#### **Databases - Introduction**

Amos Azaria, Netanel Chkroun

### Grading policy and Home Assignments

- Grading policy:
  - 4 Home assignments: Each assignment can add up to 2 bonus points to the final grade (you need at least 55 on the exam).
  - Final Grade: Test + bonus points
- Submission in pairs.
- First home assignment will be published next week
- Submission time is always at 23:55.
- No late submission.
- Only a subset of the questions in each assignment will be graded.

## Don't copy! Don't share your work! Don't work in groups!

- 5 עונשי מינימום
- א. הורשע הסטודנט בהכנסת חומר עזר אסור או החזקתו- עונשו של המורשע לא יפחת מציון 0 בקורס בו בוצעה העבירה והרחקה מהלימודים למשך סמסטר אחד לכל הפחות.
- הורשע הסטודנט בהכנסת שינוי כלשהו בבחינה עונשו של המורשע יכלול הרחקה מהלימודים למשך סמסטר אחד לכל הפחות.
- הורשע הסטודנט הונאה בעבודות סמינריוניות או כל מטלה לימודית אחרת עונשו של המורשע יכלול לפחות ציון 0 בקורס בו בוצעה העבירה והרחקה מהלימודים למשך סמסטר אחד לכל הפחות.
- ד. הורשע הסטודנט בפלילים בעבירה שיש עמה קלון הקשורה למעמדו כסטודנט באוניברסיטה - יורחק הסטודנט מלימודים למשך סמסטר אחד לפחות.

#### Github

- You may not post your assignments to a public repository until after the deadline.
- You may want to consider the usage of Bitbucket instead of Github, or use the private repository in Github.

#### Course Structure

- Relational Databases Management Systems
  - SQL
  - Building databases
  - ERD Entity-Relational-Diagrams (next semseter)
- Java
  - Connecting to MySQL
  - Streams
- Object representation languages:
  - XML (and XSD)
  - JSON
- NoSQL
  - Key-Value Store, Wide-Column Store, Document Store, Graph Store, Search Engines, RDF
- Big Data

Handling huge amounts of data

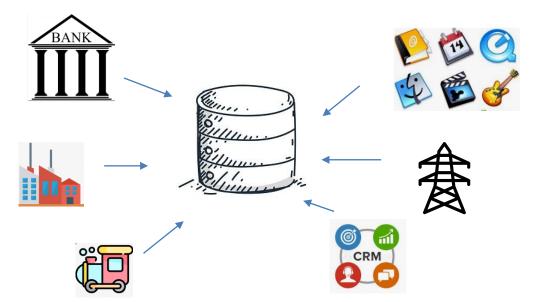
- Spark
- Naïve Bayes and Regression
  - Linear regression
  - Logistic regression

What can we do when we need to know what might happen where we don't have data

#### What is a Data Base?

A database is an organized collection of structured information, or data, typically stored electronically in a computer system. A database is usually controlled by a database management system (DBMS).

Together, the data and the DBMS, along with the applications that are associated with them, are referred to as a database system, often shortened to just database. (Oracle Site)



#### Basic classification of Databases

#### Relational data base

Admission No.	Firstname	Surname	DoB	Gender	degree
68102	Janet	Baker	18/12/1994	F	1
68103	Darren	Kirk	19/12/1994	М	1
68104	Sophie	Meadows	20/12/1993	F	1
68105	Ravi	Patel	21/12/1994	М	2
68106	Daisy	Rogers	22/12/1993	F	2
68107	Tom	Rogers	22/12/1993	М	2
67998	Brett	Akers	11/10/1994	М	2
67999	John	Deakin	03/12/1994	М	2
68000	Celia	Fisher	20/02/1995	F	2
68001	Leo	Harris	21/09/1994	М	2
68002	hankunthal	Patel	14/06/1995	F	2
67893	Sally	Brown	18/12/1995	F	1
67894	Jane	Carter	12/09/1995	F	1
67895	Paul	Edwards	20/04/1996	М	1

None relational data base (noSQL)

## **Relational Databases**

#### Relational Databases

- Relational databases are built from tables (relations), with attributes (columns) and entries (rows).
- Each table represents a certain entity.
- Rows in one table can be linked to rows in other tables (in the same database).
- Relational data-bases are based upon relational algebra and SQL.
- Most common databases in use:
  - Open Source: MySQL, SQLite, PostgreSQL
  - Proprietary: SQL Server and Oracle.

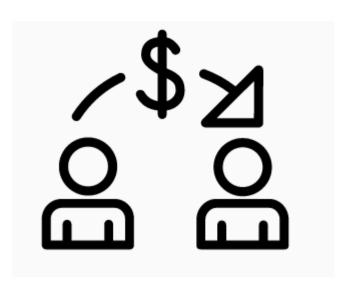
#### Relational Databases

Actions on a relational database are batched in transactions.

**Transaction -** a transaction or unit of work is a set of database operations all of which should be executed successfully in order to call the transaction successful.

