СӘТБАЕВ УНИВЕРСИТЕТІ

SATBAYEV UNIVERSITY

Статистика және деректерді талдау

Ержан Асел Ануаркызы,Ph.D., «Ақпараттық жүйелер» кафедрасының қауымдастырылған профессоры

Assel.yerzhan@satbayev.university

№3. Жиілік кестелері және деректерді визуализациялау



Ақпараттық жүйелерге кіріспе

Деректер – қазіргі цифрлық қоғамның негізі. Бұл бөлімде ақпарат пен деректің мәнін, деректер базасының рөлін және деректер модельдерінің жіктелуін қарастырамыз. Ақпараттық жүйелерді дұрыс түсіну – кез келген кәсіби маманның базалық құзыреттілігі болып табылады.



Ақпарат дегеніміз не?

Деректер (data) – өңделмеген фактілер мен цифрлар жиынтығы. Ақпарат (information) – осы деректерді мағыналы контекстке орналастырып, шешім қабылдауға негіз беретін өңделген нәтиже.

1 Деректер

Өңделмеген сандар, мәтіндер, суреттер – мағынасыз жеке элементтер.

2 Ақпарат

Деректерді талдау арқылы алынған мағыналы, пайдалы білім.

3 Қабылдау процесі

Жинау → Өңдеу → Талдау → Беру тізбегі арқылы жүзеге асады.

Деректер базасының мақсаты

Ұзақ мерзімді сақтау

Деректерді жылдар бойы сенімді, қауіпсіз және структурированный түрде сақтауды қамтамасыз етеді. Деректер жоғалмайды және бүлінбейді.

Тиімді өңдеу

Миллиондаған жазбалар арасынан қажетті ақпаратты секундтар ішінде іздеп, сұрыптап, өңдеу мүмкіндігін береді.

Көп қолданушылы жүйе

Бір мезетте жүздеген пайдаланушының деректерге қатынасуын транзакциялардың тұтастығын сақтай отырып ұйымдастырады.



Деректердің негізгі компоненттері

Логикалық байланыстар

Деректер арасындағы мағыналық қатынастар – бірінші кестедегі жазба екінші кестедегімен қалай байланысатынын анықтайды. Бұл байланыстар бизнес-ережелерді бейнелейді.

Сыртқы жадта ұйымдастыру

Деректер физикалық деңгейде дискіде блоктар, беттер және кластерлер түрінде сақталады. Индекстеу жылдам іздеуге мүмкіндік береді.

Метадеректер

Деректер туралы деректер – кестелер, бағандар, типтер мен шектеулерді сипаттайтын жүйелік каталог.



Деректер модельдерін жіктеу

Деректер базасын ұйымдастырудың үш негізгі тәсілі тарихи дамып, әрқайсысы өзіндік артықшылықтарымен ерекшеленеді.

Иерархиялық (ағаштәріздес) модель

Деректер «ата-ана → бала» қатынасымен ағаш түрінде ұйымдастырылады. IBM IMS жүйесі – классикалық мысал. Қарапайым, бірақ икемсіз.

Желілік модель

Деректер графтық құрылымда байланысады, бір баланың бірнеше ата-анасы болуы мүмкін. Иерархиялыққа қарағанда икемдірек, бірақ күрделірек.

Реляциялық модель

Деректер кестелерде сақталады, SQL арқылы басқарылады. Қазіргі кезде ең кең таралған модель – Oracle, PostgreSQL, MySQL негіздейді.



Реляциялық деректер базасы

Реляциялық модель – деректер базасының алтын стандарты. Edgar F. Codd ұсынған бұл теория 1970 жылдан бері деректер базасы индустриясының іргетасын қалайды. Кестелер, кілттер, байланыстар – бұл бөлімнің өзегі.

Реляциялық модель негіздері

Кестелік құрылым

Реляциялық модельде барлық деректер **қатарлар (жолдар)** мен **бағандардан** тұратын кестелерде сақталады. Әр кесте – бір мәні бар мән жиыны (relation).

- Қатар (tuple) – бір жазба
- Баған (attribute) – бір өріс
- Ұяшық – нақты мән

Атрибут және домен

Атрибут – кестенің бағаны, белгілі бір сипатты білдіреді. **Домен** – атрибуттың қабылдай алатын мүмкін мәндер жиыны. Мысалы, «Жас» атрибутының домені: 0–150 бүтін сан.

Домен анықтамасы деректердің дұрыстығын қамтамасыз ету үшін маңызды.

Кортеж және кардиналдылық

1:1

Бір-бірге байланыс

Бір жазба тек бір жазбамен байланысады. Мысалы: адам – жеке куәлік.

1:K

Бір-көпке байланыс

Бір жазба көптеген жазбалармен байланысады. Мысалы: тапсырыс – тауарлар.

K:K

Көп-көпке байланыс

Екі жақтан да бірнеше байланыс болуы мүмкін. Аралық кесте арқылы шешіледі.

