагас бүтэцлэгдсэн – JSON, XML гэх мэт тодорхой формат бүхий, олон улсын NoSQL стандартын дагуу бүрдсэн мэдээлэл,
- Задгай буюу тодорхой цэгцтэй хадгалсан бүтэц байхгүй. Жишээ нь цахим бичиг баримт, зураг, дүрс, текст файлууд.

Цуглуулсан өгөгдлийг хадгалах, боловсруулах (ORGANIZE)

Их өгөгдлийн сангийн дараагийн бүрэлдэхүүн хэсэг нь мэдээлэл хариуцагч байгууллагуудаас хүлээн авсан өгөгдлийг нэгтгэн зохион байгуулах үйл ажиллагаа юм. Энэ хэсэгт мэдээллийн нэгдсэн хадгалалтыг зохион байгуулахын (Data Lake) зэрэгцээ уялдаа холбоотой өгөгдлүүд буюу Өгөгдлийн агуулахыг (Data Warehouse) бий болгоно. Data warehouse зохион байгуулах нь тавигдсан хэрэгцээ шаардлагад нийцүүлэхийн зэрэгцээ, өгөгдлийг нийцүүлэх, хамаарлыг тодорхойлох зэрэг нарийн үйл ажиллагаанууд хийгддэгээрээ нарийн төвөгтэй, шийдэл, хүн хүч, хугацаа шаардсан үйл ажиллагаа болдог. Энэхүү үйл ажиллагаа нь программ хангамжийн төдийгүй сервер орчны зохион байгуулалтын оновчтой байдлыг давхар шаарддаг. Өгөгдлийн төрлөөс хамааран классик болон тархмал кластер серверийн орчинг бүрдүүлэх боломжтой.

Классик серверийн орчин (Classic / Scale-up architecture)

- Нэг эсвэл цөөн тооны өндөр хүчин чадалтай сервер дээр өгөгдлийн агуулах (Data Warehouse) болон өгөгдлийн орчинг байрлуулна.
 - Хадгалалт, тооцоолол бүгд нэг системд төвлөрсөн.
 - Дотоод сүлжээгээр маш хурдан холбогдсон, latency бага.
- Давуу тал
- Архитектур энгийн, суурилуулах, удирдах хялбар.
 - Өндөр хүчин чадалтай CPU, RAM, storage-г нэг дор төвлөрүүлж болдог.
 - Жижиг/дунд хэмжээний өгөгдөл, тогтвортой бүтэцтэй мэдээлэлд тохиромжтой.

Сул тал

- Нэг серверийн хүчин чадлын дээд хязгаарт тулдаг.
- Тэлэлт (scalability) харьцангуй өндөр өртөгтэй.
- Нэг цэгийн эвдрэл (single point of failure) эрсдэлтэй.

Жишээ

- Oracle Exadata, Microsoft SQL Server дээр 50–100 TB хэмжээтэй, тогтвортой форматтай өгөгдлийн сан ажиллуулах

Тархмал кластер серверийн орчин (Distributed / Scale-out architecture)

- Олон сервер (node) нийлж нэг систем болон ажилладаг.

- Өгөгдөл болон тооцооллыг кластерын бүх node-уудад хуваарилах.
 - Үр ашигтайгаар том хэмжээний өгөгдөл (PB түвшин) болон төрөл бүрийн өгөгдөл (structured, semi-structured, unstructured) боловсруулах боломжтой.
- Давуу тал
- Хялбар өргөтгөх боломжтой (scale horizontally)
- Их хэмжээний өгөгдөл, real-time аналитик, AI/ML ажлуудад тохиромжтой.
 - Хуваарилсан хадгалалт, тооцоолол нь өндөр найдвартай ажиллагаа (fault tolerance) хангана.
- Сул тал
- Тохиргоо, удирдлага, мониторинг илүү нарийн төвөгтэй.
 - Хурдан сүлжээ (10–40GbE+) болон кластерийн зохион байгуулалт шаардана.

Data Lake болон Datawarehouse гэсэн ойлголтуудтай хосолсон байдлаар LakeHouse гэсэн орчин шинэ тутам хөгжиж байна. Lakehouse гэдэг нь дээрх хоёр архитектурын давуу талуудыг шингээсэн буюу Data Lake-ийн уян хатан байдал, Data Warehouse-ийн найдвартай чанар, гүйцэтгэл, зохион байгуулалттай бүтэцтэй өгөгдөлтэй ажиллах боломжуудыг шингээсэн өгөгдөл хадгалалт, боловсруулалтын архитектур юм.

Хэдийгээр LakeHouse нь хосолсон шинж чанар бүхий давуу байдал бий болгох боловч Их өгөгдлийн санг бий болгох эхний үе шат буюу төрийн бүтэцлэгдсэн өгөгдлийг нэгтгэх үе шатанд төдийлөн ач холбогдолгүй гэж үзлээ. Их өгөгдлийн сангийн өргөжилтийн шатанд буюу чатбот, камерын зураг, бичлэг, аудио бичлэг зэрэг unstructured өгөгдөл нэгтгэх үед Lakehouse шийдлийг нэмж төлөвлөх боломжтой.

ШИНЖ ЧАНАР	DATA WAREHOUSE	DATA LAKE	LAKEHOUSE
Өгөгдлийн төрөл	Зөвхөн бүтэцтэй	Бүх төрлийн (raw, semi, unstructured)	Бүх төрлийн
Гүйцэтгэл	Хурдан (OLAP)	Удаан (raw формат)	Хурдан (кэшлэлт + индэкс)
Шинжилгээ	OLAP-centric	AI/ML-ready	OLAP + ML + BI
Хэрэглээ	Тайлан, BI	Судалгаа, AI/ML	Хосолсон
Жишээ системүүд	Teradata, Snowflake	Hadoop, S3 + Spark	Databricks, Delta Lake, Apache Iceberg, etc.

LakeHouse шийдлийг нэвтрүүлэхийн тулд

- Тусгай DevOps, data engineer баг
- Open Table Format (Delta, Iceberg) дээр үндэслэсэн шинэ дэд бүтэц бий болгох шаардлагатай.

Өгөгдөл ашиглан дүн шинжилгээ хийх (ANALYZE)

Өгөгдлийг үр дүнтэй ашиглах, хэрэглэгчдэд ойлгомжтой байлгах гол үйл ажиллагаа нь Их өгөгдлийн санд байгаа мэдээлэлд тулгуурлах “Өгөгдлийн каталог” үүсгэх юм. Энэхүү үйл ажиллагаа нь Datawarehouse зохион байгуулах үйл ажиллагаатай нягт холбоотой бөгөөд энэ нь өгөгдлийг ашиглагчдад үнэ цэнийг бий болгож өгдөг. Нээлттэй өгөгдлийн каталогийг бий болгосноор судлаачид, өгөгдөл хэрэглэгчид шаардлагатай боловсруулалт, судалгаа шинжилгээгээ бие даан хөгжүүлэх, шаардлагатай өгөгдлөө олж авах зэрэгт чухал ач холбогдолтой.

Мөн нээлттэй каталогуудыг ашиглахын зэрэгцээ Их өгөгдлийн сангаас тусгай дата лабораториудаар тусгайлсан өгөгдөл захиалан судалгаа шинжилгээ хийх, үр дүнгийн мэдээллийг авах боломжууд байдаг.

Өгөгдөлд олон төрлийн дата хэрэгслүүд ашиглан судалгаа шинжилгээ хийх боломжтой байна. Тухайлбал технологийн хувьд BI (business intelligence), AI/ML, Graph analytics, Advanced analytics гэсэн мэт олон төрлийн дүн шинжилгээний системийн бүрдлийг ашиглана.

Дүн шинжилгээг тархаах (DISSEMINATION)

Их өгөгдлийн сангийн дараагийн гол хэсэг нь боловсруулж, шинжилгээ хийсэн үр дүнг бусдад түгээх, хуваалцах, эргээд бусад системд холбох боломжийг нээх үйл ажиллагаа юм.

Үүнд бүх төрлийн хянах самбар (dashboard), тайлангийн хөгжүүлэлтийг хийх, зарим шаардлагатай тайлан, хянах самбаруудад тодорхой хандах эрхийн зохион байгуулалтыг хийх шаардлагатай. Ингэснээр салбарын мэргэжилтэн, шийдвэр гаргагчдыг шаардлагатай мэдээ тайлангаар цаг алдалгүй өргөн сонголттойгоор хангах, түүнийг ашиглах боломж бүрдэнэ. Мөн нэмэлт тайлан, дашбоард хөгжүүлэх захиалга хүлээн авах, өгөгдлийн шинжилгээ хийх, шинээр тайлан хөгжүүлэх болон тархаах үйл ажиллагааны зохицуулалтыг хийх шаардлагатай. Цаашид нээлттэй болон хандах эрх бүхий API (Application Programming Interface) буюу программ хангамжийн боломжийг нээж өгснөөр өгөгдлөө бусад системүүдэд нийлүүлэх цогц шийдэл болж өгнө.

Их өгөгдлийн, дэд бүтцийг удирдах, засаглах (MANAGE AND GOVERN)

Их өгөгдлийн сангийн дэд бүтэц бүхэлдээ хэвийн ажиллах, уг үйл ажиллагааг удирдах хамгаалах нь гол үйл ажиллагаа болно. Энэ хүрээнд мэдээллийн нууцлал аюулгүй байдлыг хангах, мета мэдээллийн санг бүрдүүлэх, иргэн байгууллагуудын эмзэг мэдээллийг арилгах, Их өгөгдлийн санд нэгтгэх мэдээллийн түлхүүр үзүүлэлтүүдийг шифрлэх, мэдээллийн үнэн зөв байдлыг мэдээлэл хариуцагч тал хариуцах, мэдээлэл нийлүүлсэн болон ашиглалтын лог, засвар өөрчлөл зэрэг нарийн төвөгтэй үйл ажиллагаануудыг хийж гүйцэтгэнэ. Энэ нь Их өгөгдлийн сангийн ажиллах журам, үйл ажиллагааны стандарт зэрэг бичиг баримтуудыг боловсруулан гаргах, байнга шинэчлэн сайжруулах боловсруулах мөрдлөг болгох үйл ажиллагаануудыг хамруулна.

ДЭД БҮТЦИЙГ ТӨЛӨВЛӨХ /PIAN THE INFRASTRUCTURE/ 1.1.1 ӨГӨГДЛИЙН ХЭМЖЭЭ БА ӨСӨЛТ /DATA VOLUME AND GROWTH/

Их өгөгдлийн санг зохион байгуулахад хэдий хэмжээний өгөгдөл цуглах, тэдгээрийн хэмжээг тооцоолох болон ойролцоогоор өсөлт ямар хэмжээнд байх асуудлыг судалж үзлээ. Өнөөдрийн байдлаар 4 суурь болон 152 төрөлжсөн мэдээллийн сангуудад нийт иргэн, байгууллагуудын өгөгдөл болон тэдгээртэй холбоотой үйлчилгээ, үйл ажиллагааны мэдээллүүдийг цуглуулна гэж тооцлоо. Энэхүү тооцоололд мэдээлэл хариуцагчаас шаардлагатай өгөгдлүүдийг ETL, ELT байдлаар татна гэж тооцлоо. Тухайлбал е-баримт, гаалийн мэдээ зэрэг өгөгдлүүдийг тодорхой түвшинд нэгтгэсэн байдлаар Их өгөгдлийн санд нийлүүлэх шаардлагатай. Эс тэгвээс маш том хэмжээний өгөгдлийг үр ашиг багатайгаар хуулбарласнаас өөрцгүй шийдэл болох болно.

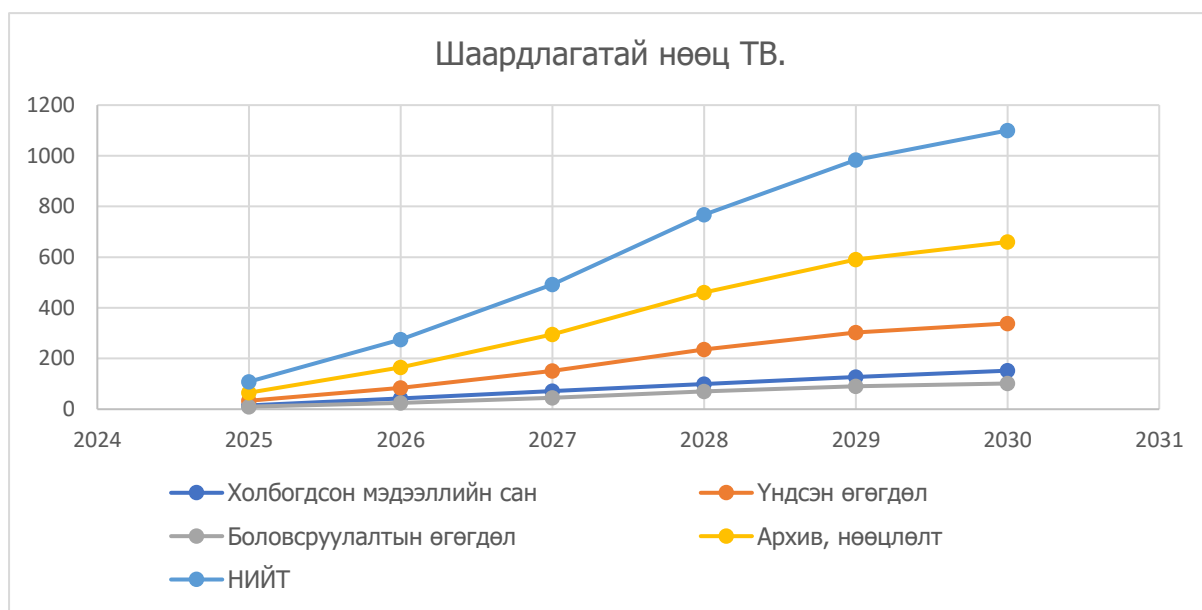
Тооцооллыг өгөгдөл агуулах болон боловсруулах гэсэн хоёр хэсэгт хуваан авч үзлээ. Мэдээлэл хариуцагч байгууллагуудаас өгөгдөл хүлээн авч хадгалах, тэдгээр үзүүлэлтүүдийн шинэчлэлтийг хадгалахын зэрэгцээгээр шинээр тайлан хүснэгт боловсруулах, datawarehouse зохион байгуулах, нэмэлт үзүүлэлт тооцох зэргээр цуглуулсан өгөгдлөөр нэмэлт өгөгдөл боловсруулах боломжтой.

Тооцооны үзүүлэлтүүд:

- Их өгөгдлийн сан нь 3 сая иргэн, 100'000 ААНБ-н мэдээлэл, жилд дунджаар 800 сая гүйлгээ, түүнтэй уялдсан ебаримтын гүйлгээний тоо, цаг уурын өгөгдөл, тэдгээрийн давтамж зэргийг агуулах бөгөөд нийт 152 төрөлжсөн мэдээллийн сангуудад тус бүрд үл давтагдах 30 үзүүлэлттэй байх ба эдгээр үзүүлэлтүүдийн 15% нь сар бүр шинэчлэгдэж байна гэж тооцлоо
- Их өгөгдлийн санд 4 суурь болон 152 төрөлжсөн мэдээллийн сангууд бүгд нэгэн зэрэг холбогдож өгөгдөл дамжуулахгүй тодорхой үе шаттайгаар нэгдэх үйл явц хэрэгжинэ гэж тооцлоо. 2025 онд 10% буюу 15 мэдээллийн санг нэгтгэх бол цаашид жил бүр 28 мэдээллийн санг нэгтгэснээр 2030 он гэхэд бүх мэдээллийн сангууд холбогдоно.
- Үндсэн өгөгдлийн хадгалалтаас гадна тайлан, data warehouse, шинжилгээ зэрэг хэрэглээнд 30% нэмэлт өгөгдөл бий болно.
- Системийн найдвартай ажиллагаа, өгөгдлийн аюулгүй байдлыг хангах зорилгоор архивлалт(cold storage), нөөцлөлт (backup) зэрэгт нийт өгөгдлийн 150% нэмэлт багтаамж шаардлагатай гэж үзлээ. Энд архивлалт буюу удаан хугацаанд хадгалах нөөцийн хэмжээ түүхий өгөгдлийн 50% байхаар, харин шаардлагатай үед системийг сэргээх нөөцийг 1:1 байдлаар тооцов.

Хүснэгт 1. Шаардлагатай нөөцийн хэмжээ. Хэмжих нэгж ТВ.

ОН	Холбогдсон мэдээллийн сан	Үндсэн өгөгдөл	Боловсруулалтын өгөгдөл	Архив, нөөцлөлт	НИЙТ
2025	15	33.39	10.02	65.16	108.57
2026	43	84.62	25.38	164.95	274.95
2027	71	151.41	45.41	295.27	492.09
2028	99	236.03	70.79	460.23	767.14
2029	127	302.81	90.82	590.55	984.18
2030	152	338.49	101.51	660.00	1100.00



1.1.2 ТОНОГ ТӨХӨӨРӨМЖ /HARDWARE/

Дээрх хэмжээний өгөгдөл боловсруулах болон тасралтгүй ажиллагааг хангахыг тулд дата төв нь Tier III болон Tier IV түвшний шаардлага хангасан байх нь оновчтой гэж үзэж байна.

Tier III Vs Tier IV Харьцуулалт

	Шалгуур үзүүлэлт	Tier III	Tier IV
1	Цахилгаан тэжээлийн эх үүсвэр	Нэг цахилгаан тэжээл, N+1 UPS + Generator	Бие даасан 2N цахилгаан тэжээл, тусгаарлагдсан замаар
2	Сүлжээний холболт	Нэг үндсэн, нэг нөөц сүлжээ	2 тусдаа ISP, 2 үндсэн сүлжээний зам
3	Хөргөлтийн систем	N+1 хөргөлтийн систем	2N хөргөлтийн систем
4	Нөөцлөлт (Redundancy)	N+1	2N (ижил хүчин чадал бүхий бүрэн давхарласан)
5	Сэлгэн ажиллах чадвар (Concurrently Maintainable)	Бүх тоног төхөөрөмжийг тасралтгүй ажиллуулах үед солих боломжтой	Бүх системийг засварлах, солиход ч ажиллагаа тасалдахгүй
6	Fault tolerance (Алдаа даах чадвар)	Хязгаарлагдмал (зөвхөн зарим систем)	Бүрэн fault-tolerant, алдаатай системийг шууд орлуулна
7	Гамшгийн үеийн нөхөн сэргээлт (DR site)	Санал болгогдсон ч шаардлагагүй	Шаардлагатай (DR төлөвлөгөө, автомат failover)
8	Үйл ажиллагааны тасралтгүй байдал	~72 цагийн дотор засварлах боломжтой	Ямар ч доголдолгүй 24/7 ажиллагаа
9	Тоног төхөөрөмж солих хугацаа	~12-24 цаг	< 1 цаг
10	Жилийн дундаж ажиллагааны хугацаа (Availability)	~99.982% (1.6 цаг тасалдал/жил)	~99.995% (26 минут тасалдал/жил)

• СЕРВЕРҮҮД /SERVERS/

Их өгөгдлийн сангийн серверүүд нь нийт системийн хэвийн ажиллагааг тасралтгүй хангахын зэрэгцээгээр өгөгдөл боловсруулалтын онцлог байдлаас хамааран өөр өөр шинж чанартай байх шаардлага үүснэ.

- Тоон өгөгдөл боловсруулах өндөр хүчин чадал бүхий blade серверүүд шаардлагатай.
 - CPU: Өндөр хурдтай процессор (Intel Xeon, AMD EPYC) ашиглах шаардлагатай. 24-48 цөмтэй, 3.0 GHz+ процессор
 - RAM: 512 GB - 1 TB RAM
 - Storage: SSD ба HDD хосолсон шийдлүүдийг ашиглах. 10-50 TB SSD, 100-500 TB HDD
- AI/Big Data хэрэглээнээс хамаарч GPU server (NVIDIA A100, H100) зэрэг тусгай зориулалтын төхөөрөмж шаардагдана.

- Тайлангийн боловсруулалт, өгөгдлийн агуулах (Datawarehouse) байгуулах, өгөгдөл татах зорилгоор ETL engine бүхий серверүүд байж болно.

Серверийн жишиг загвар

	Загвар	CPU	RAM	Storage
1	Dell PowerEdge R750	2x Intel Xeon Gold 6338	512 GB	8x 2TB SSD
2	HPE ProLiant DL380 Gen11	2x Intel Xeon Silver 4314	256 GB	4x 2TB SSD
3	Lenovo ThinkSystem SR650	2x Intel Xeon Gold 6348	512 GB	8x 2TB SSD
4	Supermicro Ultra	2x Intel Xeon Platinum 8358	1 TB	8x 4TB NVMe
5	NVIDIA DGX A100	8x NVIDIA A100 + AMD EPYC	1 TB	15 TB NVMe

Их өгөгдлийн сан байгуулахад 2025 онд 3-4 тооны сервер (32 core CPU, 128–256 GB RAM) шаардагдах ч 2030 он гэхэд 10+ серверүүд шаардлагатай болно.

- Серверийн тооцоолол:
 - CPU: (Intel Xeon, AMD EPYC) , 24-48 core, 3.0 GHz+
 - RAM: 512 GB - 1 TB RAM
 - Storage: 10-50 TB SSD, 100-500 TB HDD
- Серверийн тоо:
 - 10, үүнд виртуалчлал болон өгөгдлийн сүлжээний зориулалтаар ашиглах серверүүд орно.
- Серверийн кластер:
 - Өндөр ачаалалтай системийн гүйцэтгэл, чанар, хэвийн ажиллагааг хангахын үүднээс кластер аргыг ашиглах.
- Load Balancing:
 - Серверийн ачааллыг тэгш хуваарилах.

• ХАДГАЛАХ /STORAGE/:

Төрийн нэгдсэн сангийн өгөгдлийн хадгалалтын шийдлийг тодорхойлох нь системийн урт хугацааны тогтвортой ажиллагаанд чухал үүрэг гүйцэтгэнэ.

- Дата төв:
 - SAN (Storage Area Network): Төвлөрсөн хадгалалт бүхий том хэмжээний өгөгдөл хадгалах.
 - NAS (Network Attached Storage): Өгөгдлийг хуваарилах зориулалттай, хэрэглэгчдэд илүү ойр хадгалалт.
- Үүлэн технологи (Cloud Storage):
 - Amazon S3, Google Cloud Storage, Microsoft Azure Storage: Эдгээр нь өргөтгөх, аюулгүй байдал, хямд өртөгтэй ашиглах боломжийг бий болгодог.

Хадгалалтыг дата төв буюу уламжлалт шийдлийг үүлэн технологитой хослуулан хэрэглэх ч тохиолдол байдалд. Хадгалалтын үр ашигтай байдал, ашиглалт зэргээс хамааран өгөгдөл боловсруулах, ашиглахад зориулан SSD, харин архив, нөөцтэй холбоотой өгөгдлийг HDD ашиглан хадгалах нь түгээмэл байдаг.

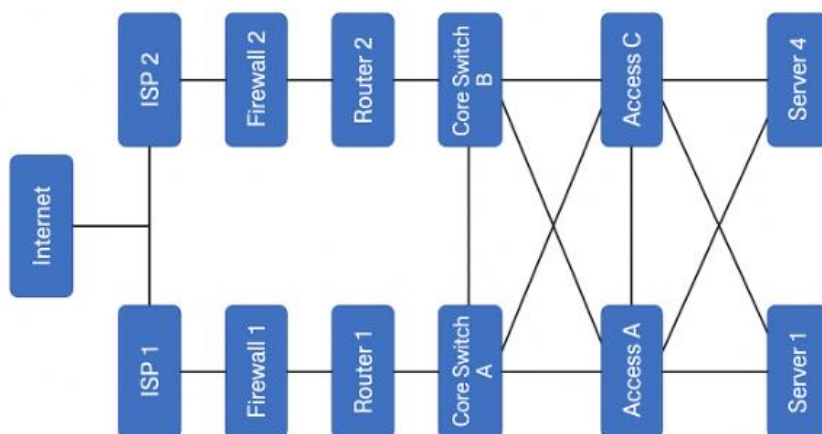
ХУВИЛБАР	ДАВУУ ТАЛУУД	СУЛ ТАЛУУД
On-Prem SAN/NAS	Хяналт, аюулгүй байдал сайн	Анхны зардал өндөр, өргөтгөл төвөгтэй
Cloud (e.g., AWS S3, Azure Blob)	Уян хатан, өргөтгөл хялбар	Зардал урт хугацаанд өснө, интернэт хамаарал
Hybrid	Хосолсон давуу талтай	Удирдлага, зохицуулалт төвөгтэй

Байгуулагдаж буй их өгөгдлийн сан буюу энэхүү систем нь хууль тогтоомжийн хүрээнд мэдээлэл солилцооны үндсэн систем болох эсэхээс хамааран өгөгдлийг хаана хадгалах талаар тусгасан байдаг. “Нээлттэй мэдээллийн ил тод байдлын тухай” хуулийн 18.12. “Үндсэн системийг Монгол Улсын нутаг дэвсгэрт байршуулна” гэсэн заалттай хамаарах эсэхийг урьдчилан тодорхойлох нь хадгалалтын оновчтой шийдвэрийг хэрэгжүүлэхэд чухал нөлөөтэй.

- СҮЛЖЭЭ /NETWORKING/

Их өгөгдлийн дата төвийн сүлжээний орчин нь найдвартай ажиллагаа, тасалдалгүй байдлыг хангасан олон улсын жишигт нийцсэн өндөр хурдны uplink бүхий дата төвийн backbone, төрөлжсөн мэдээллийн сангуудтай холбогдон ажиллах IPsec VPN, MPLS орчинтой, Load balancing, firewall, routing бүхий бүрэн шийдлийг агуулсан байна.

Зураг. Сүлжээний топологи



Core Network Layer (Гол сүлжээний давхарга)

- Үүрэг: Дата төвийн бүхий л дотоод болон гадаад удирдана
- Техник шийдэл:
 - 2 ширхэг *core switch/router* ашиглана (redundant, Аль нэг core switch доголдсон ч сүлжээ ажиллана).
 - Олон uplink буюу өндөр зурвасын өргөнтэй, 40G/100G холболттой.

Distribution Layer (Түгээлтийн давхарга)

- Үүрэг: Core switch-ээс орж ирсэн өгөгдлийг access давхаргад тараана.
- Техник шийдэл:
 - Top-of-Rack (ToR) болон End-of-Row (EoR) switch-үүдтэй холбогдоно (Distribution switch бүр core switch-тэй хоёр талын redundant uplink-тай байна).
 - Spanning Tree Protocol (STP), VRRP, HSRP зэрэг протоколыг redundancy-д ашиглана.
 - VLAN segmentation болон ACL (Access Control List) хэрэгждэг.

Access Layer (Хандалтын давхарга)

- Үүрэг: Серверүүд, өгөгдлийн санд холбогдох дотоод сүлжээний терминалуудыг зохион байгуулна.
- Техник шийдэл:
 - ToR switch бүр серверийн rack-д суурилна.
 - L2 switching + Port channel/LACP ашиглана.
 - Хос uplink connection (active-active эсвэл active-passive хэлбэрээр).

Out-of-Band Management Network (Хяналтын тусдаа сүлжээ)

- Үүрэг: Тоног төхөөрөмжийн удирдлага, log болон мониторингийн мэдээллийг дамжуулах хяналтын сүлжээ.
- Техник шийдэл:
 - IPMI, iLO, DRAC порт ашиглана.
 - Тусдаа VLAN болон физик сүлжээгээр тусгаарлагдсан байна.
 - Аюулгүй байдлын firewall, ACL хяналттай.

Internet & External Connectivity (Гадаад холболт)

- Үүрэг: Дата төвийн хэрэглэгчид, төрийн байгууллагууд, бусад дата төвүүдтэй холбогдох.
- Техник шийдэл:
 - 2 болон түүнээс дээш интернэт үйлчилгээ үзүүлэгч (ISP) ашиглана.
 - Бие даасан BGP ASN ба Public IP range-тэй байх.
 - Redundant чиглэл бүхий шилэн кабель (MPLS, DIA).

Firewall & Load Balancer давхарга

- Үүрэг: Аюулгүй байдал болон ачаалал хуваарилалтыг хангах.
- Firewall: NGFW буюу дараагийн үеийн firewall ашиглана (Layer 7 хүртэл хяналттай).
- Load Balancer: Web, API, DB зэрэг системүүдийн хооронд урсгалыг тэнцвэржүүлнэ.

Network Operations Center (NOC) ба SDN хяналт

- NOC: 24/7 хяналттай, real-time monitoring систем зохион байгуулах.
- SDN (Software Defined Networking):
 - Сүлжээг нэгдсэн удирдлагатай болгох, аюулгүй байдлыг дээшлүүлэх.
 - Сүлжээний ачааллыг удирдах.

Жишиг сүлжээний тоног төхөөрөмжийн харьцуулалт

	Тоног төхөөрөмж	Үндсэн үзүүлэлтүүд	Ойролцоогоор үнэ (USD)
1	Core Switch – Cisco Catalyst 9500	40G/100G uplinks, Layer 3, StackWise Virtual	\$15,000 – \$25,000
2	Core Switch – Arista 7280R	100G ports, deep buffers, BGP/MPLS support	\$18,000 – \$30,000
3	Distribution Switch – Cisco 9300	24/48 ports, PoE+, Layer 2/3 switching	\$6,000 – \$9,000
4	Distribution Switch – Aruba 6300M	Modular, PoE, stacking up to 10 devices	\$5,000 – \$8,000
5	Access Switch – Juniper EX4300	1/10G ports, Virtual Chassis, Layer 3	\$4,000 – \$6,500
6	Firewall – Fortinet FortiGate 100F	1 Gbps throughput, SSL inspection, SD-WAN	\$3,500 – \$5,000
7	Firewall – Palo Alto PA-3220	5 Gbps threat prevention, App-ID, URL filtering	\$10,000 – \$15,000
8	Load Balancer – F5 BIG-IP i2600	10 Gbps throughput, L7 LB, SSL offload	\$20,000 – \$35,000
9	Router – Cisco ASR 1001-X	20 Gbps throughput, MPLS, BGP/OSPF	\$12,000 – \$18,000
10	Router – Juniper MX204	80 Gbps throughput, full-featured routing	\$25,000 – \$40,000

- СЭЛГЭН АЖИЛЛАХ, БЭЛЭН БАЙДАЛ /REDUNDANCY AND HIGH AVAILABILITY, RECOVERY/

Сэлгэн ажиллах, бэлэн байдал (Redundancy and High Availability) ба Гамшгийн үеийн нөхөн сэргээлт (Disaster Recovery) нь дата төвийн хувьд заавал хангасан байх ёстой нөхцөл юм. Эдгээр нь дата төвийн тасралтгүй ажиллагааг хангах, шаардлагатай үед шуурхай нөхөн сэргээх үйл ажиллагааны цогц юм.

СЭЛГЭН АЖИЛЛАХ ЧАДВАР /REDUNDANCY/

- N+1 redundancy: Сүлжээ, цахилгаан, хөргөлтийн шийдлүүдэд ямар нэгэн төхөөрөмж доголдсон ч системийн хэвийн ажиллагаа хангагдана.
- Active-Active эсвэл Active-Passive топологи: Төв болон нөөц серверүүд, сүлжээ, өгөгдлийн сан зэрэг нь хоёр тусдаа физик цэг дээр байрлана.

ДЭД БҮТЭЦ	ШААРДЛАГА
Серверүүд	Нөөц (failover) сервертэй байх (virtualization боломжтой)
Storage	RAID 10/RAID 6 нөөцлөлт, dual power path
Сүлжээ	Хос firewall, хос switch
Цахилгаан тэжээл	UPS + дизель генератор, цахилгааны давхар эх үүсвэр
Хөргөлт	Хос агааржуулалтын систем (N+1 configuration)

БЭЛЭН БАЙДАЛ /HIGH AVAILABILITY/

Бэлэн байдлыг хангах хамгийн чухал үйл ажиллагаа нь байнгын тогтмол хяналт, автомат хяналтын систем болон системийн үйл ажиллагааны тасалдалгүй байдлыг хангасан үйл ажиллагааны стандарт юм.

- 24/7 хяналтын систем (NOC)
- Автомат мониторинг, alert систем

- SLA: $\geq 99.982\%$ uptime (Tier III түвшний дата төвийн хувьд хамгийн ихдээ жилд 1.6 цагийн тасалдал зөвшөөрөгдөнө)

НӨХӨН СЭРГЭЭЛТ /DISASTER RECOVERY/

Гэнэтийн гамшиг ослын тухайд хэрэгжүүлэх үйл ажиллагааны төлөвлөгөө, бэлэн байдлыг хангах явдал байдаг. Энэ нь тоног төхөөрөмж төдийгүй хүний нөөц, сургалт, үйл ажиллагааны заавар, журам зэрэг цогц үйл ажиллагааны хүрээнд хангагддаг. Нөхөн сэргээх үйл ажиллагаа нь байнга тохиолддог биш ч бэлэн байдлыг тогтмол хангах, заавар журмуудыг тогтмол шинэчлэх, хүний нөөцийг бэлтгэж дадлаажуулах үйл ажиллагааг сайн хангаснаар учирч болох эрсдэлийг хохирол багатай туулан гарах нөхцөл бүрддэг.

Иймд нөхөн сэргээх үйл ажиллагааны төлөвлөгөөг нарийвчлан боловсруулсан байх шаардлагатай. Төлөвлөгөө нь:

- Эрсдэлийн үнэлгээ (Risk Assessment)
 - Зорилго:

Системд учирч болох голлох аюул, эрсдэлийг тодорхойлон, тэдгээрийн нөлөө, давтамжийг тооцох.
 - Агуулга:
 - Физик эрсдэл: Газар хөдлөлт, гал, усны үер, цахилгааны тасалдал зэрэг
 - Техникийн эрсдэл: Сервер, сүлжээ, DB алдаа, виртуал орчны доголдол
 - Кибер халдлага: Ransomware, вирус, гаднын халдлага, өгөгдөл алдагдал
 - Хүний хүчин зүйл: Зохисгүй эрхтэй нэвтрэлт, санамсаргүй устгал зэрэг үйл ажиллагаа
 - Үйл ажиллагаа:
 - Эрсдэл бүрийн хувьд [магадлал] $\times$ [нөлөөлөл] оноо өгч үнэлгээ хийх
 - Critical/High/Medium/Low түвшинд ангилал хийх
 -
- Plan Maintenance & Validation
 - Зорилго:

Нөхөн сэргээх төлөвлөгөө хуучирч, хэрэгжүүлэх боломжгүй болохоос сэргийлэх.
 - Агуулга:
 - Төлөвлөгөөг тогтмол шинэчлэх (жилийн 1–2 удаа)
 - Төлөвлөгөөг тоног төхөөрөмж, системийн шинэчлэлтэй уялдуулж сайжруулах
 - Хувилбарын удирдлагыг ашиглан (version control) төлөвлөгөөний өөрчлөлтийг баримтжуулах
 - Баталгаажуулалт:
 - Төлөвлөгөөг бодитоор хэсэгчлэн турших (tabletop exercise, failover simulation)

- Нөхөн сэргээх үйл ажиллагааг хэмжих шалгуур: RTO, RPO хангаж байгаа эсэхийг тодорхойлох
- Тестлэх хуваарь
 - Зорилго:

Нөхөн сэргээх төлөвлөгөө бодит нөхцөлд хэр ажиллахыг шалгах.
 - Төрөл:
 - Tabletop exercise: Албан ёсны ярилцлага хэлбэрээр DR тохиолдлыг онолын хувьд ярилцах
 - Simulation: Бодит орчинд хэсэгчилсэн/бүтэн DR үйлдэл хийх
 - Failover Test: Үндсэн системийг унтрааж DR site руу шилжүүлэх
 - Хуваарь:
 - Хагас жил тутам зохион байгуулах
 - Туршилтын үр дүнг дүгнэлт, сайжруулалтын зөвлөмж болгон нөхөн сэргээх төлөвлөгөөнд тусгах
- Хариуцсан хүний нөөц, харилцах мэдээлэл (DR Roles & Contact List)
 - Зорилго:

DR үед хурдан хариу өгөх, мэдээллийн солилцоог жигд зохион байгуулах
 - Агуулга:

DR Team: Хариу арга хэмжээ авах баг (техникч, систем админ, сүлжээ инженер, мэдээллийн аюулгүй байдал гэх мэт)
 - Үүргийн хуваарилалт:
 - DR зохицуулагч
 - Сэргээн босгох баг
 - Харилцаа холбооны менежер
 - Харилцах сувгууд:
 - Албан и-мэйл, утасны жагсаалт, сошиал сувгууд
- Сэргээх алхам бүрийн баримтжуулалт (Step-by-Step Recovery Documentation)
 - Зорилго:

Хэрхэн, ямар дарааллаар систем, өгөгдлийг сэргээхийг дэлгэрэнгүй тайлбарласан гарын авлага байх
 - Агуулга:
 - Сэргээх алхам алхмын жагсаалт (жишээ: 1. Virtual machine restore → 2. DNS re-route → 3. Load balancer failover)

- Хугацаа (RTO), өгөгдлийн түвшин (RPO)-ийн зорилт тус бүрд тодорхойлогдсон байна
- Screen capture, CLI командыг оролцуулсан байх
- Баримт бичгийг систем тус бүрд хуваарилсан байна (DB сэргээх, app restore гэх зэрэг)

Гамшгийн нөхөн сэргээлт (Disaster Recovery):

- RTO (Recovery Time Objective): Өгөгдлийг сэргээх хугацааг 4-6 цаг байх.
- RPO (Recovery Point Objective): Өгөгдлийн алдагдлын хугацаа 1-2 цаг байх.

Сэргээх шийдлүүд:

- Offsite Backup: Архивын нөөцөөс сэргээх.
- Automated Backups: Өдөр тутам нөөц авах үйл ажиллагааг автоматжуулах.

ДҮГНЭЛТ

Энэхүү бүлэгт төрийн мэдээлэл хариуцагч байгууллагуудаас өгөгдөл татан төвлөрүүлж, хадгалах, боловсруулах, шинжлэх, тайлагнах зорилготой Их өгөгдлийн сан болон холбогдох дата төвийн дэд бүтцийг 2025–2030 онд үе шаттайгаар хэрэгжүүлэх техникийн үндэслэл, шаардлагыг тодорхойллоо.

ӨГӨГДЛИЙН ХЭМЖЭЭ БА ӨСӨЛТИЙН ТААМАГЛАЛ

- Өгөгдлийн эх үүсвэр: 4 суурь ба 152 төрөлжсөн мэдээллийн сан.
- Өсөлтийн үе шат:
 - 2025 онд эхний 15 мэдээллийн санг нэгтгэж, дараа жил бүр 28 мэдээллийн санг үе шаттайгаар нэгтгэн 2030 он гэхэд бүрэн холбох.
- Хэмжээний тооцоо:
 - 3 сая иргэн, 100,000 ААНБ-ын 30 үзүүлэлт бүхий өгөгдөл
 - Сар бүр 15% шинэчлэгдэх, жилийн гүйлгээний 800 сая бичлэг
 - 2030 он гэхэд үндсэн өгөгдөл + боловсруулалт + архив/нөөцлөлтийн багтаамж нийлээд нийт 1 PB хүрэх төсөөлөл.
- Өгөгдлийн төрөл:
 - Үндсэн өгөгдөл
 - Тайлангийн болон дата агуулах (Datawarehouse)
 - Архивлалт ба сэргээх нөөц.

Тооцоолол нь өгөгдлийг шат дараатай татаж, хуулбарлалт багатайгаар үр ашигтай нэгдүүлэхийг онцолж байна.

ТОНОГ ТӨХӨӨРӨМЖ БА АРХИТЕКТУР

Серверүүд

- Үүрэг:
 - Тоон өгөгдөл боловсруулах
 - Тайлан, AI/Big Data боловсруулалт
 - ETL процессыг гүйцэтгэх
- Техникийн шаардлага:
 - CPU: 24–48 цөмтэй Xeon/EPYC процессор, 3.0 GHz+
 - RAM: 512 GB – 1 TB
 - Storage: 84 x 14TB NL-SAS(≈1.176 PB raw)
 - GPU серверүүд: NVIDIA A100/H100
- Тоон өсөлт:
 - 2025 онд 3–4 серверээр эхэлж, 2030 онд 10+ сервер ашиглах.
- Зохион байгуулалт:
 - Кластер (Cluster)
 - Load Balancing
 - Virtualization

Хадгалалт (Storage)

- Архив ба нөөцлөлт:
 - Түүхий өгөгдлийн 50% архивлалт
 - 1:1 хэмжээтэй сэргээх нөөц

Найдвартай ажиллагаа, бэлэн байдал

- Redundancy (N+1):
 - Сервер, storage, сүлжээ, цахилгаан тэжээл бүгд нөөцтэй
- High Availability:
 - 99.982% uptime (Tier III)
 - 24/7 мониторинг, SLA
- Disaster Recovery:
 - RTO: 4–6 цаг
 - RPO: 1–2 цаг
 - Хагас жил тутам DR simulation туршилт
 - Сэргээх алхам бүрийг баримтжуулсан гарын авлага

ДАВУУ ТАЛ БА ЭРСДЭЛ

Давуу тал:

- Олон эх үүсвэрийн өгөгдлийг төвлөрүүлж шат дараатай нэгтгэх
- Өргөтгөх боломжтой архитектур
- Олон улсын Tier III-IV дата төвийн шаардлага хангасан дэд бүтэц

Эрсдэл:

- Архив, нөөц хадгалалтад төсөөлснөөс их багтаамж хэрэг болох магадлал
- Үүлэн хадгалалт ашиглавал зардлын өсөлт, интернэт хамаарал нөлөөлж болно
- Нэгтгэх мэдээллийн чанарын асуудлууд
- DR (гамшгийн нөхөн сэргээлт) төлөвлөгөө тогтмол турших шаардлага

АМЖИЛТТАЙ ХЭРЭГЖИХ ГОЛ ХҮЧИН ЗҮЙЛС:

- Өгөгдөл татан төвлөрүүлэх ETL процессыг үе шаттай хэрэгжүүлэх
- Сүлжээний бэлэн байдал, redundancy-ийг хангах
- DR төлөвлөгөөг тогтмол сайжруулж баталгаажуулах
- Хүний нөөц, мониторинг, үйлчилгээний чанарыг өндөр түвшинд барих

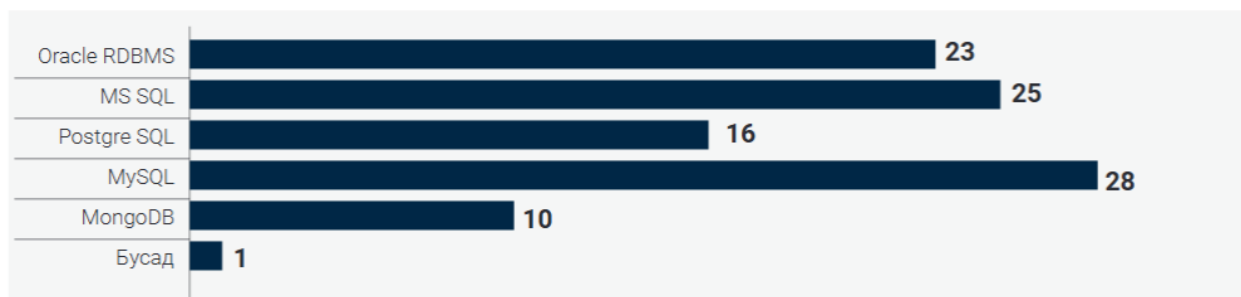
ИХ ӨГӨГДЛИЙН ПРОГРАММ ХАНГАМЖ, БАГАЖ ХЭРЭГСЛИЙН СУДАЛГАА /SOFTWARE AND TOOLS/

1.2.1 ӨГӨГДЛИЙН САНГИЙН ТЕХНОЛОГИ

Их өгөгдлийн сан байгуулах нь олон төрлийн өгөгдлийг нэгтгэн хадгалах, боловсруулах, дүн шинжилгээ хийх зорилготой бөгөөд энэхүү үйл явцыг хангах өгөгдлийн сангийн технологийг хэрхэн оновчтой сонгохоос ихээхэн хамаарах юм. Их өгөгдлийн сан нь бүтэцлэгдсэн, хагас бүтэцлэгдсэн болон бүтэцгүй өгөгдөл татах боломжтой. Хэдийгээр төрийн мэдээллийн сангууд буюу төрөлжсөн мэдээллийн 152 сангуудыг дийлэнх нь бүтэцлэгдсэн байдлаар хөгжүүлэгдсэн хэдий ч цаашид олон эх сурвалжуудаас өгөгдөл цуглуулах хэрэгцээ шаардлагын хүрээнд их өгөгдлийн санг нь ямар ч төрлийн өгөгдөл хүлээн авах бололцоог хадгалсан байх нь зүйтэй юм.

