

>>> neue fische

School and Pool for Digital Talent

What types of storage do you know ?



Types of Data

Unstructured

Data does not conform to a rigid structure.

Semi- structured

Data conforms to a schema but deviations are possible

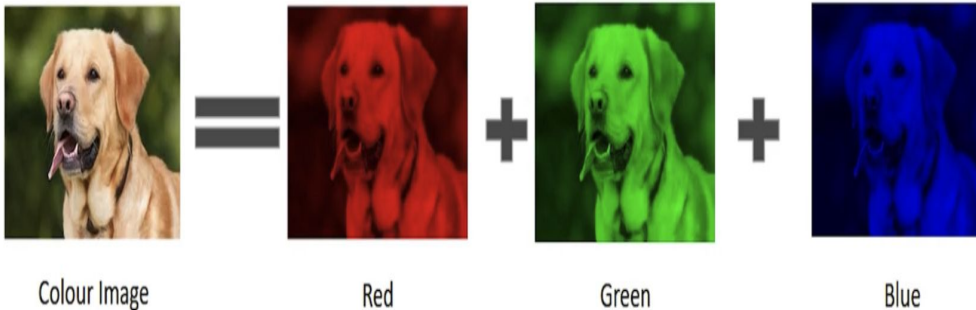
Structured

Schema and types of data are known and encoded in the system



Unstructured

- Free text - 'lorem ipsum какой прекрасный сегодня день!'
- Images
- Audio files
- Can be stored in "object store" - AWS S3

[illegible]

The diagram illustrates the arrangement of 195 numbers in a grid, grouped into three sets: A, B, and R. The numbers are arranged in a way that suggests a specific pattern or relationship between the sets.

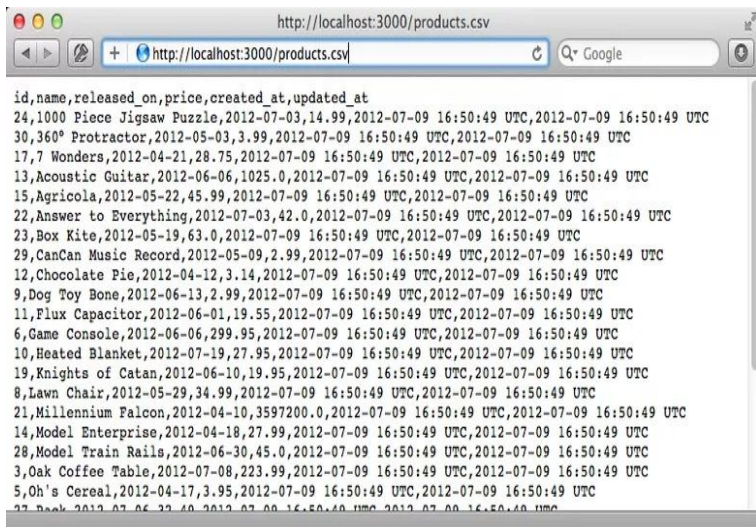
Set A (Top-Left): Contains 105 numbers. The numbers are arranged in a grid that is 10 rows high and 10 columns wide, with the last row containing only 5 numbers. The numbers are: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105.

Set B (Bottom-Left): Contains 45 numbers. The numbers are arranged in a grid that is 5 rows high and 9 columns wide. The numbers are: 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145.

Set R (Top-Right): Contains 45 numbers. The numbers are arranged in a grid that is 5 rows high and 9 columns wide. The numbers are: 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195.

