

№3 Дәріс. Деректерді жинау және сақтау әдістері

Дәрістің мақсаты:

Бұл дәрістің мақсаты — деректерді жинау көздерін, әдістерін және оларды сақтау технологияларын түсіндіру, сонымен қатар заманауи ақпараттық жүйелерде қолданылатын деректерді ұйымдастыру тәсілдері мен құралдарын таныстыру.

Кілт сөздер:

Деректерді жинау, сақтау, деректер көзі, сенсорлар, лог файлдар, дерекқор, NoSQL, SQL, Data Warehouse, Data Lake, ETL, API, Big Data, Cloud Storage, IoT, real-time data, stream processing.

Дәріс жоспары:

1. Кіріспе
2. Деректерді жинау тәсілдері
 - 2.1 Қолмен жинау
 - 2.2 Автоматтандырылған жинау
 - 2.3 Дерек көздерінің түрлері
3. Деректерді сақтау әдістері
 - 3.1 Реляциялық дерекқорлар (SQL)
 - 3.2 Реляциялық емес дерекқорлар (NoSQL)
 - 3.3 Data Warehouse жүйелері
 - 3.4 Data Lake жүйелері
 - 3.5 Бұлттық сақтау технологиялары
4. ETL және Stream Processing
5. Қауіпсіздік және деректер сенімділігі
6. Қорытынды
7. Бақылау сұрақтары
8. Пайдаланылған әдебиеттер

1. Кіріспе

Қазіргі цифрлық экономикада деректер кез келген ұйымның, мемлекеттің және ғылыми ортаның стратегиялық ресурсына айналды. «Деректер — жаңа мұнай» деген пікір ақпарат дәуірінде нақты шындыққа айналды. Қол жеткізілетін ақпарат көлемі күн сайын геометриялық прогрессиямен өсуде: әлеуметтік желілер, мобильді құрылғылар, IoT сенсорлар, бизнес жүйелер, банктер, мемлекеттік қызмет жүйелері үздіксіз дерек генерациялайды. IDC зерттеуіне сәйкес, әлемдегі деректер көлемі 2025 жылға қарай 175 зеттабайттан асады деп болжануда.

Осындай көлемдегі деректерді тиімді жинау, сақтау, құрылымдау және талдау — ұйымдардың бәсекеге қабілеттілігі мен табыстылығының негізгі факторы. Деректерді дұрыс ұйымдастырмаған жағдайда олар құнсыз әрі

жарамсыз болады, сондықтан заманауи ақпараттық жүйелерде деректерді басқару архитектурасы басты рөл атқарады.

Заманауи деректер экожүйесі тек сақтау ғана емес, сонымен қатар нақты уақыт режимінде өңдеу, қауіпсіздік, резервтеу және қолжетімділікті қамтамасыз етуді талап етеді. SQL және NoSQL жүйелері, Data Warehouse және Data Lake платформалары, ETL және Streaming технологиялары — бәрі де деректерді тиімді пайдалану стратегиясының ажырамас бөлігі.

Бұл дәрісте біз деректерді жинау тәсілдерін, дерек көздерінің түрлерін, заманауи сақтау технологияларын және үздік тәжірибелерді қарастырамыз. Сонымен қатар нақты уақыттағы деректер ағындарын өңдеу, үлкен деректер инфрақұрылымдарын құру және бұлттық платформалардың рөлін талқылаймыз. Материал цифрлық трансформация кезеңіндегі білім мен дағдыларды меңгеруге бағытталған.

2. Деректерді жинау тәсілдері

2.1 Қолмен жинау

Деректер адамның тікелей қатысуымен жиналады:

- Сауалнама, анкета, интервью
- Форма толтыру
- Әлеуметтік зерттеулер
- Қолмен енгізілетін реестрлер

Артықшылығы: нақтылық, түсіндіру мүмкіндігі

Кемшілігі: баяу, қателікке бейім, үлкен көлемге жарамайды

2.2 Автоматтандырылған жинау

Бағдарламалық немесе аппараттық құралдар арқылы:

- Сенсорлар, датчиктер (IoT)
- Жүйелік логтар
- Веб-трекинг (cookies, аналитика)
- API арқылы дерек алу
- Web-scraping
- ERP/CRM жүйелері

Артықшылықтары: жылдам, дәл, үлкен көлемде

Кемшіліктері: инфрақұрылым қажет, техникалық күрделілік

2.3 Дерек көздерінің түрлері

Дерек көзі	Мысал
Құрылымдалған	ДҚ кестелері
Жартылай құрылымдалған	JSON, XML, log файлдар
Құрылымсыз	мәтін, аудио, видео
IoT сенсорлар	температура, қозғалыс, GPS
Әлеуметтік желілер	посттар, пікірлер
Корпоративтік жүйелер	ERP, CRM