Кортеж – кестедегі бір жол (жазба). Кардиналдылық – кестедегі жолдардың жалпы саны. Дәреже (degree) – бағандардың саны. Бұл ұғымдар реляциялық алгебраның математикалық негізін құрайды.

Семантикалық модель (ER)



Субъект-байланыс моделі

ER (Entity-Relationship) модель – деректер базасын жобалаудың визуалды тілі. Peter Chen ұсынған бұл модель реалды дүниені субъектілер, атрибуттар және байланыстар арқылы бейнелейді.

01

Субъектілерді анықтау

Жүйедегі негізгі объектілер мен ұғымдар (Студент, Курс, Оқытушы).

02

Атрибуттарды белгілеу

Әр субъектінің қасиеттері мен сипаттамалары анықталады.

03

Байланыстарды белгілеу

Субъектілер арасындағы қатынастар мен кардиналдылық.

Реляциялық алгебра

Реляциялық алгебра – кестелермен жұмыс істеудің математикалық аппараты. SQL тілінің теориялық негізі осы алгебраға сүйенеді.

Таңдау (Selection)

Белгілі шартқа сай келетін жолдарды сүзгілеу операциясы. σ(шарт) символымен белгіленеді.

Проекция (Projection)

Кестеден белгілі бағандарды ғана таңдау. π(атрибуттар) символымен белгіленеді.

Біріктіру (Join)

Екі кестені ортақ баған бойынша біріктіру. Ең жиі қолданылатын операция.

Функциялық тәуелділік

X→Y: X мәні белгілі болса, Y мәні бірмәнді анықталады. Нормализацияның негізі.

Қалыпты формалар

Нормализация – деректер базасын артықшылықтардан (redundancy) және аномалиялардан тазарту процесі. Кезең-кезеңмен жүзеге асырылады.

1

1NF – Бірінші қалыпты форма

Атомдық мәндер: әр ұяшықта тек бір мән болуы керек.

2

2NF – Екінші қалыпты форма

Толық функциялық тәуелділік: кілттің бөліктеріне ішінара тәуелділік жойылады.

3

3NF – Үшінші қалыпты форма

Транзитивті тәуелділіктер жойылады: А→Б→В тізбегі бөлінеді.

4

BCNF – Бойс-Кодд формасы

3NF-тен күшті форма: барлық функциялық тәуелділіктер супер-кілттен туындауы керек.

SQL тілі – Деректерді басқару құралы

SQL (Structured Query Language) – реляциялық деректер базасымен жұмыс істеудің әмбебап тілі. 1974 жылы IBM зерттеушілері әзірлеген бұл тіл бүгінде деректер ғылымының негізгі құралы болып саналады.



SQL тіліне шолу

Декларативті тіл ерекшелігі

SQL – **декларативті** тіл: сіз *не* алғыңыз келетінін айтасыз, бірақ *қалай* алу керектігін жүйе өзі шешеді. Бұл программалық тілдерден (C++, Python) түбегейлі айырмашылығы.

Мысалы: `SELECT аты FROM қызметкерлер WHERE жалақы > 500000` – бұл сұраныс жалақысы 500 000-нан жоғары барлық қызметкерлердің атын қайтарады.

Негізгі қасиеттері

- Платформаға тәуелсіз стандарт
- Оқуға жеңіл, ағылшын тіліне жақын синтаксис
- ISO/IEC стандарты бойынша сертификатталған
- Барлық негізгі ДББЖ-де қолданылады

SQL операторларының жіктелуі



DDL

Data Definition Language – Деректерді анықтау тілі. CREATE, ALTER, DROP командалары арқылы кесте құрылымын басқарады.



DML

Data Manipulation Language – Деректерді манипуляциялау тілі. INSERT, UPDATE, DELETE, SELECT командалары.



DCL

Data Control Language – Деректерді басқару тілі. GRANT, REVOKE командалары арқылы рұқсаттарды реттейді.

Кесте құрылымын басқару (DDL)

CREATE TABLE

Жаңа кесте жасау командасы. Баған атаулары, деректер типтері және шектеулер (constraints) анықталады:

```
CREATE TABLE Студент (
 ID INT PRIMARY KEY,
 Аты VARCHAR(50),
 Туған_күні DATE,
 Курс INT DEFAULT 1
);
```

ALTER және DROP

ALTER TABLE – бар кестенің құрылымын өзгерту: жаңа баған қосу, типті өзгерту немесе шектеу орнату.

```
ALTER TABLE Студент
ADD Email VARCHAR(100);

DROP TABLE Студент;
```

⚠ DROP TABLE командасы деректерді қайтарымсыз жояды!

Деректерді таңдау (SELECT)

SELECT – SQL-дің ең жиі қолданылатын командасы. Бір немесе бірнеше кестеден деректерді сүзгілеп, сұрыптап, топтастырып шығарады.