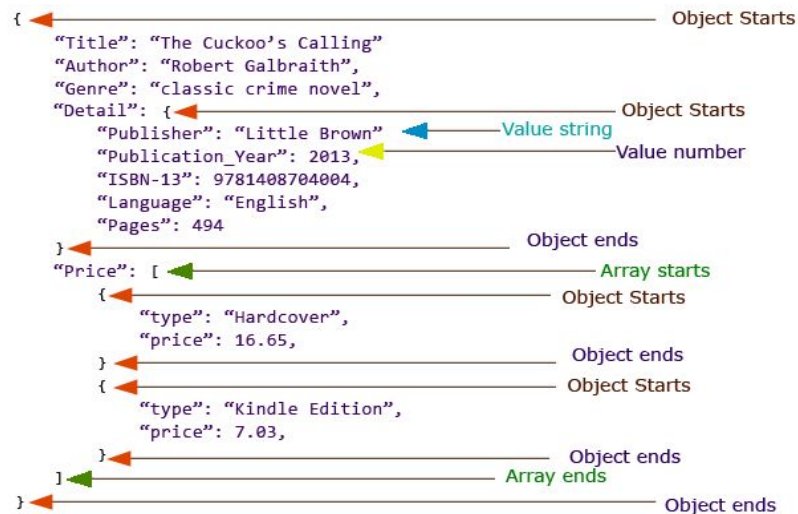
Semi-structured

- Stored in document databases - “MongoDB”, “Elastic Search”
- CSV - “comma separated values”



id	name	released_on	price	created_at	updated_at
24	1000 Piece Jigsaw Puzzle	2012-07-03	14.99	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
30	360° Protractor	2012-05-03	3.99	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
17	7 Wonders	2012-04-21	28.75	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
13	Acoustic Guitar	2012-06-06	1025.0	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
15	Agricola	2012-05-22	45.99	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
22	Answer to Everything	2012-07-03	42.0	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
23	Box Kite	2012-05-19	63.0	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
29	CanCan Music Record	2012-05-09	2.99	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
12	Chocolate Pie	2012-04-12	3.14	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
9	Dog Toy Bone	2012-06-13	2.99	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
11	Flux Capacitor	2012-06-01	19.55	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
6	Game Console	2012-06-06	299.95	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
10	Heated Blanket	2012-07-19	27.95	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
19	Knights of Catan	2012-06-10	19.95	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
8	Lawn Chair	2012-05-29	34.99	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
21	Millennium Falcon	2012-04-10	3597200.0	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
14	Model Enterprise	2012-04-18	27.99	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
28	Model Train Rails	2012-06-30	45.0	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
3	Oak Coffee Table	2012-07-08	223.99	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
5	Oh's Cereal	2012-04-17	3.95	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC
27	Book	2012-07-06	22.40	2012-07-09 16:50:49 UTC	2012-07-09 16:50:49 UTC

- JSON - JavaScript object notation



```
{
  "Title": "The Cuckoo's Calling",
  "Author": "Robert Galbraith",
  "Genre": "classic crime novel",
  "Detail": {
    "Publisher": "Little Brown",
    "Publication_Year": 2013,
    "ISBN-13": "9781408704004",
    "Language": "English",
    "Pages": 494
  },
  "Price": [
    {
      "type": "Hardcover",
      "price": 16.65,
    },
    {
      "type": "Kindle Edition",
      "price": 7.03,
    }
  ]
}
```

Object Starts

Object Starts

Value string

Value number

Object ends

Array starts

Object Starts

Object ends

Object Starts

Object ends

Array ends

Object ends

Structured

- Most of the data generated by business processes.
- Enforcing the structure reduces flexibility but fosters reliability.
- Deviations from the structure mean :
 - **The business process has changed**
 - **A bug has been introduced**
- Data that is frequently needed can be loaded from disk to memory to be accessed faster
- **Structured data can be stored and validated by machines**



Why learn about Databases?

Why?

- Most of (large) data is stored in databases
- An analyst needs to be able to connect to a database and access its data in order to unlock insights

What?

- Understand the basics of databases
- Be able to connect to a database

How?