- Бүтэцлэгдсэн өгөгдөл
 - Тодорхой бүтэцтэйгээр тодорхойлогдож хөгжүүлсэн мэдээллийн сангууд (жишээ нь, иргэдийн бүртгэл, татварын мэдээлэл) бөгөөд энэ төрлийн өгөгдлийг RDBMS өгөгдлийн сангаар удирдах боломжтой.
- Хагас бүтэцлэгдсэн өгөгдөл
 - Мэдээллийн зарим хэсэг нь XML, JSON хэлбэрээр хадгалагдах бөгөөд тэдгээрийг боловсруулахад NoSQL системүүд тохиромжтой байж болно.
- Бүтэцгүй өгөгдөл
 - Төрийн мэдээллийн санд зураг, аудио бичлэг зэрэг бүтэцгүй өгөгдөл ч байж болох тул эдгээрийг хадгалахад мэдээллийн нуур буюу Data Lake шийдэл шаардлагатай.

Төрийн байгууллагуудад ашиглагдаж буй нийтлэг өгөгдлийн сангийн системүүд



Эх сурвалж: <https://www.1212.mn/mn/statistic/file-library/view/98925351>

- RDBMS

Төрийн байгууллагуудын дунд хийгдсэн судалгаагаар MySQL, MS SQL, Oracle, PostgreSQL өгөгдлийн сан удирдах системүүд түлхүү ашиглагдаж байна. Гэхдээ энэ нь тухайн байгууллагын хүрээнд тодорхойлсон зорилт, зорилгын хүрээнд хийсэн өгөгдлийн сангийн сонголтууд гэдгийг харгалзан үзэх нь зүйтэй болов уу.

Их хэмжээний өгөгдөл удирдаж байгаа байгууллагуудын хувьд Oracle RDBMS системийг нийтлэг байдлаар ашиглаж байна.

- NoSQL DBMS

Хагас бүтэцтэй болон бүтэцгүй өгөгдлийг боловсруулахад NoSQL системүүд (MongoDB, Cassandra)-ийг ашиглах нь сайн сонголт байж болох юм. Эдгээр системүүд нь өгөгдлийн хурдтай боловсруулахад тохиромжтой бөгөөд их хэмжээний өгөгдлийг хадгалах боломжтой байдаг.

ТӨРӨЛ	НИЙТЛЭГ DBMS	ДАВУУ ТАЛУУД	СУЛ ТАЛУУД	ҮНЭ (ЖИЛИЙН)
Бүтэцлэгдсэн	PostgreSQL, Oracle, MS SQL Server	ACID transaction, SQL query, BI/Reporting	Volume scalability бага, лицензийн төлбөр өндөр (Oracle)	PostgreSQL – үнэгүй, Oracle EE ~\$47,500/CPU, MS SQL Std ~\$3,700/CPU
Хагас бүтэцтэй	MongoDB, Couchbase	Schema flexible, JSON/XML-д оновчтой	Complex query дэмжлэг муу, ACID хязгаарлагдмал	MongoDB ~ \$57+/cap, Couchbase ~\$4,000/node
Бүтэцгүй	Hadoop HDFS, Amazon S3, MinIO	Маш их дата хадгалах боломжтой, өндөр fault tolerance	Query engine шаардагдана, real-time хязгаарлагдмал	Hadoop – үнэгүй, S3 ~\$23/TB/cap, MinIO ~ Enterprise Support \$10,000+
BigData	Apache Hive, HBase, ClickHouse	Batch/streaming data analytics, scalability өндөр	Нарийн тохируулга шаарддаг, latency өндөр	Hive/HBase – үнэгүй, ClickHouse – үнэгүй (Cloud: \$0.25/GB)

Их өгөгдлийн сангийн хувьд бүх төрлийн өгөгдлийг (бүтэцтэй, хагас бүтэцтэй, бүтэцгүй) цуглуулж, хадгалах, боловсруулах, шинжилгээ хийх зорилгоор Data Warehouse (DWH) болон Data Lake системүүдийг хослуулан ашиглах нь оновчтой шийдэл болно.

Data Warehouse

- Бүтэцлэгдсэн, баталгаажсан, боловсруулагдсан өгөгдөл
- BI тайлан, дашбоард, OLAP анализ хийх зориулалттай
- Мэдээллийн загварчлал (star/snowflake schema)-д үндэслэн
- Өндөр гүйцэтгэлтэй SQL engine дээрх тусгай орчин юм.

ШИНЖ ЧАНАР	ТАЙЛБАР
Өгөгдлийн төрөл	Бүтэцлэгдсэн
Шүүлт/тайлан	Хугацаагаар(сар, улирал, жилээр г.м) дүгнэлт хийхэд тохиромжтой
Формат	SQL (RDBMS)
Гүйцэтгэл	Өндөр хурдтай (index, partition ашигладаг)

Data Lake

- Бүтэцтэй, хагас бүтэцтэй, бүтэцгүй зураг, видео, аудио, JSON, PDF, лог файл гэх мэт) бүх төрлийн өгөгдөл
- Байршуулж хадгалах зориулалттай, том хэмжээний файл төвлөрсөн сан юм.

ШИНЖ ЧАНАР	ТАЙЛБАР
Өгөгдлийн төрөл	Аль ч төрөл (structured, semi-structured, unstructured)
Боловсруулалт	ML/AI, raw data profiling, exploratory analysis
Формат	Parquet, ORC, Avro, CSV, JSON, MP4, JPEG
Гүйцэтгэл	Хурдтай;сайн тохируулга хийх шаардлагатай

Харьцуулалт

ҮЗҮҮЛЭЛТ	DATA WAREHOUSE	DATA LAKE
Өгөгдөл боловсруулалт	Боловсруулсан өгөгдөл (curated)	Түүхий өгөгдөл (raw)
Санах ой ашиглалт	Хадгалах хэмжээ багатай	Маш их хадгалалт
Боловсруулалт	BI/OLAP	AI/ML, data discovery
Зорилго	Тайлан, тогтмол хяналт	Илүү гүн шинжилгээ, R&D

Эдгээрээс үүдэн их өгөгдлийн сангийн өгөгдөлтэй ажиллах урсгал нь:

- Raw data → Data Lake
- Data Lake → ETL → Data Warehouse
- Data Warehouse → Тайлан, BI, OLAP
- Data Lake → ML/AI загварчлал байх боломжтой

ШИЙДЭЛ	ЗОРИУЛАЛТ	ЖИШЭЭ ХЭРЭГСЭЛ
Data Warehouse	Тайлан, BI	PostgreSQL, Amazon Redshift, Snowflake, Google BigQuery
Data Lake	Түүхий өгөгдөл, AI	Hadoop HDFS, AWS S3, Azure Data Lake, MinIO
ETL	Өгөгдөл нэгтгэх	Apache NiFi, Airflow, n8n, Talend

Их өгөгдлийн сан, дэд бүтцийг бий болгож үйл ажиллагаанд ашиглах нь шинэ тутам хөгжиж байгаа салбар бөгөөд үүнд хүний нөөц, тоног төхөөрөмж болон мэдээлэл хариуцагч байгууллагууд зэрэг олон талын хамтын ажиллагааг шаарддаг. Энэ хүрээнд

олон систем, технологиуд хөгжиж байгаа бөгөөд манай орны хувьд их өгөгдөлтэй ажиллаж буй сайн туршлагууд ч хуримтлагдаж эхлээд байна.

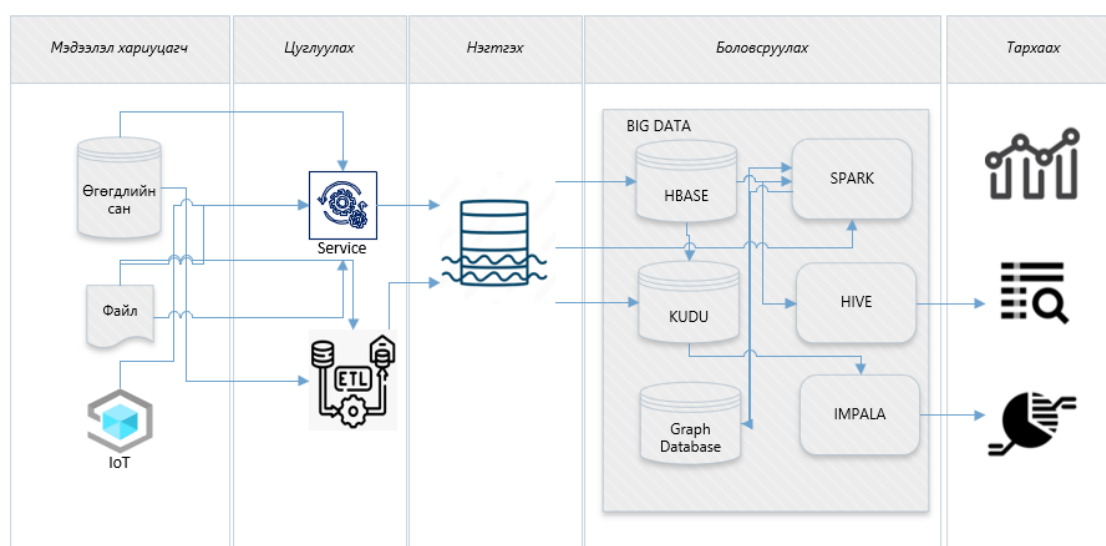
Мөн их өгөгдлийн санд холбогдон ажиллах мэдээлэл хариуцагч байгууллагуудын хувьд тодорхой нөхцөл, бэлтгэл ажлыг хангах зайлшгүй шаардлагатайг анхаарах хэрэгтэй.

Бид мэдээлэл хариуцагч байгууллагууд дээрх staging server дээрээс өгөгдлийг хоёр байдлаар татахаар төлөвлөлөө. Мэдээлэл хариуцагч байгууллагууд нь их өгөгдлийн санд нийлүүлэхээр урьдчилан тохирсон өгөгдлөө бэлтгэн өөрсдийн орчинд буюу staging server-т байршуулсны дараа их өгөгдлийн санд тухайн өгөгдлийг татна. Энэ нь их өгөгдлийн санд мэдээлэл хариуцагч байгууллагуудаас баталгаажсан буюу албажсан өгөгдөл татах, алдаа гарах эрсдэлээс давхар сэргийлсэн үйл ажиллагааны зохицуулалт юм.

Мэдээлэл хариуцагч байгууллагаас:

- Тодорхой тогтсон хуваарийн дагуу өгөгдлийг татах
- Real-time буюу бодит цагийн горимоор өгөгдөл татах. Энэ нь тогтсон сервисүүд, IoT зэргээс өгөгдөл авах боломжийг бүрдүүлнэ.

Их өгөгдлийн санд цугларсан өгөгдөлд дүн шинжилгээ, боловсруулалт хийх нь BigData хүрээний шийдэл болох бөгөөд үүнийг дараах байдлаар технологийн хүрээнд тодорхойллоо.



ӨГӨГДӨЛ ЦУГЛУУЛАХ (ACQUIRE)

Шийдлийг хоёр аргаар шийдвэрлэхээр тооцож орууллаа.

- Уламжлалт болон олон төрлийн өгөгдлийн эх үүсвэрээс цугларах өгөгдлийг Их өгөгдлийг нэгтгэх хэрэгслийг (Big Data integration tool) ашиглан багцлах (batch) аргаар буюу өгөгдлийг багц багцаар нь шатлан цуглуулж авна.
- Харин тасралтгүй горимоор маш хурдацтай үүсэх өгөгдлийг (лог, IOT өгөгдөл, streaming өгөгдөл) тасралтгүй горимоор үүсэх өгөгдлийг KAFKA технологийг ашиглан их өгөгдлийн санд нэгтгэх нь зүйтэй. Энэ технологийн тусламжтайгаар өгөгдлийн урсгалыг бодит хугацаанд нь хоцрогдолгүйгээр удирдах, өгөгдөл дээр хувиргалт хийх, өгөгдөл боловсруулах, өгөгдөл хооронд нэгтгэх гэх мэт олон үйлдэл хийх боломжтой болдог. Мөн энэ хэрэгсэл ашиглан хадгалагдсан өгөгдөл бусад системийн өгөгдлийн нэгдсэн эх сурвалж болдгоороо давуу талтай.

Өгөгдөл цуглуулах дээрх хоёр арган дээр нэмж тодорхой логик оруулж болох хэрэгслээр өгөгдлийг цэвэрлэх, нэгтгэх гэх мэт үйлдлүүд нэмж хийгдэнэ. Их өгөгдлийг нэгтгэх хэрэгслийг ашиглаж байгаа нэг шалтгаан нь уг хэрэгсэлд өгөгдөл цэвэрлэх, нэгтгэх

логикийн дагуу бусад систем (engine) дээр код үүсгэж тухайн систем дээр ажилладгаараа давуу талтай.

ХАДГАЛАХ, БОЛОВСРУУЛАХ (ORGANIZE)

Хадгалах

- Эхний шийдэл нь NoSQL буюу хагас бүтэцлэгдсэн байдлаар HBASE технологийг ашиглан хадгалах юм. NoSQL өгөгдлийн санг цуглуулах өгөгдлийн бүтэц (багана) тодорхойгүй, байнга өөрчлөгдөж өгөгдөл үүсэх давтамж их болох магадлалтай тул ачааллыг даах, бүтэц хамаарахгүйгээр хадгалах чадвартай тул сонгох нь зүйтэй. Мөн HBASE нь их өгөгдөл дээр чөлөөтэй өгөгдлийн шинжилгээг хийх боломжийг олгодгоороо бусад NoSQL өгөгдлийн сангаас давуу талтай.
- Дараагийн төрөлд хүснэгтэн бүтэц (KUDU гэх мэт) өгөгдлийн санг ашиглахаар тооцоолсон. Учир нь энэ төрөл өгөгдлийн шинжилгээг маш хурдан хийх зорилгоор бүтээгдсэн бөгөөд өргөн тархсан өгөгдлийн сан юм. Энэ нь өгөгдлийг багана хэлбэртэй хадгалдаг тул өгөгдлийг шүүж бодолт хийхэд илүү хурдан ажиллана. Мөн хурдацтай үүсэх өгөгдлийн урсгал дээр ажилладаг тул өгөгдлийн бодит цагийн горимоор хийх боломж олгодгоороо давуу талтай.
- Сүүлийн төрөлд нь график өгөгдлийн санг (Neo4J гэх мэт) ашиглахаар тооцоолсон. Учир нь өгөгдлийн олон дамжсан холбоо хамаарлын шинжилгээг хийхэд зохион байгуулалтын оновчтой шийдэл ашиглахгүй бол холбоо хамаарлын шинжилгээ хийхэд хүндрэл үүсэх эрсдэлтэй. Иймд график өгөгдлийн санг ашиглах шаардлагатай юм. Энэ шийдэл нь мөр, багана, хүснэгтэд өгөгдлийг зохион байгуулдаг уламжлалт мэдээллийн сангаас ялгаатай нь өгөгдлийн бүртгэл хоорондын хамаарлаар тодорхойлогддог уян хатан бүтэцтэй байдаг. Өөрөөр хэлбэл бүх өгөгдлийн бичлэг болон тэдгээрийн хоорондын зангилаа хамаарлыг хадгалдаг. Энэ шийдэл энгийн оновчлол дээр суурилагдсан тул бусад мэдээллийн сангаас илүү хурдтайгаар гүнзгий хайлтыг гүйцэтгэдэг.

Боловсруулах

Иймд дээрх хадгалсан өгөгдөл тус бүр дээр зохицож ажилладаг хэд хэдэн өгөгдөл боловсруулах технологийг (SPARK, HIVE, IMPALA, Graph analytics tool гэх мэт) ашиглах нь зүйтэй. Жишээ нь SPARK-г өгөгдлийг нэгтгэж цэвэрлэх үйлдэлд ашиглаж болох юм. Дээр дурдсанаар Их өгөгдлийг нэгтгэх хэрэгсэлд (Big data integration) Spark engine-г ашиглан бичсэн логикийг ажиллуулж болно. Энэ нь их өгөгдлийг боловсруулдаг нэгдсэн өгөгдөл боловсруулах хөдөлгүүр юм. Энэ шийдэл нь ихэнх технологитой зохицож ажилладаг тул хэрэглээнд өргөн ашиглагдаж байна. Харин HBASE дээр хадгалсан өгөгдлийг Apache Hive ашиглан боловсруулалт хийх нь зүйтэй. Энэ нь их өгөгдлийг боловсруулдаг хөдөлгүүр бөгөөд илүү нарийн олон төрлийн өгөгдөл дээр ажиллах зорилгоор бүтээгдсэн технологи юм. KUDU дээр хадгалсан өгөгдлийг IMPALA ашиглан маш хурдан боловсруулалт хийж болно. Impala нь их өгөгдөл дээр ажилладаг SQL хөдөлгүүр (Massively Parallel Processing) бөгөөд өгөгдлийн шинжилгээг бодит цагийн горимоор боловсруулах боломжийг олгодог.

Цэвэрлэж боловсруулсан өгөгдлийг Data science/Advanced analytics технологийг ашиглан таамаглал хийхэд ашиглаж болно. Data science/Advanced analytics нь өгөгдөлд шинжилгээ хийх, тайлан болон хяналтын самбар (dashboard) хөгжүүлэх, нэгтгэх үйлдлүүдийг хялбарчлах шийдэл юм. Их өгөгдөл, машин сургалт, хиймэл оюун ухаан болон өгөгдөл олборлолт зэрэг төрөл бүрийн бүрэлдэхүүн хэсгүүдийг кодчилолгүйгээр

(code free data science) нэгтгэдэг технологи болон график интерфейс ашиглах нь бүх түвшний өгөгдлийн санг нэгтгэн загварчлах, дүн шинжилгээ хийх, дүрслэхэд зориулагдсан урьдчилан боловсруулагдсан (ETL: Extract, Transformation, Loading) зангилаануудыг ашиглан өгөгдлийг ямар нэг кодчиллын алдаагүйгээр урсгах боломжийг олгодог.

Тархаах, дүн шинжилгээ хийх

ANALYZE/DELIVER буюу эцсийн хэрэглэгчид дүн шинжилгээг ойлгомжтой харуулахад өөртөө үйлчлэх мэдээлэл бэлтгэх, бизнесийн мэдээлэл болон дүрслэл шинжилгээ (Self-Service Data prep, BI болон Graph analyses) гэсэн хэрэгслүүд чухал үүрэг гүйцэтгэнэ. Өөртөө үйлчлэх мэдээлэл бэлтгэх, бизнесийн мэдээллийн шийдэл (Self-Service Data prep ба BI) нь орчин үеийн өөртөө үйлчлэх хэрэгсэл юм. Жишээлбэл Tableau Desktop хувилбар нь Oracle, MySQL, PostgreSQL, CSV file, Excel, XML зэрэг 90 гаруй төрлийн эх сурвалжаас өгөгдлийг нэгэн зэрэг авч маш богино хугацаанд боловсруулалт хийдэг орчин үеийн аналитикийн программ бөгөөд багц мэдээллийг задлан шинжилж, үр дүнг энгийн, ойлгомжтой дүрслэн хэрэглэгчид харуулах, компани дотор мэдээлэл солилцох зэрэг олон ажлыг хурдан гүйцэтгэх чадвартай. Дүрслэл шинжилгээ (Graph analyses) нь дэлхийн 50 орчим төрийн байгууллага ашиглаж буй хамгийн сүүлийн үеийн өгөгдөл шинжилгээний систем бөгөөд ялангуяа харилцан хамаарал бүхий өгөгдлийн шинжилгээг хийхэд хурд болон аргачлалын хувьд дэвшилтэт шийдэл болно.

1.2.2 МЭДЭЭЛЛИЙН ИНТЕГРАЦЛАЛ БА МЕНЕЖМЕНТ:

Өгөгдлийг нэгтгэх, боловсруулах, ачаалах (ETL) процесс нь нэгдсэн өгөгдлийн санг үүсгэх, өгөгдлийн чанар, аюулгүй байдлыг хангах чухал үйл ажиллагаа юм. Их өгөгдлийн сангийн системийн хувьд олон төрлийн эх сурвалжаас (төрөлжсөн мэдээллийн сан, IoT төхөөрөмжүүд) өгөгдөл цуглуулах, хадгалах, нэгтгэх шаардлагатай үүсч байна. Тухайлбал төрөлжсөн мэдээллийн сангуудаас өгөгдөл татах, data lake-с өгөгдөл татах, datawarehouse бүртгүүлэх зэрэг үйл ажиллагаануудад ETL, ELT-г ашиглах нь зохимжтой юм .

ӨГӨГДӨЛ ОЛБОРЛОХ, ХУВИРГАХ, АЧААЛАХ (ETL) ХЭРЭГСЭЛ

ETL хэрэгслийн үйл ажиллагаа

ӨГӨГДӨЛ ТАТАХ (EXTRACT)

Зорилго:

Мэдээлэл нийлүүлэгч байгууллага бүрийн төрөлжсөн системээс өгөгдлийг аюулгүй, үр дүнтэй цуглуулах.

Эх сурвалжийн төрөл(Source):

- RDBMS (PostgreSQL, Oracle, MS SQL, MySQL)
- API сервис (REST, GraphQL)
- Файлын формат (CSV, XML, JSON, Excel)
- IoT, Sensor Data (real-time feeds, MQTT, OPC-UA)

Арга:

- Push – Байгууллага өөрсдийн өгөгдлөө их өгөгдлийн санд байршуулах
- Pull – ETL систем тухайн байгууллагын зөвшөөрөгдсөн серверт холбогдон өгөгдөл татах

Нийтлэг ашиглагддаг системүүд:

- Apache NiFi, Talend, n8n, Pentaho, Airbyte

Аюулгүй байдал:

- API key/token, VPN, IP whitelist
- Шифрлэлт (TLS 1.2+, PGP encrypted transfer)

ӨГӨГДӨЛ ХУВИРГАХ (TRANSFORM)

Зорилго:

Татсан өгөгдлийг нэгдсэн системд тохирох формат, хэлбэр, стандарт болгон хувиргах.

Өгөгдөл цэвэрлэх:

- NULL утга, давхардал, format error арилгах зэрэг шаардлагатай өгөгдлийн цэвэрлэгээнүүдийг хийх

Өгөгдлийг стандартчилах:

- Огноо, код, нэр томъёо зэрэг ангилал кодын нийцтэй байдлыг хангах

Anonymization & hashing:

- Хувь хүн, албан байгууллагын эмзэг мэдээллийг арилгах, шаардлагатай мэдээлэлд нууцлалт хийх

Data mapping:

- Төрөлжсөн мэдээллийн сангийн талбаруудыг metadata dictionary-д тохируулан хувиргах

Validation:

- Бусад өгөгдөлтэй тохирч байгаа эсэх, өгөгдөл хоорондын нийцтэй байдлыг шалгах (foreign key, referential integrity)

ӨГӨГДӨЛ АЧААЛАХ (LOAD)

Зорилго:

Хувиргасан, шаардлага хангасан өгөгдлийг өгөгдлийн нэгдсэн санд, мэдээллийн агуулах (DWH) эсвэл мэдээллийн нуур (Data Lake)-т аюулгүй, хяналттай байдлаар ачаалах.

ETL нь ажиллах зарчмын хувьд Batch load буюу тодорхой тохируулсан хугацаагаар ажиллах, Incremental load буюу зөвхөн шинээр нэмэгдсэн эсвэл зөвхөн

өөрчлөгдсөн өгөгдлийг татах , Streaming load буюу real-time байдлаар өгөгдөл (Kafka, Spark Streaming) татах байж болно. Их хэмжээний өгөгдлийн хувьд Incremental load буюу зөвхөн шинээр болон өөрчлөгдсөн өгөгдлийг татах шийдэл нь ихээхэн ач холбогдолтой юм. Учир нь BigData шийдлийн тухайд боловсруулалтай харьцуулахад ETL нь хугацаа болон нөөц ихээр шаардсан үйл ажиллагаа байдаг.

ETL Системүүдийн Харьцуулалт

	ETL Систем	Гол онцлог	Давуу тал	Сул тал	Нээлттэй эсэх
1	Talend Open Studio	GUI дээр суурилсан, powerful data integration suite	ETL pipeline нарийн, багц үүсгэхэд хялбар	Enterprise edition төлбөртэй	Тийм (Community Edition)
2	Apache NiFi	Data flow-ийг визуалаар зурж удирддаг, real-time data stream дэмждэг	Real-time өгөгдөл боловсруулах маш сайн	Data quality validation муу, төвөгтэй тохиргоо шаарддаг	Тийм
3	n8n	Low-code workflow automation, REST API, JSON/YAML хөрвүүлэлт сайн	Ихэнх эх сурвалжтай холбогдох боломжтой, UI ойлгомжтой	Огцом их хэмжээний өгөгдөлд тохиромжгүй	Тийм
4	Airbyte	Open-source ELT, source-to-target mapping хялбар	Cloud & db source-уудтай auto schema sync хялбар	Data transformation хязгаарлагдмал	Тийм
5	Pentaho Data Integration (Kettle)	Enterprise reporting ба ETL integration хосолсон	Report + ETL хосолсон ашиглалтад тохиромжтой	UI хуучирсан, хөгжүүлэлт удаан	Тийм
6	Apache Spark (Structured Streaming)	Big data боловсруулалт, real-time stream болон batch mode дэмждэг	ML, big data, distributed processing-д тохиромжтой	Бага хэмжээтэй өгөгдөлд тохиромжгүй, Spark	Тийм

САЙН ТУРШЛАГА

Дэлхийн орнууд судалгаа, шинжилгээнд суурилсан бодлого, шийдвэр гаргах чадавхийг нэмэгдүүлэх зорилгоор төрийн байгууллагуудын өгөгдлийг нэгтгэсэн их өгөгдлийн санг байгуулж, цахим шилжилт хийх хандлага нэмэгдэж байна. Тухайлбал Эстони, АНУ, БНСУ, Шинэ Зеланд зэрэг орнууд төрийн болон төрийн бус байгууллагуудын салангид тоон мэдээллийг холбож уялдуулах байдлаар их өгөгдлийн сан байгуулах ажлыг амжилттай хэрэгжүүлж, түүнээс гарсан үр дүнг салбар бүрийн судалгаа, шинжилгээ, бодлогын шийдвэрт ашиглаж байна. Эдгээрээс дараах бодит үр дүнг жишээ байдлаар онцоллоо. Үүнд:

- Бага насны хүүхдийн гэр бүлийн байдал (*хүчирхийлэл, ял шийтгэл зэрэг*)-аас шалтгаалсан ирээдүйн амьдралын эрсдлийг тодорхойлох боломжтой үзүүлэлт гаргаснаар хүүхдэд чиглэсэн төрийн бодлого болон төсвийн халамжийн зардлыг оновчтой болгох нөхцлийг бүрдүүлсэн (Шинэ Зеланд болон АНУ).
- Замын хөдөлгөөний урсгалыг тасралтгүй байлгах боломжийг нэмэгдүүлснээр хотуудын дэд бүтцэд оруулж буй хөрөнгө оруулалтыг илүү оновчтой болгосон. 100 гаруй авто замын уулзвар дээрх замын хөдөлгөөний дүрс бичлэгийг нэгтгэж, тоон өгөгдөлд хөрвүүлснээр гэрлэн дохионы удирдлагыг уулзвар бүр дээрх ачааллаас хамааруулан зохицуулдаг нэгдсэн системтэй болсон. Ингэснээр жил бүр тээврийн хэрэгслүүдийн зам дээр өнгөрүүлдэг цагийг 805 мянган цагаар багасгаж 1 сая галлон нефть буюу 3 сая ам.долларыг хэмнэж чадсан байна (АНУ).

- Эстони улсын хувьд эрүүл мэндийн салбарын болон иргэний мэдээллийн сангийн мэдээллийг нэгтгэснээр маш олон давуу талуудыг авчирсан байна.

Үүнд :

- Иргэдийн өвчний түүх, эмийн жор зэрэг мэдээллийг нэгдсэн байдлаар харах боломжтой болсноор шинжилгээний давхардал буурах, эмнэлгийн байгууллага хооронд өвчтөний мэдээлэл богино хугацаанд хуваалцах боломж бүрдэх, улмаар өвчтөний онош оновчтой тодорхойлогдох боломж ихээр нэмэгдсэн;
- Төрөлт, нас баралтын талаарх мэдээлэл тухай бүр шинэчлэгдэх болсноор хүн амын мэдээлэл болон эмнэлгийн үйлчилгээний хэрэгцээ шаардлага, эрэлтийг хурдан хугацаанд тодорхойлоход дөхөм болсон;
- Нэгдсэн системийн хүрээнд цахим эмийн жорын систем бий болсон нь цаг хугацаа, эдийн засгийн хувьд асар өндөр ач холбогдолтой болсон;
- Эрүүл мэндийн мэдээллийн санг бусад мэдээллийн сангуудтай холбосноор нийгмийн халамж тэр дундаа ажилгүйдлийн тэтгэмж, өвчний чөлөө зэргийг давхар шалгах, баталгаажуулах боломжийг олгосон;

Эмнэлгийн байгууллагын үйлчилгээ болгон дугаартай болсноор үйлчилгээний чанар, эмчилгээний эцсийн үр дүнг тодорхойлох, ингэснээр эрүүл мэндийн салбарын хөгжлийн бодлогыг оновчтой тодорхойлоход том түлхэц болсон.

Их өгөгдөл (Big Data) болон нээлттэй өгөгдлийн менежмент нь улс орнуудын төрийн мэдээллийн бүтэц, бизнесийн үйл ажиллагаа болон олон нийтийн оролцоог дэмжих чухал хэрэглэгдэхүүн болж хөгжиж байна. Энэхүү бүтэц доторх Мета-өгөгдлийн менежмент нь өгөгдлийг зохион байгуулах, хадгалах, ашиглах, хуваалцах зэрэг үйл явцыг илүү зохион байгуулалттай, ашиглахад хялбар болгох зорилготой чухал бүрэлдэхүүн хэсэг юм.

АНУ

АНУ-ын их өгөгдлийн сан болон нээлттэй өгөгдлийн менежмент нь голчлон Open Data Policy, Federal Data Strategy болон National Archives and Records Administration (NARA)-ийн гаргасан удирдамжуудын хүрээнд зохицуулагддаг.

- Мета-Өгөгдлийн Стандарт:
АНУ нь их өгөгдөл болон нээлттэй өгөгдлийг засгийн газрын нээлттэй байдлыг хангах, нээлттэй өгөгдөл хуваалцах зорилгоор machine-readable data formats (жишээ нь, JSON, CSV) ашиглахыг дэмждэг. Мета-өгөгдлийг дотоод засгийн газрын хэмжээнд холбогдох кодчилал, хаяг, болон тайлбаруудыг заавал оруулж байхаар стандартчилж өгсөн байдаг.
- Захиргааны журам:
Зорилтот өгөгдөл нь олон нийтийн ашиглахад нээлттэй байхаас гадна зохицуулагч байгууллагууд, байгууллагын бүтэц, тайлангийн форматуудыг тодорхойлж өгсөн.

ЕВРОПЫН ХОЛБОО

Европын Холбоо нь өгөгдлийн менежментийн болон нээлттэй өгөгдөл хуваалцах тухай олон улсын харилцааны хувьд дэлхийн хэмжээнд хамгийн боловсронгуй тогтолцоог бий болгосон байна.

- Мета-Өгөгдлийн Стандарт:
Европын Холбоо нь INSPIRE Directive болон European Data Portal-ийн хүрээнд үндэсний болон олон улсын хэмжээнд өгөгдлийг стандартчилсан загвараар

боловсруулах, харуулах, хуваалцахыг шаарддаг. Мета-өгөгдөл нь ISO 19115 болон Dublin Core Metadata Initiative (DCMI) стандартуудыг мөрдлөг болгодог байна.

- Захиргааны журам:
Нээлттэй өгөгдөл, мета-өгөгдлийн стандартад чиглэсэн нарийн журамтай бөгөөд нээлттэй өгөгдлийг зөвхөн тодорхой хэмжээнд хяналттай байлгана. Энэхүү хяналт нь улс орны нийгэм, эдийн засагт нөлөөлөхөд шууд нөлөөлдөг гэж үздэг байна.

ЯПОН

Япон улс нь томоохон технологийн компаниуд ихтэй, их өгөгдөл болон нээлттэй өгөгдлийн чиглэлд өндөр хөгжсөн улс юм.

- Мета-Өгөгдлийн Стандарт:
Япон нь үндэсний хэмжээнд томоохон өгөгдөл болон технологийн байгууллагуудтай хамтран их өгөгдөл, мета-өгөгдлийн стандартуудыг боловсруулж тогтмол шинэчлэл хийж явдаг. Гол стандарт Japanese Industrial Standards (JIS) болон Metadata Registry -г үйл ажиллагаандаа ашигладаг.
- Захиргааны журам:
Япон улсад нээлттэй өгөгдөл нь GovData платформ дээр төвлөрч байдаг бөгөөд улс орон нь өгөгдлийн аюулгүй байдал болон нууцлалын өндөр шаардлагыг хэрэгжүүлдэг.

ШИНЭ ЗЕЛАНД

Эдгээр улсаас хүн ам болон өгөгдлийн сангийн бүтцийн хувьд Монгол улстай ижил төстэй ч төрийн их өгөгдлийн сангийн үйл ажиллагааг маш амжилттай хэрэгжүүлсэн Шинэ Зеланд улсын авч хэрэгжүүлсэн арга хэмжээ, туршлагыг дэлгэрэнгүй авч үзье.

Төрийн их өгөгдлийн сан бүрдүүлэх, ашиглах тал дээр Шинэ Зеланд улс тэргүүлэгч улс орнуудын эгнээнд явж байна. Тус улс нь 2011 оноос анхлан Нэгдсэн өгөгдлийн сан (*Integrated data infrastructure*)-г байгуулж, иргэдийн болон бизнесийн өгөгдлийн сангуудыг нэгтгэн, холбож шийдвэр гаргалтанд төдийгүй судлаачдад судалгаа шинжилгээний зориулалтаар ашиглаж эхэлсэн нь их өгөгдлийн сангийн эхлэл болжээ. Улмаар 2013 оноос эхлэн хүн ам, өрхийн мэдээллийн сан, бизнес регистрийн сангуудыг холбох үйл ажиллагааг зохион байгуулж одоогоор нийт 8 салбарын 48 төрлийн өгөгдлийн сангийн 166 тэрбум ширхэг тоон мэдээллийг нэгтгэж чадсан. Энэхүү мэдээллийн сангийн түслэмжтай төрийн байгууллага, төрийн бус байгууллага болон бусад хэрэглэгчдийг шаардлагатай тоон мэдээллээр хангах ажлыг тогтмол гүйцэтгэдэг төдийгүй улирал бүр 1-2 тоон мэдээллийн санг шинээр нэгтгэж, хамрах хүрээг тогтмол нэмэгдүүлдэг байна.

ӨГӨГДЛИЙН НЭГДСЭН САНГИЙН БҮТЭЦ, МЭДЭЭЛЭЛ ДАМЖУУЛАХ ҮЙЛ ЯВЦ

Шинэ Зеланд улс нь (i) Иргэн, өрхийн мэдээллийн сан (ii) Бизнес регистрийн сан гэсэн 2 гол санг тусад нь хөгжүүлж, салбар бүрийн мэдээллийг холбосны дараагаар тус хоёр мэдээллийн санг нэгтгэх байдлаар өгөгдлийн нэгдсэн санг бүрдүүлсэн.

Тус улс өгөгдлийн нэгдсэн дэд бүтэц(иргэн, өрхийн мэдээллийн сан), бизнес регистрийн зэрэгцээ өгөгдлийн сан гэсэн үндсэн хоёр мэдээллийн санг холбох байдлаар төрийн их өгөгдлийн санг бий болгосон. Өгөгдлийн нэгдсэн дэд бүтэц буюу иргэн, өрхийн мэдээллийн сан байгуулахдаа төрийн бусад өгөгдлийн сангуудыг өгөгдлийн сангийн гол нуруунд холбох зарчмыг баримталсан. Өөрөөр хэлбэл их өгөгдлийн сан нь иргэн болон ААНБ гэсэн хоёр гол багана дээр зангидагдсан гэсэн үг юм.

Өгөгдлийн нэгдсэн дэд бүтэц нь иргэн, өрхийн түвшний бүх мэдээллийг хамардаг бөгөөд Шинэ Зеланд улсын хувьд судлаачдыг хүн амын ялгаатай бүлгүүдийг хооронд нь харьцуулах боломжийг олгох зорилгоор тус улсад амьдарч байсан бүх иргэдийг хамруулсан өгөгдлийн санг бий болгож, уг мэдээллийн сангаа өгөгдлийн сангийн гол нуруу болгон ашигласан байна. Үүнд улсын уугуул иргэд буюу тус улсад төрсөн иргэд, гадаад улсын иргэн боловч тус улсад албан ёсний виз авч амьдарч байсан иргэд зэрэг багтана. Улмаар энэхүү ялгаатай хүн ам зүйн бүлгийн иргэдийг татварын мэдээллээр дамжуулан нэгтгэжээ.

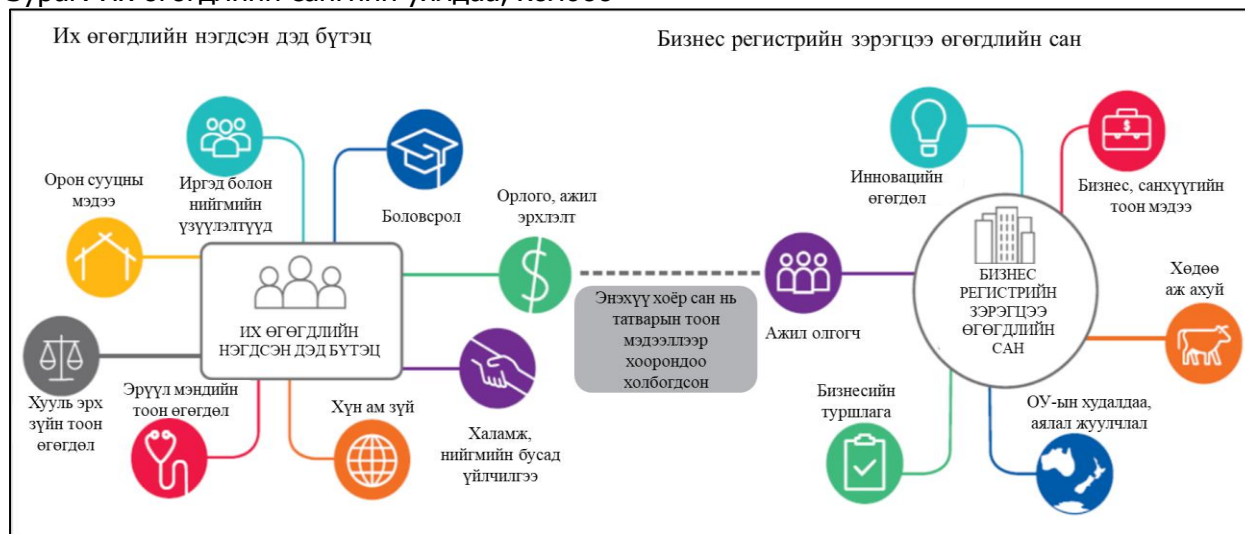
Дараагийн гол багана болон бизнес регистрийн мэдээллийн санг байгуулахдаа гол нуруунд 6 дэд сан (*ААН-ийн үйл ажиллагаа, орлого, санхүүгийн тоон мэдээ, татвар, ажиллах хүчин, гадаад худалдаа, төрийн дэмжлэг*) болон 6 түүвэр судалгааны тоон мэдээллийг (*Бизнесийн жилийн судалгаа, үйл ажиллагаа, инноваци, судалгаа хөгжүүлэлт, менежмент, санхүүгийн байдал*) холбогдсон байна.

Зураг. Шинэ Зеланд улсын Бизнес регистрийн сангийн бүтэц



Мэдээллийн сангуудыг нэгтгэх байдлаар их өгөгдлийн дэд бүтцийг байгуулахдаа өгөгдлийн нэгдсэн дэд бүтэц буюу иргэний мэдээллийн сан болон бизнес регистрийн санг тусад нь хөгжүүлж, салбар бүрийн мэдээллийг холбосны дараа тус хоёр мэдээллийн санг нэгтгэх ажлыг зохион байгуулсан. Ингэхдээ иргэний хувьд регистрийн дугаар, байгууллагын хувьд байгууллагын регистрийн дугаарыг гол эх сурвалжаа болгож регистрийн дугаар тус бүрд харгалзах татварын өгөгдлийн сангийн тусламжтайгаар нэгтгэх ажлыг гүйцэтгэсэн. Ингэснээр тухайн иргэний хувьд боловсрол, эрүүл мэнд, хүн ам зүйн үндсэн мэдээллүүдээс гадна ажил эрхлэлт, хөрөнгө, мал аж ахуй, татвар төлөлт зэрэг санхүүгийн чиглэлийн мэдээллүүдийг давхар харах боломжтой болсон бол аж ахуйн нэгжийн хувьд мөн ажиллагчдын мэдээллийг харах боломжийг бүрдүүлж өгсөн.

Зураг. Их өгөгдлийн сангийн уялдаа, холбоо



Нэгдсэн өгөгдлийн сангийн мэдээллийг дамжуулах үйл явц:

Шинэ Зеланд улсын их өгөгдлийн сангийн мэдээлэл дараах 5 үе шатыг дамждаг.

Үүнд:

1. Төрийн байгууллагуудаас ирсэн тоон мэдээллийг аюулгүй серверийн орчинд цуглуулах, шаардлагатай тохиолдолд тоон мэдээллийг засварлах;
2. Цуглуулсан өгөгдлийн санг холбох, уялдуулах;
3. Нэгтгэсэн сангаас хувь хүний холбоотой эмзэг мэдээллийг арилгах, устгах;
4. Судалгаанд ашиглах зориулалт бүхий тоон мэдээллийг бэлэн болгох;
5. Баталгаажсан судалгааны бүтээлүүдийг олон нийтэд түгээх

Ийнхүү их өгөгдлийн мэдээллийн сангийн үйл ажиллагаа явуулахдаа 5 АЮУЛГҮЙ БАЙДАЛ /5 SAFE/-ын зарчим буюу аюулгүй мэдээлэл, төсөл, хүн, орчин, бүтээгдэхүүн гэсэн 5 зарчмыг баримталдаг байна. Энэхүү зарчмын хүрээнд аюулгүй серверийн орчинд зөвхөн баталгаажсан зөвшөөрөл бүхий судлаачид судалгааны зориулалтаар хувь хүний мэдээлэл агуулаагүй тоон мэдээг ашигласнаараа олон нийтийн дунд хүлээн зөвшөөрөгдсөн, итгэл даахуйц, тогтвортой байх нөхцөлийг хангаж чадаж байна.

Зураг. Шинэ Зеланд улсын их өгөгдлийн сангийн мэдээлэл дамжуулах үйл явц

1	 Түүхий тоон мэдээлэл цуглуулах үе	- Тоон мэдээллийг байгууллагуудуус аюулгүй, шифрлэгдсэн диск ашиглах байдлаар цуглуулна; - Мэдээллийг холбоход бэлтгэл болгож тоон мэдээллээ бага хэмжээгээр цэвэрлэж засварлана.
2	 Өгөгдлийн санг холбох, уялдуулах үе	- Тоон мэдээллийг аюулгүй серверийн орчинд холбох, уялдуулах үйл ажиллагааг гүйцэтгэнэ; - Тоон мэдээллийг холбох эрх бүхий ажилтаны өгөгдлийн санд нэвтрэх эрх хязгаарлагдмал бөгөөд зөвхөн хувь хүний нууцлал шифрлэгдсэн мэдээлэлтэй харьцана; - Өгөгдлийн санг холбохдоо шууд арга буюу тухайн хүний регистрийн дугаараар мэдээллийг холбох болон магадлалын арга буюу хүн ам зүйн хувьсагчдаар холбох аргыг ашиглана.
3	 Хувийн мэдээллийг арилгах үе	Судлаачдад нэгтгэсэн тоон мэдээллийг түгээхээс өмнө хувь хүнтэй холбоотой <i>loveog нэр, төрсөн он сар өдөр, оршин суугаа хаяг гэх мэт</i> мэдээллийг арилгах үйл явц хийгдэнэ АЮУЛГҮЙ МЭДЭЭЛЭЛ
4	 Судалгаанд ашиглах зориулалт бүхий тоон мэдээлэл бэлэн болгох үе	- Тухайн төсөл нь олон нийтийн эрх ашгийн төлөө судалгааны зориулалттай байх ёстой АЮУЛГҮЙ ТӨСӨЛ - Тоон мэдээлэл нь зөвхөн албан ёсны судлаачдад нээлттэй байх бөгөөд зөвхөн судалгаанд шаардлагатай тоон мэдээлэлдээ нэвтрэх боломжтой АЮУЛГҮЙ ХҮН - Тоон мэдээлэл нь зөвхөн аюулгүй тусгай серверийн орчинд нэвтрэх болно АЮУЛГҮЙ ОРЧИН
5	 Баталгаажсан судалгааны бүтээлийг олон нийтэд түгээх үе	Судалгаа нь хувь хүний нууцлалыг чандлан хадгалсан эсэхэд хяналт, шалгалт хийсний дараагаар зөвхөн баталгаажуулсан үр дүнг олон нийтэд мэдээлдэг АЮУЛГҮЙ БҮТЭЭГДЭХҮҮН

Дээр дурьдсан улс орнууд болон бусад их өгөгдлийн санг хөгжүүлсэн улс орнуудын туршлагаас харахад мэдээлэл цуглуулах, ашиглах, хуваалцах, хамгаалах, ил тод байлгах зэрэг үндсэн чиглэлийн хүрээнд их өгөгдлийн сан (Big Data) болон нэгдсэн дата төвийн хууль эрхзүйн орчныг бүрдүүлэх бодлогын баримт бичиг, журам, стандартууд нь зайлшгүй шаардлагатай нь харагдаж байна.