Қарапайым сұраныс

```
SELECT Аты, Жалақы
FROM Қызметкерлер
ORDER BY Жалақы DESC;
```

WHERE шарты

```
SELECT *
FROM Студент
WHERE Курс = 3
  AND Баға > 3.5;
```

Жиынтық функциялар

```
SELECT Бөлім,
  COUNT(*) AS Саны,
  AVG(Жалақы) AS Орташа
FROM Қызметкер
GROUP BY Бөлім;
```

Деректерді өзгерту (DML)

INSERT – Қосу

Кестеге жаңа жол (жазба) қосады. Мәндер бағандар тәртібімен немесе атаулармен берілуі мүмкін.

```
INSERT INTO Студент
VALUES (101, 'Алия', '2003-05-10',
2);
```

UPDATE – Жаңарту

Бар жазбалардың мәндерін өзгертеді. WHERE шартынсыз барлық жолдар өзгереді – абай болыңыз!

```
UPDATE Студент
SET Курс = 3
WHERE ID = 101;
```

DELETE – Жою

Шартқа сай жолдарды кестеден жояды. Транзакция ішінде орындалса, ROLLBACK арқылы қайтаруға болады.

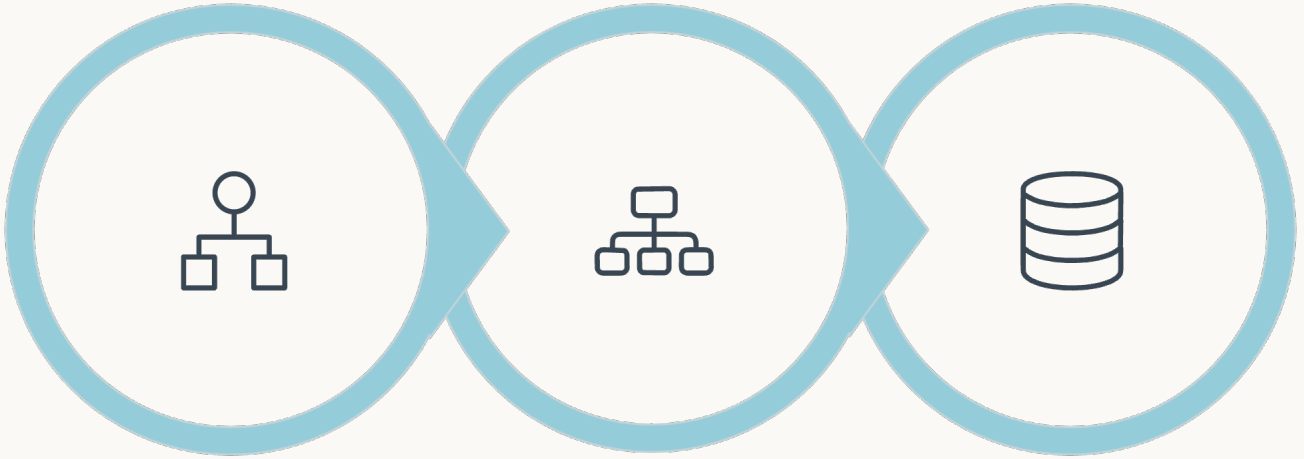
```
DELETE FROM Студент
WHERE ID = 101;
```



Жобалау және басқару

Деректер базасын жобалау – техникалық және концептуалды шеберлікті талап ететін процесс. Дұрыс жобаланған база жылдар бойы тиімді жұмыс істейді, ал қателер кейін қымбатқа түседі.

Деректер базасын жобалау циклдары



Концептуал Логикалық Физикалық

Концептуалды кезеңде бизнес талаптары жиналып, ER диаграммасы жасалады. Логикалық кезеңде реляциялық схема қалыптасып, нормализация жүргізіледі. Физикалық кезеңде нақты ДББЖ үшін кестелер, индекстер мен сақтау стратегиясы анықталады.

Физикалық жобалау

Шешілетін техникалық есептер

- Деректер типтерін нақтылау
- Индекстеу стратегиясын таңдау
- Бөлімшелеу (partitioning) жоспары
- Резервтік көшіру тәртібі
- Дискілік кеңістікті бағалау

Өнімділікті арттыру

Физикалық жобалаудың басты мақсаты – жүйенің жылдамдығы мен сенімділігін арттыру. **Индекстер** іздеуді жеделдетеді, **кластерлеу** байланысты деректерді бірге сақтайды, **буферлеу** дискіге жүктемені азайтады. Дұрыс физикалық жоба 10–100 есе жылдамдық береді.

Сақталған процедуралар



Автоматтандырылған логика

Сақталған процедура (Stored Procedure) – деректер базасы серверінде сақталатын және қайта-қайта орындалатын SQL командалары жиыны. Бизнес логикасын клиент жағынан серверге ауыстырады.

Өнімділік

Алдын ала компиляцияланған – орындалу жылдамдығы жоғары.

Қауіпсіздік

Тікелей кестеге қатынасты шектейді, рұқсатты басқарады.

Жазу тілдері

PL/SQL (Oracle), T-SQL (SQL Server), PL/pgSQL (PostgreSQL).

Деректерді өңдеудегі уақыт факторы

Өзектілік критерийлері

Деректің жарамдылық мерзімі – домен мен мақсатқа байланысты. Қор биржасы деректері миллисекундтық өзектілік талап етсе, демографиялық деректер жылдық жиіліктегі жаңартумен шектеледі.

