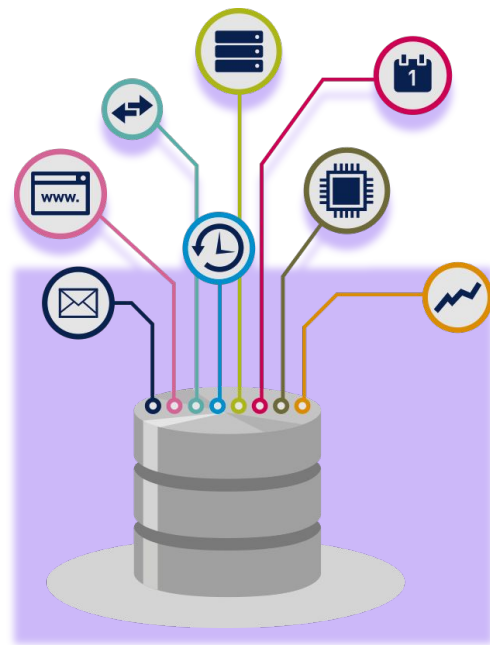


ОСНОВНЫЕ КОНЦЕПЦИИ (ПОНЯТИЯ) баз данных



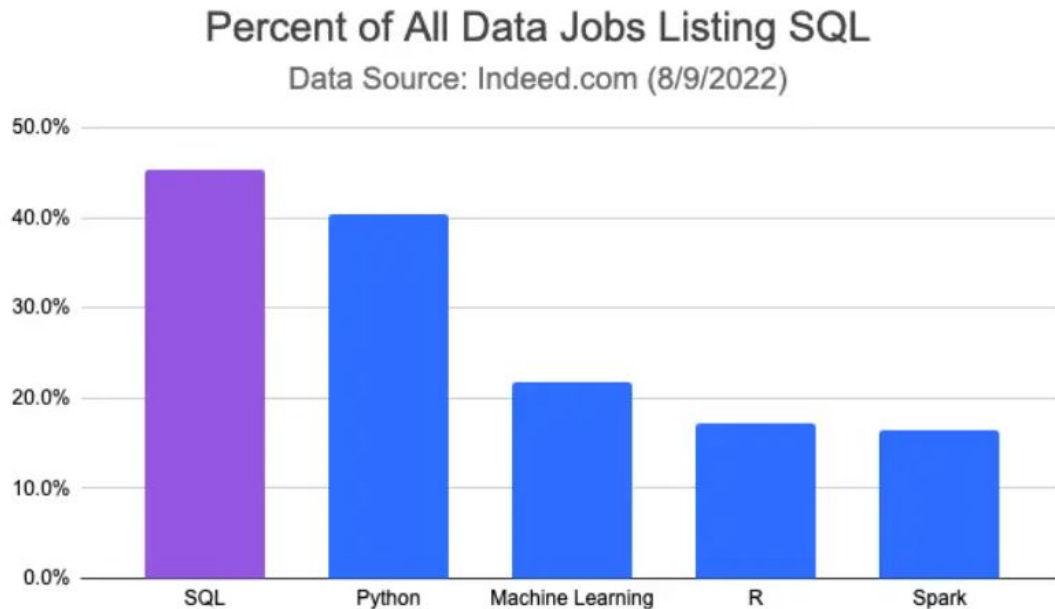


SQL

STRUCTURED QUERY LANGUAGE

Эскуэль
ЭсКьюЭль
СиКуЭль
СиКьюЭль

SQL - САМЫЙ ВОСТРЕБОВАННЫЙ НАВЫК DATA СПЕЦИАЛИСТА В МИРЕ (BREAD & BUTTER)

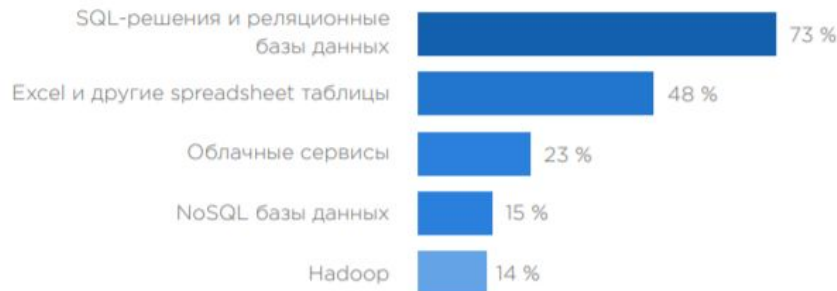


As we can see, SQL is the most in-demand skill among all jobs in data, appearing in 45.4% of job postings.

