Introduction to Databases

Content

1. Overview of databases and database management systems
2. Key concepts in relational databases
3. Tutorial: introduction to MySQL
4. Tutorial: building a simple API to operate a MySQL database using Python

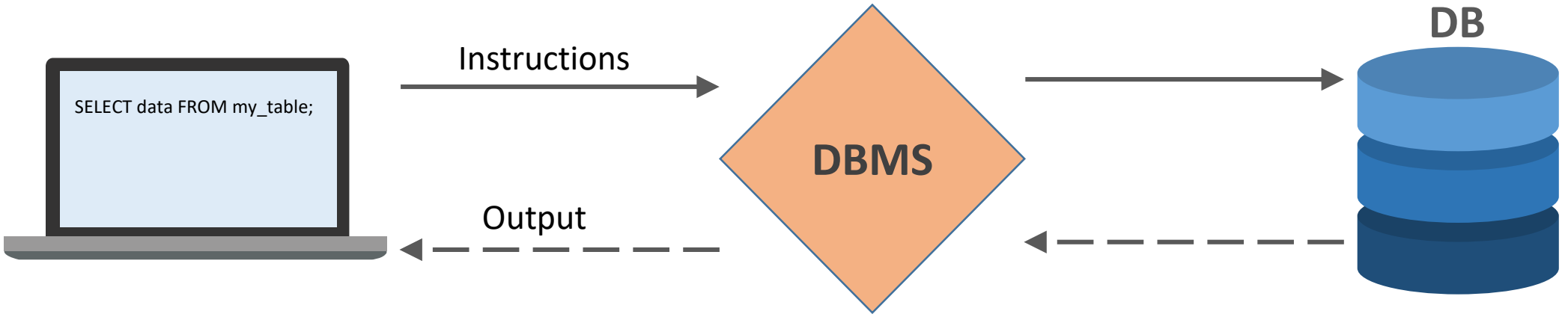
Image credit: Zyxwv99, Wikipedia

What is a Database?

- A database is an organized storage of information:
  - Shopping list
  - Phone book
  - Bloomberg terminal
  - Your memory
  - …

Image credit: Jan Antonin Kolar

Database Management Systems (DBMS)



DBMS is software providing an interface between user and data:

- Create, retrieve, update, and delete (CRUD)
- Interaction with programming languages
- Security (admins, users, backups)
- Concurrent requests from several users
- …

Relational (SQL) DBMS

- Data is stored in and represented as rectangular tables (columns and rows)
- Rows are uniquely identified: `id` (`ticker`) are unique in the left (right) table below

id	ticker	date	return	price
1	AAPL	2020-08-18	0.0083	462.25
2	AAPL	2020-08-19	0.0013	462.83
3	MSFT	2020-08-18	0.0058	211.49
4	MSFT	2020-08-19	-0.0085	209.70
5	XOM	2020-08-18	-0.0049	42.43

ticker	name	sector
AAPL	Apple Inc.	Technology
MSFT	Microsoft Corp.	Technology
XOM	ExxonMobil	Oil & Gas

- SQL = Structured Query Language, originally -- **S**tructured **E**nglish **Qu**ery **L**anguage
- Examples of RDBMS: MySQL, Microsoft SQL Server, PostgreSQL…

Non-relational (NoSQL) DBMS

- Data can be organized in any way – no language standard
- For example, key-value pairs like Python dictionaries or JSONs:

```
[
    {
        "AAPL": {
            "name": "Apple Inc.",
            "return": [0.0083, 0.0013],
            "date": ["2018-08-18", "2018-08-19"]
            },

        "MSFT": {
            "name": "Microsoft Corp.",
            "return": [0.0058,  -0.0085],
            "date": ["2019-08-18", "2018-08-19"]
            }
    }
]
```

- Examples: MongoDB for storing documents, Oracle NoSQL DB for key-value pairs…

SQL: Primary Keys

- A primary key **uniquely** identifies each row in a table

id	ticker	date	return	price
1	AAPL	2020-08-18	0.0083	462.25
2	AAPL	2020-08-19	0.0013	462.83
3	MSFT	2020-08-18	0.0058	211.49
4	MSFT	2020-08-19	-0.0085	209.70
5	XOM	2020-08-18	-0.0049	42.43
6	XOM	2020-08-18	-0.0049	42.43

ticker	name	sector
AAPL	Apple Inc.	Technology
MSFT	Microsoft Corp.	Technology
XOM	ExxonMobil	Oil & Gas

- In the **`stock_data`** table to the left the primary key is `id` allowing for duplicate rows
- In the **`stock_info`** table to the right the primary key is ticker – we only can have one row for each unique ticker

SQL: Foreign Keys

- A foreign key is a column in a table that links this table to another one

id	ticker	date	return	price
1	AAPL	2020-08-18	0.0083	462.25
2	AAPL	2020-08-19	0.0013	462.83
3	MSFT	2020-08-18	0.0058	211.49
4	MSFT	2020-08-19	-0.0085	209.70
5	XOM	2020-08-18	-0.0049	42.43

ticker	name	sector
AAPL	Apple Inc.	Technology
MSFT	Microsoft Corp.	Technology
XOM	ExxonMobil	Oil & Gas

- In the example above the **`stock_data`** table **`ticker`** is a foreign key, because it is a primary key in the **`stock_info`** table

SQL: Composite Keys

- This one is easy, we can uniquely identify a row with two columns, which totally makes sense in the case of the **`stock_info`** table:

ticker	date	return	price
AAPL	2020-08-18	0.0083	462.25
AAPL	2020-08-19	0.0013	462.83
MSFT	2020-08-18	0.0058	211.49
MSFT	2020-08-19	-0.0085	209.70
XOM	2020-08-18	-0.0049	42.43

ticker	name	sector
AAPL	Apple Inc.	Technology
MSFT	Microsoft Corp.	Technology
XOM	ExxonMobil	Oil & Gas

Summary

- DBMS provide an interface between the user and database
- Key functions of DBMS is CRUD operations + maintenance, administration, security
- Relational databases are synonymous to SQL represent the data as tables with rows and columns
- SQL is standardized, but there are different dialects
- Non-relational (NoSQL) databases admit any representation, however the language is DBMS-specific

Links

- A very good general tutorial about MySQL
- Another good tutorial