Real-time деректер ағыны

Apache Kafka, Apache Flink сияқты жүйелер секундына миллиондаған оқиғаны өңдейді. IoT құрылғылары, қаржылық транзакциялар мен әлеуметтік желілер үшін маңызды.

Тарихи деректер

Ескі деректер архивке жіберіліп, тиімді сақталады. Уақыт белгілері (timestamps) кез келген өзгерісті бақылауға мүмкіндік береді.

Деректер сапасы: Сенімділік және дәлдік

Толықтық (Completeness)

Барлық қажетті өрістер толтырылған ба? Бос мәндер (NULL) талдаудың дәлдігін төмендетеді.

Сәйкестік (Consistency)

Деректер барлық жүйелерде бірдей ме? Қайшылықтар дұрыс емес шешімдерге әкеледі.

Дәлдік (Accuracy)

Деректер реалды дүниені дұрыс бейнелей ме? Қате деректер – дұрыс шешімнің жауы.

Уақтылылық (Timeliness)

Деректер қажетті уақытта қол жетімді ме? Ескірген ақпарат зиянды болуы мүмкін.

Қате деректердің зардаптары

IBM зерттеулері бойынша АҚШ бизнесі жыл сайын нашар деректер сапасынан **3,1 триллион доллар** шығын шегеді. Медицинада – дұрыс емес диагноз, қаржыда – зиянды транзакциялар, мемлекетте – дұрыс емес саясат шешімдері.

Үлкен деректер (Big Data) әлемі

Адамзат тарихында ешқашан бұндай көлемде деректер жасалмаған. Күн сайын 2,5 квинтиллион байт жаңа деректер туындайды. Бұл ақпараттық революция бизнес, ғылым және мемлекеттік басқаруды түбегейлі өзгертіп отыр.

2.5Q

Күндізгі деректер

Байт. Күн сайын жасалатын жаңа деректер көлемі.

3V

Big Data моделі

Volume, Velocity, Variety – үлкен деректердің үш өлшемі.