*по данным Indeed.com - топовой площадки по размещению вакансий (аналог HeadHunter зарубежом)

Системы хранения данных

Интересно, что о DWH говорят около половины респондентов, но при этом около 48 % хранят данные в Excel и других spreadsheet — таблицах. Возможно, это связано с тем, что во многих компаниях большая часть данных все-таки представлена в формате small-data.



73% компаний в Казахстане используют SQL-решения для хранения данных (**2020**)

Системы хранения данных



**74% компаний
в Казахстане
используют
SQL-решения
для хранения
данных (2021)**

В 2024 году SQL - ОСТАЕТСЯ САМЫМ ВОСТРЕБОВАННЫМ НАВЫКОМ DATA СПЕЦИАЛИСТА В МИРЕ

ИСТОЧНИКИ:

1 <https://www.linkedin.com/pulse/data-analyst-most-in-demand-skills-2024-asanka-rathnayake-vybgc/>

2 <https://www.coursera.org/articles/in-demand-data-analyst-skills-to-get-hired>

Here are the Top 50 most in-demand skills for Data Analysts



coursera

1. SQL

Structured Query Language, or [SQL](#), is the standard language used to communicate with databases. Knowing SQL lets you update, organize, and query data stored in [relational databases](#), as well as modify data structures (schema).

Since almost all data analysts will need to use SQL to access data from a company's database, it's arguably the most important skill to learn to get a job. In fact, it's common for [data analyst interviews](#) to include a technical screening with SQL.

SQL Language Subsets

Разделы Языка SQL

```
graph TD; Root[SQL Language Subsets / Разделы Языка SQL] --> DML[DML]; Root --> DDL[DDL]; Root --> DCL[DCL]; Root --> TCL[TCL]; DML --> DML_keywords[SELECT, INSERT, UPDATE, DELETE]; DDL --> DDL_keywords[CREATE, ALTER, DROP]; DCL --> DCL_keywords[GRANT, REVOKE]; TCL --> TCL_keywords[BEGIN, TRAN, COMMIT, TRAN, ROLLBACK];
```

DML

SELECT
INSERT
UPDATE
DELETE

DDL

CREATE
ALTER
DROP

DCL

GRANT
REVOKE

TCL

BEGIN
TRAN
COMMIT
TRAN
ROLLBACK

DML – DATA MANIPULATION LANGUAGE

SELECT – чтение данных таблицы

INSERT INTO – вставка данных в таблицу

UPDATE – обновление данных в таблице

DELETE – удаление данных из таблицы



DML –

DATA MANIPULATION LANGUAGE

- это то, что все себе представляют, говоря о SQL
- аналитики используют DML 80% времени
- DDL, DCL, TCL реже нужны аналитикам, чаще - разработчикам



DDL – DATA DEFINITION LANGUAGE

CREATE – создание объектов

ALTER – изменение объектов

DROP – удаление объектов



DCL – DATA CONTROL LANGUAGE

GRANT – дать разрешение
пользователю

REVOKE – забрать разрешение у
пользователя



TCL – TRANSACTION CONTROL LANGUAGE

BEGIN TRAN – начало транзакции

COMMIT TRAN – конец транзакции

ROLLBACK – откат транзакции назад



SQL

структурированный
язык
запросов
для взаимодействия с

реляционными
базами данных





ЧТО ТАКОЕ БАЗА ДАННЫХ?