	БАРИМТ БИЧГИЙН ЖАГСААЛТ	ТАЙЛБАР
1	Big Data-ийн үндэсний бодлого	Их өгөгдлийн сан бүрдүүлэх, ашиглах, хөгжүүлэх стратеги
2	Мэдээлэл цуглуулах, боловсруулалтын стандарт	Ямар мэдээллийг хэрхэн, хэнээс, ямар давтамжтай цуглуулах, стандарт форматын тогтолцоо
3	Metadata стандарт ба заавар	Өгөгдлийн талбар, ангилал, кодчилал, тайлбар, хувиргалт зэрэг metadata-ийн нэгдсэн заавар
4	Өгөгдлийн нууцлал, хамгаалалтын бодлого	Өгөгдлийн хамгаалалт, зөвшөөрөлтэй хандалт, хаш код, шифрлэлт, хандалтын бүртгэл

5	Өгөгдлийн эрх зүйн нөхцөл ба өмчлөл	Өгөгдлийг хэнтэй хуваалцах, ямар нөхцөлөөр, ямар хязгаарлалттай байх зэрэг дүрэм журмууд
6	Өгөгдөл солилцооны протокол	Их өгөгдлийн сантай болон байгууллагууд хооронд өгөгдөл дамжуулах формат, API стандарт, баталгаажуулалтын протокол (OAuth2.0, JWT)
7	Ил тод байдал, нээлттэй өгөгдлийн бодлого	Нийтэд нээлттэй өгөгдлийг хэрхэн ялгах, API-г хэрхэн нээлттэй болгох, лицензийн нөхцөл
8	Архив, нөөцлөлтийн журам	Өгөгдлийг хэр хугацаанд, ямар нөхцөлтэйгөөр хадгалах, устгах горим
9	Disaster Recovery (DR) төлөвлөгөө, заавар	Хэрхэн нөхөн сэргээх, ямар шалгууртай failover хийх, шаардлагатай хугацаа (RTO, RPO)
10	Хяналт, хариуцлагын тогтолцоо	Хандалтын бүртгэл, аудит лог, тайлагналын тогтолцоо, зөрчлийн үед хүлээлгэх хариуцлага

1.2.3 МЭДЭЭЛЭЛ ЗАГВАРЧЛАХ ХЭРЭГСЭЛ /DATA MODELING TOOLS/

Өгөгдлийн сангийн схем ба өгөгдөл хоорондын харилцан хамаарлыг тодорхойлох, мэдээллийн урсгалыг зохион байгуулах нь өгөгдлийн интеграцчлалын чухал хэсэг юм. Data Lake болон их өгөгдлийн сангийн хувьд зайлшгүй шийдвэрлэх ёстой асуудал бол өгөгдөл хоорондын уялдаа, харилцан хамаарлыг (relationship, semantic links) оновчтой зохион байгуулах явдал юм. Олон эх сурвалжаас цугралсан өгөгдөл, мөн дээр нь олон төрлийн форматтай (CSV, JSON, зураг, видео, DB dumps г.м.) түүхий өгөгдөл орж ирэхэд тэдгээрийг хооронд нь ойлгомжтой, холбогдож ажиллахуйц байдлыг хангахгүй бол тэдгээр нь ашиглахад хүндрэлтэй, ашиглаж болгохгүй өгөгдлийн цуглуулга болдог. Data Swamp буюу “булингартсан нуур” -ч гэж хэлдэг.

Өгөгдлийн хоорондын уялдаа хангаж мэдээллийг загварчлаагүйгээс

- Өгөгдлийн хамаарал алдагдах (Relationship)
 - Олон системээс ирсэн дата хооронд холбоо байхгүй бол тайлан, шинжилгээ хийх боломжгүй болно
- Data lineage (Өгөгдлийн гарал үүсэл) тодорхойгүй болох
 - Өгөгдөл хэзээ ямар системээс ямар форматтай орж ирсэн, ямар болж өөрчлөгдсөн, ямар ямар хувиргалтууд хийсэн мэдээллүүд нарийн тодорхой байх ёстой. Энэ нь эргээд үр дүн буюу хүснэгт, тайлангуудын үнэн зөв байдлыг тайлбарлах гол хүчин зүйл юм.
- Business glossary (Нэр томъёо, утгын ялгаатай байдал)
 - Өгөгдлийг олон эх сурвалжаас цуглуулж байгаа нь тэр бүр нийцтэй байдлыг үүсгэж чаддаггүй. Өөрөөр хэлбэл мэдээлэл хариуцагч байгууллага бүр өөрийн үйл ажиллагааны хэрэгцээ шаардлагаас хамааран өгөгдлийн өөр өөр нэршил, өөр өөр агуулгаар бүртгэсэн байдаг.
- Өгөгдлийн давхцал
 - Ижил үзүүлэлт олон мэдээлэл хариуцагч дээр өөр өөр байдлаар, өөр өөр форматаар хадгалагдаж байдаг

Өгөгдлийг загварчлахдаа Schema-on-Write буюу өгөгдлийг ах татахдаа загварчилж хадгалах, эсвэл Schema-on-Read буюу түүхий өгөгдлийг шууд хадгадаад дараа нь боловсруулалт хийхдээ загварчлах аргууд байдаг. Дээр 2 шийдлийн хослуулж бас болдог байна.

Өгөгдлийг загварчлах үйл ажиллагааны хүрээнд хэд хэдэн арга хэмжээг авч хэрэгжүүлэх шаардлагатай:

METADATA CATALOG & DATA GOVERNANCE

- Entity нэр, тайлбар, холбоог тодорхойлох
Өгөгдлийн томъёолсон нэршил, тайлбар болон үзүүлэлт хоорондын уялдааг тодорхойлох
- Field-level lineage(гарал үүсэл)
Өгөгдөл хэзээ ямар системээс ямар форматтай орж ирсэн, ямар болж өөрчлөгдсөн, ямар ямар хувиргалтууд хийсэн мэдээллүүд нарийн тодорхойлох
- Data sensitivity level (буюу эмзэг мэдээлэл) тодорхойлох
Хувь хүн, байгууллагын эмзэг мэдээллийг тодорхойлсон байдал болон тухайн өгөгдлийн хамгаалалтын зэрэглэлийг тодорхойлсон байдал. Жишээ нь нийтэд нээлттэй, дотоод үйл ажиллагаанд ашиглах, хязгаарлагдмал эрх бүхий хэрэглэгчид хандаж болох, өндөр хамгаалалттай мэдээлэл гэх зэрэг.

Эмзэг мэдээлэлтэй холбоотой үйл ажиллагаанууд, тухайлбал эмзэг мэдээллийг арилгах, шифрлэх, аудит лог тавих гэх зэрэг.

- Нэр томьёоны уялдаа (үзүүлэлтийн mapping)-г тодорхойлох
Мэдээлэл хариуцагчдаас нэг ижил ойлголт, утга бүхий өгөгдөл өөр өөр нэршил, форматаар ирэх тохиолдол байдаг. Иймд энэ үйл ажиллагааг стандартчилж шийдвэрлэсэн байх шаардлагатай.
- Өгөгдлийн байршил
Өгөгдлийн эх сурвалжийг тодорхойлох, өгөгдлийг баталгаажуулах, дахин татах зэрэг үйл ажиллагааны шаардлагаар өгөгдлийн байршлыг нарийн тодорхойлсон байх шаардлагатай

Өргөн хэрэглэгддэг системүүд:

НЭР	ОНЦЛОГ
Apache Atlas	Hadoop экосистемтэй сайн нийцэж ажилладаг, lineage tracking, tag-based policy хэрэгжүүлэх боломжтой
AWS Glue Data Catalog	Cloud Data Lake-д metadata удирдах боломжтой
Microsoft Purview	Data mapping, data lineage, business glossary
Collibra	Enterprise data governance-д түгээмэл ашиглагддаг
Alation	Data catalog + collaboration хийх боломжтой

ENTITY RELATIONSHIP MODELING (ERD)

Өгөгдлийн хоорондын холбоог загварчлах хамгийн үндсэн арга бол Entity-Relationship Diagram (ERD):

- Entity (Object) → Хүн, компани, бараа, үйл явдал гэх мэт
- Attribute (Property) → Entity-н талбарууд
- Relationship → Entity-н хоорондын холбоо (1:1, 1:N, N:M)

Өргөн хэрэглэгддэг системүүд:

НЭР	ОНЦЛОГ
dbdiagram.io	Хялбар, шууд SQL export хийх боломжтой
Lucidchart	Online, collaboration
ER/Studio	Complex enterprise model зурагдах боломжтой
Diagrams.net	Free, олон форматтай export хийнэ

Эдгээр системүүдээс Apache Atlas-г авч үзэхэд metadata catalog, data governance, data lineage-ийн зориулалттай үйл ажиллагаануудыг хөтлөн явуулах боломжтой систем юм. Ялангуяа Data Lake буюу том хэмжээний өгөгдлийн нууртай орнуудад (Hadoop ecosystem, cloud storage) metadata зохион байгуулахад түгээмэл ашиглагддаг.

APACHE ATLAS

Боломжууд

- Metadata Management
- Classification & Tagging
- Data Lineage (Өгөгдлийн гарал үүсэл)
- Security & Integration
- Search & Discovery

Apache Atlas нь дээрх боломжуудыг агуулсан open-source систем бөгөөд зах зээл дээр түүнтэй ижил төстэй олон системүүд байдаг. Эдгээрээс metadata удирдах, өгөгдлийн гарал үүсэл (lineage) хянах, өгөгдлийн ангилал хийх, data governance(өгөгдлийн засаглал) хэрэгжүүлэх боломжтой зарим нийтлэг системүүдийг харьцуулж үзье.

СИСТЕМ	ОНЦЛОГ	ДАВУУ ТАЛ	СУЛ ТАЛ	ЖИШИГ ҮНЭ
Collibra	Enterprise Data Catalog, Data Governance, Lineage	Хүчирхэг тайлбарын сан, автомат lineage	Үнэтэй	\$100K+/жил
Alation	Search-based data catalog, Active data governance	AI-д суурилсан хайлт, хэрэглэгчийн уян хатан шийдэл	Лиценз үнэтэй, нэвтрүүлэлт нарийн төвөгтэй	\$50K – \$300K+/жил
Informatica EDC	Enterprise Data Catalog, Data Privacy	Автомат хайлт, ML(машин сургалт)-д нийцтэй	Хэт олон модультай, нарийн төвөгтэй тохиргоо шаарддаг	\$150K+/жил
Microsoft Purview	Azure экосистемд илүү нийцтэй	Үүлэн шийдэлтэй, AI-д суурилсан хайлт	Azure-д төвлөрсөн	~\$1,000 + /cap
IBM Watson Knowledge Catalog	Metadata, lineage, data quality, privacy	Машин сургалт, өгөгдлийн гарал үүсэл, өгөгдлийн хамгаалалт	IBM системд төвлөрсөн	\$30K – \$250K+/жил
Data.World	Graph-based, collaborative catalog	Үүлэн шийдэл(SaaS), график стандарт	Enterprise түвшний шийдэл хязгаарлагдмал	\$20K – \$100K+/жил

Эдгээрээс дүгнэж харьцуулалт хийвэл:

ҮЗҮҮЛЭЛТ	APACHE ATLAS	БУСАД СИСТЕМҮҮД
Үнэ	Free	Өндөр үнэтэй
UI	Энгийн	Хэрэглэхэд хялбар, нийцтэй
ML/AI нэмэлт	Боломж хязгаарлагдмал	Өргөн боломжуудтай
Business Glossary	Хязгаарлагдмал	Дэлгэрэнгүй, үйл ажиллагаатай уялддаг
Support	Хязгаарлагдмал	Cloud + On-premises үйлчилгээ сайтай
Lineage Visualization	Байгаа	Илүү өргөн боломжуудыг агуулсан
Security integration	Байгаа	AD, LDAP, SSO integration хийх боломжуудтай

Санал болгоход их өгөгдлийн сангийн нэвтрүүлэлтийн эхний тодорхой үе шатад APACHE ATLAS ашиглах, цаашид систем өргөжихтэй уялдуулан enterprise шийдэл сонгох нь илүү оновчтой гэж дүгнэж байна.

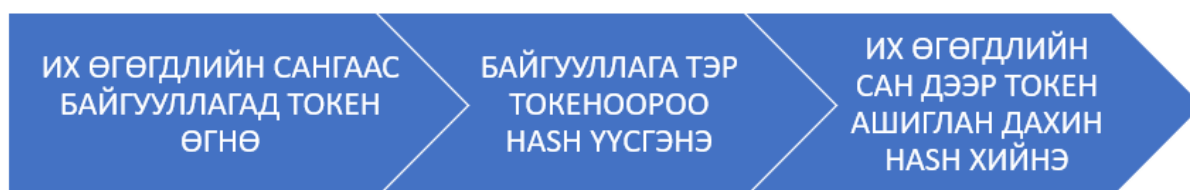
ӨГӨГДӨЛ СОЛИЛЦОХ ДҮРЭМ, НУУЦЛАЛ ХАМГААЛАЛТ

Засгийн Газрын “Их өгөгдлийн сан” үүсгэх тухай 2023 оны 403 дугаар тогтоол батлагдсанаар их өгөгдлийн сан байгуулах үйл явцыг дараагийн шатанд гаргасан. Уг тогтоолд заасанчлсан хувь хүн, байгууллагын эмзэг мэдээллийг их өгөгдлийн санд нийлүүлэхгүй байх, мөн хувь хүн, байгууллагыг тодорхойлох боломжгүй болгож их өгөгдлийн санд мэдээлийн санд өгөгдөл нийлүүлэх тухай дурьдсан байгаа.

Их өгөгдлийн санд нэг иргэн, байгууллагын мэдээлэл олон мэдээлэл хариуцагчаас ирснээр тухайн иргэн, байгууллагын мэдээлэл их өгөгдлийн санд зангидагдах шаардлагатай болно.

Үүний тулд дараах шийдлийг санал болгож байна:

ТӨВ TOKEN ДЭЭР СУУРИЛСАН HASH ШИЙДЭЛ



1. Төв сервер дээрээс байгууллага бүрт “OrgToken” үүсгэж өгнө.

- А байгууллагын OrgToken → ORG_TOK_A
- Б байгууллагын OrgToken → ORG_TOK_B

2. Байгууллага иргэн болон байгууллагын түлхүүр үзүүлэлтийн утгыг (регистер) OrgToken-оор hash хийнэ.

Жишээ (А байгууллага):

- Hash_A = SHA-256(Регистр + ORG_TOK_A)
- А байгууллага хаш утгаа төв серверт илгээнэ.

3. Б байгууллага ч мөн адил хийнэ.

- Hash_B = SHA-256(Регистр + ORG_TOK_B)
- Б байгууллага хаш утгаа төв серверт илгээнэ.

4. Их өгөгдлийн сангийн Төв сервер дээр А болон Б байгууллагын hash-ийг дахин төвийн “CentralToken” той хамт hash хийнэ.

- FinalHash_A = SHA-256(Hash_A + CENTRAL_TOKEN)
- FinalHash_B = SHA-256(Hash_B + CENTRAL_TOKEN)

5. Их өгөгдлийн санд ийнхүү шинээр hash хийж орсон ижил түлхүүр утгууд (регистер) нь ижил байна.

- FinalHash_A = FinalHash_B.
- Hash_A ≠ Hash_B

Энэхүү шийдлийн хувьд CentralToken-г найдвартай хамгаалах ёстой. Эс тэгвээс hash хийгдсэн өгөгдлүүдийг таних боломж үүснэ.

Иймд:

- CentralToken-ийг өндөр хамгаалалт дор хадгалах:
 - HSM (Hardware Security Module)
 - Vault server (Hashicorp Vault гэх мэт)
- OrgToken-г мэдээлэл хариуцагч байгууллагууд мөн өндөр түвшинд хамгаалах
- CentralToken-г тодорхой давтамжтайгаар сольж байх
- Double hashing ашиглагдаж байгаа тул сервер орчны нөөц, хүчин чадлыг шалгаж бэлэн байдлыг тогтмол

МЭДЭЭЛЭЛ ДҮРСЛЭХ, ТАЙЛАГНАХ ХЭРЭГСЭЛ /DATA VISUALIZATION AND REPORTING TOOLS/

Өгөгдлийг дүрслэх буюу тайлан, дашбоард, график байдлаар боловсруулалт хийх болон их өгөгдлийн санд дүн шинжилгээ, боловсруулалт хийх гэсэн үндсэн үйл ажиллагааг энд авч үзнэ. Өгөгдлийн боловсруулалт, дүн шинжилгээ хийж хэрэглэгчид ойлгомжтой, сонирхолтой байдлаар хүргэхэд мэдээллийн дүрслэл чухал үүрэг гүйцэтгэдэг. Маш олон төрлийн системүүд байдгаас мэдээллийг бодит цагийн горимоор боловсруулж, хялбархан тайлагнахад зөв системийг сонгож хэрэглэх нь чухал юм.

Мэдээллийг дүрслэх нийтлэг системүүд

Tableau:

Tableau нь өгөгдлийн дүрслэлийг сайн хийдэг хэрэгслүүдийн нэг бөгөөд томоохон өгөгдөлд боловсруулалт хийх, дүн шинжилгээг бодит цагийн горимоор хийх боломжийг олгодог. Хэрэглэгч өөрийн хэрэгцээ шаардлагад нийцүүлэн тусгай орчинд системийг суурилуулах боломжтой.

Давуу тал: Хэрэглэгчдэд боловсруулсан өгөгдлөө хялбархан дүн шинжилгээ хийх, тайлагнах, олон төрлийн график, диаграмм байгуулах боломжийг олгодог.

Power BI (Microsoft):

Power BI нь Microsoft-ийн бүтээгдэхүүн бөгөөд нээлттэй өгөгдөл, томоохон өгөгдлийн сангуудтай холбогдож, мэдээллийн шинжилгээ хийх, хялбархан тайлан гаргах боломжийг олгодог.

Давуу тал: MS Excel болон Microsoft-ийн бүтээгдэхүүнтэй нийцтэй ажилладаг, ашиглахад амархан, сургалттай интерфейстэй

QlikView:

QlikView нь том хэмжээний өгөгдөлтэй ажиллах, бодит цагийн горимоор дүн шинжилгээ хийх боломжийг олгодог.

Давуу тал: Хэрэглэгчдэд дүрслэлийг илүү нарийн тохируулах боломжийг олгодог, олон төрлийн уян хатан дашбоардуудыг хялбархан хөгжүүлэх боломжтой.

D3.js:

D3.js нь өгөгдлийн дүрслэлийг бий болгох боломжтой JavaScript сан юм. Өгөгдлийн интерактив дүрслэлийг уян хатан зохион байгуулах боломжтой.

Давуу тал: Хөгжүүлэгчид болон программчлалын түвшинд хөгжүүлэлт хийх боломжтой бөгөөд энэ нь өөрийн системүүдийг хөгжүүлж буй тохиолдолд давуу талыг үүсгэдэг

Google Data Studio:

Google-ийн BI төрлийн үйлчилгээ бөгөөд Google-ийн өгөгдөлтэй холбогдож, мэдээллийн самбар, тайлан хийх боломжийг олгодог.

Давуу тал: Google үйлчилгээнд хурдан холбогдох боломжтой, ашиглахад хялбар.

Их өгөгдөлд дүн шинжилгээ хийх нийтлэг хэрэгсэл

Apache Spark:

Spark нь том хэмжээний өгөгдөлд дүн шинжилгээ хийх өндөр хурдны боловсруулалтын систем юм.

Давуу тал: Их хэмжээний өгөгдлийг боловсруулж, томоохон хэмжээний datawarehouse-тай холбогдон ажиллах, дүн шинжилгээ хийхэд тохиромжтой.

Hadoop:

Hadoop нь их өгөгдлийн орчинд өгөгдлийн боловсруулалт хийх боломжтой open source платформ бөгөөд өгөгдлийг өндөр хурдтай боловсруулахад тохиромжтой.

Давуу тал: Хэмжээний хэмжээний өгөгдөлтэй ажиллахад өндөр бүтээмжтэй байдаг.

Google BigQuery:

Google Cloud-ийн өгөгдлийн шинжилгээний үйлчилгээ систем бөгөөд их өгөгдлийг боловсруулах, дүн шинжилгээ хийхэд ашиглагддаг.

Давуу тал: Их өгөгдлийн дүн шинжилгээг өндөр хурдтай хийх шийдэл.

Манай орны хувьд эдгээр системүүдээс Tableau, QlikView, Apache Spark болон Hadoop системүүд харьцангуй өргөн хэрэглэгдэж байна. Аливаа систем нэвтрүүлэлт, ашиглалтын явцад боловсон хүчний хүртээмжтэй байдал болон системүүд дээрх засвар үйлчилгээний чанар чухал нөлөөтэй байдаг.

Их өгөгдлийн сангийн хөгжүүлэлтийн хүрээнд Tableau, Hadoop шийдлийг оновчтой гэж үзэж байна.

ДҮГНЭЛТ

Энэхүү бүлэгт их өгөгдлийн санг байгуулахад шаардлагатай программ хангамжийн экосистем, өгөгдлийн сангийн төрөл, ETL хэрэгслүүд, өгөгдөл загварчлал, аюулгүй байдлын технологи, дүн шинжилгээний хэрэгслүүдийн онцлог, шаардлагыг нэгтгэн харууллаа.

Их өгөгдлийн сан байгуулах үйл ажиллагаа нь өгөгдлийг зөвхөн цуглуулах бус:

- Data Lake ба Data Warehouse хослуулах
- ETL pipeline-г оновчтой байгуулах
- Hash төв шийдлээр нууцлал хамгаалах
- Metadata catalog үүсгэж Data Swamp-ээс сэргийлэх
- Visualization tooling-ийг бодит хэрэглээнд тохируулах
- Олон улсын стандарт, сайн туршлагыг нутагшуулах шаардлагатай цогц үйл ажиллагаа юм.

ӨГӨГДЛИЙН САНГИЙН ТЕХНОЛОГИ БА ТӨРӨЛ

Их өгөгдлийн сан нь төрийн олон төрлийн эх сурвалжуудаас өгөгдөл цуглуулж, боловсруулж, шинжилгээ хийх загварчлагдсан, өргөтгөх боломжтой технологийн шийдлийг агуулсан байх шаардлагатай.

Өгөгдлийн төрлүүд

ТӨРӨЛ	ОНЦЛОГ	ТЕХНОЛОГИЙН ЖИШЭЭ
Бүтэцлэгдсэн	SQL өгөгдлийн сан, ACID transaction, BI тайлан	PostgreSQL, Oracle, MS SQL Server
Хагас бүтэцтэй	JSON, XML зэрэг flexible schema	MongoDB, Couchbase
Бүтэцгүй	Зураг, видео, аудио зэрэг том хэмжээний өгөгдөл	Hadoop HDFS, Amazon S3, MinIO

Data Warehouse vs Data Lake

- Data Warehouse → OLAP, тайлан, BI
- Data Lake → AI/ML, exploratory analysis
- Шилжилтийн урсгал:
 - Raw data → Data Lake
 - Data Lake → ETL → Data Warehouse
 - Data Warehouse → BI/OLAP
 - Data Lake → ML/AI

ӨГӨГДӨЛ ЦУГЛУУЛАХ, ETL ХЭРЭГСЭЛ

Өгөгдөл татах

- Source төрөл: RDBMS, API, файлууд, IoT, real-time feeds
- Push / Pull горим: байгууллага өгөгдлөө заасан серверт байршуулах эсвэл төв серверээс татах

Өгөгдөл хувиргах

- Data cleaning: NULL арилгах, формат засах
- Data standardization: огноо, ангилал кодыг нийцрүүлэх

- Anonymization & hashing: хувь хүн, байгууллагын эмзэг мэдээллийг арилгах, түлхүүр талбарыг нууцлах
- Validation: integrity, foreign key шалгалт

Өгөгдөл ачаалах

- Batch load → тогтсон хуваарьтай
- Incremental load → зөвхөн өөрчлөгдсөн өгөгдөл татах
- Streaming load → real-time (Kafka, Spark Streaming)

BIG DATA ТЕХНОЛОГИ БА ӨГӨГДӨЛ ХАДГАЛАЛТ

Өгөгдөл хадгалах технологиуд

- NoSQL (HBase): Бүтэцгүй, хагас бүтэцтэй өгөгдөлд тохиромжтой, хямдхан scale хийх чадвартай.
- Columnar DB (Kudu): Хурдан хайлт, real-time query-д тохиромжтой.
- Graph DB (Neo4j): Хамаарал судлахад хүчирхэг.

Өгөгдөл боловсруулах технологиуд

- Spark → ETL pipeline-д хүчирхэг
- Hive → их өгөгдөл дээр нарийн аналитик
- Impala → real-time SQL analytics
- Graph tools → гүнзгий хамаарал олж илрүүлэх

ӨГӨГДЛИЙН АЮУЛГҮЙ БАЙДАЛ БА HASH ШИЙДЭЛ

Hash-шийдэл

Төв Token ашигласан hash систем:

- Байгууллагууд регистрээ өөрийн OrgToken-оор hash хийнэ.
- Их өгөгдлийн төв сервер давхар CentralToken-оор дахин hash хийнэ.
- Ижил түлхүүр өгөгдлүүд final hash-аар нэгтгэгдэнэ.

Давуу тал:

- Эмзэг мэдээллийг шууд хадгалахгүй
- Ижил түлхүүрийг таних боломж бүрдүүлнэ

Эрсдэл:

- CentralToken алдагдвал өгөгдлийн нууцлал алдагдах эрсдэлтэй.
- Хамгаалалт шаардлагатай:
 - HSM ашиглах
 - Vault server ашиглах
 - Token-г давтамжтай солих

ӨГӨГДӨЛ ЗАГВАРЧЛАЛ БА METADATA УДИРДЛАГА

Их өгөгдлийн сангийн байгуулахад тулгарч болох гол эрсдэл нь Data Swamp буюу ангилал код, нэгж, агуулгын зөрүүтэй өгөгдлийн сав буюу “булингартсан нуур” бий болгох явдал юм. Үүнээс сэргийлэх нь их өгөгдлийн сан байгуулах суурь зарчим юм.

Metadata & Data Governance

- Entity тодорхойлолт
- Field-level lineage
- Data sensitivity classification
- Business glossary
- Data location

Өгөгдлийн дүрслэл ба шинжилгээ

СИСТЕМ	ОНЦЛОГ
Tableau	Олон төрлийн эх сурвалжаас өгөгдөл татаж боловсруулна, хялбар дашбоард
Power BI	MS экосистемтэй нийцтэй, хямд өртөгтэй
QlikView	Complex visualization
D3.js	Хөгжүүлэгч төвтэй уян хатан шийдэл
Google Data Studio	Google орчинтой хурдан холболт

Tableau → Монголд харьцангуй өргөн хэрэглэгдэж байна.

Big Data analytics tools

- Apache Spark: Маш их өгөгдлийг хурдтай боловсруулах
- Hadoop: Хэмжээний хувьд хязгааргүй өгөгдөлд тохиромжтой
- Google BigQuery: Cloud-based analytics

АЮУЛГҮЙ БАЙДЛЫН ТӨЛӨВЛӨЛТ /SECURITY PIANNING/

Их өгөгдлийн сангийн хувьд өгөгдлийн аюулгүй байдлын төлөвлөлт нь бүх төрлийн өгөгдлийг хамгаалах, хууль эрхзүйн шаардлагуудыг хангах, иргэд болон ААНБ-н эмзэг мэдээллийн аюулгүй байлгах үүрэгтэй. Өгөгдлийн аюулгүй байдлыг хангах гэдэг нь хууль эрхзүй, дүрэм журмаас гадна өгөгдөлд хандах хандалтын хяналтыг хэрэгжүүлэх, технологийн зөв шийдлийг нэвтрүүлэх болон үйл ажиллагааны зөв зохион байгуулалтыг хэрхэн төлөвлөлтөөс шууд хамаарна.

ӨГӨГДЛИЙН АЮУЛГҮЙ БАЙДАЛ, ДАГАЖ МӨРДӨХ ШААРДЛАГЫГ ТОДОРХОЙЛОХ

Их өгөгдлийн сангийн өгөгдлийн аюулгүй байдал нь хуулийн эрхзүйн шаардлагад нийцсэн, хамгаалалтын хатуу үйл ажиллагааг шаардана. Өгөгдлийн аюулгүй байдлыг хангах нь дараах хэд хэдэн үндсэн хүчин зүйлсээс бүрдэнэ:

ӨГӨГДЛИЙН НУУЦЛАЛ БА ХУВИЙН МЭДЭЭЛЛИЙН ХАМГААЛАЛТ

- Монгол Улсад хүчин төгөлдөр үйлчилж буй хууль эрхзүйн зохицуулалтууд:
 - Нээлттэй мэдээллийн ил тод байдлын тухай хууль
 - Төрийн байгууллагууд мэдээллээ олон нийтэд ил тод, нээлттэй байршуулах үүрэгтэй
 - Нээлттэй өгөгдлийг өгөгдлийн портал болон API хэлбэрээр түгээх
 - Хүний хувийн мэдээлэл хамгаалах тухай хууль
 - Хувь хүний танигдахуйц мэдээллийг (PII) зөвшөөрөлгүйгээр ашиглахыг хориглоно
 - Өгөгдөл ашиглалт, агуулах, дамжуулалтын явцад шифрлэлт, хандалтын хяналт зайлшгүй шаардагдана
 - Хувь хүний өгөгдөлд хандсан бүртгэлийг бүрэн логжуулна
 - Кибер аюулгүй байдлын тухай хууль
 - Төрийн мэдээллийн систем, дата төвийн аюулгүй байдалд тавигдах шаардлага
 - Хөндлөнгийн нэвтрэлт, халдлагаас хамгаалах техник болон зохион байгуулалтын арга хэмжээнүүд
 - Мэдээллийн технологийн үндэсний бодлого
 - Төрийн нэгдсэн дата орчныг бүрдүүлэх зорилго бүхий эрх зүйн болон институцийн бодлогын баримт бичиг

- Мэдээллийн интеграц, өгөгдөлд суурилсан шийдвэр гаргалт, цахим засаглалын үндэс суурь болно
- Монгол Улсын Засгийн газрын 2023 оны 403 дугаар тогтоол
 - “Төрийн байгууллагын мэдээллийн сангуудыг нэгтгэн, өгөгдлийн уялдаа холбоо, хүртээмжийг сайжруулах арга хэмжээний тухай” тогтоол
 - Үндэсний дата төв байгуулах, төрийн мэдээллийн сангуудыг интеграцчилах, өгөгдөлд суурилсан бодлого боловсруулах үндсэн чиглэлийг тогтоосон
 - Уг тогтоолын хүрээнд мэдээлэл нийлүүлэх байгууллагуудын үүрэг, техник зохион байгуулалтын шаардлагыг заасан
 - GDPR (EX), OECD Privacy Guidelines зэрэгтэй нийцэх бодлогын элементүүд хэрэгжүүлнэ
 - ISO/IEC 27001 (мэдээллийн аюулгүй байдал), ISO/IEC 38505 (өгөгдлийн засаглал) стандартад нийцүүлэн байгууллагын тогтолцоог төлөвлөнө
- Өгөгдлийг шифрлэх

Өгөгдөл дамжих болон хадгалагдах явцад шифрлэлт (encryption) хийх нь өгөгдлийн аюулгүй байдлыг хангах чухал арга хэрэгсэл юм. Шифрлэх нь тухайн өгөгдлийг зөвхөн зөвшөөрөгдсөн хэрэглэгчдэд ашиглах боломжийг олгодог.
- Өгөгдлийн бүрэн бүтэн байдал, үнэн зөв, найдвартай байдлыг хангах

Өгөгдөл зөв форматтай, дотоод логик алдаагүй, санаатай эсвэл санамсаргүй өөрчлөлтөд өртөхгүй байхаар системийг зохион байгуулахыг шаарддаг. Өөрөөр хэлбэл өгөгдөл өгөгдөл зөв хэлбэртэй хадгалагдах, хоорондоо уялдаатай байх, зөвхөн зөвшөөрөгдсөн хүн өөрчилж боломжтой байх, өөрчлөгдсөн эсэхийг хянах боломжтой байх нь энэхүү үйл ажиллагааны цөм юм.
- Аюулгүй байдлын олон улсын болон MNS стандартуудыг хангах
 - ISO 27001: Мэдээллийн аюулгүй байдлын удирдлагын тогтолцоог хэрэгжүүлэх, хамгаалалтын стандартыг баримтлах.
 - MNS ISO 27001, MNS ISO 27002:2017, MNS ISO 27005:2012, MNS 5724:2012, MNS 5601:2015 стандартуудыг хэрэгжүүлэх

ХАНДАЛТЫН ХЯНАЛТ /ACCESS CONTROL/

Их өгөгдлийн сангийн өгөгдлийг зөвхөн зөвшөөрөгдсөн хэрэглэгчид, байгууллагуудад хандах боломжийг олгох нь аюулгүй байдлын үндсэн чиглэл юм. Бүх оролцогч талуудын хандалтын хяналт нь дараах зарчим, хэрэгслүүдээс бүрдэнэ.

Хандалтын төрөл

- Зөвшөөрөлтэй хандалт (Authenticated Access): Өгөгдөлд хандах эрхийг зөвшөөрсөн хэрэглэгчид болон системүүдэд олгох. Энэ нь хэрэглэгчийн танилт болон нууц үг, биометрийн өгөгдлүүдийг ашиглан хандалтыг баталгаажуулах.
- Нэвтрэх эрхийг хязгаарлах (Authorization): Тохирох зөвшөөрөл бүхий хэрэглэгчид зөвхөн өөртөө зориулсан өгөгдөл болон үйлдлийг хийх эрхтэй байх. Нэвтрэх эрхийн хязгаарлалтыг тухайн хэрэглэгчийн үүрэг, чиг үүрэг, түвшинд тулгуурлан хэрэгжүүлнэ.
- Бүртгэл хөтлөх, аудит лог: Хандалт хийсэн хэрэглэгч болон системийн үйлдлийг бүртгэн хянах, аюулгүй байдлын асуудал үүссэн тохиолдолд бүрэн аудит хийх боломжийг олгох.

Хандалтын хяналтын бодлого

- Principle of Least Privilege: Хэрэглэгчийн хандах эрхийг хамгийн бага хэмжээгээр олгох. Өгөгдөлд хандах эрхийг зөвхөн тухайн хэрэглэгчийн үүрэгт заагдсан хэмжээнд л олгох.
- User Access Management (Хэрэглэгчийн хяналт): Хэрэглэгчийн бүртгэл үүсгэх, хасах, өөрчлөх үйлдлүүдийг зөвхөн системийн удирдлагаар хянах.
- Multi-factor Authentication (Давхар танилт, баталгаажуулалт): Хэрэглэгчийн хандалтын баталгаажуулалтыг сайжруулахын тулд олон шатлалаар танилт хийх.

Өгөгдлийн хамгаалалтын хэрэгсэл

- RBAC (Role-Based Access Control): өгөгдлийн хандалтын хязгаарлалтыг байгууллагын үүрэгт суурилж зохицуулдаг хандалтын удирдлагын систем юм. Энэ систем нь хэрэглэгчид тодорхой үүргийг гүйцэтгэж буй үндсэн дээр тухайн үүрэгтэй холбоотой хандалтын эрхийг олгодог.

- ABAC (Attribute-Based Access Control): Тодорхой шинж чанар дээр үндэслэн хандалтын эрхийг хязгаарлах. ABAC нь хандалтын эрхийг уламжлалт үүрэгт суурилсан хандалттай харьцуулахад илүү уян хатан бөгөөд нарийн зохицуулалт хийдэг.
- VPN (Virtual Private Network): VPN нь интернэтээр дамжих аюулгүй, хувийн холболтыг хангах технологи юм. VPN нь хэрэглэгчдийн интернэтийн хандалтыг шифрлэх болон тэдний мэдээллийг нууцлах зорилгоор ашиглагддаг бөгөөд энэ нь гаднын халдлагаас хамгаалж, аюулгүй холболтыг бий болгох зорилготой.

Хандалт, хяналтын шийдлүүд

- Identity and Access Management (IAM) системүүд: Microsoft Azure AD, Okta зэрэг системүүд нь хандалтын хяналтыг бүрэн гүйцэтгэж, хэрэглэгчийн нэвтрэх эрхийг хянах боломжийг олгодог.
- Data Loss Prevention (DLP) шийдлүүд: Өгөгдөл алдагдах эрсдэлийг бууруулах шийдлүүд, хандалтын эрхийг хязгаарлах боломжтой.

ШИЙДЭЛ	ЗОРИЛГО	СИСТЕМҮҮД
IAM системүүд	Хэрэглэгчийн хандалтын эрхийг удирдах, баталгаажуулах, эрх олголт, хэрэглэгчийн аудит хийх	Microsoft Azure Active Directory, Okta, IBM Security Identity Governance, OneLogin
DLP системүүд	Өгөгдлийг урьдчилан сэргийлэх, хулгайлах, буруу хандалт хийхээс хамгаалах	Symantec DLP, Forcepoint DLP, Digital Guardian, McAfee Total Protection for DLP

ДҮГНЭЛТ

Их өгөгдлийн сан байгуулах ажилд аюулгүй байдлын төлөвлөлт нь системийн хамгийн чухал бүрэлдэхүүн хэсэг бөгөөд хууль эрхзүй, технологийн стандарт, зохион байгуулалтын арга хэмжээг цогцоор нь хангасан байх шаардлагатай. Монгол улсын хууль тогтоомж, олон улсын стандартад нийцүүлэн боловсруулсан энэхүү төлөвлөлт нь төрийн их өгөгдлийн сангийн найдвартай, аюулгүй ажиллагааг баталгаажуулах үндэс суурь болно.

ХУУЛЬ ЭРХ ЗҮЙН ЗОХИЦУУЛАЛТ

Их өгөгдлийн сан нь дараах үндсэн хууль, тогтоол, бодлогын баримт бичгүүдэд нийцсэн байх шаардлагатай:

ХУУЛЬ / БОДЛОГО	ГОЛ АГУУЛГА
Нийтийн мэдээллийн ил тод байдлын тухай хууль	Төрийн өгөгдлийг олон нийтэд ил тод, нээлттэй болгох үүрэг
Хүний хувийн мэдээлэл хамгаалах тухай хууль	Хувь хүний PII-г зөвшөөрөлгүй ашиглахыг хориглох, шифрлэлт, аудит шаардлага
Кибер аюулгүй байдлын тухай хууль	Төрийн мэдээллийн систем, дата төвийн хамгаалалт
Мэдээллийн технологийн үндэсний бодлого	Төрийн нэгдсэн дата орчны эрх зүйн үндэс, цахим засаглал
ЗГ-ын 2023 оны 403-р тогтоол	Төрийн мэдээллийн санг нэгтгэх, интеграцчилах арга хэмжээний чиглэл
Олон улсын хэм хэмжээ	GDPR, OECD Privacy Guidelines, ISO/IEC 27001, ISO/IEC 38505

ӨГӨГДЛИЙН НУУЦЛАЛ БА ХАМГААЛАЛТ

Шифрлэлт (Encryption)

- Өгөгдлийг дамжуулах болон хадгалах явцад шифрлэж зөвхөн зөвшөөрөгдсөн хэрэглэгчид хандах боломжтой болгоно.
- TLS 1.2+, AES-256 зэрэг шифрлэлт ашиглах шаардлагатай.

Өгөгдлийн бүрэн бүтэн байдал

- Өгөгдөл өөрчлөгдөхөөс хамгаалах integrity mechanisms шаардлагатай.
- Логик алдаа, дотоод зөрчлийг хянах механизм нэвтрүүлэх.

Бүртгэл, аудит лог

- Хэн, хэзээ, ямар өгөгдөлд хандсан бүртгэлтэй байх.
- Зөрчил илэрсэн тохиолдолд мөрдөн шалгах боломжтой аудит лог систем шаардлагатай.
-

ОЛОН УЛСЫН БА ҮНДЭСНИЙ СТАНДАРТУУД

Хэрэгжүүлэх стандартууд

- ISO 27001 – Мэдээллийн аюулгүй байдлын удирдлагын тогтолцоо
- MNS ISO 27001, 27002:2017, 27005:2012
- MNS 5724:2012, MNS 5601:2015 – Мэдээллийн хамгаалалтын үндэсний стандартад нийцсэн шаардлагууд

ХАНДАЛТЫН ХЯНАЛТ (ACCESS CONTROL)

Их өгөгдлийн сангийн хандалтыг зөвхөн эрх бүхий хэрэглэгчид хийх боломжтой байдлаар зохион байгуулах нь хамгаалалтын цөм зарчим юм.

ХАНДАЛТЫН ТӨРӨЛ	ТАЙЛБАР
Authenticated Access	Хэрэглэгчийн эрх (username/password, биометр) ашиглан баталгаажуулалт хийх
Authorization	Зөвшөөрөл бүхий хэрэглэгчид л өөрийн эрхлэх үйлдлүүдийг хийх боломжтой байх
Audit Log	Хандалтын бүх үйлдлийг бүртгэх, хянах боломжтой байх

Хэрэгжүүлэх бодлогууд

- Least Privilege Principle – Хэрэглэгчид зөвхөн олгогдсон эрх, үүргийнхээ хэмжээнд хандалтын эрхтэй байна.
- User Access Management – Хэрэглэгч үүсгэх, хасах, өөрчлөх процессыг өндөр хяналттай удирдана.
- Multi-factor Authentication (MFA) – Хандалтыг олон шатлалаар баталгаажуулна.

Их өгөгдлийн сангийн аюулгүй байдлыг хангах нь зөвхөн техник, программ хангамжийн шийдлээс хамаарахгүй бөгөөд олон хүчин зүйлийн хамтын чармайлтын дүнд бий болох тасралтгүй үйл ажиллагаа юм.

- Хууль эрхзүйн шаардлага, олон улсын стандартыг нарийн мөрдөх
- Шифрлэлт, audit лог, хандалтын хяналтын бодлогыг хэрэгжүүлэх
- IAM ба DLP шийдлүүдийг нэвтрүүлэх
- Үүрэгт суурилсан (RBAC) болон шинж чанарт суурилсан (ABAC) хандалтыг уялдуулан хэрэглэх
- Хамгийн бага эрхийн зарчмыг мөрдөх
- Гаднаас холбогдох сувгийг VPN, шифрлэлтээр хамгаалах
- Байгууллагын ажилтнуудын сургалт, хяналт тасралтгүй байх

ХЯНАЛТ, ЗАСВАР ҮЙЛЧИЛГЭЭ /MONITORING AND MAINTENANCE/

Их өгөгдлийн сан нь асар их хэмжээний өгөгдлийг хурдан хугацаанд боловсруулж, хадгалах, олон тооны хандалтыг хүлээн авч тасралтгүй ажиллах шаардлагатай. Дээрх үйл ажиллагааг хангахын тулд зориулалт бүрийн олон төрлийн системүүд өндөр гүйцэтгэлтэй байх ёстойн зэрэгцээ тэдгээрийн тасралтгүй үйл ажиллагааг хангахын тулд хяналт, засвар үйлчилгээний журмыг найдвартай хэрэгжүүлэх шаардлагатай.

СИСТЕМИЙН ҮЙЛ АЖИЛЛАГААГ ХЯНАХ.

- Хурд ба ачаалал хянах (Performance Monitoring):
 - Response Time: Системийн хариу өгөх хугацааг хянах.
 - Throughput: Өгөгдөл боловсруулах хурд, өгөгдлийг нэгтгэх, ачаалах хурдыг хянах.
 - Latency: Өгөгдөл дамжих хугацаа.
 - CPU, RAM, дискийн хэрэглээ: Тогтмол хяналт тавьж, гүйцэтгэлд нөлөөлж болзошгүй ачааллыг хянах.
 - Мониторинг хэрэгсэл: Nagios, Zabbix, Datadog зэрэг системүүдийг ашиглан гүйцэтгэлийг хянах боломжтой.

- Системийн эрсдэл, алдаа хянах:
 - Лог хөтлөх: Алдаа, системийн мэдээлэл, аюулгүй байдлын ноцтой зөрчилтэй холбоотой тэмдэглэлүүдийг тогтмол бүртгэж, хяналт тавих.
 - Асуудал илрүүлэх ба хариу арга хэмжээ авах: Системийн ачаалал, хурдтай холбоотой асуудлуудыг хурдан илрүүлж шийдвэрлэх.

ӨГӨГДЛИЙН АЮУЛГҮЙ БАЙДАЛ БА НУУЦЛАЛ ХЯНАХ

Өгөгдлийн аюулгүй байдлын үүднээс дараах зүйлсийг хянах:

- Хандалтын хяналт: Хэрэглэгч бүрд тохиромжтой хандалтын эрхийн хязгаарлалтыг тогтоох, хандалтын лог хөтлөх, хандалтын шинжилгээ хийх.
- Шифрлэлт: Өгөгдлийн хамгаалалт болон шифрлэлт ба өгөгдөл дамжихдаа шифрлэх, хадгалахдаа шифрлэлт арга хэрэгслүүдийг хэрэглэх.
- Өгөгдлийн алдагдал: Өгөгдөл алдагдах, эвдрэх эсвэл гэмтэх эрсдэлээс хамгаалах механизмуудыг нэвтрүүлэх (DLP систем).
- Аюулгүй байдалд хяналт: Орчны аюулгүй байдлын хянах хэрэгслүүдийг хэрэгжүүлэх, intrusion detection, firewall, anti-virus, antivirus гэх зэрэг.