3. Деректерді сақтау әдістері

3.1 Реляциялық дерекқорлар (SQL)

- MySQL
- PostgreSQL
- Oracle
- MS SQL Server

Ерекшелігі: кесте құрылымы, SQL сұраныстар

Артықшылығы: транзакция сенімділігі, формальды құрылым

3.2 NoSQL дерекқорлар

Түрі	Мысалы	Ерекшелігі
Document DB	MongoDB, CouchDB	күжат түрінде
Key-Value DB	Redis, Riak	кэш, real-time
Column DB	Cassandra, HBase	Big Data
Graph DB	Neo4j	граф байланыстар

Артықшылығы: икемді, үлкен деректерге функционалды

3.3 Data Warehouse (DWH)

Талдау жұмыстарын орындауға арналған қойма мысалдары:

- Amazon Redshift
- Google BigQuery
- Snowflake
- Microsoft Azure Synapse

Мақсаты: өңделген деректер, аналитика

3.4 Data Lake

Түрлі форматтағы шикі деректерді сақтайды.

Мысалы:

- Hadoop HDFS
- AWS S3

Артықшылығы: Big Data, AI/ML үшін қолайлы

3.5 Бұлттық сақтау технологиялары

- Google Cloud Storage
- Amazon S3
- Azure Blob Storage

Артықшылығы: масштабтау, сенімділік, қауіпсіздік

4. ETL және Stream Processing

ETL (Extract-Transform-Load)

Деректерді шығару — түрлендіру — жүктеу.

Құралдар:

- Apache Airflow
- Talend
- Informatica
- SSIS

Ағындық өңдеу (Stream Processing)

Нақты уақыттағы деректерді өңдеу құралдары:

- Apache Kafka
- Apache Spark Streaming
- Apache Flink

5. Деректер қауіпсіздігі және сенімділігі

- Шифрлау
- Бэкап
- Деректер репликациясы
- Қол жеткізуді басқару
- GDPR/ISO стандарттары

6. Қорытынды

Деректерді тиімді жинау және сақтау — ақпараттық жүйелердің табысты жұмысының негізі. Қазіргі таңда кез келген салада: медицинада, білімде, өндірісте, қаржы секторында және мемлекеттік басқаруда — деректерге негізделген шешім қабылдау үрдісі қалыптасты. Жинақталған деректер дұрыс өңделген жағдайда ұйымдарға стратегиялық артықшылық береді, процестерді автоматтандыруға, шығынды азайтуға, тәуекелді төмендетуге және жаңа қызмет түрлерін дамытуға мүмкіндік береді.

Дәрісте қарастырған SQL және NoSQL жүйелері, Data Warehouse және Data Lake архитектуралары, бұлттық платформалар, ETL және Stream Processing технологиялары — заманауи деректер инфрақұрылымының негізгі компоненттері. Әр технология өзінің міндетіне сәйкес қолданылады:

SQL — құрылымдалған деректер үшін, NoSQL — икемді және үлкен ауқымдағы жүйелер үшін, Data Warehouse — аналитика үшін, Data Lake — машиналық оқыту және Big Data үшін тиімді.

Деректерді басқару тек техникалық процесс емес, ол сонымен қатар ақпараттық қауіпсіздік, деректердің құпиялылығы, қолжетімділігі, сенімділігі және заңнамалық талаптарды (GDPR, ISO 27001, ҚР цифрлық қауіпсіздік нормалары) сақтау міндеттерін қамтиды.

Болашақта жасанды интеллект, IoT, 5G және киберфизикалық жүйелердің дамуына байланысты деректер көлемі мен талаптары одан әрі өседі. Сондықтан деректермен жұмыс дағдылары — XXI ғасырдың маңызды кәсіби компетенцияларының бірі. Деректерді дұрыс ұйымдастырып, өңдеп, қолдана білген маман еңбек нарығында әрқашан сұранысқа ие болады.

7. Бақылау сұрақтары

1. Деректерді жинау тәсілдері қандай?
2. Автоматтандырылған деректер жинаудың артықшылықтары қандай?

3. SQL және NoSQL арасындағы айырмашылық неде?
4. Data Warehouse мен Data Lake ерекшелігін түсіндіріңіз.
5. ETL процесінің кезеңдерін атаңыз.
6. Ағындық деректерді өңдеу қай кезде қолданылады?

8. Пайдаланылған әдебиеттер

1. Silberschatz A., Korth H., Sudarshan S. *Database System Concepts*.
2. Martin Kleppmann. *Designing Data-Intensive Applications*.
3. O'Reilly. *Streaming Systems*.
4. Apache Hadoop, Kafka, Spark ресми құжаттамалары
5. ISO/IEC 27001 қауіпсіздік стандарты