БАЗА ДАННЫХ

- организованная структура для хранения и обработки данных



ТИПЫ БАЗ ДАННЫХ

РЕЛЯЦИОННЫЕ
(SQL)

01

НЕРЕЛЯЦИОННЫЕ
(NO SQL)

02



Типы баз данных

Flat file

плоский файл

Hierarchical

иерархические

Relational

реляционные

Non-Relational

нереляционные

Flat file

плоский файл

plain text file data storage
данные хранятся в текстовом виде
в “плоском” файле : CSV, TXT

- **2 dimensional, like tables -**
двунаправленные, как
таблицы
- **редко сейчас используются**
для хранения данных, но
часто - для выгрузки и обмена
- **многие системы (БД, CRM,**
ERP) имеют возможность
выгрузки и загрузки данных в
виде Flat File

22pbi_accounts - Notepad

File Edit Format View Help

```
pbi01,databoom123,Ramadinkyzy,Kymbat
pbi02,databoom123,Aigarayev,Amangeldy
pbi03,databoom123,Utepova,Dinara
pbi04,databoom123,Baidalinova,Zhansaule
pbi05,databoom123,Atymtayeva,Shynar
pbi06,databoom123,Sakiyeva,Guldana
pbi07,databoom123,Kulikova,Yuliya
pbi08,databoom123,Saukambekov,Olzhas
pbi09,databoom123,ZHAMIL,DARYN
pbi10,databoom123,Razmatova,Madina
pbi11,databoom123,Tsoy,Olga
pbi12,databoom123,GUSHTYK,ANNA
pbi13,databoom123,Kassimova,Indira
pbi14,databoom123,Joldassova,Bagila
pbi15,databoom123,Omarova,Bota
pbi16,databoom123,Pompushko,Yuriy
pbi17,databoom123,Zhunisbayeva,Aigerim
pbi18,databoom123,Amirova,Zarina
pbi19,databoom123,Baryshev,Roman
pbi20,databoom123,Mussabekova,Aina
pbi21,databoom123,Rangno,Stanislav
pbi22,databoom123,Babkin,Mikhail
pbi23,databoom123,Кучкарова,Насиба
pbi24,databoom123,Polyakova,Natalya
pbi25,databoom123,Shuvakov,Anton
pbi26,databoom123,Vavilin,Yuriy
pbi27,databoom123,Yessegazyeva,Aigerim
pbi28,databoom123,Galikhanov,Dinnur
pbi29,databoom123,Tashmetov,Komilzhon
pbi30,databoom123,Yelamanova,Akku
```

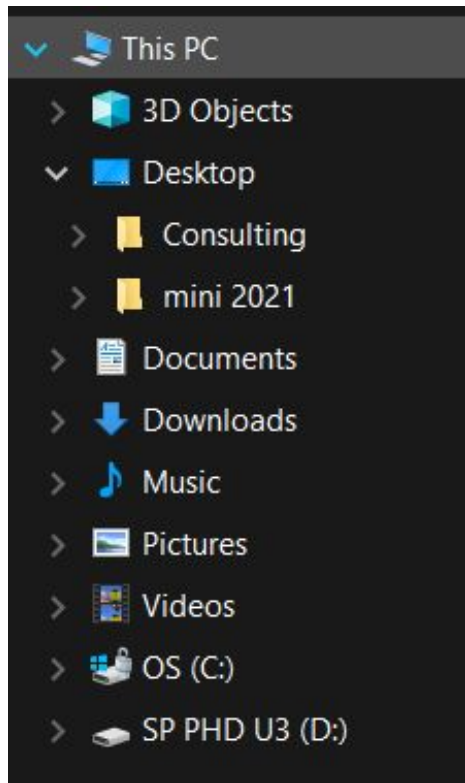
Hierarchical

иерархические

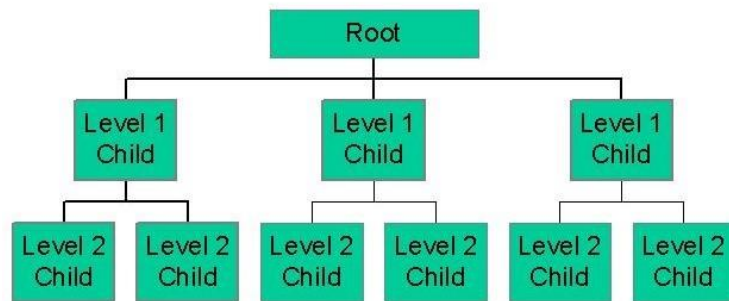
“Tree” structure, each “parent” table can have multiple “children”, but each child only one parent