СИСТЕМИЙН ХҮРТЭЭМЖТЭЙ БАЙДАЛ

Их өгөгдлийн сангийн систем нь өндөр хүртээмжтэй байх зайлшгүй шаардлагатай:

- Системийн баталгаажуулалт: Системийн тасралтгүй байдал, тогтвортой байдлыг хангахын тулд сүлжээний хяналт, серверийн бие даасан байдлыг хангах.
- Нөөцлөлт (Backup): Өгөгдлийг тогтмол хугацаанд нөөцлөх, өгөгдөл алдагдах эрсдэлийг бууруулах.
- Байнгын нөөц болон сэргээх (Disaster Recovery): Бүх төрлийн алдаа, осол гарах үед өгөгдлийг хурдан сэргээх боломжийг хангах.

ЗАСВАР ҮЙЛЧИЛГЭЭНИЙ УДИРДАМЖ

- Системийн шинэчлэлт: Тогтмол шинэчлэлт, аюулгүй байдлын багцуудыг тогтмол суулгах. Үүний тулд автомат шинэчлэлтийн процессыг нэвтрүүлэх.
- Эрсдэлтэй үйлдлүүдийг тогтоох: Системийн асуудал гарсан үед эрсдэлтэй үйлдлийг илрүүлэх болон хариу арга хэмжээг шуурхай авах.
- Засвар үйлчилгээний төлөвлөгөө:
 - Тохиромжтой хугацаа: Системийн тохиргоо болон засвар үйлчилгээ хийх хугацааг урьдчилан төлөвлөх, энэхүү үйл ажиллагааг тодорхой хяналт, зөвшөөрлийн дагуу гүйцэтгэх.
 - Системийн эрсдэлтэй асуудлуудыг шийдвэрлэх: Системийн засвар үйлчилгээг тогтмол хийх, хамгийн их хэрэглэгддэг хэсгүүдийг зөв тохируулах.

ТАЙЛАГНАЛ БА ШИНЖИЛГЭЭ

- Шинжилгээний тайлан: Өгөгдлийн сангийн ажиллагааг хянах, үнэлэх, тайлагнах үйл явцыг тогтмол хийх. Гүйцэтгэл, хандалт, аюулгүй байдалтай холбоотой статистик болон мэдээлэлд дүн шинжилгээ хийх.

- Шуурхай хариу үйлдэл: Асуудал үүсэх үед системийн хариу арга хэмжээг хурдан шуурхай, төлөвлөгөөний дагуу авч хэрэгжүүлэх, системийн нийт хэрэглэгчдэд шаардлагатай мэдээ, мэдэгдэл болон тайлбарыг хүргэх.

ЖИШЭЭ ХЯНАЛТЫН ХЭРЭГСЭЛ

- Nagios/Zabbix: Хяналт ба мониторинг хийх хэрэгслүүд нь системийн гүйцэтгэл, аюулгүй байдал, хандалтын хяналтыг хийхэд тохиромжтой.
- Prometheus + Grafana: Удирдлагын самбар дээр гүйцэтгэлийг хянах, лог анализ хийх хэрэгсэл.
- Datadog: Үүлэн үйлчилгээ болон серверүүдийн гүйцэтгэлийг хянахад зориулсан хэрэгсэл.

ТОГТМОЛ ХЯНАЛТ, ЗАСВАР ҮЙЛЧИЛГЭЭНИЙ ПРОЦЕСС

- Өгөгдлийн чанар хянах: Өгөгдлийг тогтмол хянах, хадгалалтын чанарыг үнэлэх, дүн шинжилгээ хийх.
- Алдаа засварлах, хариу арга хэмжээ авах: Системд алдаа илрэх үед төлөвлөгөөний дагуу шуурхай хариу арга хэмжээ авах, хариу арга хэмжээг бүртгэлжүүлж дүн шинжилгээ хийх ба хохирлын хэмжээг багасгахыг тулд төлөвлөгөөнд сайжруулалт хийх.
- Шинэчлэлт, баталгаажуулалт: Системийн шинэчлэлт хийх, сүүлийн үеийн багцуудыг суулгах, засвар үйлчилгээг тогтмол хийх.

ДҮГНЭЛТ

Их өгөгдлийн сан нь их хэмжээний өгөгдөл хадгалах, боловсруулах, бодит цагийн горимд дүн шинжилгээ хийх өндөр шаардлага тавигддаг систем бөгөөд түүнд тогтвортой, тасралтгүй ажиллагааг хангах “хяналт, засвар үйлчилгээний төлөвлөгөө” зайлшгүй хэрэгтэй. Энэ бүлэгт системийн гүйцэтгэл, аюулгүй байдал, хүртээмж, шинэчлэлт, шуурхай хариу арга хэмжээ зэргийг цогц байдлаар төлөвлөхийг шаардлагатайг онцолж байна.

СИСТЕМИЙН АЖИЛЛАГААГ ХЯНАХ

ҮЙЛ АЖИЛЛАГАА	ҮЗҮҮЛЭЛТҮҮД
Performance Monitoring	Response Time, Throughput, Latency, CPU/RAM/Disk хэрэглээ
System Errors & Risks	Лог хөтлөлт, алдаа илрүүлэх, хариу арга хэмжээ
Жишээ мониторинг хэрэгслүүд	Nagios, Zabbix, Prometheus + Grafana, Datadog

Performance Monitoring

- Хариу өгөх хугацаа (Response Time)
- Өгөгдөл боловсруулах хурд (Throughput)
- Дамжуулалтын хоцролт (Latency)
- Нөөц ашиглалт (CPU, RAM, Disk)

Системийн эрсдэл, алдаа хянах

- Лог бүртгэл → асуудлыг илрүүлэх үндсэн хэрэгсэл
- Шуурхай хариу арга хэмжээ → SLA хангах

ӨГӨГДЛИЙН АЮУЛГҮЙ БАЙДАЛ БА НУУЦЛАЛЫН ХЯНАЛТ

Өгөгдлийн аюулгүй байдлыг хангах үүднээс дараах хяналт, механизмуудыг хэрэгжүүлэх:

ХЯНАЛТЫН ТАЛБАР	ҮЙЛ АЖИЛЛАГАА
Хандалт хяналт	Хандалтын эрх хязгаарлалт, аудит лог, шинжилгээ
Шифрлэлт	Дамжуулах ба хадгалах өгөгдөлд encryption хийх (AES, TLS)
Өгөгдлийн алдагдлын хамгаалалт	DLP систем ашиглах
Intrusion Detection & Security Tools	IDS/IPS, firewall, antivirus хэрэгслүүд

DLP (Data Loss Prevention)

- Өгөгдөл алдагдах эрсдэлийг бууруулах технологи

Intrusion Detection Systems (IDS)

- Гаднаас халдлага илрүүлэх, урьдчилан сэргийлэх

СИСТЕМИЙН ХҮРТЭЭМЖ БА ТАСРАЛТГҮЙ АЖИЛЛАГАА

Их өгөгдлийн сан нь өндөр хүртээмжтэй (High Availability) ажиллах ёстой бөгөөд тасралтгүй ажиллагаа нь системийн амжилтын үндэс юм.

ҮЙЛ АЖИЛЛАГАА	ЗОРИЛГО
Сүлжээ, сервер хяналт	Дотоод ба гадаад сүлжээний хяналт, redundancy хангах
Нөөцлөлт (Backup)	Өгөгдлийн тогтвортой хадгалалт, data loss-аас сэргийлэх
Disaster Recovery (DR)	Осол, алдаа гарсан үед өгөгдөл, системийг хурдан сэргээх

Backup

- Тогтсон хуваарьтай нөөцлөлт хийх → RPO, RTO хангах

Disaster Recovery

- DR төлөвлөгөө → Failover, Offsite Backup, Cloud DR шийдэл

ЗАСВАР ҮЙЛЧИЛГЭЭ

Тогтмол засвар үйлчилгээ (Maintenance) нь системийг эрсдэлгүй, өндөр гүйцэтгэлтэй байлгах чухал нөхцөл юм.

Системийн автомат шинэчлэл

- Patch management систем ашиглах

Risk Identification

- Эрсдэлтэй өөрчлөлтүүдийг аудит логд бүртгэх

Төлөвлөгөөт засвар

- Цаг хугацааг зөв төлөвлөж, хэрэглэгчдэд урьдчилан мэдээлэх байдлаар хэрэгжүүлэх

Их өгөгдлийн сангийн амжилттай ажиллах нөхцөл нь тогтмол хяналт, засвар үйлчилгээний сайн төлөвлөлт юм. Гол шаардлагууд:

- Системийн гүйцэтгэлийг real-time хянах
- Аюулгүй байдлын хатуу хяналт хэрэгжүүлэх
- Байнгын нөөц, DR төлөвлөгөөг баталгаажуулах
- Засвар үйлчилгээ, шинэчлэлтийг тогтмол хийх
- Статистик шинжилгээ, тайлангаар шийдвэр гаргалтыг дэмжих

Энэ нь тасралтгүй ажиллагаа, өгөгдлийн найдвартай байдал, хэрэглэгчийн сэтгэл ханамжийг хангах цогц арга хэмжээний цөм болно.

УР ЧАДВАРЫГ БИЙ БОЛГОХ, СУРГАЛТ

Их өгөгдлийн сангийн хувьд өгөгдлийн шинжлэх ухаан, аналитик, мэдээллийн менежментэд шаардлагатай ур чадварыг бий болгох нь маш чухал юм. Энэ нь төрийн албан хаагчид болон олон нийтэд их өгөгдлийн ач холбогдлыг ойлгуулж, их өгөгдлийг үр ашигтай ашиглахад шаардлагатай мэдлэг, ур чадварыг эзэмшүүлэхэд чухал түлхэц болно. Мөн их өгөгдлийн санг хариуцан ажиллах мэргэжилтнүүдийг бэлтгэх, тэдний ур чадварыг тогтмол дээшлүүлж байх нь их өгөгдлийн сангийн хэвийн үйл ажиллагааг хангах, тасралтгүй сайжруулах хөгжүүлэх суурь үндэс болно. Нөгөө талаас нотолгоонд суурилсан шийдвэр гаргалтыг дэмжих, их өгөгдлийн хэрэглээг нэмэгдүүлэхийн тулд олон нийтэд их өгөгдлийг сурталчлан таниулах, сургах нь их өгөгдлийн эко системийн нэг чухал хэсэг юм.

ӨГӨГДЛИЙН ШИНЖЛЭХ УХААН, АНАЛИТИК, МЭДЭЭЛЛИЙН МЕНЕЖМЕНТЭД ШААРДЛАГАТАЙ УР ЧАДВАРЫГ БИЙ БОЛГОХ.

Өгөгдлийн шинжлэх ухаан ба аналитикийн ур чадварууд

- Өгөгдлийн шинжлэх ухаан:
 - Программчлал: Python, R хэлний тусламжтай өгөгдлийн боловсруулалт, дүн шинжилгээ хийх
 - Өгөгдлийн боловсруулалт: Өгөгдлийг цэвэрлэх, хөрвүүлэх, боловсруулалт хийх ур чадвар эзэмших.
 - Машин сургалт: Машин сургалтын алгоритмуудыг эзэмших, өгөгдөлд дүн шинжилгээ хийх, таамаглал гаргах.
 - Тооцоолол: Статистик анализ хийх, өгөгдлийг дүрслэх (Data Visualization), тайлбарлах.
- Аналитик ур чадвар:
 - Өгөгдлийн шинжилгээ: Excel, SQL, Power BI болон Tableau ашиглан өгөгдлийн шинжилгээ хийх.
 - Их өгөгдлийн технологи: Hadoop, Spark зэрэг их хэмжээний өгөгдөлтэй ажилладаг системүүдийг эзэмших
 - Тайлан гаргах: Өгөгдөлд суурилсан тайлан, санал дүгнэлт боловсруулах чадвар эзэмших.

Сургалтын хөтөлбөрийн бүтэц

- Анхан шатны сургалт: Өгөгдлийн шинжлэх ухаан, статистик, машин сургалт болон дүн шинжилгээний суурь ойлголтуудыг эзэмшүүлэх.
- Дунд шатны сургалт: Өгөгдлийг боловсруулж, нарийвчилсан шинжилгээ хийх аргачлалууд, мэдээллийн менежментийн системийн ашиглалтад сургах.
- Ахисан шатны сургалт: Их өгөгдлийн орчинд ажиллах, үр дүнтэй тайлан боловсруулах, өгөгдөлд суурилсан шийдвэр гаргалт хийх чадвар эзэмших.

Сургалтын аргачлалууд

- Онлайн сургалт: Coursera, edX, LinkedIn Learning зэрэг платформуудын өгөгдлийн шинжлэх ухааны чиглэлээр бэлтгэсэн чанартай онлайн сургалтуудыг судлах. Дээрх

системүүдээр дамжуулах өөрсдийн онлайн сургалтын хөтөлбөрийг бэлтгэж сургалтыг зохион байгуулах.

- Ангийн сургалт:
Төрийн албан хаагчдад зориулсан их өгөгдлийн чиглэлийн эрдэм шинжилгээний хурал, хэлэлцүүлгийг зохион байгуулах.
- Туршилт, дадлага:
Их өгөгдлийн орчинд ажиллах, өгөгдлийн боловсруулалт хийх, дүн шинжилгээ хийх боломжуудыг бодит орчинд бий болгох.

ТӨРИЙН АЛБАН ХААГЧИД БОЛОН ОЛОН НИЙТИЙН ДУНД ИХ ӨГӨГДЛИЙН ХЭРЭГЛЭЭ, АЧ ХОЛБОГДЛЫГ ТАЙЛБАРЛАН ТАНИУЛАХ СУРГАЛТЫН ХӨТӨЛБӨР, ТӨЛӨВЛӨГӨӨ

Төрийн албан хаагчдад зориулсан сургалт

- Их өгөгдлийн хэрэглээ:
 - Их өгөгдлийн сангууд, катологиудын тухай, мета мэдээллийн сангийн тухай болон тэдгээрийн ач холбогдлыг тодорхойлох.
 - Их өгөгдлийн санд нийтүүлж буй өгөгдлийн стандарт, нийцтэй байдал, өгөгдлийн үнэ цэнэ, ач холбогдол, хэрэгцээ шаардлагыг нарийвчлан тодорхойлох
 - Өгөгдөл боловсруулах, хамгаалах, тайлагнах үндсэн мэдлэг, ур чадварыг эзэмшүүлэх.
- Гүйцэтгэл ба шийдвэр гаргалт:
 - Нотолгоонд суурилсан шийдвэр гаргах ур чадварыг олгох.
 - Төрийн байгууллагуудад өгөгдлийн шинжилгээг ашиглан стратеги төлөвлөлт боловсруулах.

Олон нийтэд зориулсан сургалт

- Нээлттэй өгөгдөл:
 - Олон нийтийн дунд нээлттэй өгөгдлийн ач холбогдлыг таниулах.
 - Иргэний мэдээлэл төрийн байгууллагуудад үнэн зөв, хуучраагүй бүртгэгдсэн байх нь төрийн бодлого шийдвэрлэлтэд хэрхэн нөлөө үзүүлэх талаар сурталчлан таниулах, ирэн өөрийн мэдээллийн үнэн зөвийг хянах талаар мэдлэг мэдээлэл өгөх(emongolia системээр өөрийн тодорхой мэдээллүүдийг хянах, үнэн зөвд анхаарах боломжтой)
 - Их өгөгдлийг хэрхэн ашиглах, их өгөгдлийг судалснаар, хэрэглэснээр ямар ямар шинэ боломж бий болж болох, тэдгээрийг хэрхэн өдөр тутмын амьдралдаа үр бүтээлтэй ашиглаж болох талаар сургалт зохион байгуулах.
- Сургалтын аргачлалууд:
 - Уулзалт, нээлттэй өдөрлөг: Төрийн байгууллага болон олон нийтийг хамарсан сургалт, танилцуулга, өдөрлөг зохион байгуулж их өгөгдлийн хэрэглээг сурталчлан таниулах, сургалт зохион байгуулах
 - Онлайн хичээл: Их өгөгдлийн хэрэглээ болон түүний ач холбогдлыг нээлттэй онлайн сургалтуудаар дамжуулан олон нийтэд хүртээмжтэй болгох.
- Олон нийтийн компанит ажил:

- Олон нийтэд сонирхлын түвшингүүдэд хуваасан болон төрлийн их өгөгдлийг сурталчлан таниулах үйл ажиллагааг зохион байгуулах. Жишээ нь оюутнуудын дунд, стартап компаниудын хүрээнд гэх зэргээр бүлэглэсэн компани ажлуудын хүрээнд их өгөгдлийн хэрэглээ, ач холбогдлыг ойлгуулах, сурталчлах нь зүйтэй.

МЭДЭЭЛЭЛ ХАРИУЦАГЧ БАЙГУУЛЛАГУУД БОЛОН ИХ ӨГӨГДЛИЙН САНГ ХАРИУЦАН АЖИЛЛАХ ХҮНИЙ НӨӨЦИЙН ТООЦОО СУДАЛГАА, МЭРГЭЖИЛТЭН БЭЛТГЭХ ЗАГВАР, ТӨЛӨВЛӨГӨӨГ БОЛОВСРУУЛАХ;

Их өгөгдлийн сан байгуулах төсөл бол зөвхөн технологи биш харин хүний нөөц, технологи, үйл ажиллагааны нарийн төлөвлөлтийн үр дүнд бий болох томоохон үйл ажиллагаа юм. Эдгээрээс хамгийн том хүчин зүйл нь хүний нөөц буюу бэлтгэгдсэн ур чадвар бүхий ажилтан нь амжилтын гол түлхүүр болно.

Иймд их өгөгдлийн сан, дата төв болон өгөгдөл нийлүүлэгч байгууллагуудын хувьд хүний нөөцийн тооцоо, шаардлагатай ур чадвар болон мэргэжилтэн бэлтгэх загварыг авч үзье.

ӨГӨГДӨЛ ХАРИУЦАГЧ(НИЙЛҮҮЛЭГЧ) БАЙГУУЛЛАГУУДЫН ХҮНИЙ НӨӨЦИЙН ХЭРЭГЦЭЭ

Өгөгдөл хариуцагч байгууллага гэдэг нь:

- Төрийн 152 байгууллага
- Томоохон төрийн өмчит аж ахуйн нэгжүүд
- Онцгой чиг үүргийн агентлагууд
- Цаашид ААНБ, их дээд сургуулиуд, судалгаа шинжилгээний байгууллагууд.

Эдгээр байгууллагууд нь их өгөгдлийн сан буюу дата төвтэй холбогдож, өгөгдлөө нийлүүлэх үүрэг бүхий мэргэшсэн багтай байх шаардлагатай. Тухайн байгууллагаас нийлүүлж буй өгөгдлийн үнэн зөв, нийцтэй байдал, түүнийг тогтсон дүрмийн дагуу алдаа мадаггүй дамжуулах, өгөгдлийг шифрлэх, харилцагчийн эмзэг мэдээллийг арилгах зэрэг нарийн төвөгтэй үйл ажиллагааг эдгээр багууд хариуцан ажиллана. Анхдагч эх үүсвэрээс нийлүүлж буй өгөгдөл чанартай байх нь нийт их өгөгдлийн сангийн чанарт шууд нөлөө үзүүлнэ.

Өгөгдөл хариуцагч байгууллагад шаардагдах хүний нөөц (хамгийн багадаа)

АЛБАН ТУШААЛ	ТОО	ГОЛ ҮҮРЭГ
Өгөгдлийн сангийн админ (DBA)	1+	DB query, extract, format
ETL/Интеграцийн мэргэжилтэн	1+	CSV, JSON, XML гадагш гаргах
Өгөгдлийн чанарын хянагч	1+	Data validation, metadata mapping
Мэдээллийн аюулгүй байдлын ажилтан	0.5 (давхар эрхлэх боломжтой)	Hashing, encryption, эрхийн хяналт

Жич: Төрийн байгууллагуудад “Мэдээллийн аюулгүй байдлын ажилтан” орон тоо бий болсон нь давуу тал бөгөөд тус ажилтны ажил үүргийн хуваарьт их өгөгдлийн санд өгөгдөл нийлүүлэхтэй холбоотой тодорхой эрх үүргүүдийг шингээж өгөх шаардлагатай. Нэг байгууллагад дунджаар 3–4 мэргэжилтэн шаардлагатай гэж үзвэл ойролцоогоор 530 орчим бэлтгэгдсэн, ур чадвар бүхий хүний нөөц шаардагдахаар байна.

Нийт өгөгдөл нийлүүлэгч байгууллагын хүний тооцоо

ТОО ХЭМЖЭЭ	ХЭМЖЭЭ ТООЦОО
Байгууллагын тоо	152
Дундаж шаардлагатай мэргэжилтэн	3.5 хүн
Нийт хүний нөөцийн хэрэгцээ	~530 хүн

ИХ ӨГӨГДЛИЙН САН ХАРИУЦСАН ДАТА ТӨВИЙН ХҮНИЙ НӨӨЦИЙН ХЭРЭГЦЭЭ

Хүний нөөцийн задаргаа:

АЛБАН ТУШААЛ	ШААРДЛАГАТАЙ ТОО	ҮҮРЭГ
Data Architect	1–2	Архитектурын зураглал, зохион байгуулалт
Data Engineer	2–3	ETL pipeline, Data Lake, DWH
ETL Developer	2–3	Өгөгдөл цэвэрлэх, transform хийх
DBA	2–3	PostgreSQL, MongoDB, Hadoop cluster management
BI/Reporting Expert	1–2	Power BI, Tableau, Superset
AI/ML Engineer	1	ML pipeline, data modeling
DevOps/Monitoring Engineer	1–2	Kubernetes, logging, monitoring
Security Specialist	1	RBAC, encryption, security policy

Их өгөгдлийн сан хариуцсан дата төвийн үндсэн баг нь ойролцоогоор 12-18 нарийн мэргэжлийн хүний бүрэлдэхүүнтэй байна. Үүнд байгууллагын бусад үйл ажиллагаа хариуцсан ажиллагсад нэмэгдэнэ.

Шаардлагатай мэдлэг, ур чадварууд

ЧИГЛЭЛ	УР ЧАДВАРУУД
DBA	PostgreSQL, MySQL, MongoDB, HDFS, SQL, backup/restore
Data Engineer	Python, Spark, NiFi, Airflow, JSON/XML
ETL Developer	Talend, Pentaho, n8n, API Integration
Security Specialist	ISO 27001, Hashing, Encryption, RBAC, Firewall
BI Expert	Power BI, Tableau, Superset, SQL query
AI/ML Engineer	Python, Scikit-Learn, TensorFlow, MLFlow

МЭРГЭЖИЛТЭН БЭЛТГЭХ ЗАГВАР, ТӨЛӨВЛӨГӨӨ

Сургалтыг гурван үе шатад хуваан авч үзлээ:

I үе шат (Суурь ойлголт) | 2 долоо хоног

- Open Data, Big Data ойлголт
- Database concepts
- Data governance basics

- ISO 27001 суурь ойлголт

II үе шат (Техник сургалт) | 3-4 долоо хоног

- ETL хэрэгслүүд (Talend, NiFi, Airflow, n8n)
- SQL advanced
- Data Quality tools
- Data lineage, Metadata catalog
- Encryption, Hashing

III үе шат (Практик сургалт) | 4 долоо хоног

- Real dataset extraction
- Data cleansing workshop
- Data mapping, code conversion
- Tableau, Power BI, Superset дээр тайлан боловсруулах
- AI/ML workshop

Сургалтын багц төлөвлөгөө

СУРГАЛТ	ХУГАЦАА	ОРОЛЦОГЧИД
DBA Training	21 хоног	152 байгууллагын DBA-ууд
ETL Development	28 хоног	ETL developer-ууд
Security Workshop	14 хоног	Security staff
BI & Analytics	14 хоног	BI Analysts
AI/ML Bootcamp	14 хоног	AI Engineers

ДҮГНЭЛТ

Их өгөгдлийн сангийн амжилттай ажиллагаанд зөвхөн техник, технологи бус харин үр чадварлаг хүний нөөц хамгийн чухал хүчин зүйл болохыг энэ бүлэгт тусгаж өглөө. Өгөгдлийн шинжлэх ухаан, аналитик, мэдээллийн менежмент, аюулгүй байдлын чиглэлээр мэргэжилтнүүдийг бэлтгэх нь үндсэн зорилт бөгөөд Python, R, SQL, Hadoop, Spark зэрэг технологиос гадна өгөгдлийг цэвэрлэх, боловсруулалт хийх, тайлан боловсруулах үр чадвар чухал шаардлагатай юм.

Сургалт гурван түвшинд (анхан, дунд, ахисан) явагдаж, онлайн болон танхимын хэлбэр хослон хэрэгжихээр төлөвлөсөн. Төрийн албан хаагчдад их өгөгдлийн ач холбогдлыг ойлгуулах, өгөгдлийг стандартын дагуу бэлтгэх, тайлагнах чадвар олгох бол олон нийтэд нээлттэй өгөгдлийн ач холбогдлыг таниулах, өдөр тутмын амьдралдаа ашиглах мэдлэг өгөх сургалтууд хийгдэнэ.

Өгөгдөл нийлүүлэгч байгууллагууд нэг бүрд өгөгдлийн сангийн админ, ETL мэргэжилтэн, өгөгдлийн чанарын хянагч, мэдээллийн аюулгүй байдлын ажилтан зэрэг мэргэжилтнүүд зайлшгүй хэрэгтэй бөгөөд нийтдээ 500 гаруй хүний нөөц шаардлагатай гэж тооцсон. Их өгөгдлийн санг хариуцан ажиллах дата төвд ч мөн архитектор, инженер, ETL хөгжүүлэгч, BI шинжээч, аюулгүй байдлын мэргэжилтэн гээд 12–18 орчим мэргэжилтэн шаардлагатай байна.

Ингэснээр их өгөгдлийн санг үр ашигтай, аюулгүй, тасралтгүй ажиллуулах хүний нөөцийн суурь бүрдэж, төрийн байгууллагын нотолгоонд суурилсан бодлого шийдвэр гаргалт, олон нийтийн оролцоо, өгөгдлийн үр ашигтай хэрэглээ нэмэгдэх нөхцөл хангагдана.

ТӨСӨВ: ШААРДЛАГАТАЙ ЗАРДЛЫН ТӨСВИЙН ТООЦОО

Дээр дурдсан тоног төхөөрөмж, программ хангамжуудын хүрээнд төрийн байгууллагуудаас өгөгдөл цуглуулах, хадгалах, тооцоолол хийх хүчин чадал бүхий дата төвийн зардлын тооцоог хоёр хувилбараар тооцлоо.

- А Хувилбар: Үйл ажиллагаа явуулахад хангалттай, зардлыг бага байх шийдэл
- Б Хувилбар: Тоног төхөөрөмж, программ хангамжуудыг хангалттай түвшинд байхаар сонгов.

А ХУВИЛБАР

Дэд бүтэц (Hardware)

ТОНОГ ТӨХӨӨРӨМЖ	ТОО ШИРХЭГ	НЭГЖИЙН ҮНЭ	НИЙТ (USD)
Rack Server (32 core, 256GB RAM, NVMe 4TB)	10	\$15,000	\$150,000
Enterprise Storage (SAN/NAS) ~300TB usable	1	\$250,000	\$250,000
Enterprise Core Switch (Layer 3)	4	\$8,000	\$32,000
Enterprise Firewall	2	\$20,000	\$40,000
SD-WAN Appliance	2	\$12,000	\$24,000
UPS, Rack, cooling	-	\$40,000	\$40,000
HARDWARE НИЙТ			\$536,000

Программ хангамж (Software & Licenses)

ПРОГРАММ ХАНГАМЖ	ЛИЦЕНЗ / ЖИЛ	ТОО	ДҮН (USD)
Informatica PowerCenter (ETL)	\$100,000	1	\$100,000
Collibra Data Catalog	\$80,000	1	\$80,000
Cloudera CDP Enterprise (Hadoop + Spark)	\$150,000	1	\$150,000
Tableau Server (10 user)	\$20,000	1	\$20,000
Oracle DB Enterprise Edition	\$47,500 / CPU	4 CPU	\$190,000
Linux Enterprise Support (Red Hat)	\$2,500	10	\$25,000
Splunk Enterprise (monitoring)	\$20,000	1	\$20,000
Backup enterprise software (Veeam, Veritas)	\$15,000	1	\$15,000
SOFTWARE НИЙТ			\$600,000

Cloud Hybrid үйлчилгээ

ҮЙЛЧИЛГЭЭ	ТОО ХЭМЖЭЭ	НЭГЖИЙН ҮНЭ	ДҮН (USD)
Cloud Storage (S3 archive, 200TB)	200TB	\$23/TB/сар	\$55,200/жил
Cloud Compute for AI/ML	100,000 CPU цаг	\$0.10/цаг	\$10,000
Data Transfer	50TB	\$0.05/GB	\$2,500
CLOUD НИЙТ			\$67,700

Хүний нөөцийн сургалт

ХҮНИЙ НӨӨЦ	ХЭМЖЭЭ	НЭГЖИЙН ҮНЭ	ДҮН (USD)
DBA, Data Engineer, ETL, Security, BI, AI сургалт	50 хүн	\$3,000	\$150,000

А ХУВИЛБАР, НИЙТ ТӨСӨВ: \$1,353,700

Б ХУВИЛБАР

Дэд бүтэц (Hardware)

ТОНОГ ТӨХӨӨРӨМЖ	ТОО ШИРХЭГ	НЭГЖИЙН ҮНЭ	ДҮН (USD)
High-end Server (64 core, 512GB RAM, NVMe 8TB)	12	\$30,000	\$360,000
Enterprise Storage SAN/NAS (600TB usable)	1	\$500,000	\$500,000
Core Switch (Layer 3, 40G ports)	6	\$15,000	\$90,000
SD-WAN Appliance	2	\$20,000	\$40,000
Enterprise Firewall	2	\$35,000	\$70,000
Cooling, Power, UPS	-	\$80,000	\$80,000
HARDWARE НИЙТ			\$1,140,000

Программ хангамж (Software & Licenses)

ПРОГРАММ ХАНГАМЖ	ЛИЦЕНЗ / ЖИЛ	ТОО	ДҮН (USD)
Informatica PowerCenter Advanced	\$250,000	1	\$250,000
Collibra Data Governance Suite	\$150,000	1	\$150,000
Cloudera CDP Private Cloud	\$300,000	1	\$300,000
Tableau Server (Enterprise, 50 user)	\$60,000	1	\$60,000
Oracle DB Enterprise Edition + Partitioning + Advanced Security	\$100,000 / CPU	4 CPU	\$400,000
Red Hat Enterprise Linux + Satellite	\$3,500	12	\$42,000
Splunk Enterprise (Advanced)	\$50,000	1	\$50,000
Veritas Enterprise Backup	\$50,000	1	\$50,000
SOFTWARE НИЙТ			\$1,302,000

Cloud Hybrid үйлчилгээ

ҮЙЛЧИЛГЭЭ	ТОО ХЭМЖЭЭ	НЭГЖИЙН ҮНЭ	ДҮН (USD)
Cloud Storage (S3 archive, 500TB)	500TB	\$23/TB/сар	\$138,000/жил

Cloud Compute (AI/ML workload)	300,000 CPU цаг	\$0.10/цаг	\$30,000
Data Transfer	150TB	\$0.05/GB	\$7,500
CLOUD НИЙТ			\$175,500

Хүний нөөцийн сургалт

ХҮНИЙ НӨӨЦ	ХЭМЖЭЭ	НЭГЖИЙН ҮНЭ	ДҮН (USD)
DBA, Data Architect, Data Engineer, Security, AI сургалт	80 хүн	\$4,500	\$360,000

Б ХУВИЛБАР, НИЙТ ТӨСӨВ: \$2,977,500

Хоёр хувилбарыг харьцуулбал

ЗАРДЛЫН БҮЛЭГ	А ХУВИЛБАР	Б ХУВИЛБАР
Hardware	\$536,000	\$1,140,000
Software	\$600,000	\$1,302,000
Cloud	\$67,700	\$175,500
Training	\$150,000	\$360,000
НИЙТ	\$1,353,700	\$2,977,500

Анхны хөрөнгө оруулалтаас гадна тоног төхөөрөмж, лиценз шинэчлэлийн тооцооллыг 5 жилийн хугацаанд, хоёр хувилбарын хувьд тусад нь гаргалаа.

Энэхүү тооцоололд:

- Тоног төхөөрөмжийн шинэчлэлт → жил бүр 10-15%
- Лиценз, программ хангамжийн шинэчлэлт (хэрэглээ нэмэгдэх, AI/ML өснө) → жил бүр 20% өсөх
- Cloud (data volume нэмэгдэнэ) → жил бүр 10% өснө
- Сургалтын зардлыг дараа дараагийн жилүүдэд багасах тул тооцооноос хаслаа.

А ХУВИЛБАР, НИЙТ

ЗАРДАЛ	2 ЖИЛ	3 ЖИЛ	4 ЖИЛ	5 ЖИЛ	НИЙТ
Hardware renewal	\$64,320	\$64,320	\$64,320	\$64,320	\$257,280
Software renewal	\$600,000	\$720,000	\$864,000	\$1,036,800	\$3,220,800
Cloud зардал	\$74,470	\$81,917	\$90,109	\$99,120	\$345,616
НИЙТ	\$738,790	\$866,237	\$1,018,429	\$1,200,240	\$3,823,696

Б ХУВИЛБАР, НИЙТ

ЗАРДАЛ	2 ЖИЛ	3 ЖИЛ	4 ЖИЛ	5 ЖИЛ	НИЙТ
Hardware renewal	\$136,800	\$136,800	\$136,800	\$136,800	\$547,200
Software renewal	\$1,302,000	\$1,562,400	\$1,874,880	\$2,249,856	\$6,989,136
Cloud зардал	\$193,050	\$212,355	\$233,591	\$257,048	\$896,044
НИЙТ	\$1,631,850	\$1,911,555	\$2,245,271	\$2,643,704	\$8,432,380

Хоёр хувилбарыг харьцуулбал

ТӨСВИЙН БҮЛЭГ	А ХУВИЛБАР	Б ХУВИЛБАР
Анхны хөрөнгө	\$1,353,700	\$2,977,500
Урсгал зардал (5 жил)	\$3,823,696	\$8,432,380
НИЙТ	\$5,177,396	\$11,409,880

БУСАД УЛС ОРНУУДЫН ИХ ӨГӨГДЛИЙН ДАТА ТӨВ БАЙГУУЛСАН ЖИШИГ

Үндэсний их өгөгдлийн дата төв байгуулсан болон байгуулж буй олон улс орнуудын жишээг харж болохоор байна. Тухайлбал БНХАУ 2015-2020 онуудад “Big Data Valley” төслийн хүрээнд Guizhou Big Data Center их өгөгдлийн төвийн ойролцоогоор 2.3 тэрбум ам.доллар, Сингапур 2016-2019 онуудад “Singapore Government Data Center” төвийг ойролцоогоор 145 сая ам.долларын өртгөөр байгуулж төрийн байгууллагуудыг бүх өгөгдөл шилжүүлсэн байна.

Бид жишээ болгон хөгжлийн түвшин манай улстай оролцоо улсууд болох Казахстан, Гүрж, Узбекистан улсуудад хэрэгжиж буй төслүүдийг авч үзлээ. Эдгээр улсууд нь Монголтой ДНБ, хүн амын хэмжээ, дижитал хөгжил ойролцоо дунд орлоготой, хөгжиж буй улсууд юм.

КАЗАХСТАН – ASTANA HUB DATA CENTER

- Хугацаа: 2019–2021
- Зардал: ~45 сая USD
- Тайлбар:
 - “Astana Hub” бол төрөөс санаачилсан инновацийн экосистем.
 - Дата төв нь төрийн байгууллагын системүүд болон стартапуудын клауд платформыг дэмждэг.
 - Tier III стандарттай.
 - Эрчим хүч хэмнэх технологи ашигласан.

ҮЗҮҮЛЭЛТ	ДЭЛГЭРЭНГҮЙ
Tier Level	Tier III
Серверүүд	- HPE ProLiant DL360 Gen10 - Dell PowerEdge R740 - Cisco UCS B200 M5 blades
Хадгалах (Storage)	- NetApp AFF A300 All-Flash массив - Dell EMC Unity 300
Сүлжээ (Network)	- Cisco Nexus 9000- Juniper EX4300
Виртуалчлал	VMware vSphere 7
Үйлдлийн системүүд	Linux (RHEL, Ubuntu), Windows Server 2019
Системүүд / үйлчилгээ	- E-Government портал - Open Data платформ - Digital Startup Ecosystem платформ - ERP системүүд (SAP S/4HANA)
Аюулгүй байдал	Fortinet NGFW, Cisco ASA
Хөргөх систем	Precision Air Cooling (InRow)
Эрчим хүч	2N UPS систем, Diesel генератор

- Эх сурвалж: Astana Hub | <https://astanahub.com/en/>

ГҮРЖ – GEORGIA DATA EXCHANGE PLATFORM

- Хугацаа: 2018–2020
- Зардал: ~25 сая USD
- Тайлбар:
 - ЕАБХАБ (EBRD) болон Дэлхийн Банкны санхүүжилттэй төсөл.

- Төрийн байгууллагын хооронд өгөгдөл солилцох үндэсний платформ.
- Дата төв, платформ хоёр хамтад нь байгуулагдсан.
- E-Governance, нэг цонхны үйлчилгээний дата багтаамжийг шийдвэрлэсэн.

ҮЗҮҮЛЭЛТ	ДЭЛГЭРЭНГҮЙ
Tier Level	Tier II (зарим хэсэг Tier III)
Серверүүд	- HPE ProLiant DL380 Gen10 - Lenovo ThinkSystem SR650
Хадгалах (Storage)	- IBM Storwize V7000 - Lenovo Storage D3284 JBOD
Сүлжээ (Network)	- Cisco Catalyst 9300 - MikroTik Cloud Router Switch CRS317
Виртуалчлал	VMware vSphere 6.7
Үйлдлийн системүүд	Linux (Debian, CentOS), Windows Server 2016/2019
Системүүд / үйлчилгээ	- E-Governance платформ - Government Data Exchange (X-Road аналог) - Open Data portal - ERP (Oracle EBS)
Аюулгүй байдал	Palo Alto Networks Firewall, MikroTik VPN
Хөргөх систем	Precision Air Cooling, чийгшил хянах систем
Эрчим хүч	N+1 UPS, дизель генератор

- Эх сурвалж: EBRD Georgia | <https://www.gita.gov.ge/>

УЗБЕКИСТАН – DIGITAL UZBEKISTAN DATA CENTER

- Хугацаа: 2020–2023
- Зардал: ~50–60 сая USD
- Тайлбар:
 - Huawei-тай хамтран байгуулсан.
 - Үндэсний хэмжээний төрийн мэдээллийн систем, клауд үйлчилгээ ажиллуулж эхэлсэн.
 - 99.98% хүртээмжтэй, Tier III түвшинд хүрсэн.
 - AI, Big Data технологи ашиглахаар төлөвлөсөн.

ҮЗҮҮЛЭЛТ	ДЭЛГЭРЭНГҮЙ
Tier Level	Tier III Certified
Серверүүд	- Huawei FusionServer Pro 2288H V5 - Inspur NF5280M5
Хадгалах (Storage)	- Huawei OceanStor Dorado 5000 V6 AI I-Flash- Huawei OceanStor 2600 V3
Сүлжээ (Network)	- Huawei CloudEngine S12700E - Cisco Catalyst 9500 (төрийн системд)
Виртуалчлал	Huawei FusionCompute (Hypervisor) VMware vSphere (хослуулсан)

Үйлдлийн системүүд	Linux (EulerOS, Ubuntu), Windows Server 2019
Системүүд / үйлчилгээ	- E-Government портал - Cloud платформ (Huawei Cloud Stack) - Citizen Services Portal - Digital Identity Management
Аюулгүй байдал	Huawei USG6000V Next-Gen Firewall- Cisco Firepower (төрийн байгууллагын зарим системд)
Хөргөх систем	Huawei Smart Cooling System
Эрчим хүч	2N UPS, генератор, Green Energy Integration

- Эх сурвалж: Uzbekistan ICT Ministry | <https://mitc.uz>

В ХУВИЛБАР

Манай орны тухайд бид өнөөдрийн байдлаар төрийн мэдээллийн дэд бүтэц байгуулах, төрийн үйлчилгээг цахимжуулах, мэдээллийг боловсруулах, аюулгүй байдлыг хангах болон цахим засаглалыг хэрэгжүүлэх чиглэлээр томоохон төслүүдийг амжилттай хэрэгжүүлээд байгаа билээ.

ҮНДЭСНИЙ ДАТА ТӨВ

- ҮДТ нь Монгол Улсын Засгийн газрын харьяа байгууллага юм.
- Байгуулагдсан гол зорилго:
 - Төрийн болон төрийн өмчит байгууллагуудын мэдээллийн технологийн дэд бүтцийг төвлөрүүлэх
 - Мэдээллийн аюулгүй байдлыг хангах
 - Засгийн газрын цахим үйлчилгээг дэмжих

Өмнө нь Үндэсний мэдээллийн технологийн парк (ҮМТП), Харилцаа холбоо, мэдээллийн технологийн газар зэрэг байгууллагуудын дата төвийн үүргийг хувааж гүйцэтгэж байсан бол өнөөдөр ҮДТ нь тэдгээрийг төвлөрүүлэн гүйцэтгэж байна.

ҮНДСЭН ҮҮРЭГ, ЗОРИЛТ

- Төрийн болон төрийн өмчит байгууллагуудын сервер, мэдээллийн сан, мэдээллийн системийг байрлуулах, ажиллуулах
- Засгийн газрын мэдээллийн сүлжээний дэд бүтцийг хөгжүүлэх, хамгаалах
- Өндөр хүртээмжтэй, найдвартай, аюулгүй дата төвийн орчин бүрдүүлэх
- Цахим засаглалын платформуудыг дэмжих
- Засгийн газрын үүлэн тооцооллын үйлчилгээг санал болгох
- Гамшгаас хамгаалах, сэргээн босгох төлөвлөгөө (Disaster Recovery Plan) хэрэгжүүлэх
- Цахим гарын үсгийн дэд бүтцийг хангах (PKI)

ҮНДСЭН ҮЙЛЧИЛГЭЭНИЙ ТӨРЛҮҮД

Colocation (Сервер байрлуулах үйлчилгээ)

- Байгууллагууд өөрсдийн серверүүдээ ҮДТ-д байрлуулж, цахилгаан, хөргөлт, сүлжээний найдвартай байдлаар хангагдана.
- 24/7 хяналт, хамгаалалттай.

Үүлэн үйлчилгээ (Cloud Services)

- Виртуал сервер (Virtual Machine)
- Хадгалах (Storage)
- Үүлэн платформ
- Backup болон DRaaS (Disaster Recovery as a Service)
- Container service (Docker, Kubernetes гэх мэт)

Data Backup & Recovery

- Өгөгдөл алдагдахаас сэргийлж нөөцлөх
- Гамшгийн нөхцөлд өгөгдлийг сэргээх үйлчилгээ

Төрийн цахим шуудангийн үйлчилгээ

- Засгийн газрын албан ёсны имэйл хаяг олгох
- Өндөр хамгаалалттай имэйл систем

Сүлжээний үйлчилгээ

- Засгийн газрын сүлжээ (GovNet)
- VPN, хос холболт
- Нэгдсэн сүлжээний менежмент

Цахим гарын үсгийн дэд бүтэц (PKI)

- Тоон гарын үсгийн гэрчилгээ олгох, баталгаажуулах дэд бүтэц

ДЭД БҮТЭЦ, ТЕХНИК ХАНГАМЖИЙН ХҮЧИН ЧАДАЛ

- ISO стандартын шаардлага хангасан
- Хос цахилгаан хангамжтай
- Гал унтраах хийн системтэй (FM200 зэрэг)
- 24/7 хамгаалалт, хяналттай
- Олон чиглэлийн интернэт холболттой
- Хөргөлтийн өндөр хүчин чадалтай системтэй
- Эрчим хүчний эх үүсвэрийн хослол (UPS, Generator)
- Өндөр нөөцлөлтийн багтаамжтай сүлжээ
- Tier III түвшин

ТЕХНОЛОГИЙН ЧИГ ХАНДЛАГА

- Cloud computing (Public, Private, Hybrid cloud)
- Container-based технологи (Kubernetes, Docker)
- Big Data платформууд
- Өгөгдлийн нөөцлөлт, сэргээх автоматжуулалт
- Zero Trust Security
- Green Data Center (Эрчим хүчний хэмнэлттэй шийдлүүд)

АЮУЛГҮЙ БАЙДАЛ, НУУЦЛАЛ

- Өндөр төвшний нэвтрэх хяналт (Access control)
- Биометр хандалт (Fingerprint, Face ID)
- CCTV хяналт
- DDoS хамгаалалт
- IDS/IPS системүүд
- Firewall болон сүлжээний олон давхар хамгаалалт
- Төрийн мэдээллийн аюулгүй байдлын стандартыг чанд мөрддөг (ISO/IEC 27001 гэх мэт)

ХЭРЭГЛЭГЧИД

- Төрийн яам, агентлагууд
- Орон нутгийн засаг захиргаа

- Төрийн өмчит байгууллагууд
- Цахим засаглалын үйлчилгээ үзүүлэгчид
- Хууль сахиулах байгууллагууд
- Зарим төрлийн хувийн хэвшлийн байгууллагууд (төртэй хамтран ажилладаг)

Манай орны хувьд Үндэсний дата төв гэсэн төрийн мэдлийн бэлэн найдвартай дэд бүтэц, бэлтгэгдсэн туршлагатай баг байгаа тул их өгөгдлийн санг энэхүү дэд бүтэц дээрээ тулгуурлан хөгжүүлэх нь зардлын хувьд хэмнэлттэй, үйл ажиллагааны хувьд оновчтой, хэрэгжүүлэх хугацаа багатай гэж үзэж байна.