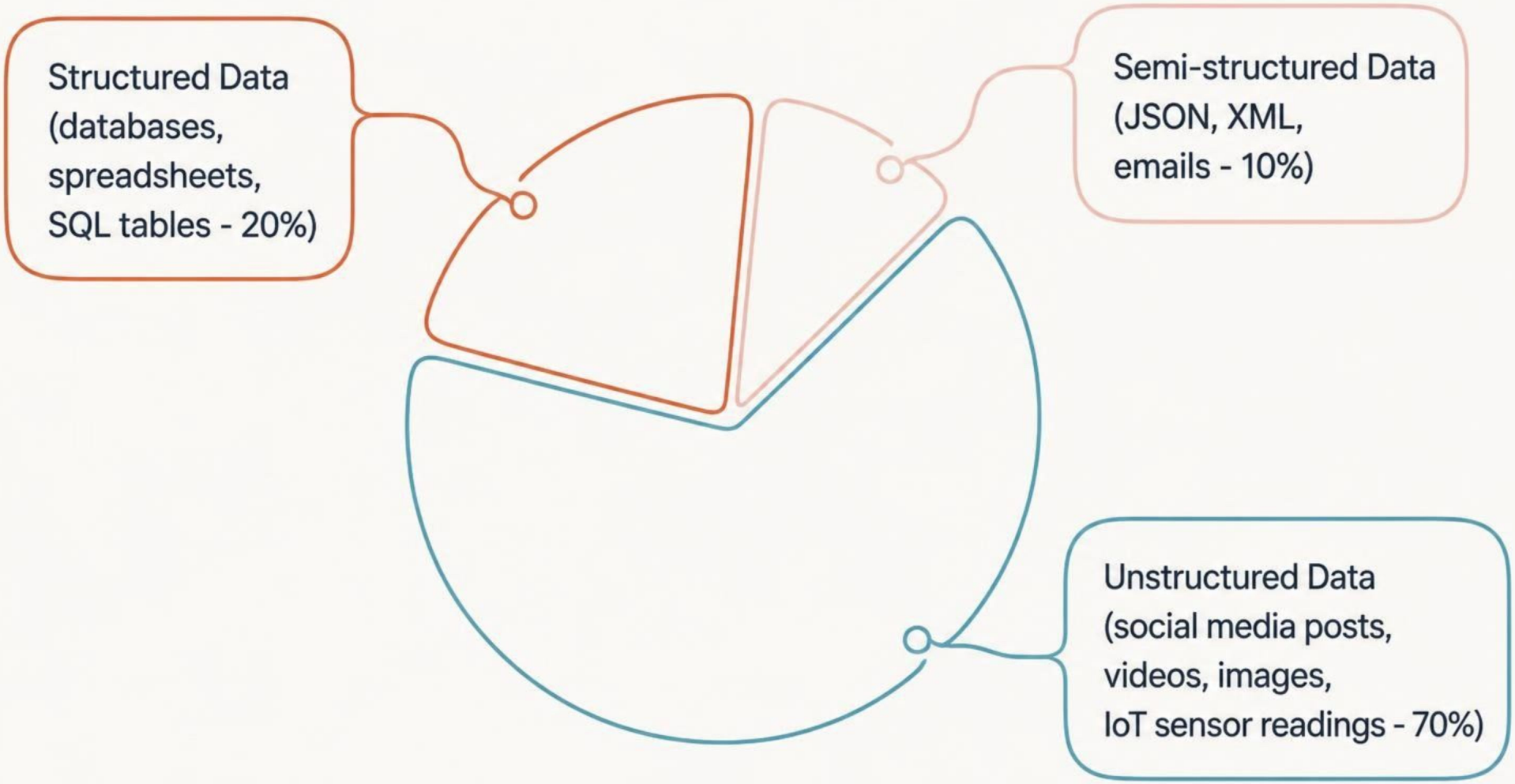
90%

Жаңа деректер

Соңғы 2 жылда жасалған деректер жалпы деректердің 90%-ын құрайды.



Үлкен деректердің құрылымы



Дүниежүзілік деректердің тек **20%-ы** ғана құрылымдалған (SQL кестелері). Қалған 80% – жартылай немесе толығымен құрылымдалмаған: суреттер, бейнелер, хабарламалар, сенсор деректері. Big Data технологиялары осы алуан форматтарды өңдеуге арналған.

Деректерді сақтау архитектурасы

1 Деректер қоймасы (Data Warehouse)

Құрылымдалған, тазартылған деректерді аналитика үшін сақтайды. Жоғары сапалы, бірақ баяу жаңартылады. Amazon Redshift, Google BigQuery – мысалдар.

- Тазартылған, сапалы деректер
- BI және есеп беруге оңтайланған
- Бағасы жоғары, бірақ сенімді

2 Деректер көлі (Data Lake)

Кез келген форматтағы деректерді өңделмеген күйінде сақтайды. Икемді, арзан, бірақ ретсіз болуы мүмкін. Apache Hadoop, AWS S3 – мысалдар.

- Барлық форматтар: JSON, CSV, видео
- ML моделін оқытуға қолайлы
- Арзан масштабталатын сақтау

Бұлттық технологиялар (AWS, Azure, GCP) деректерді физикалық серверлерсіз сақтауға және кез келген нүктеден қатынасуға мүмкіндік береді.

Деректерді қорғау және қауіпсіздік



Шифрлау (Encryption)

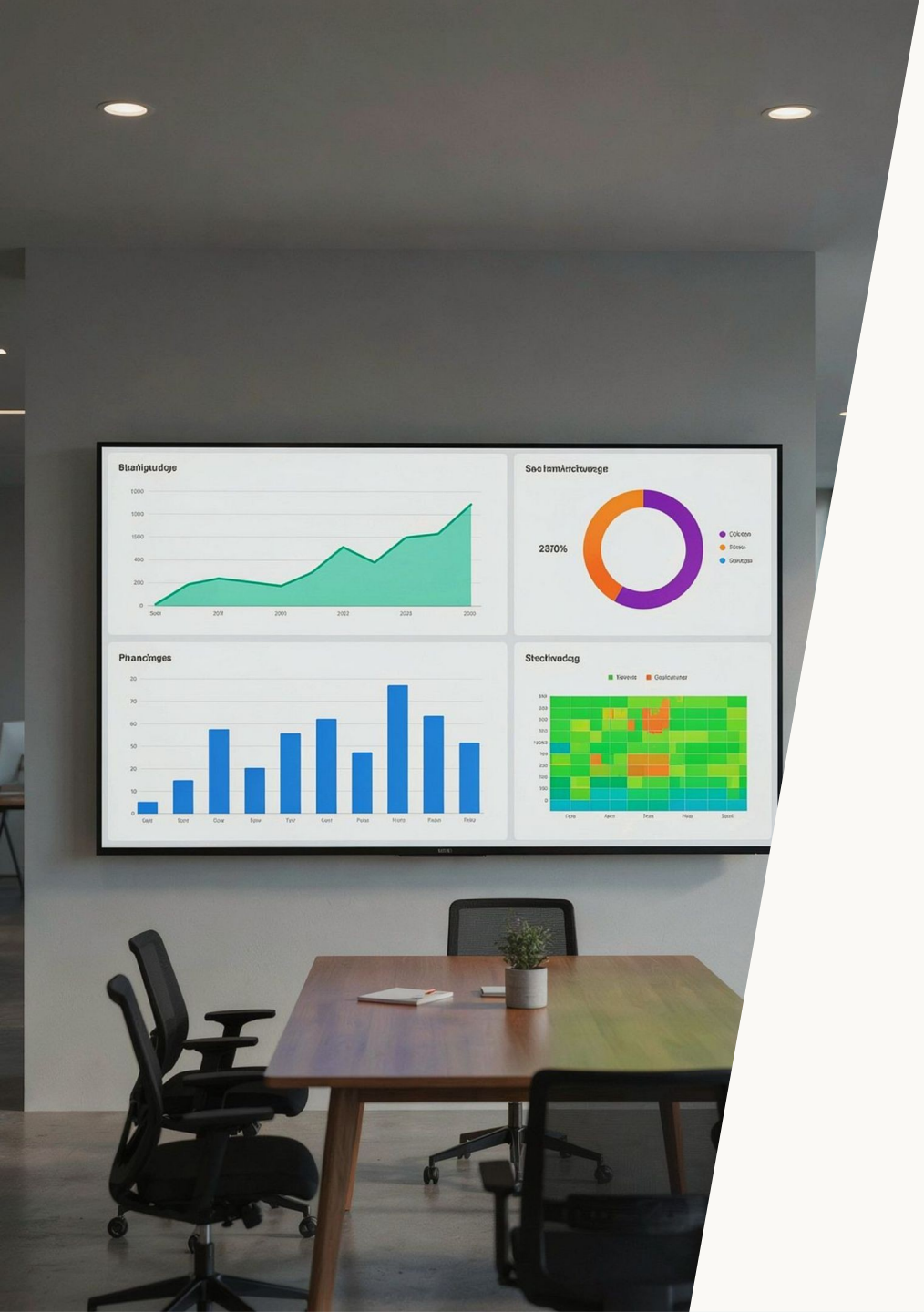
AES-256 стандарты – деректерді сақтауда және тасымалдауда қолданылады. Кілтсіз ашу мүмкін емес.