Структура “дерево”, каждая “родительская” таблица может иметь много “детей”, но у каждого “дитя” только один “родитель”

- были популярны до того, как появились реляционные (relational), сейчас уже не используются
- Parent-Child relationships (Дерево, Родитель и ребенок)



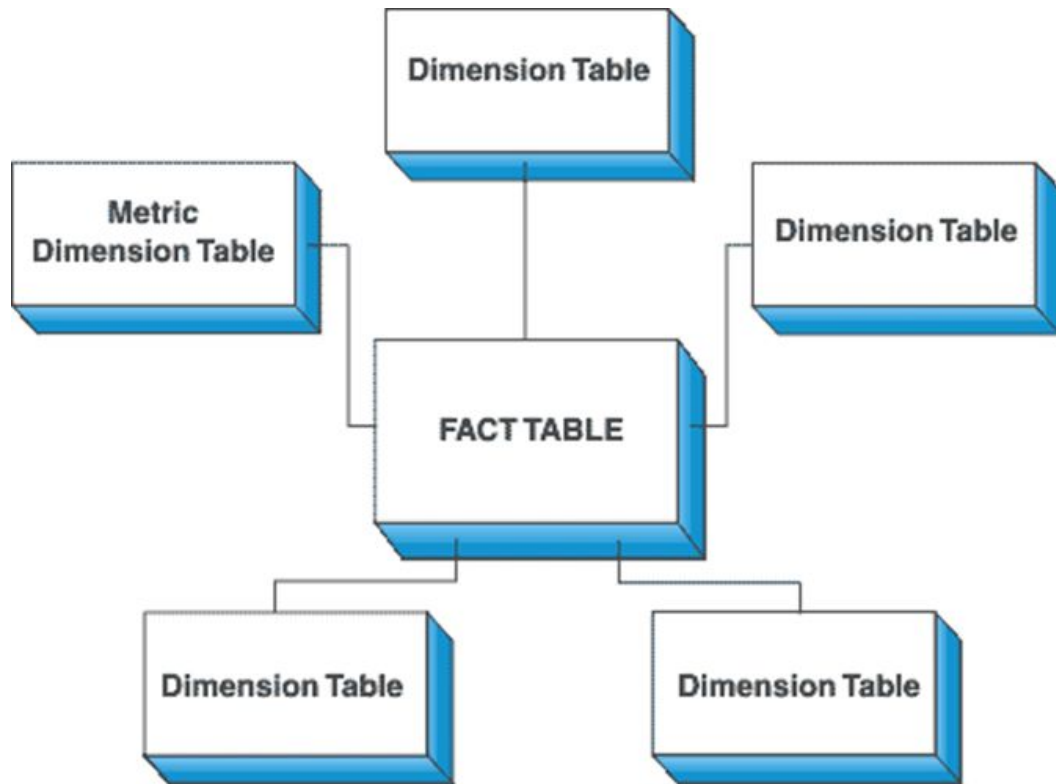
Hierarchical database model



Relational

highly-structured tabular data storage
структурированный табличный вид
хранения данных