Энэ тохиолдолд их өгөгдөлд ашиглагдаж буй орчныг өргөтгөх, шаардлагатай хөрөнгө оруулалтыг дараах байдлаар хийх нь оновчтой юм.

- Сервер, орчны хүчин чадлыг нэмэгдүүлэх
- Шаардлагатай системүүдийн лицензүүдийг сунгах, шинээр авах
- Хадгалах (Storage) хүчин чадлыг нэмэгдүүлэх
- Мэргэжилтнүүдийг сургах
- Их өгөгдлийн санд шинээр байгууллагуудыг нэмэх, энэ чиглэлээр хөгжүүлэлт, тохиргоо, шийдлүүдийг хийх
- Их өгөгдлийн хэрэглээг нэмэгдүүлэх, тайлан, дашбоард, API хөгжүүлэх, OpenDate.gov.mn системийн дараагийн хувилбарыг хөгжүүлэх
- Data as a Service буюу өгөгдлийн лаборатори байгуулах, эрдэмтэн судлаачдад судалгаа, боловсруулалт хийх боломжийг нэмэгдүүлэх гэж үзлээ.

СУРГАЛТ

Ерөнхий сургалт (152 байгууллага) зохион байгуулах. Энэхүү сургалтын зорилго нь Их өгөгдлийн талаарх үндсэн ойлголтыг өгөх, төрийн байгууллагуудыг их өгөгдлийн санд нэгдэх бэлэн байдлыг нэмэгдүүлэх, байгууллагуудын дотоодын өгөгдлийн засаглал, ангилал кодын стандартуудыг хэрэгжүүлэхэд дэмжлэг болох.

- Сэдэв: Ерөнхий танилцуулга, үндсэн ойлголтууд
- Давтамж: 1 удаа
- Оролцогчид: 152 байгууллагаас дунджаар 2 хүн = 304 хүн
- Төсөв: 40,000 USD, Танхим, тоног төхөөрөмж түрээс, материалын зардал

Гүнзгий сургалт (10 шинэ байгууллага). Их өгөгдлийн санд шинээр нэгдэх 10 байгууллагын хүрээнд илүү гүнзгийрүүлсэн сургалт хийнэ. Энэ сургалтаар Их өгөгдлийн санд өгөгдөл нийлүүлэх, өгөгдлийн засаглалыг хэрэгжүүлэх, эргээд их өгөгдлийн сангаас тайлан хүснэгт боловсруулах, үйл ажиллагаандаа их өгөгдлийг ашиглах ур чадвар олгох зорилготой 4 удаагийн сургалт зохион байгуулна гэж тооцлоо.

- Сэдэв: ETL, Data Governance, Dashboard хөгжүүлэлт
- Давтамж: 4 (салбар сургалт)
- Оролцогчид: Тухайн жилд нэмэгдэж буй 10 байгууллагын 50-100 мэргэжилтэн.
- Төсөв: 40,000 USD, Танхим, сургалтын материал

ДАТА ЛАБОРАТОРИ

Дата лаборатори нь нээлттэй өгөгдлийн үнэ цэнэ, хэрэглээг дэмжих, хүний нөөцийн чадавхыг өгөх, бизнесийн байгууллага, иргэдийг бодит тоон орчинд судалгаа шинжилгээ хийх боломжийг олгодгоороо ихээхэн чухал юм. Дата лабораторийн үйл ажиллагаа нь бүрэн хаалттай орчин байх бөгөөд судлаачид, байгууллагууд лабораторид ирж ажиллана, түүхий өгөгдөл гарахгүй, data security хангагдана, судлаач зөвхөн Зөвхөн эцсийн үр дүнгээ гадагш авч гарна, машины сургалт, дата анализ хийх боломжтой байх юм. Төсвийг нэг сервер, 10 судлаачийн компьютертэй бие даасан орчин байхаар төсөвлөлөө.

ТОНОГ ТӨХӨӨРӨМЖ / СИСТЕМ	ҮЗҮҮЛЭЛТ, ТАЙЛБАР	НЭГЖ ҮНЭ (USD)	ТОО	НИЙТ ҮНЭ (USD)
Workstation PC	Xeon W CPU, 64–128GB RAM, Nvidia GPU 16–24GB, 2TB SSD	6,500	10	65,000
Сүлжээний тоноглол	Cisco Switch, Firewall	7,000	1 багц	7,000
Server for Lab Data	HPE DL360 / 512GB RAM	20,000	1	20,000
Storage for Lab	NetApp 10TB Flash Storage	35,000	1	35,000
BI / Analytics Software	Tableau Desktop (10 лиценз)	24,000	1 багц	24,000
Data Science Toolkits	Anaconda Enterprise, Python, RStudio Server (opensource болон лицензтэй хослол)	5,000	-	5,000
Security & Access Control	LDAP integration, audit logs	4,000	-	4,000
Хөргөлт, тавилга	Стандарт тохижилт, ширээ, сандал	5,000	-	5,000
НИЙТ				165,000

Сургалт болон дата лабораторийг дээр байдлаар төсөвлөвөл В ХУВИЛБАР нь дараах байдлаар төсөвлөгдөж байна.

№	ҮЙЛ АЖИЛЛАГАА	ТОО ШИРХЭГ / ТАЙЛБАР	НИЙТ (USD)
1	Серверийн хүчин чадал нэмэгдүүлэх	HPE DL380 Gen11 – 4 ширхэг	92,000
2	Storage өргөтгөх	NetApp AFF A150 – 1 ширхэг (20TB)	110,000
3	Лиценз сунгах, программ хангамж	Cloudera, Tableau, Backup licenses	100,000
4	Сургалт	Ерөнхий сургалт (152 байгууллага), 10 байгууллагын гүнзгийрүүлсэн сургалт	80,000

5	Олон нийт, сурталчилгаа	Видео контент, ТВ, радио, олон нийтийн арга хэмжээ	80,000
6	Тайлан, дашбоард хөгжүүлэлт	Backend, Dashboard, ETL, Open Data Portal	120,000
7	Шинээр байгууллага нэмэх интеграц	10 байгууллагын өгөгдөл нэгтгэх үйл ажиллагаа	100,000
8	Data Lab байгуулах	Судлаачид ажиллах орчин, тоног төхөөрөмж	165,000
9	Technical Documentation	Архитектур, Config бичиглэл	8,000
10	DR (Disaster Recovery) Planning	DR тест, Runbook	5,000
НИЙТ			860,000

Их өгөгдлийн дата төвийг байгуулахдаа Үндэсний дата төвийг суурь болгох өргөтөх байдлаар үйл ажиллагаа явуулах нь оновчтой гэж үзэж байна. Мөн Их өгөгдлийн хэрэглээг нэмэгдүүлэх, иргэд олон нийтэд сурталчлан таниулах үйл ажиллагааны хүрээнд Дата лаборатори байгуулах, Opendata.gov.mn н их өгөгдөлд суурилсан шинэ хувилбарыг хөгжүүлэх нь зүйтэй юм.

Их өгөгдлийн санд өгөгдөл нийлүүлж буй байгууллагууд нь эргээд их өгөгдлийн үндсэн хэрэглэгч бол их өгөгдлийн үр дүнг үйл ажиллагаандаа нэвтрүүлэгч болох ёстой. Иймд шинээр өгөгдөл нийлүүлэгч байгууллага нь өөрийн нийлүүлэх өгөгдлийг тодорхойлохын зэрэгцээгээр их өгөгдлөөс авах өгөгдлийн жагсаалт, хүсэлтийг бэлтгэн хөгжүүлэлтийн багт захиалга өгдөг байна.

Дата лаборатори байгуулах болон opendata.gov.mn болон түүнтэй дүйцэхүйц системүүдийг дотоодын систем хөгжүүлэгч байгууллагуудаар хийлгэх, тэдний боловсруулсан системүүдийг үйл ажиллагаандаа ашиглах нь системийн шинэчлэлт, зардлын хэмнэлт болон салбараа дэмжих чиглэлд чухал дэмжлэг болно.

ДҮГНЭЛТ

Их өгөгдлийн сангийн амжилттай ажиллагаанд зөвхөн техник, технологи бус харин ур чадварлаг хүний нөөц хамгийн чухал хүчин зүйл болохыг энэ бүлэгт тусгаж өглөө. Өгөгдлийн шинжлэх ухаан, аналитик, мэдээллийн менежмент, аюулгүй байдлын чиглэлээр мэргэжилтнүүдийг бэлтгэх нь үндсэн зорилт бөгөөд Python, R, SQL, Hadoop, Spark зэрэг технологиос гадна өгөгдлийг цэвэрлэх, боловсруулалт хийх, тайлан боловсруулах ур чадвар чухал шаардлагатай юм. Их өгөгдлийн дата төв байгуулах, хөгжүүлэх төсвийг гурван хувилбараар санал болгож байна:

А ХУВИЛБАР

- Анхны хөрөнгө оруулалт – \$1.35 сая
 - Hardware: \$536,000
 - Software: \$600,000
 - Cloud: \$67,700
 - Сургалт: \$150,000
- 5 жилийн урсгал зардал – \$3.82 сая
- Нийт өртөг (5 жилд) – \$5.17 сая
- Зардал багатай боловч хүчин чадал, өргөтгөх боломж хязгаарлагдмал.

Б ХУВИЛБАР

- Анхны хөрөнгө оруулалт – \$2.98 сая
 - Hardware: \$1.14 сая
 - Software: \$1.30 сая
 - Cloud: \$175,500
 - Сургалт: \$360,000
- 5 жилийн урсгал зардал – \$8.43 сая
- Нийт өртөг (5 жилд) – \$11.41 сая
- Дээд зэргийн тоног төхөөрөмж, лиценз, өндөр хүчин чадалтай системийг төлөвлөсөн. Ирээдүйн өгөгдлийн өсөлт, нарийн аналитик, AI/ML ачааллыг даах чадвартай.

Бусад улс орны түршлага(манай оронтой ойролцоо эдийн засаг, хөгжлийн түвшинтэй улс)

- Казахстан – 45 сая USD (Tier III, олон төрлийн төрийн болон стартапп үйлчилгээг дэмждэг)
- Гүрж – 25 сая USD (Төрийн өгөгдлийн солилцооны үндэсний платформ)
- Узбекистан – 55 сая USD (Huawei-тай хамтарсан, AI/Big Data нэвтрүүлсэн)

В ХУВИЛБАР

- Үндэсний дата төвийг суурь болгох:
 - Tier III төв, өндөр найдвартай дэд бүтэцтэй.
 - Зардлын хувьд хэмнэлттэй, хэрэгжүүлэх хугацаа богино.
- Дотоод нөөцөө ашиглах:
 - Үндэсний дата төвийн баг, системийн чадавх ашиглах.
 - Олон нийтэд сургалт, сурталчилгаа зохион байгуулах.
- Төсөв – \$860,000
 - Сервер, storage өргөтгөл.
 - Шинэ лиценз, программ хангамж.
 - Data Lab байгуулах (\$165,000).
 - Opendata порталын шинэчлэл.
 - Шинээр нэгдэх байгууллагуудад интеграц.
 - Сургалт, олон нийтийн арга хэмжээ.

В хувилбарын давуу талууд:

- Одоо байгаа дэд бүтэц ашиглан зардал хэмнэх
- Хурдан хугацаанд хэрэгжүүлэх боломжтой
- Дотоод систем хөгжүүлэгчдийг дэмжих
- Олон нийтийн ойлголт, оролцоог нэмэгдүүлэх
- Data Lab-аар дамжуулан шинжлэх ухаан, инновацыг дэмжих.

Монгол Улсын хувьд үндэсний дата төвийг түшиглэн В хувилбарыг сонгох нь эдийн засгийн хувьд оновчтой, хурдан хэрэгжих боломжтой бөгөөд урт хугацаанд их өгөгдлийн эко системийг тогтвортой хөгжүүлэх үндэс суурь болно. Мөн Data Lab байгуулалт, олон нийтэд сурталчлан таниулах ажлууд нь өгөгдлийн хэрэглээг өргөжүүлэх, инновац, судалгааг дэмжих чиглэлд чухал шинэлэг алхам юм.

ИХ ӨГӨГДЛИЙН САНГИЙН АЖЛЫН ДААЛГАВАР

Үндэсний хэмжээний их өгөгдлийн сан үүсгэн бүрдүүлэх, хадгалах, ашиглах үйл явц бүхий системийн архитектурыг дараах давхаргуудтай байна гэж үзлээ.

1. Data Sources Layer (Өгөгдлийн эх үүсвэрүүд)

ЭХ СУРВАЛЖ	ӨГӨГДЛИЙН ТӨРӨЛ
Төрийн мэдээллийн сангууд (152+)	Structured Data (SQL)
IoT төхөөрөмжүүд	Streaming Data
Баримт бичгийн сангууд	Semi-structured (XML, JSON)
Фото/Видео архив	Unstructured Data
Social Media, Web Logs	Unstructured/Streaming

2. Data Ingestion Layer (Өгөгдөл татах давхарга)

- ETL / ELT системүүд:
 - Informatica PowerCenter
 - Talend Data Integration
 - Apache NiFi
 - n8n (workflow automation)
- Streaming ingestion:
 - Apache Kafka
 - Flink
 - Spark Streaming
- Data validation, cleansing
- Security checks during ingestion

3. Staging Layer (Өгөгдөл хүлээн авах|хадгалах давхарга)

- Raw data lake
- Initial schema mapping
- Temporary storage for data transformations

4. Data Storage Layer (Өгөгдөл хадгалах давхарга)

- Data Warehouse
 - Structured data for reporting, BI
 - Technologies:
 - Oracle Exadata
 - PostgreSQL
 - Snowflake
- Data Lake
 - Huge volume of mixed data (semi-structured, unstructured)
 - Technologies:
 - Hadoop HDFS

- Cloudera CDP
- Amazon S3-compatible storage
- MinIO
- Cold Storage (Archive)
 - Long-term cheap storage
 - Amazon Glacier
 - Backblaze B2

5. Data processing layer(өгөгдөл боловсруулах давхарга)

- Batch Processing:
 - Spark, Hive
- Real-Time Processing:
 - Kafka Streams
 - Druid
- Machine Learning Pipelines:
 - MLflow
 - TensorFlow
 - Kubeflow

6. Data catalog, metadata layer(метадата болон дата каталогийн давхарга)

- Collibra Data Catalog
- Apache Atlas
- Data lineage
- Data classification (PII, non-PII)
- Business Glossary
- Access rights for datasets

7. Data governance, security layer (өгөгдлийн засаглал, нууцлал аюулгүй байдлын давхарга)

- Access Control (RBAC, ABAC)
- Encryption at rest & in transit
- Audit logging
- Hashing / pseudonymization for personal IDs
- Data exchange rules
- Tokenization for unified ID mapping

8. API & data sharing layer (өгөгдөл тархаах, хуваалцах давхарга)

- RESTful API services
- GraphQL for complex queries
- Open Data portals
- API Management (Kong, Apigee)
- JSON/XML/CSV output formats

9. User access & visualization layer (их өгөгдлийн хэрэглээний давхарга)

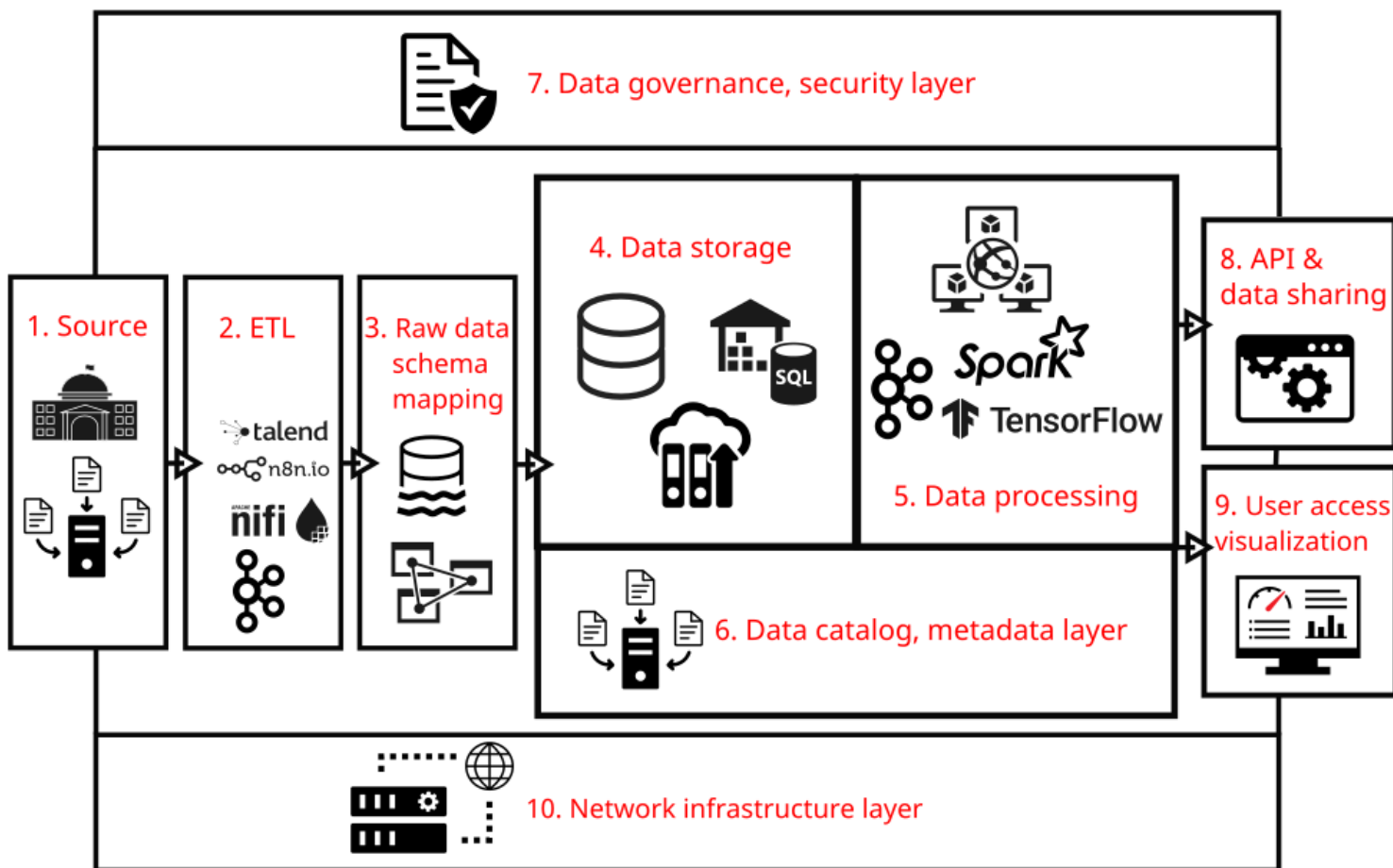
- BI Reporting:

- Power BI
 - Tableau
 - Superset
- Dashboards
- Data sandbox for researchers
- Data portals for citizens
- Хүр, emongolia...

10. Network infrastructure layer (дэд бүтцийн давхарга)

- Data Center Tier III or IV
- MPLS / VPN connectivity to ministries
- SD-WAN
- High availability routing
- Redundant paths
- Firewall, WAF protection
- NOC/SOC centers

Их өгөгдлийн сангийн дэлгэрэнгүй зураглал



НЭВТРҮҮЛЭЛТИЙН ТӨЛӨВЛӨГӨӨ

№	ҮЕ ШАТ	ҮЙЛ АЖИЛЛАГАА
I	УРЬДЧИЛСАН БЭЛТГЭЛ ҮЕ ШАТ	
1	Төслийн удирдлага байгуулах	PMO байгуулах, санхүү, худалдан авалтын зөвлөлийг томилох
2	Оролцогч талууд тодорхойлох	Байгууллагуудын өгөгдлийн төрөл, хэмжээ, бэлэн байдлыг судлах
II	ХЭРЭГЖИЛТИЙН ЭХНИЙ ШАТ	
3	Техникийн шаардлага тодорхойлох	Data volume, AI хэрэгцээ, сүлжээний хүчин чадал тооцох
4	Технологийн архитектур батлах	Hardware, software сонгох, network topology боловсруулах
5	Худалдан авалтын бэлтгэл	TOR бичих, тендер зарлах, vendor сонгох
III	СИСТЕМ БАЙГУУЛАХ ҮЕ ШАТ	
6	Тоног төхөөрөмж суурилуулах	Сервер, storage, switch, firewall байрлуулах, тест хийх
7	Сүлжээний холболт эхлүүлэх	VPN/MPLS холболт тохируулах, redundancy шалгах
8	Data Lake байгуулах	Hadoop/Cloudera суулгах, replication тохируулах
9	ETL pipeline угсрах	ETL tool суурилуулах, mapping layer үүсгэх, validation хийх
IV	ИНТЕГРАЦ БА ТУРШИЛТЫН ШАТ	
10	Эхний холболт хийх байгууллага сонгох	Эхний холболтын байгууллагууд сонгох, тест холболт хийх
11	Өгөгдөл таталт, шинжилгээ хийх	Raw data stage хийх, cleansing, туршилтын тайлан гаргах
12	Аюулгүй байдлын тест хийх	Vulnerability scan, encryption test, access шалгах
V	НЭВТРҮҮЛЭЛТ БА СУРГАЛТ	
13	Сургалт зохион байгуулах	Data center engineer, ETL developer, BI analyst сургалт
14	Production environment руу шилжих	Live data feed эхлүүлэх, portal нээх
15	SLA болон operation тохируулах	SLA бичих, backup/DR бодлого тогтоох, monitoring system тавих
VI	ХӨГЖҮҮЛЭЛТ БА ӨРГӨТГӨЛ	
16	AI, ML ашиглалтыг нэмэгдүүлэх	ML pipeline хөгжүүлэх, use case тодорхойлох
17	Нээлттэй өгөгдөл түгээх	API хөгжүүлэх, open data portal үүсгэх
18	Жил бүрийн hardware/software шинэчлэлт	Hardware upgrade, лиценз renew хийх, cloud scale хийх

Дээрх төлөвлөгөөний дагуу нэвтрүүлэлтийг хийж гүйцэтгэх бөгөөд эхний байгууллагуудад нэвтрүүлэлт хийсэн сайн туршлага дээрээ тулгуурлан их өгөгдлийн санг өргөжүүлэх, дараа дараагийн байгууллагуудыг нэгтгэх үйл ажиллагааг хэрэгжүүлнэ.

ИХ ӨГӨГДЛИЙН САНГ ҮҮСГЭН БҮРДҮҮЛЭХ, ТӨРӨЛЖСӨН МЭДЭЭЛЛИЙН САНГИЙН ӨГӨГДЛИЙГ ИХ ӨГӨГДЛИЙН САНД НЭГТГЭХ, МЭДЭЭЛЭЛ ХАРИУЦАГЧ БАЙГУУЛЛАГУУД БОЛОН ОЛОН НИЙТЭД АШИГЛУУЛАХ НӨХЦӨЛИЙГ БҮРДҮҮЛСЭН ПЛАТФОРМЫН АЖИЛ

Төслийн зорилго

Монгол Улсын төрийн мэдээллийн нээлттэй байдлыг хангах, олон нийт, бизнесийн байгууллага, судлаачид, төрийн байгууллагуудын өгөгдлийн хэрэглээний цар хүрээг өргөтгөх зорилгоор opendata.gov.mn порталыг их өгөгдлийн технологид суурилан шинэчлэх, өгөгдлийн өгөөжийг нэмэгдүүлэх.

Зорилтууд

- Үндэсний дата төвийн дэд бүтэц дээр тулгуурласан их өгөгдлийн платформ бий болгох.
- Олон нийтэд өгөгдөлтэй харьцах интерактив хэрэгслүүд (дашбоард, API, хайлт, шүүлт) хөгжүүлэх.
- Өгөгдлийг хялбар, ойлгомжтой формат (график, visual, AI chatbot) хэлбэрээр түгээх.
- Шинээр өгөгдөл нийлүүлэх байгууллагуудын өгөгдлийг системтэй нийлүүлэх.
- Дата лабораторийн орчныг нээлттэй өгөгдлийн үр ашиг нэмэгдүүлэхэд ашиглах.
- Олон нийтийн сурталчилгаа, сургалтын хөтөлбөр хэрэгжүүлэх.
- Өгөгдлийн аюулгүй байдал, хувийн мэдээлэл хамгаалалтын шаардлагыг хангах.

Хэрэгжүүлэх үндэслэл

- Монгол Улсын хууль, бодлого:
 - Монгол Улсын Үндэсний аюулгүй байдлын үзэл баримтлал
 - Алсын хараа- 2050
 - Монгол Улсын Тогтвортой хөгжлийн зорилт -2030
 - Нээлттэй мэдээллийн ил тод байдлын тухай хууль
 - Хувь хүний мэдээлэл хамгаалах тухай хууль
 - Засгийн газрын 2023 оны 403 дугаар тогтоол
 - Мэдээллийн технологийн үндэсний бодлого
- Үндэсний дата төвийн Tier III төвшний дэд бүтэц бэлэн байгаа нь зардал хэмнэх, хурдан хэрэгжүүлэх давуу талтай.
- Олон нийтэд өгөгдөлд суурилсан шийдвэр гаргалтыг түгээн дэлгэрүүлэх зайлшгүй шаардлага бий.

Өнөөгийн хүрсэн түвшин

- **МОНГОЛ УЛСЫН НЭЭЛТТЭЙ ӨГӨГДЛИЙН ПОРТАЛ (OPENDATA.GOV.MN)**
Монгол Улсын нээлттэй өгөгдлийн нэгдсэн платформ болох opendata.gov.mn нь төрийн байгууллагуудын цуглуулсан өгөгдлийг олон нийтэд ил тод, хүртээмжтэй болгох зорилготой үндэсний хэмжээний цахим систем юм. Энэхүү портал нь өгөгдлийн ил тод байдал, иргэдийн хяналт, дижитал засаглалыг дэмжих үндсэн хэрэгсэл болж байна.

Төрийн байгууллагуудын өдөр тутмын үйл ажиллагааны явцад үүссэн олон төрлийн өгөгдлийг төвлөрүүлэн хадгалах, нийтэд нээлттэй болгох зорилгоор 2015 онд энэхүү платформыг хөгжүүлж эхэлсэн бөгөөд одоо 100 гаруй байгууллагын өгөгдөлтэй болж, нээлттэй өгөгдлийн экосистемийг бий болгоод байна.

ЗОРИЛГО, АЧ ХОЛБОГДОЛ

- Төрийн ил тод байдлыг хангах

Иргэд төрийн байгууллагуудын мэдээллийг чөлөөтэй авах боломжтой болсноор төрийн шийдвэр гаргалт илүү хариуцлагатай, нээлттэй болно.

- Иргэний оролцоог нэмэгдүүлэх

Нээлттэй өгөгдөл нь иргэд, сэтгүүлчид, судлаачид, хувийн хэвшилд төрийн мэдээллийг судлах, дүн шинжилгээ хийх, өөрийн санаачилга бүхий апп, үйлчилгээ хөгжүүлэхэд тусалдаг.

- Дижитал эдийн засгийг дэмжих

Стартап, хувийн хэвшлийн байгууллагууд өгөгдөлд суурилсан шинэ бүтээгдэхүүн, үйлчилгээ, инновацыг хөгжүүлэх боломж бүрддэг.

- Нотолгоонд суурилсан шийдвэр гаргалтыг дэмжих

Судлаач, бодлого боловсруулагчид, төрийн байгууллагууд бодлого боловсруулахдаа бодит өгөгдөлд суурилж, илүү үр дүнтэй шийдвэр гаргадаг.

ПОРТАЛЫН БҮТЭЦ, БОЛОМЖУУД

Порталд нийтэлсэн өгөгдлийн багц (dataset) нь дараах салбарт хамаарна:

- Эдийн засаг, хөрөнгө оруулалт
- Байгаль орчин
- Эрүүл мэнд
- Нийгмийн хамгаалал
- Хууль, эрх зүй
- Боловсрол
- Барилга, хот байгуулалт
- Статистик мэдээлэл
- Газар ашиглалт, геодези, зураг зүй
- Тээвэр, зам тээвэр
- Хүн амын тоо, хөдөлгөөн
- Гадаад худалдаа
- Уул уурхай
- Гадаадын иргэдийн хөдөлгөөн

Техникийн функцүүд

- Өгөгдөл татах боломж
CSV, XLS, JSON, XML зэрэг олон төрлийн форматаар өгөгдлийг татаж авах боломжтой.
- API ашиглах боломж
Машин хооронд өгөгдөл солилцох RESTful API-тэй, ингэснээр

хөгжүүлэгчид автомат өгөгдлийн таталт, шинжилгээ хийх боломжтой.

- Хайлтын систем
Портал дотор түлхүүр үгээр хайлт хийх, салбар болон агентлагаар шүүх боломжтой.
- Харагдах байдал
Зарим өгөгдлийг портал дотор шууд график, хүснэгт, газрын зураг дээр дүрслэх боломжтой.
- Лицензийн мэдээлэл
Өгөгдөл бүрд лицензийн нөхцөл (Creative Commons, нээлттэй ашиглалт гэх мэт) хавсаргасан байдаг.

ОРОЛЦОГЧ БАЙГУУЛЛАГУУД

Opendata.gov.mn-д 100 гаруй төрийн байгууллага өгөгдлөө нийтэлж байна. Үүнд:

- Үндэсний статистикийн хороо
- Архивын ерөнхий газар
- Ашигт малтмал, газрын тосны газар
- Барилга, хот байгуулалтын яам
- Байгаль орчин, аялал жуулчлалын яам
- Гаалийн ерөнхий газар
- Хууль зүй, дотоод хэргийн яам
- Эрүүл мэндийн яам
- Боловсрол, шинжлэх ухааны яам гэх мэт.

Эдгээр байгууллагууд өөр өөр төрлийн өгөгдлийг цуглуулж, боловсруулах бөгөөд түүнийгээ нээлттэй өгөгдлийн порталд байршуулдаг.

• ИХ ӨГӨГДЛИЙН САН БАЙГУУЛАХ ҮЙЛ АЖИЛЛАГАА

2023 онд Засгийн газрын 403р тогтоол батлагдсанаар их өгөгдлийн сан байгуулах үйл ажиллагаанд томоохон түлхэц болсон.

Уг тогтоолоор мэдээлэл хариуцагч байгууллагууд их өгөгдлийн санд өгөгдөл нийлүүлэхдээ харилцагчийн эмзэг мэдээллийг арилгах, дамжуулах өгөгдлийн түлхүүр талбарын нууцлалыг хангах, өгөгдөл солилцоог хэрхэн шийдэх зэрэг асуудлуудад хариулт өгсөн. Ингэснээр төрийн байгууллагуудаас их өгөгдлийн санд мэдээлэл нийлүүлэх үйл ажиллагааг хэрэгжүүлж эхлээд байна.

Өнөөдрийн байдлаар

- ҮДТ-д их өгөгдлийн санг байршуулах дэд бүтцийг бий болгосон
- Байгууллагуудаас түлхүүр талбарын нууцлалыг хангах шийдэл хийгдсэн
- Өгөгдөл дамжуулах үйл ажиллагаа ХҮР системээр дамжуулан хийгдэж байгаа
- Их өгөгдлийн сангийн суурийг тавих ажил хийгдэж Үндэсний статистикийн хороо, Хөдөлмөр халамж үйлчилгээний ерөнхий газар, Нийгмийн даатгалын ерөнхий газраас өрх, иргэн, үйлчилгээний суурь мэдээллийг их өгөгдлийн санд дүрмийн дагуу байршуулаад байна.

Цаашид хийгдэж алхмууд

Их өгөгдлийн сангийн сервер, тоног төхөөрөмж, лиценз болон бусад шаардлагатай дэд бүтцийн шийдлүүд, хөрөнгө оруулалтын зэрэгцээгээр их өгөгдлийн сангийн баяжилтыг нэмэгдүүлэх, их өгөгдлийн сангийн хэрэглээг нэмэгдүүлэх, олон нийтэд зохиулсан өгөгдөлтэй харьцах интерактив хэрэгслүүд (дашбоард, API, хайлт, шүүлт) хөгжүүлэх шаардлага урган гарч байна. Энэхүү зорилтын хүрээнд OpenData.gov.mn системийн дараагийн шинэчилсэн хувилбар хөгжүүлэгдэх шаардлагатай.

Оpendata.gov.mn системийн дараагийн хувилбарт дараах шийдэл, үйл ажиллагаануудыг хэрэгжүүлэх шаардлагатай гэж үзлээ:

1. BIG DATA ШИЙДЛҮҮДТЭЙ УЯЛДУУЛАХ

Одоогийн нөхцөл:

- Opendata.gov.mn-д байгаа өгөгдөл харьцангуй жижиг хэмжээтэй, csv эсвэл excel файлууд хэлбэрээр байна.
- Их өгөгдлийн сан (Data Lake, Data Warehouse) болон хүлээж авах API, real-time өгөгдлийн урсгалыг дэмжихгүй.

Шинээр хийгдэх ажлууд:

- Их өгөгдлийн сангийн архитектуртай холболт хийх.
- Big Data Query боломж нэмэх
- Real-time сенсороос өгөгдөл татах, түгээх

2. МЕТА ӨГӨГДЛИЙН СТАНДАРТЧИЛАЛ

Одоогийн нөхцөл:

- Portal дахь датасет бүрийн meta data харилцан адилгүй, бүрэн бус.
- Өгөгдлийн засаглал бэхжээгүй тул хайлт, өгөгдлийн утга ойлгомж тодорхой бус

Шинээр хийгдэх ажлууд:

- ISO 19115 болон DCAT метадата стандартууд нэвтрүүлэх
- Collibra зэрэг дата катологи системтэй нэвтрүүлэх
- Хайлтын үр дүнг мета өгөгдөл дээр суурилсан байдлаар нарийвчлах

3. REAL-TIME ӨГӨГДЛИЙН ТҮГЭЭЛТИЙГ НЭМЭГДҮҮЛЭХ

Одоогийн нөхцөл:

- Portal-д нийтэлсэн өгөгдөл ихэвчлэн static файлууд байна.
- Иргэд, байгууллага real-time өгөгдлийг сонирхдог ч тэр нь байхгүй.

Шинээр хийгдэх ажлууд:

- Real-time data API хөгжүүлэх(filter, aggregation, paging).
- Түргэн өөрчлөгддөг(үнэ ханш, цаг агаар, замын түгжрэл, осол гэх мэт) өгөгдлийн төрлүүдийг real-time болгох
- Хэрэглэгчийн интерактив dashboard хөгжүүлэх

4. ӨГӨГДЛИЙГ ШИНЭЭР ДҮРСЛЭХ

Одоогийн нөхцөл:

- Portal дотор visualizer хязгаарлагдмал, ихэнх өгөгдөл хүснэгт хэлбэрээр л харагддаг.
- Иргэд, болон шийдвэр гаргагчид график болон бусад дүрслэлээр мэдээлэл ойлгоход хялбар болгох шаардлагатай.

Шинээр хийгдэх ажлууд:

- Интерактив дашбоард хөгжүүлэх(Opendata.gov.mn портал дотор BI iframe embed хийх)

- Газрын зураг дээр өгөгдөл харагдуулах
- Custom Chart хөгжүүлэх

5. ИРГЭДИЙН ХЭРЭГЛЭЭНИЙ ТҮВШИНГ НЭМЭГДҮҮЛЭХ

Одоогийн нөхцөл:

- Иргэд, байгууллагууд opendata.gov.mn-г ашиглах мэдлэг багатай.
- Бизнес, стартап компаниуд нээлттэй өгөгдлийн үнэ цэнийг сайн ойлгоогүй.

Шинээр хийгдэх ажлууд:

- Нээлттэй өгөгдлийг нийтэд сурталчлах, сургалт зохион байгуулах
- Нээлттэй өгөгдлийн hackathon зохион байгуулах
- Иргэдэд зориулсан хэрэглээний кейс нийтлэх
- PR, олон нийтийн мэдээллийн суваг ашиглах

OPENDATA.GOV.MN, ШИНЭЧИЛСЭН ХУВИЛБАРЫН СИСТЕМ ХӨГЖҮҮЛЭЛТ

Дээр дурдсан шинээр хийгдэх ажлуудтай холбогдуулан системд дараах модулиудыг хөгжүүлэх шаардлагатай гэж үзэж байна:

- Одоогийн opendata.gov.mn системд хийгдэж байгаа үйлдлүүд буюу хуулийн хүрээнд төрийн байгууллагуудаас системд өгөгдөл нийлүүлэх, удирдах, олон төрлөөр тархах, API үйлчилгээнүүд зэрэг.
- Их өгөгдлийн сантай холбогдон ажиллах, шинээр тайлан хүснэгт болон , график дашбоардууд гаргах
- Их өгөгдлийн сангаас хэрэглэгчдэд зориулсан сервисүүд(API) нэмж хөгжүүлэгдэх
- Метадата болон их өгөгдлийн катологи хөгжүүлэх, нэвтрүүлэх

Их өгөгдлийн санд мэдээлэл хариуцагч байгууллагуудыг нэгтгэх, шинэ шинэ өгөгдлөөр санг баяжуулахын зэрэгцээгээр их өгөгдлийн санд цугларсан мэдээллийн катологи, метадата системийг хөгжүүлэх, үйл ажиллагаанд нэвтрүүлэх, эргээд олон нийтэд нээлттэй болгох их өгөгдлийн сангийн хэрэглээг нэмэгдүүлэх, олон нийтэд их өгөгдлийн санг сурталчлан таниулахад чухал алхам болно. Иймд дараагийн бүлэг хөгжүүлэлт, нэвтрүүлэлт нь энэхүү метадата систем болон их өгөгдлийн сангийн катологи байх болно.

СИСТЕМИЙН ХЭРЭГЦЭЭ ШААРДЛАГЫГ ТОДОРХОЙЛОХ

Их өгөгдлийн санд цугларсан өгөгдлүүд, катологиудыг үндэслэн олон нийт, захиалагч талуудын хэрэгцээ шаардлагад нийцсэн тайлан хүснэгт, график болон сервисүүдийг хөгжүүлэх шаардлагатайг өмнө дурдсан. Тэгвэл эдгээр гаралтуудын хэрэгцээ шаардлагыг хэрхэн тодорхойлж захиалгыг хөгжүүлэгчдийн багт өгөх вэ гэдэг асуудал тавигдана.

Санал:

Их өгөгдлийн санд мэдээлэл нийлүүлсэн байгууллагуудын төлөөллийг оролцуулсан Өгөгдлийн зөвлөлийг (Дата менежментийн баг) зохион байгуулна. Тухайн зөвлөл нь их өгөгдлийн сангаас ямар ямар хүснэгт тайлан, график, сервисүүдийг хөгжүүлж нийтэд болон шаардлагатай албан хэрэгцээнд ашиглаж болохыг тодорхойлж тэрхүү хөгжүүлэлтийн захиалгыг боловсруулахад дэмжлэг үзүүлж ажиллана. Аливаа захиалгыг шинжээчдийн баг хүлээн авч шаардлагатай судалгааг хийж гүйцэтгэн бодлогын даалгавар боловсруулан систем хөгжүүлэгчдийн багт шилжүүлэх дарааллаар тайлан дашбоард, сервисийн хөгжүүлэлт хийгдэнэ.

Их өгөгдлийн сангийн хөгжлийн эхэн үе буюу цөөн тооны мэдээлэл хариуцагчдын өгөгдлийг нэгтгэсэн үед тухайн өгөгдлөө сайн мэдэх төлөөллийг оролцуулах нь илүү үр дүнтэй байдаг байна. Зарим тохиолдолд гол захиалагчаар статистикийн байгууллага оролцсон туршлага байдаг.(шинэ зеланд, казакстан)

Энэхүү захиалга буюу хэрэгцээ шаардлага тодорхойлох, даалгавар боловсруулах баг нь цаашид бие даасан байгууллагын дотоод бүтэц болж их өгөгдлийн сангаас шинээр стандарт хүснэгт боловсруулах, сервис гаргах зэрэг үйл ажиллагаануудын захиалга, бодлогын даалгаврыг боловсруулах шинжээчдийн багтайгаар үйл ажиллагаагаа явуулах шаардлагатай. Мөн энэхүү баг өгөгдлийн катологи, метадата системүүдийн үйл ажиллагааг хариуцах нь оновчтой.

СИСТЕМ ХӨГЖҮҮЛЭЛТИЙГ БАГИЙГ СОНГОХ

Их өгөгдлийн сангийн хамрах хүрээ, түүний өргөжилтийн бодлого зэрэгтэй уялдуулан хөгжүүлэлтийн багийг оновчтой сонгох нь ач холбогдолтой.

- Emongolia академийн хөгжүүлэлтийн баг
- Дотоодын программ хангамжийн компаниуд
- Гадаадын болон хамтарсан программ хангамжийн компаниуд энэхүү төсөл дээр ажиллах боломжтой гэж үзэж байна.

Санал:

Их өгөгдлийн сангийн суурь үе шатанд (эхний жил буюу 10 хүртэлх төрийн байгууллагуудын өгөгдлийг нэгтгэсэн байх) Emongolia академи, эсвэл opendata.gov.mn системийг хөгжүүлэгч компанийг сонгон Opendata.gov.mn платформын шинэчилсэн эхний хувилбарыг боловсруулах нь оновчтой. Цаашид хөгжүүлэлтийн багийг өргөжүүлэх, нээлттэй сонгон шалгаруулалтаар гүйцэтгэгчийг сонгох зэргээр дараа дараагийн хувилбаруудыг хөгжүүлж болох юм.