Қолжетімділікті басқару

RBAC (Role-Based Access Control) – рөлге негізделген рұқсат жүйесі. Әр пайдаланушы тек өз функциясына сай деректерді көреді.

Заңнамалық талаптар

ЕО-ның GDPR, Қазақстанның «Жеке деректер туралы» заңы – деректерді заңды басқарудың негізі.



Деректерді визуализациялаудың маңызы

Адам миы сандар мен мәтінге қарағанда суреттер мен графиктерді **60 000 есе** жылдам өңдейді. Дұрыс визуализация деректерден жасырылған заңдылықтарды, трендтер мен аномалияларды бірден анық етеді.

Data Storytelling – деректерді тек графикке айналдырып қоймай, нарративпен байланыстырып, эмоция тудыратын, шешімге жетелейтін баяндау өнері.

01

Деректерді жинау

02

Тазарту және талдау

03

Визуализация жасау

04

Әңгіме айту

Аналитика түрлері

Дескриптивті аналитика

«Не болды?» деген сұраққа жауап береді. Өткен оқиғаларды жиынтықтайды: сатылым есептері, трафик статистикасы, қаржылық қорытындылар. Ең кең таралған аналитика түрі.

Диагностикалық аналитика

«Неге болды?» деген сұраққа жауап береді. Себеп-салдар байланысын іздейді. Сатылым неге төмендеді? Тұтынушы неге кетті? Drill-down талдауы қолданылады.

Болжамды аналитика (Predictive Analytics)



Болашақ трендтерді анықтау

Болжамды аналитика статистикалық модельдер мен машиналық оқытуды қолданып, болашақ нәтижелерді болжайды. Тарихи деректердегі заңдылықтарды анықтап, алдағы оқиғаларға дайындалуға мүмкіндік береді.

Қаржы секторы

Несие тәуекелін болжау, алаяқтықты анықтау.

Бөлшек сауда

Сұраныс болжамы, қор басқару оңтайландырылуы.

Денсаулық сақтау

Аурулар тарауын болжау, пациент тәуекелін бағалау.



Жасанды интеллект және деректер

Машиналық оқыту негіздері

ML алгоритмдері деректерден өз бетімен заңдылықтарды үйренеді. Үш негізгі бағыт: жетекшілікпен оқыту (supervised), жетекшіліксіз оқыту (unsupervised) және нығайту арқылы оқыту (reinforcement learning).

Автоматты топтастыру (Clustering)

K-means, DBSCAN алгоритмдері деректерді ұқсастық бойынша автоматты топтарға бөледі. Тұтынушы сегментациясы, аномалия анықтау – жиі қолданылатын саласы.

Терең оқыту (Deep Learning)

Жасанды нейрондық желілер суретті тану, сөйлеуді түрлендіру және табиғи тілді өңдеу салаларында адамдық деңгейге жетті немесе асып түсті.

Деректер базасындағы жаңа трендтер

1

Реляциялық DB (1970-жж.)

SQL, нормализация, ACID транзакциялары – классикалық іргетас.

2

NoSQL (2000-жж.)

MongoDB, Cassandra, Redis – масштабтауға оңтайланған икемді модельдер.

3

NewSQL (2010-жж.)

SQL-дің сенімділігі + NoSQL-дің жылдамдығы. Google Spanner, CockroachDB.

4

Графтық DB (Қазір)

Neo4j – байланыстар ең маңызды болғанда: әлеуметтік желілер, ұсыныстар жүйесі.

NoSQL «Not Only SQL» дегенді білдіреді. Бұл деректер базалары горизонталды масштабталуды, икемді схеманы және жоғары жылдамдықты қамтамасыз етеді. Графтық модельдер байланыстарды орталық объект ретінде қарастырады.