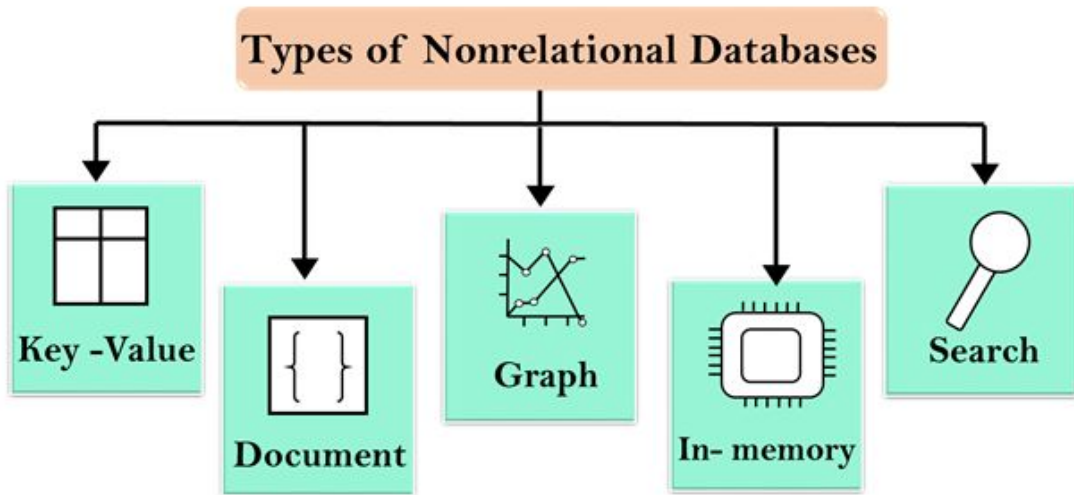
- появились в 1970 х годах и популярны с тех пор
- Таблицы и связи между ними
- scalability issues: не всегда удобно/выгодно масштабировать, увеличивать память или новые серверы заводить



Non-Relational

non-tabular data storage
НЕтабличный вид хранения данных

- NO SQL
- extremely large amounts of data - экстремально большие объемы данных, действительно BIG DATA
- используются тогда, когда реляционные уже не вмещают и/или их тяжело и дорого масштабировать (scalability issues)
- данные хранятся НЕ в таблицах



Типы баз данных

Flat file

плоский файл

данные хранятся в текстовом виде в “плоском” файле : CSV, TXT

Hierarchical

иерархические

Структура “дерево”, каждая “родительская” таблица может иметь много “детей”, но у каждого “дитя” только один “родитель”

Relational

реляционные

структурированный
табличный вид хранения
данных

Non-Relational

нереляционные

НЕтабличный вид хранения данных

The Relational Database

- данные хранятся в множестве связанных между собой таблиц, а также имеются дополнительные объекты (например, views) для доступа и работы с данными / stores data in a multiple related tables and provides supporting database objects for accessing and working with that data
- набор связанных между собой таблиц / set of tables related together
- Святой Грааль (Идеал) для такой базы данных - Третья Форма Нормальности - наивысшая целостность данных, минимальные аномалии такие как неполнота данных и избыточность (многократное сохранение одинаковых) данных / The holy grail of DB design is getting your DB into 3NF 3rd normal form – highest in terms of data integrity, reduced anomalies like redundancy and incomplete data



Что такое СУБД? What is DBMS?



Система Управления Базой Данных, это программа, которая используется для работы с базой данных. A software system used to maintain relational databases is a relational database management system (RDBMS)



Примеры СУБД: Microsoft SQL Server, MySQL, PostgreSQL



Примеры программ или GUI (graphical user interface) для работы с СУБД: Microsoft SQL Server Management Studio, HeidiSQL, Workbench

На этом курсе мы работаем с

СУБД:
Microsoft SQL Server

GUI (graphical user interface) для
работы с СУБД:
**Microsoft SQL Server Management
Studio**

SQL стандарт и диалекты

Стандарт:

- выполнять запросы чтобы получить данные
- вставлять, обновлять, удалять строки в таблицах
- создавать, изменять, удалять объекты в базе данных
- контроль доступа к



Диалекты:

• У SQL несколько диалектов. Сам язык везде один, а диалекты немного отличаются друг от друга.

• По аналогии с казахским языком: на западе Казахстана один диалект, на юге – другой, но язык везде один и тот же.



• Все диалекты SQL поддерживают этот стандарт (Transact-SQL or T-SQL, PL/SQL (PL-procedural language used in PostgreSQL database))