Хөгжүүлэлтийн үйл ажиллагааны төлөвлөгөө:

1. Дата менежментийн багийн бүрдүүлэх(их өгөгдлийн санд өгөгдөл нийлүүлсэн мэдээлэл хариуцагч байгууллагуудын төлөөллийг оролцуулах)
2. Дата менежментийн багаас их өгөгдлийн санд шинээр өгөгдөл, түүнийг нийлүүлэх байгууллагуудыг сонгох, эрэмбэлэх
3. Хамрах хүрээг тодорхойлох
 - Шинээр нэмэгдэх 10 төрийн байгууллагын мэдээллийн сан, үзүүлэлтүүд
 - IoT төхөөрөмжүүдээс ирэх өгөгдөл
 - Фото, видео, текст, аудио өгөгдөл
 - Open Data порталтай нийцэх өгөгдөл түгээх систем
4. Гүйцэтгэх ажлуудын дэлгэрэнгүй

4.1 ХЯЗГААРЛАЛТТАЙ МЭДЭЭЛЛИЙН ХАМГААЛАЛТ

- Хувь хүнийг таних боломжгүй болгох
- Хуулийн нийцлийг шалгах:
 - Хувь хүний мэдээлэл хамгаалах тухай хууль
 - GDPR, ISO/IEC 27001
- Central logging хийх:
 - Хэн, хэзээ, ямар өгөгдөлд хандсан бүртгэлтэй байх

4.2 ӨГӨГДӨЛ ТАТАХ БА СОЛИЛЦООНЫ СИСТЕМ

- ХУР системийн өгөгдөл солилцооны протокол ашиглах
- Байгууллага бүрийн өгөгдлийг standardized format руу хөрвүүлэх:
 - JSON
 - XML
 - CSV
- ETL/ELT систем ашиглан өгөгдлийг дамжуулах:
 - Batch
 - Real-time streaming
- Хяналтын процесс бий болгох:
 - Data validation
 - Error reporting
 - Retry mechanism

4.3 ХАДГАЛАХ ШИЙДЭЛ

Data Lake

- Hadoop, Cloudera
- Unstructured өгөгдөл хадгалах:
 - Лог файл
- Metadata catalog нэвтрүүлэх:
 - Apache Atlas
 - Collibra
- Data Lake security:
 - Encryption-at-rest
 - ACL (Access Control Lists)
 - Kerberos authentication

Data Warehouse

- Structured өгөгдлийн цэвэр хувилбар хадгалах
- Тайлан, BI, ML шинжилгээнд зориулсан optimized storage
- Star, Snowflake schema design хийнэ
- Columnar storage систем ашиглах:
 - Oracle Exadata
 - Snowflake
 - ClickHouse
- Query performance tuning хийх

4.4 УРЬДЧИЛАН БОЛОВСРУУЛАХ, ДҮРСЛЭХ ҮЕ ШАТ

- ETL pipeline хөгжүүлнэ:
 - Data cleansing
 - Null handling
 - Код стандартчилал
 - Entity matching
- Data lineage хөтлөх:
 - Flow diagrams
 - Automated lineage capture
- BI хэрэгслүүд:
 - Power BI
 - Tableau
 - Apache Superset
- Real-time dashboard:
 - Apache Druid
 - Kafka Streams

4.5 МЭДЭЭЛЛИЙН САНГИЙН УЯЛДАА, ИНТЕГРАЦ

- Unified Metadata Dictionary боловсруулна:
 - Entity нэр
 - Атрибут нэр
 - Өгөгдлийн төрөл
 - Стандарт код
 - Business Glossary
- Mapping table үүсгэнэ:
 - Байгууллагын код ↔ Unified код
- ERD, Logical Data Model зураг гаргана:
 - ER/Studio
 - dbdiagram.io
 - Lucidchart
- Data Quality rules тогтооно:
 - Consistency check
 - Duplicate elimination
 - Referential integrity

4.6 OPENDATA.GOV.MN СИСТЕМИЙН ШИНЭЧЛЭЛТ

- Их өгөгдлийн сангаас гарах үзүүлэлтүүд, хүснэгт, сервисүүдийн жагсаалт, хөгжүүлэлтийн бодлогын даалгаврыг боловсруулах
- Систем хөгжүүлэх багийг сонгох шалгаруулах
- Систем хөгжүүлэлтийг гүйцэтгүүлэх, явцад хяналт тавих, хүлээн авах

4.7 ИХ ӨГӨГДЛИЙН САНГИЙН АШИГЛАЛТ

- Citizen Dashboard:
 - Нээлттэй өгөгдөл дүрслэх
 - Drill-down capability
 - Mobile-friendly UI
- API services:
 - GraphQL
 - RESTful
 - Swagger Documentation
- Data Sandbox:
 - Судлаачдад data subset үзүүлэх орчин
- AI, ML ашиглалт:
 - Predictive modeling
 - Sentiment analysis
 - Forecasting dashboards

НЭГДСЭН ДҮГНЭЛТ

Өнөөдрийн байдлаар төрийн байгууллагууд тус тусын үйл ажиллагааны чиглэлээр иргэд болон байгууллагуудад үзүүлж буй бүтээгдэхүүн, үйлчилгээнийхээ хүрээнд олон төрлийн өгөгдлийн санг бий болоод байна. Хэдийгээр байгууллагуудын үйл ажиллагаа цахимжиж, бүртгэл мэдээллийн үйл явц өргөжиж байгаа хэдий ч бий болсон мэдээллийн сангуудын уялдаа холбоо, ашиглалт, өгөгдлийн хамрах хүрээ, чанар харилцан адилгүй байна. Энэ цаг үед шаардлага хангасан өгөгдөл, мэдээллийг бага зардлаар бий болгох, дахин ашиглах, нарийвчилсан дүн шинжилгээг цаг алдалгүй хийх, нотолгоонд суурилсан шийдвэр гаргах хэрэгцээ, шаардлага өсөн нэмэгдэж байна. Энэхүү хэрэгцээ шаардлагыг хангах Их өгөгдлийн сан байгуулах, түүнд өгөгдөл нийлүүлэх болон өгөгдлийг ашиглах тогтолцоог яаралтай бий болгох шаардлагатай байна. Энэ хүрээнд УИХ мэдээллийн технологийн багц хуулиуд, Засгийн газраас "Их өгөгдөл" тогтоолыг баталснаар төрийн мэдээлэл солилцооны суурь эрх зүйн орчин бүрдэж, E-Mongolia, ХҮР системийн үйл ажиллагаа өргөжиж, төрийн байгууллагуудаас иргэдийг мэдээллээр хангах үйлчилгээний хүрээ тэлж байна. Ийнхүү E-Mongolia, ХҮР системд Төрийн байгууллагуудын мэдээлэл төвлөрч иргэдэд төрийн үйлчилгээг түргэн, шуурхай хүргэх боломж нэмэгдэж байгаа нь сайшаалтай юм.

ӨНӨӨГИЙН ХҮРСЭН ҮР ДҮН

Дээр дурдсан хууль эрхзүйн хүрээнд Их өгөгдлийн сан байгуулах үйл ажиллагаанууд хийгдэж эхлээд байна. Юуны өмнө салбарын дотоод өгөгдлийн сангуудаа байгуулсан, өгөгдөлтэй ажиллах үйл ажиллагааны туршлагатай, хөгжүүлэлтийн багтай, олон тооны өгөгдөлтэй ажилладаг, иргэний болон аж ахуйн нэгж байгууллагуудын өгөгдөл бүхий байгууллагууд болох Үндэсний статистикийн хороо, Хөдөлмөр халамж үйлчилгээний ерөнхий газар, Нийгмийн даатгалын ерөнхий газруудын тодорхой өгөгдлүүдийг нэгтгэх үйл ажиллагаа хийгдээд байна.

Энэхүү анхан шатны үйл ажиллагаа хийгдсэнээр өгөгдлийн нууцлал, өгөгдөл дамжуулах үйл ажиллагаа болон дэд бүтэц, их өгөгдлийн сангийн орчин зэрэг суурь дэд бүтцүүдийн туршилт тохируулга хийгдээд байна.

Өнөөдрийн байдлаар

- ҮДТ-д их өгөгдлийн санг байршуулах дэд бүтцийг бий болгосон
 - Их өгөгдлийн байгуулах анхан шатны үйл ажиллагааг явуулахад шаардлагатай сервер, нөөцийг байгуулсан. Цаашид их өгөгдлийн сангийн хэрэглээ, өргөжилт зэргээс хамааран өргөтгөх боломжтой.
- Байгууллагуудаас түлхүүр талбарын нууцлалыг хангах шийдэл хийгдсэн
 - 403-р тогтоолд заасанчлан өгөгдөл нийлүүлэгч байгууллагууд Их өгөгдлийн санд нийлүүлэх өгөгдлийнхөө нууцлал, аюулгүй байдлыг хангах үйл ажиллагааг хэрэгжүүлж байгаа. Иргэн, байгууллагын эмзэг мэдээллийг устгах, түлхүүр талбарыг нууцлах(encryption) хийх үйл ажиллагааг өгөгдөл дамжуулахаасаа өмнө хийсний дараа их өгөгдлийн санд нийлүүлж байгаа.
- Өгөгдөл дамжуулах үйл ажиллагааг ХҮР системээр дамжуулан хийгдэж байгаа

- Өгөгдөл дамжуулах үйл ажиллагааг дээрх тогтоолд заасанчлан өгөгдлийн нууцлал аюулгүй байдлыг хангасан ХУР системээр дамжуулан үндэсний дата төвд дамжуулж байгаа.
-

Ийнхүү өгөгдөл нийлүүлэгч байгууллага Их өгөгдлийн санд нийлүүлэх өгөгдлийнхөө бэлэн байдлыг хангаж эмзэг мэдээллийг арилгах, нууцлалыг стандартын дагуу кодлох, өгөгдлийг ХУР системээр дамжуулан их өгөгдлийн санд илгээх, их өгөгдлийн санд цугларсан өгөгдлийн харилцан уялдаатай байдал хангагдсан байх туршилтууд амжилттай хийгдээд байна.

Одоо бид цаашид дээрх үйл ажиллагааг хэрхэн өргөтгөх, их өгөгдлийн санд нийлүүлж буй өгөгдлүүдийн харилцан хамаарал, ангилал кодын нийцтэй байдлыг хэрхэн хангах, их өгөгдлийн сангаас ямар бүтээгдэхүүн үйлчилгээг хэрхэн олон нийт, хэрэглэгчдэд хүргэх, их өгөгдлийн санг ямар ямар байгууллагуудын өгөгдлөөр хэрхэн баяжуулж байхыг тодорхойлох нь зүйтэй юм.

СУРГАМЖ

Их өгөгдлийн сан байгуулах үйл ажиллагааг олон улс орнууд амжилттай хэрэгжүүлж байна. Өмнөх судалгааны бүлгүүдэд бид Их өгөгдлийн сан байгуулах үйл ажиллагааг амжилттай хэрэгжүүлж буй улс орнуудын жишээг авч үзсэн. Тэгвэл Их өгөгдлийн сан байгуулах үйл явцад алдаа гарсан, эсвэл зарим тодорхой хүчин зүйл, бэрхшээлийн улмаас хугацаандаа хэрэгжүүлж амжаагүй, зардлын хэтрэлттэй тулгарсан зэрэг төслүүд байгааг анхаарах нь зүйтэй юм.

АНГЛИ – CARE.DATA ТӨСӨЛ (2013–2016)

Зорилго:

- NHS (Үндэсний эрүүл мэндийн алба)-ийн дэргэд өвчтөний мэдээллийг төвлөрүүлсэн care.data их өгөгдлийн сан байгуулах
- Өгөгдлийг судалгаа, бодлого боловсруулахад чиглэлээр ашиглах

Тулгарсан бэрхшээл:

- Иргэдэд хангалттай мэдээлэл өгөөгүй, олон нийтээс дэмжлэг авч чадаагүй (иргэдэд Opt-out механизмыг ойлгуулаагүй. Opt-out гэдэг нь иргэн, хэрэглэгчийн мэдээллийн өөрөө зөвшөөрсөн гэж үзэж шууд цуглуулах боловсруулалт хийх бөгөөд хэрэв иргэн хүсвэл өгөгдлөө ашиглуулахаас татгалзах, эсвэл зөвшөөрөлгүйгээр хамрагдсан байснаа болиулах эрх, арга хэрэгсэл юм).
- Хувийн мэдээлэл хамгаалах талаарх үйл ажиллагаа тодорхой бус.
- Мэдээлэл ашиглах байгууллагуудын мэдээллийн ил тод байдал муу (жишээ нь: хувийн компаниуд өгөгдөл авах эрхтэй байсан).

Сургамж:

- Иргэдийн итгэлцлийг олохгүй бол их өгөгдөл үр дүнгүй.
- Үйл ажиллагааны ил тод байдал маш чухал.

ЭНЭТХЭГ – AADHAAR DATA ТӨВ

Зорилго:

- Aadhaar нь бүх иргэдийн биометрийн мэдээлэл (нүдний солир, хурууны хээ)-ийг агуулсан үндэсний ID систем юм
- Төвлөрсөн их өгөгдлийн сан байгуулах, төрийн олон үйлчилгээг холбох

Тулгарсан бэрхшээл:

- Мэдээллийн аюулгүй байдал алдагдал. Хэд хэдэн тохиолдлоор Aadhaar системээс мэдээлэл задарсан тохиолдол гарсан.
- Үр ашиггүй интеграци. Олон эх сурвалжаас өгөгдөл нэгтгэх үйл ажиллагааг явуулсан бөгөөд зарим үйлчилгээний өгөгдөл татах, холболт хийх ажиллагаа доголдсон.
- Иргэдийн эсэргүүцэлтэй тулгарсан. Нууцлал зөрчигдөж байна гэж үзсэн.

Сургамж:

- Data security & encryption хамгийн түрүүнд тавигдах ёстой.
- Интеграцийн шийдэлд урьдчилсан стандарт, API зохицуулалт шаардлагатай.
- Зөвшөөрлийн систем (consent management) байх хэрэгтэй.

КЕНИ – HUDUMA NAMBA (2020 ОНООС)

Зорилго:

- Үндэсний ID болон нийгмийн үйлчилгээний мэдээллийг нэгтгэх зорилгоор Huduma Namba системийг нэвтрүүлсэн.

Тулгарсан бэрхшээл:

- Хууль эрх зүйн зохицуулалт хангалттай байж чадаагүй.
- Шүүхийн шийдвэрээр их өгөгдлийн сан байгуулах үйл ажиллагааг “Өгөгдөл хамгаалах” хууль хэрэгжээгүй байхад мэдээлэл цуглуулсан гэж үзэж зогсоосон.
- Санхүүжилт, техник хангамж хангалттай байж чадаагүй.

Сургамж:

- Их өгөгдөлтэй холбоотой аливаа төслийн хувьд хууль эрх зүйн зохицуулалтууд шийдэгдсэн байх шаардлагатай.
- “Өгөгдлийн засаглалын” бодлогыг урьдчилж боловсруулсан байх шаардлагатай.

Дээрх туршлагуудаас нэгтгэж дүгнэхэд Их өгөгдлийн сан байгуулах үйл ажиллагаа нь зөвхөн техник, технологийн (систем, тоног төхөөрөмж, сервис болон өгөгдөлтэй ажиллах, дамжуулах) асуудал биш харин удирдлага, хууль эрх зүй, өгөгдлийн чанар, хамтын ажиллагаа, иргэдийн итгэлцэл зэрэг олон хүчин зүйлтэй холбоотой байдаг нь харагдаж байна.

Их өгөгдлийн сан байгуулахад тулгарч болох бэрхшээл

БЭРХШЭЭЛ	СУРГАМЖ
Иргэдтэй хангалттай зөвшилцөөгүй, нэгдсэн ойлголт өгөөгүй байх	Итгэлцэл бүрэн бий болоогүй тохиолдолд өгөгдөл цуглуулах үйл ажиллагаанд хүндрэл, бэрхшээл, эсэргүүцэл бий болдог
Өгөгдлийн нууцлал болон хууль эрх зүйн зохицуулалт муу байх	Хууль эрхзүйн хамгаалал, зохицуулалтгүйгээр өгөгдөл цуглуулах үйл ажиллагаа шууд эсэргүүцэлтэй тулгардаг
Технологид хэт анхаарал тавьж хэрэглэгчийн талаас дутуу бодолцсон байдал	Их өгөгдлийн сангийн гол хэрэглээ нь эргээд иргэдийн амьдралд хэрэг болох үйлчилгээ, бүтээгдэхүүн, шийдэл, шийдвэр болдог. Иймд хүн төвтэй загварчлал зайлшгүй шаардлагатай юм.
Мэдээллийн чанар муу, байгууллага хоорондын нийцтэй байдал хангалтгүй байх	Өгөгдлийн засаглал хэрэгжүүлсэн байдал, байгууллагуудын хамтын ажиллагааны чадвар чухал байдаг
Богино хугацааны эдийн засгийн үр ашигтай байдалд хэт автах	Их өгөгдлийн сангийн үйл ажиллагаа нь урт хугацааны үр дүнтэй, хүний эрхийн тэнцвэртэй байдлыг хангасан байх хэрэгтэй

Их өгөгдлийн сан байгуулах үйл ажиллагааг амжилттай хэрэгжүүлж байгаа, хууль эрхзүй болон өгөгдлийн чанар, байгууллага хоорондын хамтын ажиллагаа болон иргэдийн итгэлцлийг хангах талаар туршлага болохуйц орон болох БНСУ н жишээг авч үзье. БНСУ н хувьд 2020 онд “Three Data Laws”- Өгөгдлийн гурван хууль буюу “Data3 Act” эрх зүйн шинэчлэлийн хүрээнд “privacy” (нийтийн өгөгдлийн нууцлал) болон “utility” (өгөгдөл ашиглалтыг дэмжих) зохицуулалтуудыг тусгаж өгсөн нь Их өгөгдлийн сангийн суурь эрхзүйн баримт болсон.

БНСУ-н Их өгөгдлийн сан байгуулах үйл явц нь хөгжлийн үе шатуудад олон сорилтуудыг даван туулж өнөөдрийн түвшинд хүрчээ.

- 1980-1990: “Framework Act on Informatization Promotion” эрхзүйн баримт бичгийг баталсан нь системтэй e-Government(цахим засаглал) -н суурь үндэс болсон
- 2000-2010: “Promotion of Digitalization of Administrative Work for E-Government Realization Act” хуулиудын хүрээнд цахим шилжилтийг амжилттай хэрэгжүүлсэн
- 2013: Нээлттэй өгөгдлийн (Open Data) бодлого хэрэгжиж эхэлсэн нь бизнесийн байгууллагууд болон иргэдэд төрийн өгөгдлийн хүртээмжийг нэмэгдүүлэх, өгөгдөлд суурилсан бизнесийн чиг хандлагыг дэмжихэд түлхэц болсон
- 2019 оноос “data economy” үзэл санааны суурь тавигдаж эхэлсэн ба, 2020 онд Data-3 Act / “Three Data Laws” хууль батлагдсан нь Их өгөгдлийг бүрдүүлэх, ашиглах эрхзүйн зохицуулалт болсон.

Эдгээр эрхзүйн зохицуулалтуудыг бий болгох, байсан хууль тогтоомжуудыг тогтмол цаг үетэйгээ нийцүүлэн сайжруулсаар ирсэн нь өгөгдлийг нээлттэй болгох, өгөгдөлд суурилсан шийдлүүд, бизнесүүд хөгжих таатай хөрс суурь

болсон гэж үздэг. Энэ ч үүднээс БНСУ-д өгөгдөлд суурилсан төслүүд амжилттай хэрэгжсээр байна.

- E-Government систем 2000 оноос эхлэн хөгжиж өнөөдөр БНСУ дэлхийн өндөр түвшний дижитал засаглалтай улсуудын жишээ болсон.
- Open Data Strategy: 2013–2019 онуудад төрийн бүх шатанд өгөгдлийг нээлттэй болгох бодлогыг амжилттай хэрэгжүүлснээр өгөгдөлд суурилсан системүүд, аппууд ихээр бий болж тэр хирээр иргэд үзүүлж буй бүтээгдэхүүн, үйлчилгээний хүртээмж нэмэгджээ.
- BigData market: 2019 онд “data economy” үзэл санааны эхлэл тавигдаж улмаар 2025 он гэхэд AI дата төв нь экспорт, хөрөнгө оруулагчдын анхаарлын төв болоод байна.(Жишээ нь SK+AWS хамтарсан 5 тэрбум ам.долларын дата төвийг Улсан хотод барих төсөл хэрэгжиж эхлээд байна.)
- Smart City, AI хэрэглээ: Сөүлийн хотын бодит цагийн горимд өгөгдөл цуглуулах системүүд ухаалаг хот шийдлийн үлгэр жишээ болон хөгжиж байна.

ӨНӨӨГИЙН БАЙДАЛ: ИХ ӨГӨГДЛИЙН САНГИЙН ЭРХ ЗҮЙН ОРЧИН

Хууль эрхзүйн зохицуулалт

- Personal Information Protection Act (PIPA) – 2011 онд батлагдсан, 2020 болон 2023 онд өөрчлөлт орсон. Их өгөгдлийн санд иргэн, байгууллагын хувийн мэдээллийг цуглуулах, ашиглах, дамжуулах, хамгаалах бүх үйл ажиллагааны хууль эрхзүйн үндэс болж байна | <https://www.dlapiperdataprotection.com/index.html>
- Network Act, Credit Information Act зэрэг эрхзүйн зохицуулалтууд нь PIPA-д нэгтгэгдэж, өгөгдлийн ашиглалт, зээлийн мэдээллийн сангийн өгөгдлийг зохицуулах болсон. | globalcompetitionreview.com.
- 2023–2024 онуудад Их өгөгдлийн сангийн өргөтгөл, үйл ажиллагаатйх холбоотой шинэ зохицуулалтуудыг тусгаж өгсөн байна. Тухайлбал:
 - 14 хүртэлх насны хүүхдийн хувийн мэдээллийг хамгаалах зохицуулалтыг бий болгож, шаардлагатай бол хууль ёсны асран хамгаалагчийн зөвшөөрлийг хэрхэн авах зааврыг тодорхойлж өгсөн.
 - Гадаадад хувийн мэдээллийг дамжуулах нөхцөлийг нарийвчлан зааж өгсөн. Мөн онцгой нөхцөлд PIPA нь мэдээлэл дамжуулахыг түр зогсоох эрхтэй гэсэн заалтууд нэмэгдсэн.
 - Мэдээлэл алдагдсан тохиолдолд өгөгдөл эзэмшигчид 72 цагийн дотор мэдэгдэх заалт нэмэгдсэн.

Мөн иргэдэд Их өгөгдлийн ач холбогдлыг сурталчлан таниулах, иргэдийн санал хүсэлт, хандлагыг тодорхойлох зорилгоор олон нийтээс судалгаа авах ажлууд хийгддэг байна. 2024 онд иргэдийн санал асуулгад тулгуурлан

- Хувийн мэдээллийн нууцлалыг хангах бодлогын үнэлгээний удирдамж шинэчлэгдсэн: Хэрхэн нээлттэй, ойлгомжтой,

хэрэглэгчийн эрхийг хүндэтгэх стратеги бүхий нууцлалын бодлогыг үнэлэх удирдамж.

- Хувийн мэдээллийн нөлөөллийн үнэлгээ (PIA) Performance Guide буюу өгөгдөл хамгааллыг үнэлэх, төлөвлөх журмыг тодорхой болгох удирдамж.
- Гадаадын бизнесүүдэд PIPA хэрхэн хамаарах талаарх удирдамж: БНСУ-н хуулиудад гадаадын бизнесийн байгууллагуудын зайлшгүй дагаж мөрдөх шаардлагатай нөхцөлийг тайлбарласан удирдамж.

Хариуцагч байгууллага.

БНСУ-н тухайд Их өгөгдлийн сан байгуулах, түүний дэд бүтэц технологийн хэвтийн ажиллагааг хангах болон их өгөгдлийг хэрэглээнд нэвтрүүлэх, энэ бүгдийг нэгтгэн зохион байгуулах чиглэлээр хэд хэдэн байгууллага явуулдаг байна.

- National Information Society Agency (NIA)

NIA нь их өгөгдлийн платформ ба төвүүдийг байгуулан, үйл ажиллагааг иж бүрнээр дэмжих үүрэгтэй.

Гол чиг үүргүүд:

- Үндэсний дата дэд бүтэц байгуулах
- K-ICT Big Data Center (мэдээллийн encryption & pseudonymization дэмжих) болон Data Security Zone байгуулан үйл ажиллагаанд оруулах,
- Их өгөгдлийн дэд төвүүдийг байгуулах, үйл ажиллагааг нь зохион байгуулах
- Дата үйлдвэрлэл, цуглуулалт, дүн шинжилгээ, түгээлт, ашиглалт бүрийг дэмжих систем бүрдүүлэх зэрэг үйл ажиллагаанууд

- Korea Institute of Science and Technology Information (KISTI)

KISTI нь өндөр гүйцэтгэлтэй тооцоолол болон эрдэм шинжилгээ, судлалын өгөгдөл (science & technology data platform)-ын бэлэн байдлыг хангадаг төрийн байгууллага юм. Өгөгдлийн судалгаа, анализ, дата платформуудын үйлчилгээ зэргийг эрдэм шинжилгээний хүрээнд удирдана.

- Korea Information Society Development Institute (KISDI)

KISDI нь ICT болон мэдээллийн нийгмийн хөгжлийн судалгаа, бодлого боловсруулах чиглэлээр ажилладаг судалгааны институт бөгөөд BigData-г ашиглан ICT статистик, зах зээл, бодлого зохицуулалт, өрсөлдөөний зэрэг олон чиглэлээр судалгаа хийдэг байна.

- Institute for Information and Communications Technology Planning and Evaluation

Энэ байгууллага нь ICT салбарын судалгаа хөгжлийн бодлого, төслийн санхүүжилт, дэмжлэг, хөгжүүлэлт (DataOps, AI model зэрэг) зэргийг зохион байгуулдаг бөгөөд BigData болон AI хөрөнгө оруулалт, судалгаатай холбоотой үйл ажиллагаануудыг удирдлагаар хангадаг(жишээ нь дотоод аялалын AI загвар хөгжүүлэлт гэх мэт).

Их өгөгдлийн санг амжилттай байгуулсан орнуудын туршлагаас харахад Их өгөгдлийн санг бий болгосноор өгөгдөлд суурилсан шийдвэр гаргалт сайжирч төрийн хэмжээний буюу олон салбарын үйл ажиллагааг өгөгдлийн үндсэн дээр шинжлэх ухаанч, бодит баримтад тулгуурлан сайжруулах, өөрчлөх үйл явцыг хэрэгжүүлэх гол зорилготой юм. Мөн Их өгөгдлийн хэрэглээ, ашиглалтыг нэмэгдүүлснээр судалгаа шинжилгээнд суурилсан шинэ шинэ бүтээгдэхүүн үйлчилгээ, бизнес боломжууд бий болсноор олон нийт, иргэдийн амьдралын чанарт эерэг нөлөө үзүүлж байна.

Их өгөгдөл гэдэг ойлголттой холбоотой нэг жишээг дурдах нь зүйтэй гэж үзлээ. Их өгөгдлийн сан (Big Data) гэдэг нь зөвхөн “хэмжээний хувьд их хэмжээтэй өгөгдөл” гэсэн утгатай биш, харин өгөгдлийн хэмжээ (Volume), өөрчлөгдөх хурд (Velocity), төрөл зүйл (Variety) болон үнэн зөв байдал (Veracity) зэрэг олон шинж чанарыг багтаасан ойлголт юм.

Их өгөгдөлтэй ажиллана гэдэг нь маш их хэмжээтэй, хурдан өөрчлөгддөг, төрөл бүрийн бүтэцтэй өгөгдлийг хадгалах, боловсруулах, дүн шинжилгээ хийхэд уламжлалт өгөгдлийн сангийн аргаар шууд шийдэхэд хүндрэлтэй байдаг өгөгдлийн бүтэц юм. Иймд бид Их өгөгдөлтэй илүү хялбар ажиллах системүүд, орчны тухай сэдвийг авч үздэг.

Үндсэн шинж чанар буюу “4V” гэдэг ойлголт нь :

- Volume – Их хэмжээтэй өгөгдөл (терабайт, петабайт түвшинд).
- Velocity – Өгөгдөл маш хурдан орж ирдэг, бодит цагт боловсруулалт шаарддаг.
- Variety – Текст, зураг, видео, сүлжээний лог, IoT мэдрэгчийн өгөгдөл гэх мэт төрөл бүрийн формат.
- Veracity – Өгөгдлийн үнэн зөв, чанарын асуудал.

Зарим тохиолдолд олон өгөгдлийн сангуудын өгөгдлийг нэг дор цуглуулахгүйгээр, сан тус бүр дээр нь нэгтгэх процесс хийгээд, тэдгээрийн үр дүнг өөр нэг санд цуглуулах замаар тайлан боловсруулах туршлага байдаг. Энэ нь их өгөгдлийн “Data Lake” загвар биш, гэхдээ BigData архитектурын нэг төрөл болох “Federated Data Processing” загвар гэж үзэж болох юм.

Өгөгдлийг эх сан бүр дээр нь боловсруулж(distributed data processing) байгаа болон үр дүнг нэгтгэж байгаа сан нь aggregated data mart хэлбэрийг үүсгэж байгаа зэрэг нь Data Virtualization / Data Federation загварын гол шинж чанар юм. Гэхдээ өгөгдлийн хэмжээ маш их бол → BigData гэж үзэх боломжтой. Эсрэгээрээ энэ нь уламжлалт DBMS-ээр амархан боловсруулж байвал → ердийн OLAP/Data Mart архитектур гэж үзнэ.

Гэсэн хэдий ч ийм шийдэл дээр амжилттай хэрэгжүүлсэн төслийн жишээг олж харж болно.

Сингапур – Government Enterprise Architecture (Singapore Government Enterprise Architecture, SGEA)

- Сингапур улс “нэгдсэн төрийн архитектур” төслийн хүрээнд энэхүү загварыг ашиглаж байгууллагууд хоорондоо хамтын ажиллагаа, өгөгдлийн стандартчиллыг дэмжих систем хөгжүүлсэн
- Энэ нь Business, Information, Solution болон Technical Reference Model-уудаас бүрдсэн бөгөөд байгууллагуудын өгөгдөл, систем, процессуудыг холбон, эцсийн хэрэглэгчдэд нэг интерфейсээр үйлчлэх зохион байгуулалтай систем юм.

Япон, Их Британи, Испани, Швед зэрэг улс орнуудын үндэсний өгөгдөл болон Smart City бодлогодоо дээрх шийдлийг шингээж өгсөн.

- Испанийн Platform for Data Intermediation, Шведийн National Geodata Strategy, мөн Smart City-д зориулсан стандарт (жишээ нь: UK PD 8100 стандарт) нь өгөгдлийн олон эх үүсвэр ба түвшний хооронд стандартчилал, өгөгдлийн нэгдлийн суурь стандарт болдог байна.

ШИНЖ ЧАНАР	DATA VIRTUALIZATION / FEDERATION	BIG DATA ARCHITECTURE
Өгөгдөл хадгалалт	Өгөгдлийн мэдээлэл хариуцагч дээр хадгална	Төвлөрсөн Data Lake / DWH-д хуулж хадгална
Өгөгдлийн шинэчлэлт	Real-time буюу шууд хийгдэнэ	Шинэчлэл нь ETL/ELT-аар дамждаг тул бага зэрэг хоцордог
Хэрэгжүүлэх хугацаа	Хурдан	Удаан (Data ingestion, cleaning, modelling шаардлагатай)
Их хэмжээний өгөгдөлд	Хязгаарлагдмал	Илүү тохиромжтой
Машин сургалт / AI	Боломжгүй	Илүү тохиромжтой
Зардал	Эхэндээ бага	Эхэндээ өндөр, гэхдээ урт хугацаанд үр ашигтай

Их өгөгдлийн санг ямар арга замаар бий болгохоос үл хамааран өгөгдөл цуглуулж дүн шинжилгээ хийх үйл ажиллагаа нь нийтлэг тохиолдолд дараах байдлаар зохион байгуулагддаг.

- Өгөгдөл цуглуулах:
 - Өгөгдөл хариуцагчдаас өгөгдөл татах (IoT, CRM, ERP, судалгаа, системүүд)
- Өгөгдөл цэвэрлэх, боловсруулалт:
 - Чанартай, найдвартай өгөгдлийг бий болгох
- Өгөгдөлд анализ хийх :
 - BI, AI, ML, статистик арга ашиглах
- Шийдвэр гаргалтад ашиглах:
 - Өгөгдөлд үндэслэн бодлого, үйлчилгээ, стратеги боловсруулах
- Өөрчлөлт хийх:
 - Бүтэц, процесс, системийн шинэчлэл хийх

Энэхүү үйл ажиллагааны явцад яригдах дараагийн нэг гол ойлголт нь ямар мэдээллийг Их өгөгдлийн санд нэгтгэх вэ гэдэг асуудал юм. Өөрөөр хэлбэл ямар нэгэн зорилго, шийдлийг хангахын тулд тухайн өгөгдлийг нэгтгэх үү, эсвэл

боломжит өгөгдлүүдийг цуглуулсны дараа боломжит шийдлүүдийг бий болгох уу гэсэн аргачлал юм. Эдгээр нь тус тусын давуу болон сул талуудтай байдаг.

- Зорилго тодорхой, үр дүнгээ урьдчилан тооцоолсон, асуудалд суурилсан байдлаар өгөгдлийг ашиглах
- Өгөгдлийг урьдчилан нэгтгэж, дараа нь үр дүн, боломжоо тодорхойлох (data-first approach)

Data-first approach хандлагаар амжилттай хэрэгжүүлсэн Их өгөгдлийн төслүүдээс дурдвал:

ЭСТОНИ – X-ROAD ПЛАТФОРМ

Эстони нь эхлээд бүх төрийн байгууллагын мэдээллийг солилцох боломжтой X-Road нэртэй өгөгдөл дамжуулах түгээлтийн системийг байгуулсан. Анхны зорилго нь иргэний бүртгэл, татвар зэрэг суурь өгөгдлийг холбох байсан ч эрүүл мэнд, боловсрол, бизнесийн салбарын өгөгдлүүдийг нэмж холбосон.

Нэгтгэлийн явцад:

- Автоматжуулсан татварын тайлан
- Иргэний мэдээлэлд хандах эрхийг хянах систем (Transparency log)
- e-Residency, блокчейн суурьтай баталгаажуулалтын системүүдийг хөгжүүлсэн

СИНГАПУР – SMART NATION INITIATIVE

Төрөөс Smart City буюу өгөгдөлд суурилсан ухаалаг хотын шийдлийг эхлүүлсэн. Энэ хүрээнд зам тээврийн хөдөлгөөн, хотын агаарын чанар, гэрэлтүүлэг, камерын бичлэг, мэдрэгчүүд, IoT төхөөрөмжүүдээс өгөгдлийг өргөн хүрээнд цуглуулсан. Ингэхдээ ямар систем яг ямар үр дүн өгөхийг урьдчилан нарийн тооцоолоогүй.

Үр дүн

- Зам тээврийн системийг автомат болсон
- Ухаалаг гэрэлтүүлгийн систем бий болж эрчим хүчний хэрэглээ буурсан
- Smart Elderly Care буюу өндөр настнуудын бие бялдрын хяналтын систем туршигдсан

Ийнхүү Data-first approach байдлаар Их өгөгдлийн сан байгуулах нь дараах нөхцөлд л амжилттай хэрэгжих боломжтой байдаг байна:

- Хууль эрх зүйн зохицуулалт хангалттай сайн байх
- Өгөгдлийн нууцлал ба хамгаалалт шийдэгдсэн байх
- Иргэдийн итгэлцэл өндөр байх
- Техникийн архитектур (API, metadata, security) сайн байх

Эсрэгээрээ, зорилго тодорхойгүй өгөгдөл цуглуулалт нь

- Өгөгдлийн итгэлцэл нийцтэй байдал алдагдах, улмаар үр дүнгүй болж хувирах магадлалтай
- “Data swamp” буюу хэрэгцээгүй өгөгдлийн овоолго үүсэх эрсдэлтэй

Дээрх байдлаас дүгнэхэд манай орны хувьд зорилго тодорхой, үр дүнгээ урьдчилан тооцоолсон, асуудалд суурилсан байдлаар Их өгөгдлийн санг байгуулах үйл явцыг хэрэгжүүлэх нь зохимжтой.

Зорилгод суурилсан байдлаар Их өгөгдлийн санг бий болгох нь дараах давуу талуудыг бий болгодог:

ДАВУУ ТАЛ	ТАЙЛБАР
1. Тодорхой үр дүнтэй, богино хугацаанд үр дүнд хүрэх боломжтой	Хэрэгцээ нь тодорхой тул өгөгдлийг шууд ашиглаж, үр дүн гаргах боломжтой (жишээ: нийгмийн халамжийг оновчтой болгох, хууль зөрчлийн гол зангилааг тодорхойлох гэх мэт)
2. Зардал бага	Зөвхөн хэрэгтэй өгөгдлийг цуглуулдаг. Хэрэгцээгүй өгөгдөлд мөнгө үрэхгүй.
3. Удирдлагын түвшинд дэмжлэг авахад амархан	“Энэ асуудлыг энэ өгөгдлөөр ингэж шийднэ” гэдэг логик ойлгомжтой тул шийдвэр гаргагчид дэмжих магадлал өндөр
4. Туршиж, шаталсан байдлаар өргөжүүлэхэд хялбар	Жижиг шийдлээс эхэлж, алхам алхмаар сайжруулах байдлаар өргөжүүлэх боломжтой
5. Өгөгдлийн нууцлал, хууль эрхзүйн эрсдэл бага	Бүх төрлийн өгөгдлийг нэгтгэж, цуглуулдаггүй тул иргэдийн хувийн мэдээллийн эрсдэл багасна
6. Хүний нөөцийн чадавх	Нэг асуудал дээр төвлөрөх тул тухайн асуудалтай холбоотой дата, систем, шинжээчдийг хамруулах, тэдгээрийг хялбар сурах, хөгжүүлэх боломжтой

Дээрх судалгаануудад үндэслэн манай орны хувьд Их өгөгдлийн байгуулах төслийн хоёр үе шаттайгаар хэрэгжүүлэх бөгөөд өгөгдлийг “зорилтот үр дүнгээ урьдчилан тооцоолсон, асуудалд суурилсан” байдлаар цуглуулж бий болгох нь зүйтэй. Ингэхдээ өгөгдлийг нэгдсэн байдлаар цуглуулах, гэхдээ зарим тохиолдолд буюу, жишээ нь e-barimt н сангаас өгөгдлийг задгайгаар нь татахгүйгээр тодорхой хэмжээнд буюу өдөр бүрийн гүйлгээг багцалсан байдлаар их өгөгдлийн санд нийлүүлэх нь илүү оновчтой юм.

1р үе шат: 2025-2027 он.

- Зорилт:

- Их өгөгдлийн санг бий болгоход шаардагдах хууль эрхзүйн шинэтгэлүүдийг хийх
- Их өгөгдлийн санг хариуцах байгууллага, агентлагийг байгуулж үүрэгжүүлэх
- Их өгөгдлийн сангийн суурь дэд бүтцийг бий болгох
- Суурь болон төрөлжсөн мэдээллийн сангуудад суурилсан “зорилтот үр дүнгээ урьдчилан тооцоолсон” загварын Их өгөгдлийн санг бий болгох
- Их өгөгдлийн санд суурилсан Нээлттэй өгөгдлийн дэд бүтцийг хөгжүүлэх

- Нээлттэй өгөгдөлд суурилсан бизнесийг дэмжих

2р үе шат: 2028-2030 он.

- Зорилт:

- Хууль эрхзүйн зохицуулалтыг эргэн харж сайжруулах, бизнесийн байгууллагын өгөгдлийг Их өгөгдлийн санд нэгтгэх зохицуулалтыг оруулах
- Төрийн болон бизнесийн байгууллагын өгөгдлийг Их өгөгдлийн санд цуглуулах
- IoT, бүтэцгүй болон хагас бүтэцтэй өгөгдөл цуглуулах.
- Их өгөгдлийн дэд бүтцийг өргөтгөх, GPU-д суурилсан хүчин чадлыг нэмэгдүүлэх, LakeHouse технологи нэвтрүүлэх
- ML болон Ai үйл ажиллагаануудыг нэвтрүүлэх
- Нээлттэй өгөгдөлд суурилсан ML, Ai үйл ажиллагаа, бизнесийг дэмжих

ҮЙЛ АЖИЛЛАГАА, 2025-2027.

- Бид өгөгдөл, түүний нууцлал аюулгүй байдалтай холбоотой хууль, тогтоол, дагасан журмуудыг тодорхой хэмжээнд боловсруулан гаргаад байгаа боловч зарим нэмэлт, сайжруулалтуудыг хийх шаардлагатай гэж үзэж байна.

Үүнд:

- 2023 оны Засгийн газрын 403-р тогтоолд Их өгөгдлийн сан байгуулах үйл ажиллагааг хамтран хэрэгжүүлэх байгууллагуудыг зааж өгсөн. Гэвч эдгээр байгууллагууд нь тус тусын чиг үүрэг бүхий төрийн байгууллагууд бөгөөд Их өгөгдлийн санг байгуулах, түүний өдөр тутмын хэвийн үйл ажиллагааг хангах болон цаашдын хөгжлийн асуудлыг хариуцах нарийн зохицуулалт дээрх тогтоолд агуулагдаагүй байна. Мөн Их өгөгдлийн санд өгөгдөл дамжуулах үйл ажиллагааг ХУР системээр дамжуулах боломжийг бий болгох тухай заалт орсон байна. Төрийн мэдээлэл солилцооны ХУР систем одоогийн байдлаар байгууллага хоорондын өгөгдөл дамжуулах, өгөгдлийн нууцлал аюулгүй байдлыг хангах үйл ажиллагааг амжилттай хэрэгжүүлж байгаа боловч Их өгөгдлийн санд багц өгөгдөл дамжуулах, мэдрэгч төхөөрөмжүүдээс өндөр давтамж бүхий өгөгдөл татах зэрэг их өгөгдлийн онцлогуудад нийцсэн өгөгдөл дамжуулах шийдлийг агуулаагүй.

Иймд:

- Нийтийн мэдээллийн ил тод байдлын тухай хуульд Их өгөгдлийн сангийн зохицуулалтыг тусгасан заалтуудыг тусгах
 - Их өгөгдлийн сангийн тодорхойлолт, зориулалтыг тусгах
 - Их өгөгдлийн санг хариуцан ажиллах байгууллагыг тодорхой болгох
 - Их өгөгдлийн санд өгөгдөл нийлүүлэх тусгай дэд бүтцийг төрийн мэдээлэл солилцооны ХУР системд бий болгох
 - Төрийн байгууллагуудыг их өгөгдлийн санд өгөгдөл нийлүүлэх үйл ажиллагааг үүрэгжүүлэх
- Хүний хувийн мэдээлэл хамгаалах тухай хуульд
 - Их өгөгдлийн сантай холбоотой заалт оруулах, Их өгөгдлийн санд мэдээлэл хариуцагчаас иргэний мэдээллийг 403 р

тогтоолын хүрээнд нийлүүлж болох тухай, тэрхүү мэдээллийг Их өгөгдлийн сангийн хариуцагч байгууллага хамгаалах

- Иргэн өөрөө хүсвэл Их өгөгдлийн санд хувийн мэдээллээ нийлүүлэхгүй байж болох тухай заалт, түүнтэй холбоотой харилцааг хэрхэн зохицуулах заалт
- Мэдээлэл хариуцагч байгууллага, Их өгөгдлийн сан зэрэг мэдээлэл хариуцаж байгаа байгууллага иргэний мэдээллийг алдсан тохиолдолд иргэнд заавал мэдэгддэг байх заалт.(Их өгөгдлийн сангийн дараагийн үе шат буюу ухаалаг төхөөрөмжүүдээс иргэний дуу, бичлэг, зураг зэрэг өгөгдлийг татдаг болох үед мэдээлэл хариуцагчаас иргэн, байгууллагын эмзэг мэдээллийг арилгах, хязгаарлалттай мэдээллийг танигдахгүй болгож их өгөгдлийн санд мэдээллийг дамжуулах заалттай зөрчил үүсгэх боломжтойг анхаарах)
- Засгийн газрын 2023 оны 403-р тогтоол
 - Тогтоолд Их өгөгдлийн санд өгөгдөл нийлүүлэхийг зөвлөмж байдлаар тусгасан байна. Энэ заалтыг илүү үүрэгжүүлэх, өгөгдөл нийлүүлэх үйл ажиллагаанд шаардлагатай бүтэц, зохион байгуулалтыг мэдээлэл хариуцагч байгууллагууд шийдсэн байх шаардлагатай.

Эдгээр хууль, тогтоолуудын хэрэгжилтийг хангахад холбогдох үйл ажиллагааны журмууд шаардлагатай болно. Тухайлбал:

- Мэдээлэл хариуцагч байгууллага Их өгөгдлийн санд дамжуулах өгөгдлийг хэрхэн бэлтгэх, ямар түвшний эмзэг мэдээллийг дамжуулахгүй байх, ямар хязгаарлалттай мэдээллийг хэрхэн танигдахгүй болгох. Хязгаарлалттай мэдээллийг танигдахгүй болгох үйл ажиллагааг хэрхэн зохион байгуулах, түүний нууцлалыг хэрхэн хангах зэрэг үйл ажиллагааг нарийвчлан заасан өгөгдөл бэлтгэх талаарх журам
- Их өгөгдлийн санд дамжуулж буй өгөгдлийн үнэн зөв байдлыг хэрхэн хангах, мэдээллийг ямар давтамжтайгаар шинэчлэх, хэрхэн аюулгүй дамжуулах үйл ажиллагааны журам
- Их өгөгдлийн санг хариуцан ажиллах байгууллага, нэгжийн үйл ажиллагааны журам, хууль эрхзүйн зохицуулалтууд
- Их өгөгдлийн санд шинээр нэмэгдэх байгууллагуудад тавигдах шаардлага, тэдгээрийг үнэлэх аргачлал
- Их өгөгдлийн санд өгөгдөл нийлүүлэх байгууллагууд нь Их өгөгдлийн санд зөвшөөрөгдсөн ангилал, кодыг нэвтрүүлсэн байх шаардлагатай бөгөөд энэ үйл ажиллагааг хэрхэн хэрэгжүүлэх талаарх зөвлөмж, стандарт батлах
- Мета мэдээллийн санг зохион байгуулах, хэрэгжүүлэх, ангилал кодын нийцлийг хангах журам

- Их өгөгдлийн сангаас өгөгдөл авах, Их өгөгдлийн сангийн мэдээлэлтэй ажиллах зэрэг үйл ажиллагаануудыг журамлах шаардлагатай.

Мөн түүнчлэн Их өгөгдлийн удирдах хороо буюу их өгөгдөлтэй холбоотой асуудлуудыг шийдвэрлэх, холбогдох журам, заавар батлах, өгөгдөл засаглахтай холбоотой үйл ажиллагаанд шийдвэр гаргах эрх бүхий бүтцийг бий болгож үйл ажиллагаанд оруулах шаардлагатай.

- Хэрэгцээ шаардлага тодорхойлох.

Их өгөгдлийн санг хэрэглээнд нэвтрүүлэх, их өгөгдлийн тоон мэдээллийг олон нийтэд түгээхэд Opendata.gov.mn платформыг ашиглах нь зохистой гэж үзэж байна. Энэхүү платформ нь өнөөдрийн байдлаа төрийн байгууллагууд хууль заасан үүргийнхээ дагуу нээлттэй мэдээллээ олон нийтэд түгээдэг гол суваг болоод байгаа билээ. Тэгвэл ийнхүү бэлэн болсон сууриа ашиглан Opendata.gov.mn платформыг өргөтгөж Их өгөгдлийн сангаас хүснэгт, тайлан, дашбоард боловсруулан олон нийтэд хүргэх үйл ажиллагааг хэрэгжүүлэх нь оновчтой.