- Learn about database types and how they are structured
- Connect to a database and explore its content



Introduction to Databases



Databases

A systematic collection of data

Data is either stored on disk or in-memory (faster)

Support electronic storage and manipulation of data



Types of Databases

Structured data - **Relational Database Management Systems**
("RDBMS" or "SQL" databases)

- **Use "SQL" (Structured query language)** to query RDBMS
- Have a predefined schema
- Data is stored in tabular form of columns and rows
- The relationship between data is **relational**
- Examples: Postgres, MySql, Oracle, SQLite

NoSQL Databases - "Not only SQL"

- **don't use SQL** as the primary language
- have no predefined schema
- Are there to deal with use cases that SQL databases are not good at
- Examples: Neo4j, Elasticsearch, MongoDB



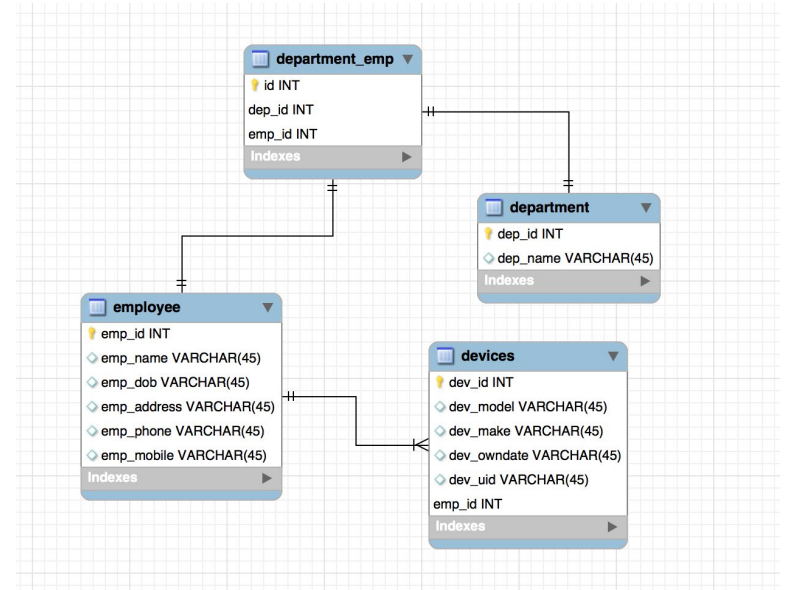
RDBMS

- Many different types of databases exist and each uses a different flavour of SQL
- Their syntax can differ, but the core concepts are the same
- Some databases will implement a subset of the functionality
- Some DB will be optimized for speed of read, others for speed of write



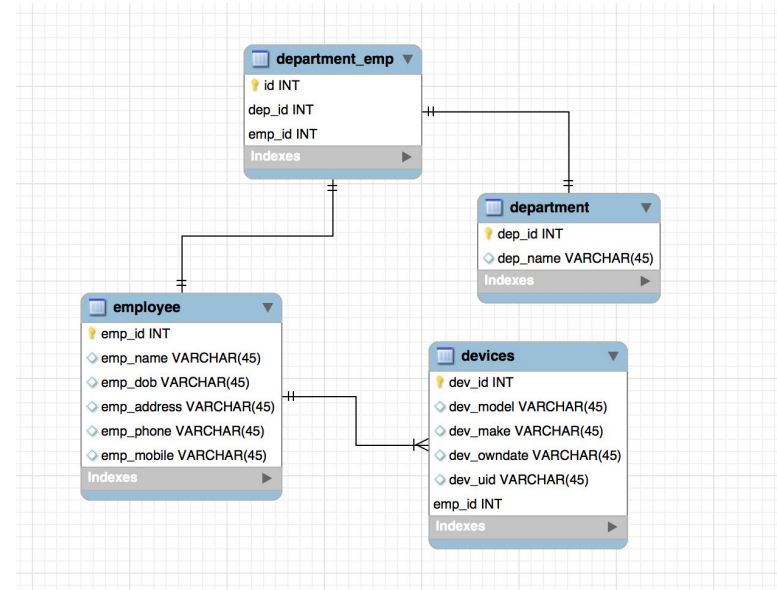
Database Structure

- A database consists of one/multiple schemas
- Schemas consist of tables
- Tables consist of columns and rows
- A column is a variable and has a unique name
- A row is an observation
- Every cell is a single value