Мемлекеттік және қоғамдық деректер

Цифрлық үкімет архитектурасы

Цифрлық мемлекет – азаматтарға қызметтерді онлайн ұсынып, мемлекеттік деректерді бірыңғай платформада жинақтайтын жүйе. eGov, электрондық үкімет порталдары – заманауи мысалдар.

Ашық деректер (Open Data)

Мемлекеттік деректерді азаматтар мен зерттеушілерге еркін қол жетімді ету бастамасы. Бюджет деректері, санақ нәтижелері, жол апаттары статистикасы – қоғамдық пайда үшін ашылады. Транспаренттілікті арттырып, инновацияларды ынталандырады.



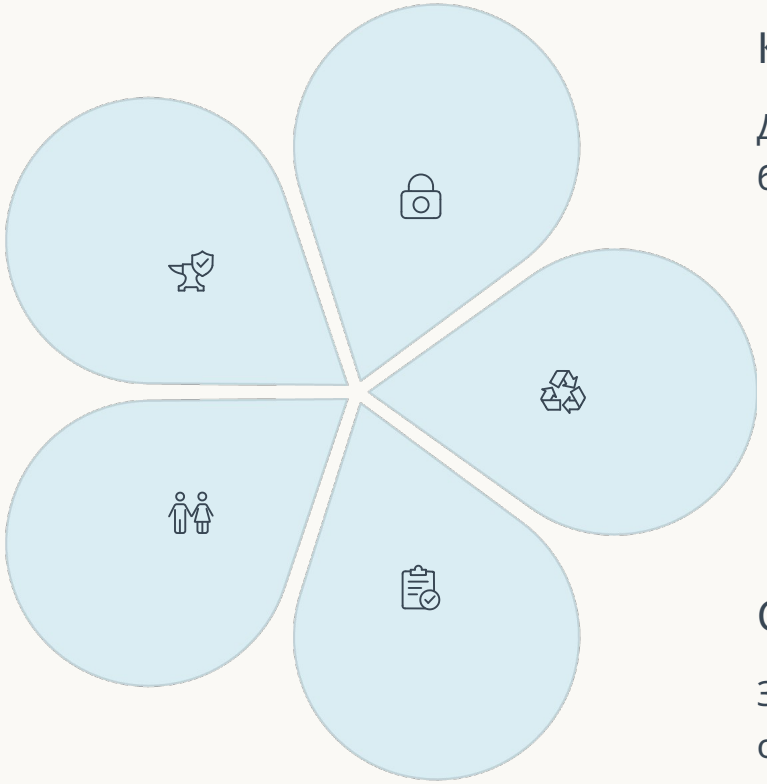
Деректерді басқару стратегиясы

Сапаны басқару

Деректер сапасының стандарттары мен метрикалары

Деректер иесі

Корпоративтік деректердің жауапты иелерін анықтау



Қауіпсіздік

Деректерге рұқсатсыз қатынасты болдырмау

Өмірлік цикл

Жасалудан жойылуға дейінгі деректер жолы

Сәйкестік (Compliance)

Заңнамалық және нормативтік талаптарды орындау

Деректермен жұмыс істеу мәдениеті

Деректерге негізделген шешімдер

Data-driven компаниялар интуицияға емес, нақты деректерге сүйенеді. McKinsey зерттеулері бойынша мұндай компаниялар бәсекелестерінен **23 есе** көп тұтынушы тартып, **6 есе** жоғары табыс табады.

Деректер сауаттылығы

Деректер сауаттылығы (Data Literacy) – ұйымдағы барлық қызметкерлердің деректерді оқып, түсініп, сыни тұрғыдан бағалай алу қабілеті. Бұл тек IT мамандарының ғана емес, бүкіл команданың құзыреттілігі болуы керек.

- Деректерді оқу және түсіну
- Визуализацияны интерпретациялау
- Деректерге сыни тұрғыдан қарау

Кәсіби құзыреттілік: Деректер сарапшысы



Техникалық дағдылар

SQL, Python/R, статистика, ML негіздері, визуализация құралдары (Tableau, Power BI).

Деректерді дайындау

Нақты деректердің 80%-ы лас, толық емес. Тазарту (data cleaning) – сарапшы уақытының басым бөлігін алады.

Коммуникация

Нәтижелерді техникалық емес аудиторияға түсінікті жеткізу – жетекші дағды.

Деректердің этикалық аспектілері

Алгоритмдік біржақтылық (Algorithmic Bias)

ML модельдері жаттығу деректеріндегі алым-теріс пиғылды мұраға алады. Жалдау алгоритмдері белгілі топтарға қарсы кемсіту жасаған нақты мысалдар тіркелген. Bias-ті анықтау және жою – этикалық міндет.

Жеке деректердің этикалық пайдалануы

Цифрлық із (digital footprint) – адамдар өздері білмей қалдыратын ақпарат. Бұл деректерді жинаушылардың жауапкершілігі – ашықтық, келісім алу және минималды жинау (data minimization) принциптеріне сүйену.