Тэгвэл Их өгөгдлийн санд ямар өгөгдөл цуглаад байгаа вэ, ямар мэдээллийг ямар давтамжтайгаар ямар хэлбэрээр олон нийтэд хүргэх вэ, ямар API сервисүүдийг нээж ашиглалтад оруулах вэ гэсэн асуулт урган гарна.

Одоогийн байдлаар буюу Их өгөгдлийн эхэн үеийн туршилт, хөгжүүлэлтийн шатанд өгөгдлийг их өгөгдлийн санд нэгтгэхдээ ямар нэгэн зорилт, шийдвэрлэх асуудал дэвшүүлэлгүйгээр өгөгдөл нэгтгэх үйл ажиллагааны технологийн шийдлийн бэлэн байдлыг хангах зорилгоор цуглуулсан байна. Иймд энэхүү Их өгөгдлийн сангаас гарах боломжтой судалгаа, тооцоолол, тайлан хүснэгтийн боломжийг судалгааны үндсэн дээр тодорхойлж ажлын даалгавар боловсруулах ажил нэн шаардлагатай байна. Ийнхүү Их өгөгдлийн сангаас гарах боловсруулалтын даалгавар бэлэн болсны дараа түүний хөгжүүлэлтийг дотоодын болон гадаадын хөгжүүлэгч компаниудын дунд тендер зохион байгуулах, эсвэл e-mongolia academy-г түшиглэх хөгжүүлэлтийг хийлгэх байдлаар opendata.gov.mn/BIGDATA төслийг хэрэгжүүлэх боломжтой.

СУДАЛГААНЫ ГҮЙЦЭТГЭГЧ СОНГОН ШАЛГАРУУЛАХ ЗАР

ИХ ӨГӨГДЛИЙН САНГИЙН ХЭРЭГЛЭЭГ НЭМЭГДҮҮЛЭХ ЗОРИЛГООР ИХ ӨГӨГДЛИЙН САНГААС ГАРАХ ҮЗҮҮЛЭЛТИЙГ ТООЦОХ СУДАЛГААНЫ АЖИЛД ГЭРЭЭТ БАЙГУУЛЛАГА ШАЛГАРУУЛЖ АВНА.

Үндэсний статистикийн хороо нь Монгол Улсын хүн амын төрөлтийн төсөөлөл, гэр бүлд гарч буй өөрчлөлт, тэдгээрт нөлөөлж буй хүчин зүйлсийг тодорхойлох зорилгоор түүвэр судалгааг зохион байгуулахаар төлөвлөн ажиллаж байна. Судалгааны мэдээлэл цуглуулах багийн өдөр тутмын ажилд хяналт тавих, удирдан зохион байгуулах 1 сарын хугацаатай гэрээт ажилтныг сонгон шалгаруулж авна.

ЦХИХХЯ нь Их өгөгдлийн сан байгуулах үйл ажиллагааг хэрэгжүүлж эхлээд байгаа билээ. Энэ ажлын хүрээнд төрийн байгууллагуудаас иргэн, байгууллагын эмзэг мэдээллийг агуулаагүй, хязгаарлалттай мэдээллийг зохих журмын дагуу нууцлалыг нь хангасан өгөгдлийг Төрийн өгөгдөл дамжуулах ХУР системээр дамжуулан Үндэсний дата төвд байршуулж эхлээ байна. Өнөөдрийн байдлаар ... төрийн байгууллагын ... хэмжээний өгөгдлийг нэгтгээд байна.

Төрийн Их өгөгдлийн сангийн хэрэглээг нэмэгдүүлэх, олон нийт, бизнесийн байгууллагуудад хүртээмжтэй өгөгдөл, мэдээллийг хүргэх зорилгоор бий болсон Их өгөгдлийн сангаас гаргах боломжтой үзүүлэлт, тайлан, сервисүүдийг тодорхойлж тэдгээрийг гүйцэтгэх ажлын даалгаврыг гаргах зорилт тавиад байна. Энэхүү ажлыг дотоодын мэдээллийн технологи, судалгаа шинжилгээний байгууллагуудаар гүйцэтгүүлэхээр ажиллаж байна.

Энэхүү гэрээт ажлыг гүйцэтгэх хуулийн этгээд нь мэдээллийн технологи, судалгаа шинжилгээний чиглэлээр үйл ажиллагаа явуулдаг, 10+ жилийн туршлагатай байхыг шаардах ба төслийг хэрэгжүүлэх баг 7 буюу түүнээс дээш гишүүнтэй байхыг шаардана. Баг нь:

- Төслийн удирдагч - 1
- Өгөгдлийн сангийн шинжээч - 2
- Системийн шинжээч - 2
- Статистикч - 2

Төслийн удирдагч - 1:

- Мэдээллийн технологи, өгөгдлийн шинжлэх ухаан эсвэл математик-статистикийн чиглэлээр их дээд сургууль төгссөн, магистр болон түүнээс дээш зэрэгтэй байх;
- Тухайн чиглэл бүрд хамаарах ажлын хүрээгээр зөвлөх үйлчилгээ үзүүлэхэд шаардагдах зохих мэргэжил, ур чадвар эзэмшсэн
- Мэргэжлийн чиглэлээр 10 ба түүнээс дээш жил ажилласан туршлагатай байх/;
- Баримт бичиг боловсруулахад шаардлагатай судалгаа, дүн шинжилгээ хийх, хөтөлбөр боловсруулах ажилд оролцож байсан туршлагатай байх;
- Харилцааны соёлтой, багаар ажиллах чадвартай, хувийн зохион байгуулалт сайтай байх;

Өгөгдлийн шинжээч - 2:

- Мэдээллийн технологи, өгөгдлийн шинжлэх ухаан эсвэл математик-статистикийн чиглэлээр их дээд сургууль төгссөн байх, 1-ээс багагүй гишүүн нь өгөгдлийн шинжлэх ухаан чиглэлийн магистр зэрэгтэй байх;
- Тухайн чиглэл бүрд хамаарах ажлын хүрээгээр зөвлөх үйлчилгээ үзүүлэхэд шаардагдах зохих мэргэжил, ур чадвар (өгөгдлийн сан дээр ажиллах ахисан түвшний мэдлэг, туршлага, ур чадвар (NoSQL, RDS, Datawarehouse, BigData) эзэмшсэн; эзэмшсэн, өгөгдлийн чиглэлээр 3с доошгүй төсөл хэрэгжүүлсэн байх;
- Мэргэжлийн чиглэлээр 5 ба түүнээс дээш жил ажилласан туршлагатай байх;
- Баримт бичиг боловсруулахад шаардлагатай судалгаа, дүн шинжилгээ хийх ажилд оролцож байсан туршлагатай байх.
- Харилцааны соёлтой, багаар ажиллах чадвартай, хувийн зохион байгуулалт сайтай байх;

Системийн шинжээч – 2:

- Мэдээллийн технологи, өгөгдлийн шинжлэх ухааны чиглэлээр дээд боловсрол эзэмшсэн;
- Тухайн чиглэл бүрд хамаарах ажлын хүрээгээр зөвлөх үйлчилгээ үзүүлэхэд шаардагдах зохих мэргэжил, ур чадвар эзэмшсэн;
- 2-оос доошгүй системийн шинжилгээ, зохиомж, шийдэл, архитектурын ажил гүйцэтгэсэн байх;
- Мэргэжлийн чиглэлээр 5 ба түүнээс дээш жил ажилласан байх шаардлагатай
- Баримт бичиг боловсруулахад шаардлагатай судалгаа, дүн шинжилгээ хийх ажилд оролцож байсан туршлагатай байх;
- Харилцааны соёлтой, багаар ажиллах чадвартай, хувийн зохион байгуулалт сайтай байх;

Статистикч -2 :

- Статистикийн чиглэлээр дээд боловсрол эзэмшсэн;
- Тухайн чиглэл бүрд хамаарах ажлын хүрээгээр зөвлөх үйлчилгээ үзүүлэхэд шаардагдах зохих мэргэжил, ур чадвар эзэмшсэн
- Мэргэжлийн чиглэлээр 5 ба түүнээс дээш жил ажилласан туршлагатай байх;
- Баримт бичиг боловсруулахад шаардлагатай судалгаа, дүн шинжилгээ хийх, төрийн байгууллагуудын хооронд өгөгдөл солилцох, нууцлалыг хангах үйл ажиллагааны талаар мэдлэг, туршлагатай байх.
- Харилцааны соёлтой, багаар ажиллах чадвартай, хувийн зохион байгуулалт сайтай байх;

Гэрээт ажлыг гүйцэтгэх чиг үүрэг:

- Төрийн их өгөгдлийн санд цуглуулсан өгөгдлийн судалгааг хийж өгөгдлийн нийцтэй байдлын үнэлгээг хийх
- Ангилал, код, хэмжих нэгжийн болон өгөгдөл шинэчлэгдэх давтамжийн нийцэл тогтоох
- Их өгөгдлийн сангийн катологи бүрдүүлэх үйл ажиллагааны зураглал боловсруулах, ажлын даалгавар боловсруулах
- Их өгөгдлийн сангаас гарах өгөгдлийн тооцоо, судалгаа хийх
 - Үзүүлэлтүүд. Жишээ нь ядуурлын индекс, SDG-н үзүүлэлтүүд гэх зэрэг. Эдгээрээс SDG буюу тогтвортой хөгжлийн

зорилтын биелэлтүүдээс тодорхой үзүүлэлтүүдийг сонгож Их өгөгдлийн сангаас гаргадаг болох нь илүү оновчтой юм. Тухайлбал HDI (Хүний хөгжлийн индекс) → Эрүүл мэнд, боловсрол, орлогын нийлмэл хэмжүүр, MPI (Олон талт ядуурлын индекс) → Орлогоос гадна боловсрол, эрүүл мэнд, нийгмийн үйлчилгээний хүртээмж зэрэг нийлмэл үзүүлэлтүүдийг тооцох, мөн тэдгээрээс гадна

- SDG-5.4.1: Төлбөргүй гэр бүлийн хөдөлмөр (асаргаа, халамжийн ажил)-ийг хэмжих үзүүлэлт
- SDG-8.6.1: 15–24 насны боловсрол, сургалт, хөдөлмөрт хамрагдаагүй залуусын хувь
- SDG-10.4.2: Нийгмийн хамгаалал ба татварын бодлогоор ядуурлыг бууруулж буй үр нөлөө зэрэг салбар хоорондын үзүүлэлтүүдийг тооцох, үр дүнг олон хэлбэрээр тархаах
- Хүснэгтүүд. Олон салбарыг хамарсан нийлмэл салбаруудын онцлогийг шингээсэн тайлан, хүснэгтүүд
- График болон дашбоардууд
- API сервисүүд.
- Дээрх судалгааны үндсэн дээр opendata.gov.mn системийн Их өгөгдлийн үзүүлэлтүүдийн хэсэгт оруулах тайлан хүснэгт, тооцсон үзүүлэлтүүд, график дашбоардууд болон API сервисүүдийг хөгжүүлэх ажлын даалгавар боловсруулах
- Их өгөгдөлтэй ажиллах явцдаа хууль, тогтоомж болон холбогдох журмын дагуу нууцлах, аливаа хэлбэрээр өгөгдлийн нууцлалыг задруулахгүй байх;
- Ажлын явцын тайлан, гүйцэтгэлийг тухай бүрд нь танилцуулах, шаардлагатай тохиолдолд нэмэлт тодруулга өгөх;

Ажил гүйцэтгэх хугацаа, төсөв:

Гэрээт ажил нь 2022 оны ... дугаар сарын ...-оос 2022 оны ... дүгээр сарын ...-г дуустал 7 хүн 60 өдрийн буюу $7 * 60 * 200'000 + 20\%(7 * 60 * 200'000) = 100'800'000.00$ төсөвтэй ажил байна.

Багийн гишүүдийн бүрдүүлэх материал:

- Төрийн албан хаагчийн анкет (ТАЗ-ийн 2022 оны 600 дугаар тогтоолын 1-р хавсралтын дагуу)
- Мэргэжлийн сургууль (Их, дээд сургууль, коллеж гэх мэт) төгссөн дипломын хуулбар
- Иргэний үнэмлэхийн хуулбар

Энэхүү ажлын даалгаврыг дагуу opendata.gov.mn системд Их өгөгдлийн тайлангийн бүлэгт нэмэгдэх хүснэгт, үзүүлэлт, графикууд болон API сервисүүд, тэдгээрийг хэрхэн хөгжүүлэх талаарх ажлын даалгавар бэлэн болсон байна.

Уг ажлын даалгаврыг үндэслэх тендер зарлах, эсвэл e-mongolia академийн хөгжүүлэгчийн сонгож хөгжүүлэлтийг хийж гүйцэтгэнэ. Энэхүү хөгжүүлэлтийг хийж гүйцэтгэхийн зэрэгцээгээр график дашбоардуудыг opendata.gov.mn системд

харуулах, хэрэглэгчид хялбар байдлаар өөрсдийн хэрэгцээнд нийцсэн хүснэгт, дашбоардуудыг байгуулах боломж олгох системийн худалдан авалтыг хийх шаардлагатай. (Шаардлагатай программ хангамжуудыг дараагийн бүлэгт дурдлаа.)

Их өгөгдлийн сан байгуулах эхний үе шатны үйл ажиллагааны эцэст Их өгөгдлийн сангийн өгөгдөл баяжихын хирээр түүний хэрэглээг хэрхэн нэмэгдүүлэх вэ, шинэ үзүүлэлт, тайлан болон сервисийн хөгжүүлэлтийг хэрхэн хийх вэ зэрэг үйл ажиллагаанууд журамлагдаж, үйл ажиллагаа зүгширсэн байна. Тэр хирээр Их өгөгдлийн хэрэглээ нэмэгдэж их өгөгдөлд суурилсан шинэ шинэ бизнесүүд үүд хаалгаа нээх боломжтой болно.

ҮЙЛ АЖИЛЛАГАА, 2028-2030.

Дараагийн үе шатны хамгийн том сорилт нь Их өгөгдлийн санг хагас бүтэцтэй болон бүтэцгүй өгөгдлөөр баяжуулах боломжтой болгох, энэ үйл ажиллагааны хүрээнд шинэ техник технологийг эзэмших явдал юм. Энэ нь Их өгөгдөлд суурилсан ML, Ai зэрэг дараагийн хэрэглээний үүд хаалгыг нээж өгөх юм.

Мөн бизнесийн байгууллагуудаас өгөгдөл татах эрхзүйн орчныг бий болгож тэдэнтэй өгөгдөл солилцох үйл ажиллагааг эхлүүлнэ.

- Их өгөгдлийн байгуулах үйл явцад шаардагдаж буй хууль эрхзүйн сайжруулалтыг тогтмол хийж байх шаардлагатай. Үүний зэрэгцээгээр Их өгөгдлийн сангийн дараагийн үе шатны бэлэн байдлыг хангахын тулд Их өгөгдлийн санг тойрсон хууль, холбогдох тогтоол, журмуудад нэмэлт зохицуулалтуудыг хийх шаардлагатай.

Үүнд:

- IoT, мэдрэгч төхөөрөмжүүдээс өгөгдөл татах үйл ажиллагааг хуульчлах. Хэрвээ тэдгээр өгөгдлүүд нь Хүний хувийн нууцыг хамгаалах хуулийн заалтуудтай зөрчил үүсгэхээр бол шаардлагатай өөрчлөлтүүдийг хуульд тусгах. (Жишээ нь замын хөдөлгөөний камерын зураг, эмнэлгийн төхөөрөмжийн өгөгдөл зэрэг нь хувь хүний эмзэг мэдээллийг шууд агуулах боломжтой гэж үздэг)
- Бизнесийн байгууллагуудаас өгөгдөл татах эрхзүйн зохицуулалтыг холбогдох хууль, журмуудад шингээж өгөх шаардлагатай. Иргэн, байгууллагууд бизнесийн харилцаанд орж аливаа бүтээгдэхүүн, үйлчилгээ авах болон үзүүлэх явцдаа харилцагчийн хамгийн сүүлийн шинэчилсэн мэдээллийг авч байдаг. Жишээ нь үүрэн холбооны салбарын өгөгдлийг нэгтгэснээр Монгол орны аялал жуулчлалын маршрут, ачаалал, шилжилт хөдөлгөөний дэлгэрэнгүй шинжилгээ хийх боломж бий болно.
- Их өгөгдлийн санд ML, Ai -д суурилсан бүтээгдэхүүн хөгжүүлэлт, дүн шинжилгээ хийх үйл ажиллагааг журамлах шаардлагатай.
- LakeHouse технологийг Их өгөгдлийн санд нэвтрүүлэх
- Их өгөгдлийн дэд бүтцийг өргөтгөх, шинээр GPU-д суурилсан өргөтгөлүүд хийх.
- Их өгөгдлийн санд суурилсан машин сургалтууд, хиймэл оюун ухаанд суурилсан дүн шинжилгээ хийх, шинээр бүтээгдэхүүн хөгжүүлэх
- Opendata.gov.mn системийг ML, Ai д суурилсан шийдлүүдээр өргөтгөх

ТОНОГ ТӨХӨӨРӨМЖИЙН ШИНЭЧЛЭЛ

- Их өгөгдлийн сангийн нэг том хэрэглээ нь ML, Ai хөгжүүлэх үйл ажиллагаа болон хөгжиж байна. Энэхүү үйл ажиллагаа нь зайлшгүй хэрэгжүүлэх шаардлагатай үе шат тул эртнээс бэлтгэл ажлыг хангаж байх нь оновчтой юм. Их өгөгдлийн сан байгуулах үйл явц нь хувийн хэвшлийн өгөгдлийг нэгтгэх, IoT болон мэдрэгч төхөөрөмжүүдээс өгөгдөл татах үйл явцтай зэрэгцэн ML хөгжүүлэх боломжтой болсон байна. GPU (Graphics Processing Unit) нь компьютерын техник хангамжийн нэг төрөл бөгөөд голчлон график боловсруулалт, параллель тооцоолол хийхэд зориулагдсан. Энэ нь олон тооны жижиг тооцооллыг нэгэн зэрэг гүйцэтгэх чадвартай бөгөөд ялангуяа машин сургалт (ML), гүнзгий сургалт

(Deep Learning), тоглоомын график, шинжлэх ухааны туршилт зэрэгт өргөн ашиглагддаг.

GPU-ИЙН ГОЛ ОНЦЛОГУУД:

1. Параллель тооцоолол:
 - GPU нь хэдэн мянган жижиг цөм (cores)-тэй бөгөөд энэ нь нэгэн зэрэг олон тооцоолол хийх боломжийг олгодог.
 - Жишээ: CPU-д 4-16 цөм байдаг бол NVIDIA A100 GPU-д 6,912 CUDA цөм байдаг.
2. Өндөр хурдны санах ой:
 - GPU нь өндөр хурдны VRAM (Video RAM)-тай бөгөөд өгөгдлийг хурдан боловсруулахад ашиглагддаг (жишээ: 40GB HBM3 санах ой).
3. ML-д ашиглах:
 - Машин сургалтын загварууд, ялангуяа нейроны сүлжээний матрицын үйлдлүүдийг (matrix multiplications) хурдасгахад тохиромжтой.
 - Жишээ: TensorFlow, PyTorch зэрэг фреймворкууд GPU-ийн хүчийг ашигладаг.
4. Гол хэрэглээ:
 - Машин сургалт: Том хэмжээний загваруудыг сурчихах (training) болон таамаглал (inference) хийх.
 - График боловсруулалт: Тоглоом, 3D загварчлал, видео рендеринг.
 - Шинжлэх ухааны тооцоолол: Физикийн симуляци, молекулын загварчлал.

GPU vs. CPU:

- CPU: Цөөн тооны хүчирхэг цөмтэй, ерөнхий зориулалтын тооцоололд (serial tasks) тохиромжтой.
- GPU: Олон тооны жижиг цөмтэй, параллель тооцоололд (parallel tasks) илүү тохиромжтой.

ЖИШЭЭ GPU-УУД:

- NVIDIA: A100, H100, RTX 4090
- AMD: Radeon Instinct
- Intel: Arc GPU

МАШИН СУРГАЛТЫН (ML) ХЭРЭГЛЭЭ БА ТООЦООЛОЛ

Энд яригддаг гол ойлголт нь ямар хэрэгцээг ямар хэмжээний өгөгдлөөр ямар тооцоолол загвар ашиглаж шийдэх вэ гэдэг гол асуулт болдог. Өөрөөр хэлбэл эхлээд шийдвэрлэх ёстой асуудлаа, өгөгдлийн бэлэн байдал болон тооцоолол загваруудын сонголтуудтай уялдуулж ашиглагдах GPU-г тооцоолох шаардлагатай байдаг.

ML нь их хэмжээний өгөгдлийн санг боловсруулах, загварчлах, урьдчилан таамаглахад ашиглагдана. ML-ийн тооцооллын хэрэгцээг тодорхойлохдоо дараах зүйлсийг анхаарах шаардлагатай:

- Өгөгдлийн хэмжээ ба төрөл:

- Том хэмжээний өгөгдөл (бүтэцтэй болон хагас бүтэцтэй эсвэл бүтэцгүй) боловсруулахад өндөр хурдны тооцоолол шаардлагатай.
 - Жишээ нь 1 петабайт (PB) хэмжээтэй өгөгдөлтэй ажиллаж байгаа бол өгөгдлийн хуваалт (data partitioning) болон параллель боловсруулалт (parallel processing) хийх зайлшгүй шаардлагатай тулгарна.
- ML загварын төрөл:
 - Deep Learning (гүнзгий сургалт) загварууд (CNN, RNN, Transformer гэх мэт) нь өндөр тооцооллын хүчин чадал шаарддаг бол уламжлалт алгоритмууд (Linear Regression, Decision Trees) бага нөөц шаардана.
 - Загварын нарийн төвөгтэй байдал (жишээ: нейроны тоо, давхаргын гүн) нь тооцооллын хэрэгцээг нэмэгдүүлнэ.
- Сургалт ба таамаглал:
 - Сургалтын (training) үе шат нь их хэмжээний тооцоолол шаарддаг, ялангуяа гүнзгий сургалтын загваруудын хувьд.
 - Inference (таамаглал) хийхэд бага нөөц шаардлагатай ч хурд чухал байдаг.
- Тооцооллын шаардлага:
 - CPU эсвэл GPU-ийн хэрэглээг тодорхойлохын тулд загварын параметрууд (жишээ: 1 сая параметр vs. 1 тэрбум параметр) болон өгөгдлийн багцын хэмжээг тооцоолно. Жишээ нь BERT загварыг сургахад олон GPU эсвэл TPU шаардлагатай бөгөөд их хугацаа шаардагддаг.

GPU-ИЙН ХЭРЭГЛЭЭ БА ТООЦООЛОЛ

GPU нь ML-ийн сургалт болон боловсруулалтад өндөр хурдны параллель тооцоолол хийхэд ашиглагддаг. GPU-ийн тооцооллын хэрэгцээг тодорхойлохдоо:

- GPU-ийн төрөл ба хүчин чадал:
 - Жишээ: NVIDIA A100 эсвэл H100 GPU нь өндөр гүйцэтгэлтэй бөгөөд том хэмжээний ML загваруудыг сургахад тохиромжтой.
 - GPU-ийн санах ой (VRAM): Том загварууд (жишээ: GPT-3, LLaMA) 40GB ба түүнээс дээш VRAM шаарддаг.
- Параллель боловсруулалт:
 - GPU нь матрицын үйлдэл (matrix operations) нь тооцооллыг хурдасгадаг.
 - Жишээ: Deep Learning-д ашигладаг batch processing нь GPU-ийн олон цөмийг ашиглан өгөгдлийг хуваан боловсруулна.
- Тооцооллын хэмжээ:
 - GPU-ийн тооцооллын хэрэгцээг FLOPS (Floating Point Operations Per Second) хэмжээгээр тооцдог.
 - Жишээ: NVIDIA A100 нь ~312 TFLOPS (FP16) хүчин чадалтай бөгөөд энэ нь том загваруудыг сургах боломжтой.
- GPU-ийн тоо ба кластер:
 - Том өгөгдлийн сан боловсруулахад олон GPU бүхий кластер шаардлагатай. Жишээ: 100 GPU-ийн кластер нь 1 PB өгөгдлийг хэдэн өдрийн дотор боловсруулж чадна.

ПРАКТИК ТООЦООЛЛЫН АЛХАМ

Их хэмжээний өгөгдлийн сан үүсгэхэд ML болон GPU-ийн хэрэглээг тооцоолохын тулд:

1. Өгөгдлийн хэмжээг тодорхойлох:
 - Жишээ: 1 PB өгөгдөл, 100 сая бичлэг, тус бүр 10 KB хэмжээтэй.
2. ML загварын шаардлагыг тодорхойлох:
 - Загварын төрөл (жишээ: Transformer), параметруудийн тоо, сургалтын давталтын тоо (epochs).
3. Тооцооллын хэрэгцээг тооцоолох:
 - Жишээ: 1 тэрбум параметр бүхий загвар сургахад $\sim 10^{18}$ FLOPS шаардлагатай.
 - Хэрэв 10 GPU (тус бүр 300 TFLOPS) ашиглавал, ойролцоогоор 1-2 долоо хоног шаардлагатай.
4. GPU-ийн тоог сонгох:
 - Хэрэгцээндээ тохируулан GPU-ийн тоо, төрлийг сонго. Жишээ: 4x NVIDIA A100 (80GB) эсвэл AWS/GCP-ийн кластер.
5. Санах ой ба хадгалах систем:
 - Том өгөгдлийн санг хадгалахад өндөр хурдны SSD эсвэл Hadoop/Spark зэрэг хуваарилагдсан систем шаардлагатай.

GPU ДЭМЖИХ СИСТЕМҮҮД

- Фреймворк ба хэрэгсэл: TensorFlow, PyTorch, Apache Spark зэрэг хэрэгслүүд GPU-ийн тооцооллыг дэмждэг.
- Клаудын үйлчилгээ: AWS (EC2 P4d), Google Cloud (TPU), Azure (NDv4) зэрэг клаудын үйлчилгээг ашиглан GPU-ийн хүчин чадлыг хялбархан өргөтгөж болно.
- Оновчлол: Data sharding, model parallelism, distributed training зэрэг техникүүдийг ашиглан тооцооллын хурдыг нэмэгдүүлнэ.

ЖИШЭЭ ТООЦООЛОЛ

- Өгөгдөл: 1 PB, 100 сая бичлэг, текст болон зураг агуулсан.
- Загвар: Transformer-based NLP загвар (1 тэрбум параметр).
- GPU: 8x NVIDIA A100 (40GB).
- Тооцоолол: Ойролцоогоор 1-2 долоо хоног сургалт, 10^{18} FLOPS.
- Хадгалах сан: NVMe SSD эсвэл S3-based хадгалах сан.
- Зардал: AWS дээр 8x A100 инстанс ($\sim \$32$ /цаг) ашиглавал, 2 долоо хоногт $\sim \$10,000$ -\$15,000.

СЕРВЕР

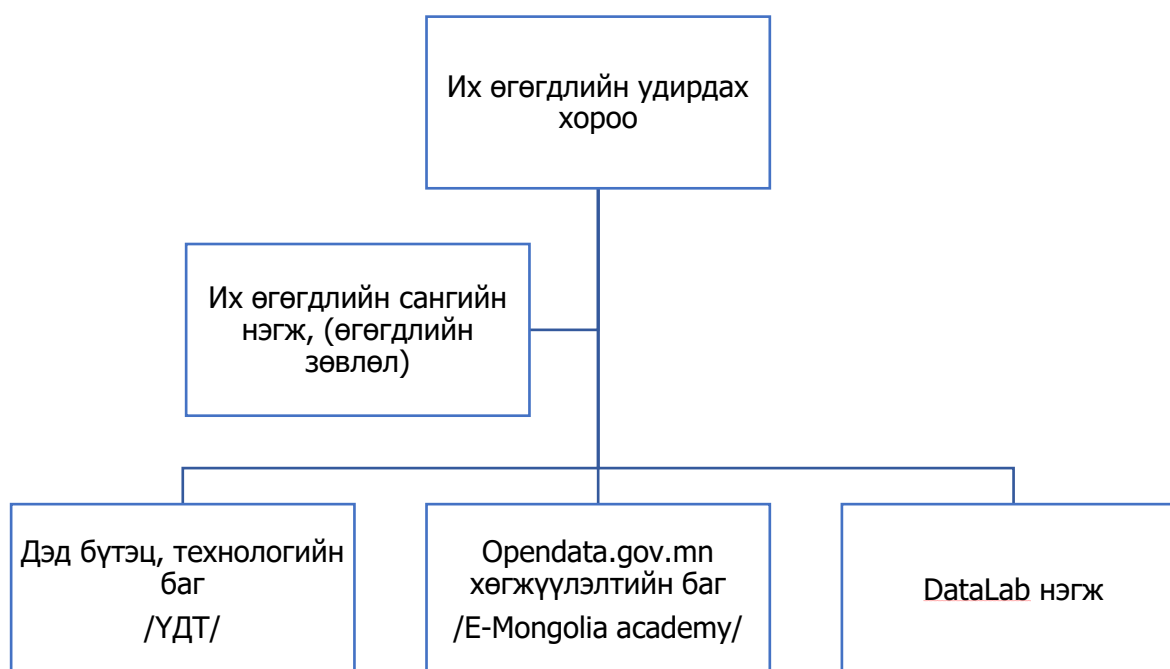
NVIDIA H100 / A100 / L40S GPU суурилуулсан Supermicro, Dell, HPE, Inspur зэрэг олон үйлдвэрлэгчийн серверүүд зах зээлд байгаа. Эдгээрээс DELL загварын серверийг жишиг болгон авч үзлээ:

	ҮЗҮҮЛЭЛТ
Брэнд	Dell PowerEdge XE9680 (6U)
GPU	8× NVIDIA A100 80 GB SXM4, NVLink
CPU	2× Intel Xeon Platinum 8470
Memory	64 GB DDR5 (өргөтгөх боломжтой)
Storage	8× NVMe U.2 bays
Network	100 GbE + 200 Gb/s HDR Infiniband
Үнэ	Ойролцоогоор 380 000 USD

БҮТЭЦ ЗОХИОН БАЙГУУЛАЛТ

Их өгөгдлийн санг бий болгох, өргөтгөх, хэрэглээнд нэвтрүүлэх нь нэг удаагийн үйл ажиллагаа бус байнга буюу тогтмол хийгдэх үйл ажиллагаа юм. Иймд энэхүү үйл ажиллагааг хариуцан ажиллах бүтэц, үйл ажиллагааны нэгж зайлгүй шаардлагатай юм.

Өмнөх бүлгүүдэд дурдсан судалгаанд үндэслэн Их өгөгдлийн сангийн үйл ажиллагааг хангахад дараах нэгжүүд, бүтэц зохион байгуулалт шаардлагатай.



Их өгөгдлийн удирдах хороо

Их өгөгдлийн санг бий болгох нь дан ганц яам, эсвэл аль нэг төрийн байгууллагын чиг үүргийн үйл ажиллагаа биш харин төрийн олон байгууллагуудыг хамарсан хамтын шийдэл, шийдвэр шаардсан өргөн хүрээний үйл ажиллагаа юм. Иймд Их өгөгдлийн сангийн үйл ажиллагааны хүрээнд шаардлагатай байгаа хамтын шийдвэр, дүрэм журам, үйл ажиллагааг нэгтгэн баталгаажуулах эрх бүхий бүтэц зайлшгүй шаардлагатай гэж үзэж байна.

Их өгөгдлийн удирдах хороо нь их өгөгдлийн сан болон мета мэдээллийн санг удирдах, хамгаалах, ашиглахтай холбоотой тулгамдсан асуудлыг шийдвэрлэх, бодлого журам батлах, түүний хэрэгжилтэд хяналт тавих чиг үүрэг бүхий орон тооны болон орон тооны бус гишүүдтэй, байнгын ажиллагаатай, салбар дундын зохицуулалт хийх чиг үүрэг бүхий бүтэц юм.

Цахим хөгжил, харилцаа холбооны яам тус Хороог байгуулах бөгөөд Их өгөгдлийн санд өгөгдөл нийлүүлж байгаа байгууллагууд болон их хэмжээний өгөгдөлтэй ажилладаг төрийн болон хувийн хэвшлийн байгууллагуудын төлөөллийг хорооны бүрэлдэхүүнд оруулж болно.

Гүйцэтгэх үүрэг:

- Өгөгдөл цуглуулах, нэгтгэх, хадгалах, ашиглах, чанарт хяналт тавих дүрэм, удирдамж зэрэг өгөгдлийн засаглалын бодлого, журмыг хэлэлцэх, шаардлагатай бол батлах
- Өгөгдлийн үнэн зөв, бүрэн бүтэн, найдвартай байдлыг хангуулахтай холбоотой шийдвэрийн хэрэгжилтэд хяналт тавих, хариуцлага тооцох асуудлаар холбогдох байгууллагад санал хүргүүлэх
- Өгөгдөл эзэмшигчдийн мэдээллийн санг удирдах, хадгалах, солилцоход гүйцэтгэх үүргийг тодорхойлж, түүнийг хэрэгжүүлэхэд дэмжлэг үзүүлэх
- Өгөгдлийн нэгдсэн стандарт, ангилал кодыг батлах
- Их өгөгдлийн сангийн нэгж болон холбогдох төрийн байгууллагуудаас өгөгдлийн удирдлага, зохицуулалттай холбоотой гаргасан санал хүсэлтийг судалж шийдвэрлэх
- Мета мэдээллийн сангийн удирдлага, зохион байгуулалтыг боловсронгуй болгох
- Өгөгдөл засаглахтай холбоотой гарч буй эрсдэлийг бууруулах, зөрүүтэй байдлыг шийдвэрлэх, нэгдсэн сангийн удирдлагын процессыг сайжруулах зэрэг асуудлуудыг шийдвэрлэх

Их өгөгдлийн сангийн нэгж

Их өгөгдлийн сангийн нэгж нь Их өгөгдлийн сангийн өдөр тутмын үйл ажиллагааг хангах чиглэл бүхий бие даасан нэгж байна. Цахим хөгжил, харилцаа холбооны яам уг нэгжийг байгуулах бөгөөд ҮДТ-н их өгөгдлийн дэд бүтэц, технологи хариуцсан баг, e-mongolia academy -н opendata.gov.mn хөгжүүлэлтийн баг болон дата лаб нэгжийг шууд удирдлагаар хангана.

Их өгөгдлийн сангийн нэгж нь захиалагчийн хэрэгцээ, хүсэлтэд үндэслэн судалгаа хийх, тайлан мэдээлэл гаргах, өгөгдлийн эх үүсвэрийг тодорхойлох, чанарыг баталгаажуулах, төрийн байгууллагуудаас өгөгдөл хүлээн авах, нэгтгэх үйл ажиллагааг зохион байгуулах зэрэг их өгөгдлийн санг удирдахад чиглэсэн үйл ажиллагаа явуулна.

Тус нэгж нь дэргэдээ Их өгөгдлийн санд өгөгдөл нийлүүлэгч байгууллагуудын төлөөллийг багтаасан Өгөгдлийн зөвлөлийг ажиллуулна. Энэхүү зөвлөлийн үүрэг нь Их өгөгдлийн санд цугларсан салбар бүрийн нийлмэл өгөгдлүүдээс нийтэд хүргэх боломжтой тайлан, график, дашбоардууд болон сервисүүдийг тодорхойлох, тэдгээрийг боловсруулах бодлогын даалгаварт санал өгөх байдлаар хамтран ажиллана. Их өгөгдлийн санд өгөгдлөө нийлүүлсэн байгууллагуудын төлөөлөл энэхүү зөвлөлд багтсан байгаа нь олон салбарыг хамарсан нийлмэл

тайлан болон сервисүүдийг хөгжүүлэх үйл явцад үнэтэй санал, зөвлөмж өгнө гэж үзэж байна.

Нэгжийн гүйцэтгэх үүрэг:

- Их өгөгдлийн сангийн үйл ажиллагаатай холбоотой холбогдох хууль, журмын төслийг боловсруулан батлуулах, хэрэгжүүлэх, үүнд:
 - Өгөгдөл цуглуулах, нэгтгэх, хадгалах, ашиглах дүрэм, удирдамж зэрэг өгөгдлийн засаглал
 - Өгөгдлийн үнэн зөв, бүрэн бүтэн, найдвартай байдлыг хангуулах, өгөгдлийн чанарт хяналт тавих
 - Их өгөгдлийн санд нэгтгэх байгууллага, өгөгдлийн санд тавих шалгуур
 - Бусад шаардлагатай журмууд
 - Их өгөгдлийн сан ашиглах эрх зүйн зохицуулалтууд
- Шалгуур хангасан өгөгдлийг журмын дагуу Их өгөгдлийн санд нэгтгэх ажлыг зохион байгуулах
- Өгөгдөл эзэмшигчдэд Их өгөгдлийн санд өгөгдөл нийлүүлэхэд дэмжлэг үзүүлэх
- Их өгөгдлийн сангийн нууцлал, аюулгүй байдлыг хамгаалах, хяналт тавих
 - Өгөгдөл засаглахтай холбоотой гарч буй эрсдэлийг тодорхойлох, түүнийг бууруулах, зөрүүтэй байдлыг илрүүлэх, арилгах
 - Шинэ эх үүсвэрээс ирсэн өгөгдлийг нэгтгэх МТ-ийн аргуудыг тодорхойлох, шаардлагатай бол өгөгдлийн хувиргалтыг хийх
 - Өгөгдөл нэгтгэх үйл явцыг дэмжихэд шаардлагатай архитектур, дэд бүтцийг төлөвлөх, нэвтрүүлэх (ашиглах систем болон технологи, өгөгдөл дамжуулах үйл явц, өгөгдлийн хадгалалт болон хамгаалалт гэх мэт)
 - Өгөгдөл нэгтгэх явцад гарсан алдаа, үл нийцэл, тэдгээрийг шийдвэрлэх арга замыг тодорхойлох, цэвэрлэх, хуулбарлах, баталгаажуулах зэрэг өгөгдлийн чанарыг хангах арга хэмжээг хэрэгжүүлэх
 - Өгөгдөл нэгтгэх үйл явцыг оновчтой болгох, шинэчлэх, сайжруулах зэрэг байнгын хяналт, засвар үйлчилгээний ажлыг гүйцэтгэх /шинэ технологи нэвтрүүлэх, ажлын урсгалыг боловсронгуй болгох, хэрэглэгчдийн санал хүсэлтийг нэгтгэх зэрэг/
 - Хэрэглэгчид, мэдээллийн технологийн баг, өгөгдлийн шинжээчид, удирдлагууд зэрэг оролцогч талуудтай хамтран Их өгөгдлийн санд ирсэн захиалгын дагуу өгөгдлийн эх үүсвэрийг тодорхойлж загварчлан, үнэлгээ дүгнэлт хийж өгөгдөлд суурилсан шийдвэр гаргалтыг дэмжих, BI тайлангууд хөгжүүлэх захиалгыг opendata.gov.mn хөгжүүлэлтийн багт өгөх, хэрэгжилтэд хяналт тавих
 - Дата лаб нэгжийн үйл ажиллагаанд дэмжлэг үзүүлэх
 - Их өгөгдлийн сангийн хэрэглээг нэмэгдүүлэх, үйл ажиллагааг олон нийтэд, төрийн байгууллагуудад түгээн таниулах
 - Мета өгөгдлийн стандарт, стратегийн зорилго, зорилтын хүрээнд хэрэгжүүлэх, бодлого, журмын төслийг боловсруулан батлуулах, хэрэгжүүлэх
 - Төрийн байгууллагуудаас нийлүүлсэн мета өгөгдлийн чанарт хяналт тавих
 - Мета өгөгдлийн санд хандах, ашиглах, өөрчлөх, менежментийн бодлогыг хэрэгжүүлэх

- Мета өгөгдлийн талаарх ойлголтыг нэмэгдүүлэх чиглэлээр хэрэглэгчид, сонирхогч талуудад сургалт зохион байгуулах, сурталчлах

Дэд бүтэц, технологийн баг

ҮДТ-д их өгөгдлийн сангийн үйл ажиллагаанд дэмжлэг үзүүлэх багийг байгуулах шаардлагатай. Тус баг нь Их өгөгдлийн сангийн дэд бүтцийн хэвийн найдвартай ажиллагааг хангах, тоног төхөөрөмжүүдэд өргөтгөл, засвар үйлчилгээ хийх болон шинээр нэмэгдэж буй өгөгдөл нийлүүлэгч байгууллагуудын холболт, өгөгдөл дамжуулахтай холбоотой бүхий л үйл ажиллагаанд дэмжлэг үзүүлэн ажиллана. Хүний нөөцийн хувьд хамгийн багадаа 6 мэргэжилтэн шаардлагатай.

АЛБАН ТУШААЛ / ЧИГЛЭЛ	TOO	АЖИЛ ҮҮРГИЙН ХУВААРЬ
Data Engineer	2	ETL, өгөгдөл дамжуулалт, цэвэрлэгээ
DBA	2	Oracle, Cloudera, Hadoop зэрэг систем хариуцна
Data Architect	1	Архитектур, өгөгдлийн зохион байгуулалт
Security Specialist	1	DLP, Firewall, IAM, SIEM хариуцна

Их өгөгдлийн сангийн хэмжээ, хэрэглээ нэмэгдэхтэй уялдуулан технологийн багийг өргөтгөх, шаардлагатай бол биш даасан нэгж газар болгон өргөтгөх шаардлагатай. Энэхүү төсөлд Их өгөгдлийн санг байгуулахдаа төрд одоо бэлэн байгаа дэд бүтцийг ашиглах, түүнийг өргөтгөх замаар буюу бага зардлаар үйл ажиллагааг эхлүүлэх, бэлэн байгаа хүний нөөцөд тулгуурлан хэрэгжүүлэх арга замуудыг авч үзэж байгаа болно.

Opendata.gov.mn хөгжүүлэлтийн баг

Их өгөгдлийн санд цугларсан өгөгдлийг хэрэглээнд оруулах, олон нийт, судлаачдад ашиглах боломжийг бий болгох зайлшгүй шаардлагатай. Энэ хүрээнд Opendata.gov.mn системийг өргөтгөх, Их өгөгдлийн сангийн өгөгдлүүдийг тархаах гол суваг болгон ашиглах нь оновчтой.

Монгол Улсын нээлттэй өгөгдлийн нэгдсэн платформ болох opendata.gov.mn нь төрийн байгууллагуудын цуглуулсан өгөгдлийг олон нийтэд ил тод, хүртээмжтэй болгох зорилготой үндэсний хэмжээний цахим систем юм. Энэхүү портал нь өгөгдлийн ил тод байдал, иргэдийн хяналт, дижитал засаглалыг дэмжих үндсэн хэрэгсэл болж байна.

Төрийн байгууллагуудын өдөр тутмын үйл ажиллагааны явцад үүссэн олон төрлийн өгөгдлийг төвлөрүүлэн хадгалах, нийтэд нээлттэй болгох зорилгоор 2015 онд энэхүү платформыг хөгжүүлж эхэлсэн бөгөөд одоо 100 гаруй байгууллагын өгөгдөлтэй болж, нээлттэй өгөгдлийн экосистемийг бий болгоод байна. Тэгвэл Opendata.gov.mn одоогийн үндсэн үйлчилгээ буюу хуулийн хүрээнд төрийн байгууллагуудын нээлттэй өгөгдлөө байршуулах нэгдсэн систем байхын зэрэгцээгээр их өгөгдлийн сан буюу төрийн байгууллагуудын нэгдсэн нээлттэй өгөгдлийн сангаас тайлан хүснэгт, дашбоардууд, нээлттэй сервисүүдийг хөгжүүлж олон нийт болон судлаачдад нийлүүлэх шаардлага бий болж байна.

Энэхүү системийн өргөтгөл, хөгжүүлэлтийг хийх багийг E-Mongolia academy-г түшиглэн байгуулах нь Их өгөгдлийн санг байгуулах, хэрэглээнд оруулах эхэн үеийн зөв шийдэл гэж үзэж байна. Цаашид дотоодын компаниудтай хамтран хөгжүүлэлтийг хийх, бүрэн хэмжээгээр аутсорс хийх зэргээр өргөжих боломжтой.

Хөгжүүлэлтийн багийг 11 мэргэжилтэнтэй байхаар тооцлоо.

АЛБАН ТУШААЛ / ЧИГЛЭЛ	TOO	АЖИЛ ҮҮРГИЙН ХУВААРЬ
-----------------------	-----	----------------------

Систем шинжээч	2	Системийн судалгаа, захиалга хүлээн авах, бодлогын даалгавар боловсруулах
System developer	4	Систем хөгжүүлэлт
Data Architect	1	Архитектур, өгөгдлийн зохион байгуулалт
BI Analyst / Dashboard Developer	2	Tableau, Power BI, Superset
Metadata / Data Catalog Specialist	1	Collibra / Apache Atlas
AI/ML Specialist	1	Python, R, машин сургалт

DataLab нэгж

Их өгөгдлийн сангийн хэрэглээг нэмэгдүүлэх, судлаачид болон олон нийтэд нээлттэй болгож мэдээллээр үйлчлэх үйлчилгээг DataLab байгуулах замаар хүргэх нь ач холбогдолтой юм. Үйл ажиллагааны хувьд хэрэглэгч өөрийн судалгаа, шинжилгээнд ашиглагдах тусгайлсан өгөгдлийг захиалах, тусгайлсан орчинд өөрийн судалгаа шинжилгээ, загварын туршилтыг хийж гүйцэтгэн үр дүнгээ авах боломжуудыг олгодгоороо ач холбогдолтой юм.