Entity-Relationship model

- Data consists of entities of type object, class, person or place
- The property of an entity is described through their attribute(s)
- Relationships describe the relation between entities
- Different types of relationship exist



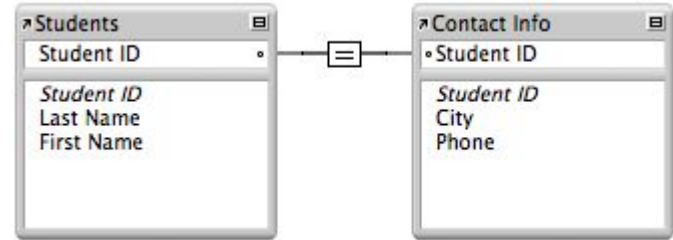
Entity-Relationship types

- One-to-one (1:1)
- One-to-many (1:n) / Many-to-one (n:1)
- Many-to-many (n:n)



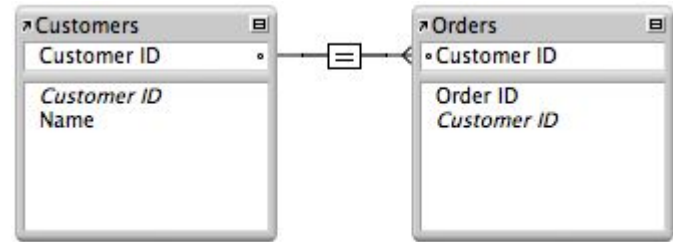
Entity-Relationship types

- One-to-one (1:1)
- One-to-many (1:n) / Many-to-one (n:1)
- Many-to-many (n:n)



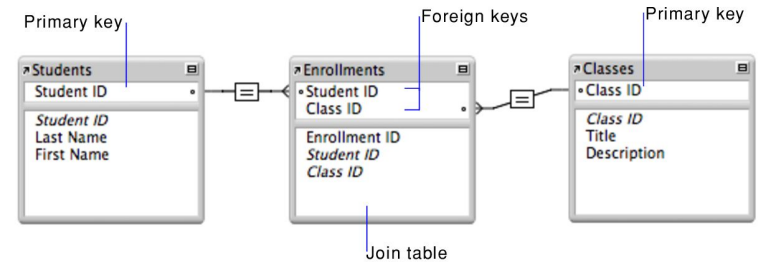
Entity-Relationship types

- One-to-one (1:1)
- One-to-many (1:n) / Many-to-one (n:1)
- Many-to-many (n:n)



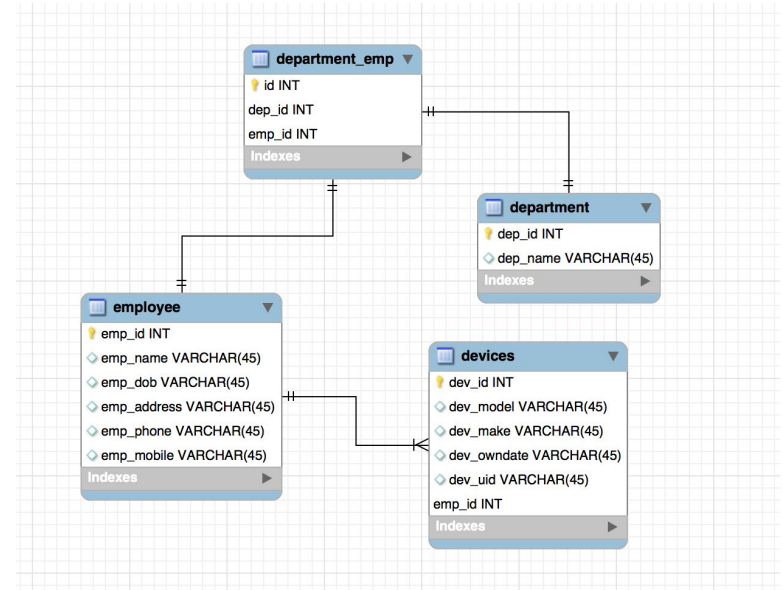
Entity-Relationship types

- One-to-one (1:1)
- One-to-many (1:n) / Many-to-one (n:1)
- Many-to-many (n:n)