Ашықтық және есеп беру

Алгоритм шешімдерін түсіндіре алу (Explainability) – сенімді AI-дың негізі. «Неге несиеңіз қабылданбады?» деген сұраққа жүйе жауап бере алуы керек.

Деректерді басқарудағы тарихи тәжірибе



Қасым ханның Қасқа жолы

XV–XVI ғасырлардағы Қазақ хандығында Қасым хан (1511–1523 жж.) жариялаған «Қасқа жол» – қазақ тарихындағы алғашқы жүйелендірілген заң нормалары жинағы. Бұл құжат мүлік, қылмыс, неке, жер және соғыс заңдарын реттеп, ақпаратты жүйелі сақтаудың алғашқы үлгісін ұсынды.

Заманауи тілмен айтсақ, Қасқа жол – **деректер базасының прото-үлгісі**: нормаланған, жіктелген, пайдаланушыға арналған ақпарат жүйесі.

Кеңестік кезеңдегі деректерді цифрландыру

1920-жж.

Латын графикасына көшу – ақпараттық кодтау жүйесін стандарттау.

1930–1950-жж.

Ұжымдастыру кезеңінде статистикалық есеп беру жүйесі қалыптасты.

1940-жж.

Кириллицаға өту – ақпарат алмасудағы жаңа кодтық жүйе.

1960–1980-жж.

Орталықтандырылған мұрағат жүйесі мен цифрлық есеп беру стандарттары енгізілді.

Кеңестік кезеңде ақпаратты кодтау жүйесінің өзгеруі – алфавиттік реформа деректерді сақтау, іздеу және бергі ұрпаққа жеткізу мәселесін күрделендіріп, мұрағаттану ғылымының қалыптасуына итермеледі.

Тәуелсіз Қазақстанның ақпараттық инфрақұрылымы

1991 жыл – Тәуелсіздік

Тәуелсіздік декларациясы жаңа мемлекеттік деректер жүйесінің басталуын белгіледі: азаматтық тіркеу, жер кадастры, мемлекеттік бюджет жүйелері нөлден құрылды.

Астана – цифрлық стратегия

1997 жылғы астананы ауыстыру – жаңа мемлекеттік инфрақұрылым жоспарлауда деректер мен аналитиканың рөлін арттырды. Smart City концепциясы Астана арқылы дамытылды.

Digital Kazakhstan

2018 жылы қабылданған «Цифрлық Қазақстан» бағдарламасы мемлекеттік деректерді басқарудың жаңа дәуірін ашты.



Деректер сапасы және аналитикалық стратегиялар

Деректердің сенімділігі

Қателіктерді азайту үшін деректер валидациясы, дубликаттарды жою, кіру нүктесінде сапаны бақылау (data quality gates) стратегиялары қолданылады. Сенімді деректер – дұрыс шешімнің іргетасы.

Ақпараттық қауіпсіздік

Деректерді кибершабуылдардан, ағып кетуден және рұқсатсыз өзгертуден қорғау. ISO 27001 стандарты – халықаралық ақпараттық қауіпсіздік жүйесінің іргетасы.

Болжамды аналитика

Predictive Analytics арқылы экономикалық өсімді жоспарлау: ЖІӨ болжамы, салық түсімдерін бағалау, инвестиция тартымдылығын модельдеу – мемлекеттік жоспарлаудың жаңа құралдары.

Жасанды интеллект және болашақ деректер ландшафты

Машиналық оқытудың үлесі

Үлкен деректер мен ML – бір-бірін күшейтетін симбиоз. Деректер көлемі өскен сайын модельдің дәлдігі артады. GPT, Gemini сияқты үлкен тіл модельдері триллиондаған параметрді петабайт деректерде үйренеді.

AutoML

Модель таңдауды автоматтандыру

MLOps

ML модельдерін өндіріске шығару

Болашақ деректер сарапшысының дағдылары

- Python / R бағдарламалау
- ML/AI негіздерін түсіну
- Бұлттық платформалармен жұмыс
- Деректер визуализациясы
- Этикалық деректер пайдалану
- Бизнес процестерін түсіну
- Коммуникация және презентация

Деректер әлемі: Этика және қоғамдық мәдениет

Этикалық аспектілер

Деректерді басқару – тек техникалық емес, моральдық жауапкершілік. Жеке құпиялылықты қорғау мен ашықтықтың тепе-теңдігін сақтау – заманауи ұйымның міндеті. Деректер – адамдар туралы, сондықтан оларды адамгершілікпен пайдалану керек.

Болашаққа бағытталған жол

Деректерге негізделген шешім қабылдау – жай тренд емес, болашақтың ажырамас бөлігі. Деректер сауаттылығы, этикалық пайдалану мен инновациялық ойлау бірігіп, адамзатты күрделі мәселелерді шешуге жетелейді.

Деректер – болашақтың мұнайы. Оны дұрыс өндіру, тазарту, сақтау және пайдалану – біздің ұрпақтың миссиясы.