Их өгөгдлийн сан баяжихын хирээр их өгөгдлийн санд хийж болох судалгаа шинжилгээний боломж асар ихээр нэмэгдэнэ. Их өгөгдлийн санд хуримтлагдсан өгөгдөл, үзүүлэлт бүрийг тайлан хүснэгт, дашбоард болгон хөгжүүлж олон нийтэд түгээх боломжгүй тул хэрэглэгчдэд их өгөгдлийн сангийн катологиос өөрт шаардлагатай датасетийг бэлтгүүлэн шаардлагатай судалгаа шинжилгээ хийх, үр дүнг боловсруулах боломжийг бий болгох шаардлагатай. Англи, солонгос зэрэг олон улс орнууд төрийн өгөгдлийг ашиглах DataLab байгуулсан сайн туршлагууд олон байдаг.

Их өгөгдлийн сангийн дэргэд байгуулагдах DataLab нь бие даасан нэгж байх бөгөөд 7 мэргэжилтэнтэйгээр үйл ажиллагаа явуулна.

ЧИГЛЭЛ	АЛБАН ТУШААЛ	АЖИЛ ҮҮРГИЙН ХУВААРЬ
1. Удирдлага, зохион байгуулалт	Manager -1 Project Coordinator -1	Судлаачдын хүсэлтийг хүлээн авах, ажил зохион байгуулах
2. Судалгаа, дүн шинжилгээ	Data Analyst -2 Data Scientist -2	Бие дааж болон судлаачидтай хамтран анализ хийх, туршилтын загвар боловсруулах
3. Өгөгдлийн хандалт, инженерчлэл	Data Engineer -2	Судлаачдын ажиллах dataset-үүдийг бэлдэх
4. Нууцлал, аюулгүй байдал	Security Officer / Data Privacy Officer -1	Хандалтын хяналт, зөвшөөрөл олгох, өгөгдөл гадагшлахгүй байхыг хянах
5. Систем дэмжлэг, дэд бүтэц	System Admin / Lab IT Technician –2	Тусгаарлагдсан орчныг (sandbox VM, docker, authentication, firewall) бэлдэх, судалгаанд ашиглах (R, Python, SPSS г.м) системүүдийн бэлэн байдлыг хангах
6. Хэрэглэгчийн дэмжлэг	Research Assistant / Tech Support -1	Судлаачдад заавар өгөх, системийн хэрэглээний тусламж үзүүлэх

7. Хууль эрх зүй	Legal Advisor - 1	Хандалтын гэрээ, өгөгдлийн эзэмшигчийн эрхийн хамгаалалт хэрэгжүүлэх
------------------	-------------------	--

ХУУЛЬ ЭРХ ЗҮЙ, БОДЛОГО ШИЙДВЭР

Их өгөгдлийн сангийн үйл ажиллагааг хариуцах бүтцийн нэгжүүдийг байгуулж албажуулах, тэдгээрийн чиг үүргийг зааж өгнө.

- Их өгөгдлийн удирдах хороог байгуулах, үйл ажиллагааг тогтмолжуулах
- Их өгөгдлийн сангийн нэгжийг байгуулах
 - Өгөгдлийн зөвлөлийг байгуулах, үйл ажиллагааг тогтмолжуулах
- Үндэсний дата төв болон e-mongolia academy их өгөгдлийн багийг ажиллуулах эрхзүйн зохицуулалтыг хийх
- DataLab байгуулах, орон тоо, ажил үүргийн хуваарь, төсвийг батлах

Их өгөгдлийн сан байгуулах нь тасралтгүй үргэлжлэх үйл явц бөгөөд энэхүү үйл ажиллагаанд шаардлагатай дараах үйл ажиллагааны журмуудыг бий болгох шаардлагатай. Энэ хүрээнд их өгөгдлийн санд ашиглагдаж буй ангилал кодын нийцтэй байдлыг хангах, мета өгөгдлүүдийн нийтлэг стандартыг мөрдүүлэх, болон өгөгдлийн нууцлал аюулгүй байдлыг хангаж Их өгөгдлийн санд өгөгдөл дамжуулах үйл ажиллагааг журамлах хэрэгтэй.

- Төрийн байгууллагууд Их өгөгдлийн санд иргэн, ААНБ-ын эмзэг мэдээллийг илгээхгүй байх, түлхүүр талбарын утгыг de-Identication хийж, танигдахгүй болгох үйл ажиллагааны журам батлах, хэрэгжүүлэх
- Мета өгөгдлийн каталог, ангиллыг бий болгож Мета өгөгдлийн сангийн стандартыг боловсруулж батлуулах, төрийн байгууллагуудад хэрэгжүүлэх
- Их өгөгдлийн санд нэгдэх байгууллагуудын бэлэн байдлыг үнэлэх аргачлал боловсруулж батлуулах, төрийг байгууллагуудыг их өгөгдлийн санд нэгдэх бэлэн байдлыг хангуулах
- Их өгөгдлийн санд ашиглаж байгаа ангилал код, стандартыг эргэн харж сайжруулах, хэрэгжилтэд хяналт тавих
- Их өгөгдлийн сангийн үйл ажиллагаа, ашиглалтын журам баталж мөрдүүлэх, сангийн ашиглалт, хэвийн найдвартай ажиллагааг хангах
- Их өгөгдлийн сангийн ашиглалтыг нэмэгдүүлэх, олон нийтэд сурталчлан таниулах. Мөн тодорхой хэмжээний үйлчилгээг олон нийт болон төрийн бусад байгууллагуудад нээх эрхзүйн орчин бий болгох

ҮЙЛ АЖИЛЛАГАА

Бэлэн байдлыг хангах

- Тоног төхөөрөмж
Хамгийн оновчтой хувилбар буюу Үндэсний дата төвийн үйл ажиллагааг түшиглэн Их өгөгдлийн сангийн техник орчныг байгуулах, цаашид өргөтгөх ажлыг үе шаттайгаар хэрэгжүүлнэ. Үүнд сүлжээ болон шаардлагатай дэд бүтцийн асуудал шийдэгдсэн ба сервер, хадгалах төхөөрөмж болон программ хангамжуудын лицензийн төлбөрийг нэмэлтээр авч үзлээ.

Одоогийн байдлаар 3 байгууллагын raw data цуглуулахад Үндэсний дата төвд 500GB орчим зай бэлтгээд байгаа бөгөөд цаашид өгөгдөл цуглуулах, тайлан дашбоард хөгжүүлэх болон, datawarehouse зохион байгуулах, bigdata шийдлүүд, машин сургалт зэрэг үйл ажиллагаануудын хэрэглээ, хэрэгцээ шаардлагаас

ЦЕГ, Мэдээлэл шуурхай удирдлагын төв	258 GB
Усны газар	1 TB
Цаг уур, орчны шинжилгээний газар	80 TB
Төрийн өмчийн бодлого, зохицуулалтын азар	500 GB
Татварын ерөнхий газар	20 TB
Газар зохион байгуулалт, геодези, зураг зүйн газар	620 GB
Эрүүл мэндийн хөгжлийн төв	1 TB
Улсын бүртгэлийн ерөнхий газар	8TB
Үндэсний статистикийн хороо	3TB
Хөдөлмөр, халамжийн үйлчилгээний ерөнхий газар	800 GB
ebarimt	80 TB

хамааран өгөгдөл болон серверийн хүчин чадлын өргөтгөл шууд хамаарна.

Суурь болон төрөлжсөн мэдээллийн сангуудаас Их өгөгдлийн санг бүрдүүлэх шатны ажил эхлэх бөгөөд бидэнд ямар хэмжээний нөөц бэлтгэх шаардлагатай вэ гэдэг асуулт тулгарна.

Төрд байгаа мэдээллүүдийг ерөнхийд нь авч үзвэл мастер болон transaction (бичилт) суурьтай төрлийн сангуудаас бүрдэнэ гэж үзэж болохоор байна. Мастер гэдэг нь суурь бүртгэлүүд буюу хаяг, утас, хариуцлагын хэлбэр, байдал, захирал, үйлчилгээний салбар, гэрчилгээний сунгалт, үнэмлэх шинэчлэл гэх мэт үзүүлэлтүүдийг агуулсан харьцангуй тогтвортой сангууд байх юм. Жишээ нь ААНБ бүртгэл, иргэний мэдээллийн сан. Харин transaction суурьтай сангууд нь үйлчилгээ авах бүрд бичилт хийгдэх буюу илүү хурдтай

өргөжих шинж чанартай сангууд юм. Жишээ нь ebarimt, татварын ерөнхий газар, гаалийн сангуудыг үүнд хамруулж болно.

Бид дараах том сангууд, тэдгээрийн хэмжээг ойролцоогоор тооцоолсны үндсэн дээр Их өгөгдлийн санг байгуулахад эхний байдлаар шаардагдах нөөцийг гаргалаа:

* төрийн байгууллагуудын мэдээллийн санг үнэлэх судалгаа, YCX

ҮЗҮҮЛЭЛТ	ТОО ХЭМЖЭЭ
Анхны өгөгдөл (одоо)	195 TB
5 жилийн өсөлт (~20%/жил)	~300 TB
Васкир давхар хадгалалт (3 хуулбар)	900 TB
Архив, лог, системийн өгөгдөл	~100 TB
Нийт хэрэгцээ (5 жил)	~1 PB

Жилийн өсөлтийг 20% гэж үзвэл бидэнд ойролцоогоор 1 petabyte орчим өгөгдөлтэй ажиллах орчныг бүрдүүлэх шаардлагатай болно.

Гэхдээ өгөгдөл цуглуулах дүрмийг хэрэгцээ шаардлагад нийцүүлэн урьдчилан тооцож ETL-г хөгжүүлэх шаардлагатай. Жишээ нь ебаримт системийг өгөгдлүүдийг ELT буюу шууд задгайгаар нь Их өгөгдлийн санд нэгтгээд дараа нь datalake\datawarehouse-с тайлан, боловсруулалт хийх нь их хэмжээний өгөгдлийг үр ашиг багатайгаар хуулбарлах, нөөцийг муу ашиглах байдалд хүргэж болзошгүй. Иймд ебаримт зэрэг их хэмжээний өгөгдлийг Их өгөгдлийн санд нийлүүлэхийн өмнө хэрэгцээ шаардлагыг урьдчилан сайн тооцоолсны үндсэн дээр шаардлагатай боловсруулалтуудыг хийж Их өгөгдлийн санд нийлүүлэх хэрэгтэй. Жишээ нь ебаримт системээс өдөр бүрийн задгай тоо татахын оронд байгууллага бүрээр борлуулсан бүтээгдэхүүн бүрийн нийт борлуулсан тоо, нийт үнийн дүнг татах нь илүү ач холбогдолтой байж болох юм.

Зохион байгуулалтын хувьд үндсэн дата төв дээр master node, data node-н 5 сервер байх бол DR(Жишээ нь дарханы дата төв)-т 2 compute node байхаар тооцлоо. Энэ нь хэрвээ үндсэн төвд тасалдал үүссэн тохиолдолд сэлгэн ажиллах боломжтой юм.

Primary Site(Үндэсний дата төв)

ТӨРӨЛ	ЗАГВАР (DELL)	ТОО	НЭГЖ ҮНЭ	НИЙТ
Master Node	Dell PowerEdge R750, 2×Xeon Gold 6330, 256GB RAM, 2×960GB NVMe, 4×4TB SAS	2	\$22,000	\$44,000
Data Node (Storage сервер)	Dell PowerEdge R750xd, 2×Xeon Gold 6330, 256GB RAM, 12×14TB SAS	3	\$28,000	\$84,000
Primary Storage Array	Dell PowerVault ME4084, 84×14TB NL-SAS (≈1.176 PB raw)	1	\$180,000	\$180,000
Backup Storage Array	Dell PowerVault ME4024, 24×14TB NL-SAS (≈336 TB raw)	2	\$55,000	\$110,000

DR Site(Дархан нөөц төв)

ТӨРӨЛ	ЗАГВАР (DELL)	ТОО	НЭГЖ ҮНЭ	НИЙТ
Compute Node	Dell PowerEdge R750, 2×Xeon Gold 6330, 256GB RAM, 2×960GB NVMe	2	\$21,000	\$42,000
DR Storage Array	Dell PowerVault ME4084, 84×14TB NL-SAS	1	\$180,000	\$180,000
Backup Storage	Dell PowerVault ME4024, 24×14TB NL-SAS	1	\$55,000	\$55,000

Сервер.

Сервер нь дата төвийн зориулалтаар ашиглагддаг, өндөр ачаалал даах зориулалт бүхий ямар ч брэндийн сервер байж болох байх бөгөөд бид жишээ болгон Dell PowerEdge R750 серверийг онцоллоо. Төслийн хүрээнд 7 сервер ашиглагдана гэж тооцсон ба үнийн хувьд нэг сервер ойролцоогоор 22'000 USD.

Processor	Up to two 3rd Generation Intel Xeon Scalable processors with up to 32 cores	
Memory	• 16 DDR4 DIMM slots, supports RDIMM 1 TB max, speeds up to 3200 MT/s • Supports registered ECC DDR4 DIMMs only	
Drive Bays	Front bays: • 0 drive bay • Up to 8 x 3.5-inch SAS/SATA (HDD/SSD) max 128 TB • Up to 12 x 3.5-inch SAS/SATA (HDD/SSD) max 192 TB • Up to 8 x 2.5-inch SAS/SATA/NVMe (HDD/SSD) max 61.44 TB • Up to 16 x 2.5-inch SAS/SATA (HDD/SSD) max 122.88 TB • Up to 16 x 2.5-inch (SAS/SATA) + 8 x 2.5-inch (NVMe) (HDD/SSD) max 184.32 TB Rear bays: • Up to 2 x 2.5-inch SAS/SATA/NVMe (HDD/SSD) max 15.36 TB	
Power Supplies	• 600W Platinum Mixed Mode (100-240Vac or 240Vdc) hot swap redundant • 800W Platinum Mixed Mode (100-240Vac or 240Vdc) hot swap redundant • 1100W Titanium Mixed Mode (100-240Vac or 240Vdc) hot swap redundant • 1400W Platinum Mixed Mode (100-240Vac or 240Vdc) hot swap redundant • 1100W -48Vdc hot swap redundant (CAUTION: only works with -48Vdc to -60Vdc power input)	
Cooling options	Air cooling	
Fans	• Standard fans/High performance Silver fans/High performance Gold fans • Up to six hot swap fans	
Dimensions	• Height – 86.8 mm (3.41 inches) • Width – 482.0 mm (18.97 inches) • Depth – 707.78 mm (27.85 inches) – without bezel 721.62 mm (28.4 inches) – with bezel	
Form Factor	2U rack server	
Embedded Management	• iDRAC9 • iDRAC Service Module • iDRAC Direct • Quick Sync 2 wireless module	
OpenManage Software	• OpenManage Enterprise • OpenManage Power Manager plugin • OpenManage SupportAssist plugin • OpenManage Update Manager plugin	
Mobility	OpenManage Mobile	
Embedded NIC	2 x 1 GbE LOM	
Network Options	1 x OCP 3.0	
GPU Options	Not supported	
PCIe	Up to 5 x PCIe Gen4 slots + 1 x PCIe Gen3 slots • 3 x16 Gen4 low profile • 2 x16 Gen4 low profile (optional) • 1 x8 Gen3 (x4 lane) low profile	
Operating System and Hypervisors	Canonical Ubuntu Server LTS • Citrix Hypervisor • Microsoft Windows Server with Hyper-V • Red Hat Enterprise Linux • SUSE Linux Ent	
OEM-ready version available	From bezel to BIOS to packaging, your servers can look and feel as if they were designed and built by you.	
Integrations and Connections	• BMC Truesight • Microsoft System Center • OpenManage Integrations • Red Hat Ansible Modules • VMware vCenter and vRealize Operations Manager	• IBM Tivoli Netcool/OMNIBus • IBM Tivoli Network Manager IP Edition • Micro Focus Operations Manager • OpenManage Connections • Nagios Core • Nagios XI
Ports	Front Ports • 1 x iDRAC Direct (Micro-AB USB) port • 1 x USB 2.0 • 1 x VGA	Rear Ports • 1 x USB 2.0 • 1 x Serial (optional) • 1 x USB 3.0 • 2 x Ethernet • 1 x VGA

Хадгалах төхөөрөмж.

Жишиг хадгалах төхөөрөмж:

Dell PowerVault ME4084, 84×14TB NL-SAS (≈1.176 PB raw)

CATEGORY	SPECIFICATION
Модел	Dell PowerVault ME4084
Хадгалах багтаамж (Raw)	84 × 14TB = 1.176 PB
Дискний төрөл	NL-SAS 7.2K RPM, 3.5"
Диск интерфейс	12Gb/s SAS
Удирдлагын модуль	Dual Controller (High Availability)
Cache санах ой	16–32 GB per controller (моделийн хувилбараас хамаарна)
RAID дэмжлэг	RAID 0, 1, 5, 6, 10, 50, 60
Холболт	SAS 12Gb/s Host Connectivity (up to 8 ports per controller)
Өргөтгөх боломж	Нэг системд 336 диск хүртэл (ME484 shelf ашиглан)
Form Factor	5U rack mount
Эрчим хүчний хэрэглээ	~1500W (Fully loaded)
Жин	~110 кг (дисктэй)
MTBF	1.2–1.6 сая цаг (дисктэй)
Удирдлага ба мониторинг	Dell PowerVault Manager (Web UI), CLI, REST API
Data Services	Thin provisioning, snapshots, replication (Sync/Async), tiering
Дэмжлэгтэй OS/Platform	Windows, Linux, VMware, Hyper-V, XenServer

Программ хангамж

Программ хангамж, лицензтэй холбоотой тооцоог хийхдээ 2025 онд 5 хүртэлх төрийн байгууллагын өгөгдлийн санд нэгтгэнэ. 2025 онд bigdata шийдлийг суурилуулж эхлэх бөгөөд өгөгдөл, тайлан, дашбоардын хөгжүүлэлтийг дараах байдлаар тооцлоо.

Системүүд

№	СИСТЕМ / ПРОГРАММ ХАНГАМЖ	ЗОРИУЛАЛТ
1	Cloudera Data Platform (CDP Private Cloud Base)	Hadoop/Spark/Data Lake кластер зохион байгуулах
2	Tableau Server (on-prem)	Data visualization / Dashboard
3	Backup Solution (Commvault эсвэл Veeam)	Backup, replication
4	Apache Airflow (OSS + Enterprise Support)	ETL pipeline orchestration
5	Red Hat Enterprise Linux + Kubernetes	OS ба кластер менежмент
6	Splunk Enterprise	Monitoring tool
7	Collibra	Data governance tooling
8	DevOps ба Аналитик багийн хэрэгсэл	Систем хөгжүүлэлтийн хэрэгслүүд

ОН	ТӨСӨВ (USD)	АШИГЛАГДАХ ЛИЦЕНЗ, СИСТЕМҮҮД
2025	25,000	- Tableau Server (5 Creator, 20 Viewer) - Cloudera CDP (4 core/node) - Airflow OSS (support plan) - Red Hat (5 node)
2026	60,000	- Tableau Viewer нэмэгдэж 80 хүрнэ - Cloudera 2x node нэмэгдэнэ (8 core/node) - Backup solution (Veeam) эхлүүлнэ - Red Hat нэмэгдэнэ
2027	90,000	- Tableau Viewer 120, Creator 10 - Cloudera кластер тэлэлт - Backup өргөтгөл - DevOps хэрэгсэл (Jira, Git, IDE) ашиглаж эхэлнэ
2028	130,000	- Tableau Viewer 160, Creator 15 - Full-scale Cloudera - Veeam full DR copy - Red Hat & Kubernetes өргөжсөн дэд бүтэц дээр
2029	165,000	- Tableau Viewer 180+, Creator 20 - Cloudera кластер 3x болсон байна - DevOps tooling өргөтгөл - Monitoring (Splunk Enterprise) + DR tooling
2030	190,000	- Tableau Viewer 200, Creator 25 - Бүх байгууллага нэгдсэн Cloudera кластерт холбогдсон байна - Backup, DR бүрэн хосолсон - Data governance tooling (Collibra) нэмэгдсэн байна

Үүнд:

DevOps нь программ хангамжийг илүү хурдан, чанартай, найдвартай хөгжүүлж, холбогдох туршилтуудыг хийн ашиглалтад оруулах (deploy) үйл ажиллагааг зохион байгуулах шийдэл юм. Энэхүү шийдлийг хэрэгжүүлэх технологийн олон сонголтуудыг хөгжүүлэгчийн багийн зүгээс хийж болно.

Жишээ нь:

ЗОРИУЛАЛТ	СИСТЕМҮҮД
Source Code Management	Git, GitHub, GitLab
CI/CD (тасралтгүй интеграци, хүргэлт)	Jenkins, GitLab CI, CircleCI
Containerization	Docker
Container Orchestration	Kubernetes
Monitoring	Prometheus, Grafana
Configuration Management	Ansible, Puppet, Chef
Cloud Platforms	AWS, Azure, Google Cloud

DevOps-г амжилттай хэрэгжүүлснээр хөгжүүлэгч кодын өөрчлөлтийг оруулахад л түүний тест, системийн шинэчлэлтүүд

илүү автоматажсан орчинд системийн хяналт дор хийгдэж хэрэглээнд нэвтрэх боломжтой болно.

Нийт тоног төхөөрөмж болон лицензийн зардал(2025-2030)

ЗОРИУЛАЛТ	ҮНЭ USD
Лиценз, системүүд	660,000
Master site: Сервер, storage, тоног төхөөрөмжүүд	418,000
DR site: Сервер, storage, тоног төхөөрөмжүүд	277,000
DataLab: Сервер, storage, тоног төхөөрөмжүүд	165,000
GPU	380,000
НИЙТ	1,900,000

*Их өгөгдлийн санг байгуулах үе шат, нэгтгэх өгөгдлийн хэмжээ зэргээс хамааран тоног төхөөрөмж болон системүүдийн худалдан авах үйл ажиллагааг үе шаттайгаар зохион байгуулж болох юм. Бид 1PB өгөгдлийг хэрэглээнд нэвтрүүлэхэд шаардагдах жишиг тоног төхөөрөмжийн тооцооллыг боловсруулсан болно.

- Хүний нөөц, түүнийг чадавхжуулах

Их өгөгдлийн сангийн нэгж

Их өгөгдлийн санг байгуулахдаа Үндэсний дата төвөөс системийн админ, кластер инженерүүдийг гаргах бол e-Mongolia академид дашбоард тайлан хөгжүүлэгчид, тайлан шинжээчдийг ажиллуулна гэж төлөвлөсөн нь төслийг хамтарсан мэргэшсэн багаар богино хугацаанд хэрэгжүүлэх зорилготой юм. Энэ тохиолдолд Их өгөгдлийн сангийн нэгж нь эрхзүйн зохицуулалт, дотоод зохион байгуулалт, стандартчилал, дата менежмент, чанар, засаглал гэсэн чиглэлд илүү төвлөрч ажиллах юм.

Удирдлага (2 хүн)

АЛБАН ТУШААЛ	АЖИЛ ҮҮРГИЙН ХУВААРЬ
Нэгжийн дарга	Хамтын ажиллагааны уялдаа, байгууллагуудтай дата солилцоо, стратеги төлөвлөлт
Төслийн зохицуулагч / РМ	Ажлын хуваарь, төлөвлөгөө, байгууллагын хоорондын гүйцэтгэл, тайлагнал

Өгөгдлийн засаглал, чанар, зохион байгуулалт (4 хүн)

АЛБАН ТУШААЛ	АЖИЛ ҮҮРГИЙН ХУВААРЬ
Өгөгдлийн засаглал хариуцсан мэргэжилтэн	Collibra ашиглан data catalog, owner, steward тогтоох
Өгөгдлийн чанар хариуцсан мэргэжилтэн	Өгөгдлийн profiling, cleansing, хуучин/шинэ өгөгдлийн стандарт

Metadata хариуцсан мэргэжилтэн	Метадата бүрдүүлэлт, lineage mapping, entity relationship
Хууль, нууцлал хариуцсан мэргэжилтэн	Хуулийн нийцэл, нууцлал, хандалтын бодлого хянах

Шинжээч (3 хүн, e-Mongolia академийн BI developer, шинжээчидтэй хамтран ажиллана, захиалгат ажил хүлээн авахаас opendata.gov.mn системд байрших хүртэлх харилцааг авч явна)

АЛБАН ТУШААЛ	АЖИЛ ҮҮРГИЙН ХУВААРЬ
Business Data Analyst	Байгууллагуудын дашбоард/тайлангийн хэрэгцээг тодорхойлж ажлын захиалга болгох
Data Integration Analyst	Үндэсний дата төвийн дата урсгалыг хянаж, ingestion pipeline-г зохицуулах
Report Request Coordinator	Захиалгын бүртгэлжүүлэх, хөгжүүлэлттэй холбоотой үзүүлэлт хоорондын уялдааны хяналт тавих

Monitoring ба дүн шинжилгээ (2 хүн, Splunk, Tableau usage, system health зэргийг хянах)

АЛБАН ТУШААЛ	АЖИЛ ҮҮРГИЙН ХУВААРЬ
Monitoring Analyst (Splunk)	Usage, query load, security event хяналт
BI Quality Analyst	Tableau дээр үүссэн тайлангуудын стандарт, үзүүлэлт хянах

Дата лаборатори

Дата лаборатори нь олон нийтэд нээлттэй үйлчилгээ үзүүлэх бие даасан нэгж байна. Тус нэгж нь Их өгөгдлийн сангаас хэрэглэгчдийн хэрэгцээ шаардлагад нийцсэн датасетүүдийг бэлтгэх, тэрхүү өгөгдөлд судлаачдад судалгаа шинжилгээ хийх лаборатори орчноор хангах болон судалгааны үр дүнг харин судлаачид эзэмших боломжтой их өгөгдлийн сангийн дэргэдэх үйлчилгээний нэгж юм.

Тус нэгж нь иргэн судлаачдад шаардлагатай өгөгдөл бэлтгэж нийлүүлэхийн зэрэгцээгээр өөрийн мэргэшсэн багт тулгуурлан Их өгөгдлийн сангийн нэгж болон бусад төрийн байгууллагуудын хүсэлтийн дагуу шаардлагатай судалгаа шинжилгээ хийх, туршилт хийх, ML болон AI хөгжүүлэх зэрэг үйл ажиллагаануудыг явуулдаг байна.

ЧИГЛЭЛ	АЛБАН ТУШААЛ	АЖИЛ ҮҮРГИЙН ХУВААРЬ
1. Удирдлага, зохион байгуулалт	Data Lab Lead -1 Project Coordinator -1	Судлаачдын хүсэлтийг хүлээн авах, ажил зохион байгуулах
2. Судалгаа, дүн шинжилгээ	Data Analyst -2 Data Scientist -2	Бие дааж болон судлаачидтай хамтран

		анализ хийх, туршилтын загвар боловсруулах
3. Өгөгдлийн хандалт, инженерчлэл	Data Engineer -2	Судлаачдын ажиллах dataset-үүдийг бэлдэх
4. Нууцлал, аюулгүй байдал	Security Officer / Data Privacy Officer -1	Хандалтын хяналт, зөвшөөрөл олгох, өгөгдөл гадагшлахгүй байхыг хянах
5. Систем дэмжлэг, дэд бүтэц	System Admin / Lab IT Technician –2	Тусгаарлагдсан орчныг (sandbox VM, docker, authentication, firewall) бэлдэх, судалгаанд ашиглах (R, Python, SPSS г.м) системүүдийн бэлэн байдлыг хангах
6. Хэрэглэгчийн дэмжлэг	Research Assistant / Tech Support -1	Судлаачдад заавар өгөх, системийн хэрэглээний тусламж үзүүлэх
7. Хууль эрх зүй	Legal Advisor - 1	Хандалтын гэрээ, өгөгдлийн эзэмшигчийн эрхийн хамгаалалт хэрэгжүүлэх

Систем админ, кластер инженерүүдийн баг(Үндэсний дата төв)

ЧИГЛЭЛ	TOO	АЖИЛ ҮҮРГИЙН ХУВААРЬ
Data Engineer	2	ETL, өгөгдөл дамжуулалт, цэвэрлэгээ
DBA	2	Oracle, Cloudera, Hadoop зэрэг систем хариуцна
Data Architect	1	Архитектур, өгөгдлийн зохион байгуулалт
Security Specialist	1	DLP, Firewall, IAM, SIEM хариуцна

Систем хөгжүүлэгч болон шинжээчдийн баг(e-mongolia академи)

ЧИГЛЭЛ	TOO	АЖИЛ ҮҮРГИЙН ХУВААРЬ
Систем шинжээч	2	Системийн судалгаа, захиалга хүлээн авах, бодлогын даалгавар боловсруулах
System developer	4	Систем хөгжүүлэлт
Data Architect	1	Архитектур, өгөгдлийн зохион байгуулалт
BI Analyst / Dashboard Developer	2	Tableau, Power BI, Superset
Metadata / Data Catalog Specialist	1	Collibra / Apache Atlas
AI/ML Specialist	1	Python, R, машин сургалт

Сургалт, чадавхжуулах арга хэмжээ.

Хүний нөөцийн бэлэн байдлыг хангах, тэдгээрийн мэргэшсэн байдлыг хангах, сургалт хөтөлбөрүүдэд хамрагдуулах нь технологи хурдацтай хувьсан өөрчлөгдөж буй энэ цаг үед зайлгүй анхаарах ёстой асуудал мөн юм.

Мэргэжилтнүүдийг 2030 он хүртэлх төлөвлөгөөтэй уялдуулан дараах байдлаар сургалтад хамруулахаар тооцлоо.

Үүнд:

- Үндэсний дата төвийн админууд (SysAdmin, DevOps)
- Их өгөгдлийн нэгжийн гишүүд (Data Engineer, Data Steward, Analyst)
- e-Mongolia Academy – тайлан, шинжилгээ, автоматжуулалт хариуцсан хөгжүүлэгчид
- Data Lab – судлаачдыг дэмжих багууд хамрагдана.

2025 он

СУРГАЛТ	МЭРГЭЖИЛТЭН	СЭДЭВ
Cloudera CDP Essentials	Data engineer, Admin	Cloudera Manager, HDFS, Hive, Impala
Tableau Fundamentals	Dashboard developer	Tableau Creator role, data blending
RHEL Administration Basics	SysAdmin	Systemctl, SELinux, user/group, firewall
Airflow OSS Introduction	Data engineer	DAG-үүд, таск зохиох, execution monitoring

2026

СУРГАЛТ	МЭРГЭЖИЛТЭН	СЭДЭВ
Cloudera Admin Advanced	SysAdmin	Kerberos, Ranger, кластер scale-up
Tableau Server Scaling	Tableau admin	Viewer expansion, performance tuning
Veeam Backup Basics	SysAdmin	Job planning, restore, retention
Red Hat Intermediate	SysAdmin	Network config, storage, SELinux advanced

2027

СУРГАЛТ	МЭРГЭЖИЛТЭН	СЭДЭВ
Tableau Advanced Visualizations	Dashboard developer	Parameters, LOD, actions
Git + Jira for Data Teams	Analyst, DevOps	Agile backlog, branching workflow
Python for Data Pipelines	Data engineer	Pandas, Airflow таск development
Veeam DR Workshop	SysAdmin	DR scenario, offsite copy, testing

2028

СУРГАЛТ	МЭРГЭЖИЛТЭН	СЭДЭВ
Kubernetes for Admins	SysAdmin	Pod, deployment, ingress
Full-scale Cloudera Tuning	Data engineer, Admin	Performance optimization
Tableau Dashboard UX	Dashboard developer	Dashboard design, user feedback
RHEL + Kubernetes Security	SysAdmin	RBAC, audit log, pod/network policies

2029

СУРГАЛТ	МЭРГЭЖИЛТЭН	СЭДЭВ
Splunk Enterprise Admin	SysAdmin	Data input, forwarders, alerting
Monitoring for Analysts	Analyst, DevOps	Log analysis, KPI alert
DevOps Toolchain Deep Dive	DevOps engineer	CI/CD, issue automation
DR Readiness Simulation	SysAdmin	Full backup/restore exercise

2030

СУРГАЛТ	МЭРГЭЖИЛТЭН	СЭДЭВ
Collibra for Data Stewards	Data steward, Analyst	Business glossary, data catalog
Data Governance & Privacy (ISO/GDPR)	Бүх бар	Data lifecycle, access policy
Tableau Governance & Security	Tableau admin	Row-level security, permissions
Cloudera Multi-tenant Management	Admin	Shared clusters, access control

ИХ ӨГӨГДЛИЙГ ХЭРЭГЛЭЭНД НЭВТРҮҮЛЭХ

- Opendata.gov.mn дээр их өгөгдлийн тайлангуудыг тусгах

Дээр дурдсанчлан Opendata.gov.mn системд шинээр нэмэгдэх тайлан дашбоард, сервисүүдийн ажлын даалгаврыг тендер шалгаруулах замаар буюу гүйцэтгэгчтэй гэрээлэн эхний удаад зохион байгуулна. Хүснэгт, сервис зэрэг их өгөгдлийн хэрэглээ нь цаашид тасралтгүй өсөн нэмэгдэх үйл ажиллагаа тул тэр бүр хөндлөнгийн байгууллагаар гүйцэтгүүлэх нь оновчгүй юм. Иймд шинээр өгөгдлийг хэрэглээнд оруулах буюу Opendata.gov.mn системд их өгөгдлийн сангийн тайлан хүснэгт, дашбоард болон их өгөгдлийн нээлттэй сервисүүд байршуулах дэд бүлгийг хөгжүүлэх ажлыг дотооддоо захиалга хүлээн авах, судлах, хөгжүүлэх, сурталчлан таниулах үйл ажиллагааг хэрэгжүүлэх хэрэгтэй. Энэ хөгжүүлэлтийн хүрээнд их өгөгдлийн сангийн нэгж, өгөгдлийн зөвлөл хамтран Их өгөгдлийн санд цугларсан байгаа өгөгдлүүдээс салбар хоорондын нийлмэл

хүснэгт, тайлан болон дашбоардууд боловсруулах ажлын захиалгыг боловсруулах шаардлагатай.

Тус захиалгыг e-mongolia академийн шинжээчдийн багтай хамтран бодлогын даалгавар болгон эцэслэж хөгжүүлэлтийг хийлгэнэ. Ийнхүү их өгөгдлийн сангаас гарах тайлан хөгжүүлэлтийн тасралтгүй үйл ажиллагааны эхлэл тавигдах бөгөөд энэ үйл ажиллагааны хүрээнд:

- Их өгөгдлийн сангаас шинээр тайлан, мэдээ боловсруулах үйл ажиллагааны заавар, үйл ажиллагааны журам боловсруулагдан гарна.
 - Захиалгыг ямар маягтын дагуу хаана хэрхэн хүсэлт гаргах
 - Захиалгын хүсэлтийг боловсруулах боломжтой эсэх баталгааг шинжээчдийн баг хэрхэн албажуулах
 - Захиалгыг бодлогын даалгаварт хэрхэн тусгах
 - Системийн хөгжүүлэлт, туршилтыг хэрхэн гүйцэтгэх
 - Opendata.gov.mn системд шинээр нэмэгдсэн их өгөгдлийн шинэ тайланг хэрхэн захиалагчид болон олон нийтэд мэдэгдэх, сурталчлан таниулах гэсэн үйл ажиллагаануудыг хэрхэн гүйцэтгэх талаар үйл ажиллагааны журамд тусгасан байх шаардлагатай.

- Дата лаборатори байгуулах

Их өгөгдлийн хэрэглээг нэмэгдүүлэх, түүний ач холбогдлыг олон нийтэд сурталчлан таниулах үйл ажиллагаанд судлаачдын оролцоо, тэдний их өгөгдлийн санд тулгуурлан хийсэн аливаа судалгаа, шинжилгээ болон иновацууд ихээхэн ач холбогдолтой юм.

Судлаачдыг ажиллах бололцоогоор хангасан их өгөгдлийн задгай өгөгдөлд шинжилгээ хийх зориулалтын дата лаборатори байгуулах шаардлагатай. Энэ ажлын хүрээнд:

- Их өгөгдлийн санд нийлүүлсэн өгөгдлийн катологи
- Өгөгдлийн нууцлал аюулгүй байдлыг хангах үйл ажиллагааны болон өгөгдөл бэлтгэх үйл ажиллагааны журам
- Өгөгдөл хүсэх маягт, судалгаа шинжилгээний орчны хүсэлт зэрэг шинжээчид өгөгдөл хүсэх үйл ажиллагаа, орчин бэлэн болгох зэрэг дотоод үйл ажиллагааны журам
- Судалгааны үр дүнг тархаах, судлаач зөвшөөрсөн тохиолдолд opendata.gov.mn системийн их өгөгдлийн тайлангийн хэсэгт хэвлэн нийтлэх зэргээр олон нийтийн хүртээл болгох
- Төрийн болон бусад байгууллагуудаас захиалгат судалгааны хүсэлт хүлээн авах, ML болон AI системүүдийг сургах, хөгжүүлэх талаарх үйл ажиллагаанууд хийгдэж, журамлагдана.

- Олон нийтэд сурталчлан таниулах

Их өгөгдлийн сангийн хэрэглээг олон нийтэд сурталчлан таниулах нь энэхүү төслийн чухал бүрэлдэхүүн хэсэг юм. Сайн чанартай, өргөн салбарыг хамарсан өгөгдөл хэзээд хэрэгцээтэй байсаар ирсэн. Иймд өгөгдлийн хэрэглээ нэмэгдэнэ гэдэг нь мэдээлэл цуглуулах, нэгтгэх,

тайлагнах бүхий л үе шатуудын тасалдалгүй зөв үйл ажиллагааны үр дүнд бий болох нь дамжиггүй. Нэгдсэн өгөгдлийн сан нь хэд хэдэн онцлог шинжүүдийг агуулсан байна:

- Олон салбарын мэдээллийг агуулсан байна
- Салбар хоорондын уялдаа, баланс хангагдсан байна.
- Ангилал кодуудын нийцтэй байдал хангагдсан байна.
- Өгөгдлийн тухай метадата бүрэн бүрдсэн байна.

Их өгөгдлийн сангийн хэрэглээ нэмэгдэхийн хирээр өгөгдлийн анхан шатны бүртгэлийн чанар сайжрах, салбар хоорондын өгөгдлийн уялдаа сайжрах, үзүүлэлтийн давхардал арилах, үзүүлэлтүүдийн ангилал кодын стандарт хангагдах зэрэг эерэг үйл явц бүхий л өгөгдөл хариуцагч талуудад бий болно.

Хэд хэдэн чиглэлээр Их өгөгдлийн санг олон нийтэд сурталчлан таниулах нь үр дүнтэй гэж үзлээ:

Иргэдэд чиглэсэн үйл ажиллагаа

ҮЙЛ АЖИЛЛАГАА	ТАЙЛБАР
"Өгөгдлийн нээлттэй өдөр" (Open Data Day)	Жил бүрийн 1 удаа зохион байгуулж, олон нийтэд дата лаб, аналитик дашбоардуудыг танилцуулах, судлаачдын туршлага хуваалцах
Инфографик + дашбоард сурталчилгаа	Нийтлэг хэрэглээтэй өгөгдлийг (жишээ нь ажилгүйдэл, агаарын чанар, боловсролын чанар гэх мэт) визуалчлан олон нийтэд ойлгомжтой байдлаар хүргэх
Podcast / YouTube контент	"Өгөгдөл юу өгүүлдэг вэ?" сэдэвтэй цуврал подкаст, аналитик хийсэн судалгааг тайлбарласан видеонууд хийх
Сургуулиудад чиглэсэн хөтөлбөр	Их, дээд сургуулийн оюутнуудад дата шинжилгээний боломжийг таниулах мастер класс, тэмцээн

Судлаачдад чиглэсэн үйл ажиллагаа

ҮЙЛ АЖИЛЛАГАА	ТАЙЛБАР
"Data Science Challenge"	Тодорхой асуудлыг (жишээ нь орон нутгийн шилжилт хөдөлгөөн зэрэг) өгөгдөлд үндэслэн судалж, шилдэг шийдлийг уралдуулах
Dataset-аар ажиллах богино сургалтууд	Судлаачдад зориулсан 1–2 өдрийн "data lab"-ын хэрэглээний сургалт зохион байгуулах
Мэдээлэл түгээх цуврал нийтлэлүүд	"Өгөгдлөөр өгүүлэх нь" сэдэвтэй 2 долоо хоног тутмын нийтлэл, кейс судалгаа, жишээ түүхүүд хэвлэл, блог, сошиал медиагаар хүргэх
Судалгааны тэтгэлэгт уралдаан	Их өгөгдлийн санг ашиглаж хийсэн судалгаанд урамшуулал олгож, үр дүнг олон нийтэд сурталчлах

Олон нийтийн оролцоог дэмжихэд чиглэсэн үйл ажиллагаа

ҮЙЛ АЖИЛЛАГАА	ТАЙЛБАР
---------------	---------

Hackathon буюу өгөгдлийн марафон	Залуу дата аналит, хөгжүүлэгчдийг татан оролцуулж, асуудал шийдвэрлэх уралдаант ажил
Гишүүнчлэл бүхий "Open Data Community"	Сайн дурын судлаачид, дата сонирхогчдыг хамарсан бүлгэм байгуулж, хамтын хөгжлийг дэмжих
QR код бүхий дата постерууд	Хотын гудамж, сургууль, төрийн байгууллагуудад дата-г ойлгомжтой тайлбарласан постерууд байрлуулж, дэлгэрэнгүйг QR кодоор үзүүлдэг болгох

АШИГЛАСАН БАРИМТ, ЭХ СУРВАЛЖУУДЫН ЖАГСААЛТ

I. Монгол Улсын хууль, эрх зүйн баримт бичиг

1. Хүний хувийн мэдээлэл хамгаалах тухай хууль (Шинэчилсэн найруулга, 2021 он)
2. Нээлттэй мэдээллийн ил тод байдлын тухай хууль (2011 он)
3. Монгол Улсын Засгийн газрын 2023 оны 403 дугаар тогтоол
"Төрийн байгууллагын мэдээллийн сангуудыг нэгтгэн, өгөгдлийн уялдаа холбоо, хүртээмжийг сайжруулах арга хэмжээний тухай"
4. Монгол Улсын Мэдээллийн технологийн үндэсний бодлого (2018 он)

II. Олон улсын стандарт, удирдамж

5. ISO/IEC 27001:2013 – Information security management systems
6. ISO/IEC 38505-1:2017 – Governance of Data
7. General Data Protection Regulation (GDPR) – European Union Regulation 2016/679
8. OECD Privacy Guidelines (2013)
9. DCAT (Data Catalog Vocabulary) стандарт – W3C Recommendation
10. Open Government License (Canada)

III. Олон улсын туршлага, бодлогын баримт бичиг

Эстони

11. X-Road Data Exchange Framework
<https://x-road.global>
12. Public Information Act (RT I 2000, 92, 597)
<https://www.riigiteataja.ee/en/eli/ee/Riigikogu/act/530102013007/consolide>
13. Estonia Open Data Portal
<https://avaandmed.eesti.ee>

Канад

14. Government of Canada Data Strategy
<https://www.canada.ca/en/government/system/digital-government/data/government-canada-data-strategy.html>
15. Open Government Portal Canada
<https://open.canada.ca/en>
16. Privacy Act (Canada)
<https://laws-lois.justice.gc.ca/eng/acts/p-21/>

Сингапур

17. Personal Data Protection Act (PDPA)
<https://www.pdpc.gov.sg/Overview-of-PDPA>
18. Singapore Government Tech Stack (SGTS)
<https://tech.gov.sg>
19. Model AI Governance Framework (Singapore)
<https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework>

IV. Технологийн баримт бичиг, бүтээгдэхүүний reference

20. Apache Atlas Documentation
<https://atlas.apache.org>
21. Collibra Data Governance Center
<https://www.collibra.com>
22. Apache NiFi Documentation
<https://nifi.apache.org/docs.html>
23. Talend Data Integration Documentation
<https://help.talend.com>
24. Apache Airflow Documentation
<https://airflow.apache.org/docs/>

25. Informatica PowerCenter Documentation
<https://www.informatica.com>
 26. Apache Kafka Documentation
<https://kafka.apache.org/documentation/>
 27. MinIO Documentation
<https://min.io/docs/>
 28. Microsoft Power BI Documentation
<https://learn.microsoft.com/power-bi>
 29. Tableau Product Documentation
<https://help.tableau.com>
 30. Cloudera Data Platform Documentation
<https://docs.cloudera.com>
- V. Судалгаа, онолын материалууд
31. Data Lake vs Data Warehouse Architecture – Snowflake Whitepaper
 32. The Data Lake Manifesto – Cloudera
 33. Privacy Preserving Data Mining – ACM Computing Surveys
 34. Metadata Management Best Practices – Gartner
 35. Cloud Cost Benchmarking 2023 – IDC