Relational Databases

- Tables are related via primary and foreign keys
- Each table has **one** primary key that is unique for each record
- A foreign key is a field (or collection of fields) in one table, that refers to the primary key in another table



ERM Exercise



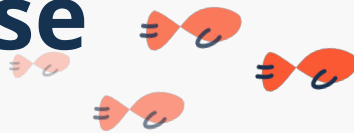
Group Work

1. Split up in 4 groups
2. Open the Miro Board we send in Slack
3. Create between 4 and 8 entities
4. Write down properties (columns) of entities
5. Model dependencies (1:1, 1:m, m:n)
6. Present your final diagram and explain decisions
7. You have 30 Minutes

Topic
Airport
Chocolate Factory
Roller Coaster Park
Restaurant



Connecting to a Database



SQL Client / Database IDE

- IDE = **I**ntegrated **D**evelopment **E**nvironment
- Powerful software that can be used to connect to a database and retrieve and visualise data (and more!)
- Local or in the cloud
- Collection of open-source, free and paid software available



Local SQL Clients

- Installed and run locally on your machine

Examples:



DBeaver



DG DataGrip



+a|b|e|a|u®

Cloud SQL Clients

- Deployed in the cloud and accessed via a web-interface

Examples:



Metabase

re dash



Google Cloud

+a|b|e|a|u®
SERVER



Looker

DBeaver



- In this course we will use DBeaver

Why?

- Free
- Easy to use
- Works for many different types of databases
- Cross platform (Windows, Linux, Mac OS, Solaris)



Installing DBeaver



Choose one:

1. Run the following command in your console:

```
brew install --cask dbeaver-community
```

OR

2. Download from their official website [DBeaver Mac OS X \(dmg\)](#) and install to your Applications folder



Setting up DBeaver



Open DBeaver > Preferences > Editors

- Enable upper case: SQL Editor > Formatting > Keyword Case > Set to: Upper
- Add line numbers: Text Editors > Show line numbers > Tick box



Connect to a PostgreSQL database

1. Click on “New Database Connection”
2. Search for and select PostgreSQL
3. Enter the connection details below



Host ds-sql-playground.c8g8r1deus2v.eu-central-1.rds.amazonaws.com
Port 5432
Database postgres
Username Will be posted in Slack/Zoom Chat
Password Will be posted in Slack/Zoom Chat

The screenshot shows the 'Connection Settings' dialog for PostgreSQL. The 'Main' tab is selected. The 'Host' field contains 'ds-sql-playground.c8g8r1deus2v.eu-central-1.rds.amazonaws.com' and the 'Port' is '5432'. The 'Database' field contains 'postgres'. Under 'Authentication', 'Database Native' is selected. The 'Username' field is empty, and the 'Password' field is masked with dots. The 'Save password locally' checkbox is checked. Under 'Advanced', the 'User role' field is empty, and the 'Local Client' dropdown is set to '<not present>'. At the bottom, there are buttons for 'Test Connection ...', '< Back', 'Next >', 'Cancel', and 'Finish'.

Exploring the database



Find the database connection in your “Database Navigator” pane on the left
Expand it to postgres > postgres > Schemas > public > Tables
Check out the introduction’s tables and answer the following questions:

1. What happens if you double click on a table?
2. What is the first and last entry listed in the records table?
3. What data type is the column ‘record_date’ in the record table?