## Transfer 1000\$ from Account X to account Y

- 1.read(X)
- 2.X = X 1000
- 3. write(X)
- 4. read(Y)
- 5. Y = Y + 1000
- 6. write(Y)



#### **ACID**

- Atomicity: each transaction is either executed in full, or not executed at all.
- Consistency: database remains consistent. Transaction does not violate any integrity constraints during its execution (if a transaction leaves the database in an illegal state, it is unrolled).
- Isolation: Transaction should be executed in isolation from other transactions. During concurrent transaction execution, intermediate transaction results from simultaneously executed transactions should not be made available to each other..
- Durability: a committed transaction always remains in database even in case of sudden power shortage or hardware/software failure etc.

# Database management system usage (http://db-engines.com/en/ranking)

364 systems in ranking, March 2021

			50+ Systems in ranking, March 2021							
	Rank				Score					
Mar 2021	Feb 2021	Mar 2020	DBMS	Database Model	Mar 2021	Feb 2021	Mar 2020			
1.	1.	1.	Oracle 😷	Relational, Multi-model 🚺	1321.73	+5.06	-18.91			
2.	2.	2.	MySQL []	Relational, Multi-model 🚺	1254.83	+11.46	-4.90			
3.	3.	3.	Microsoft SQL Server 😷	Relational, Multi-model 🚺	1015.30	-7.63	-82.55			
4.	4.	4.	PostgreSQL #	Relational, Multi-model 🚺	549.29	-1.67	+35.37			
5.	5.	5.	MongoDB 🚹	Document, Multi-model 🚺	462.39	+3.44	+24.78			
6.	6.	6.	IBM Db2 [1]	Relational, Multi-model 🚺	156.01	-1.60	-6.55			
7.	7.	<b>1</b> 8.	Redis 😷	Key-value, Multi-model 🚺	154.15	+1.58	+6.57			
8.	8.	<b>4</b> 7.	Elasticsearch 🚹	Search engine, Multi-model 🔃	152.34	+1.34	+3.17			
9.	9.	<b>1</b> 0.	SQLite [1	Relational	122.64	-0.53	+0.69			
10.	<b>1</b> 11.	<b>4</b> 9.	Microsoft Access	Relational	118.14	+3.97	-7.00			
11.	<b>4</b> 10.	11.	Cassandra 🚹	Wide column	113.63	-0.99	-7.32			
12.	12.	<b>1</b> 3.	MariaDB 🚼	Relational, Multi-model 🚺	94.45	+0.56	+6.10			
13.	13.	<b>↓</b> 12.	Splunk	Search engine	86.93	-1.61	-1.59			
14.	14.	14.	Hive	Relational	76.04	+3.72	-9.34			
15.	<b>1</b> 6.	15.	Teradata 🞛	Relational, Multi-model 🚺	71.43	+0.53	-6.41			
16.	<b>4</b> 15.	<b>1</b> 23.	Microsoft Azure SQL Database	Relational, Multi-model 🚺	70.88	-0.41	+35.44			
17.	17.	<b>4</b> 16.	Amazon DynamoDB 🖽	Multi-model 👔	68.89	-0.25	13 <sup>+6.38</sup>			
18.	<b>1</b> 9.	<b>1</b> 21.	Neo4j €	Graph	52.32	+0.16	+0.54			

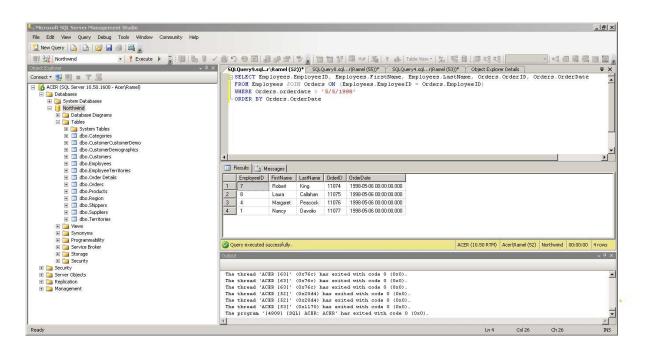
## Data base providers | Oracle

Proprietary, most widely used database in the industry.



## Data base providers | SQL Server

- Proprietary (Microsoft) but used a lot in the industry as well.
- SQL Management GUI



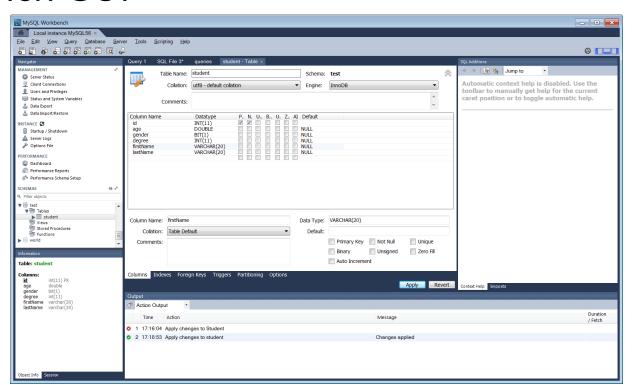


### Data base providers | My SQL



O

- Open source
- Used a lot in the industry
- Workbench GUI



## MySQL Installation

- Go to <a href="https://dev.mysql.com/downloads/installer/">https://dev.mysql.com/downloads/installer/</a>, download the installer (scroll down and click the download button) and install:
  - MySQL Server
  - MySQL Workbench: A GUI that allows us to query the DB and presents the results.
  - MySQL Connectors:
    - JDBC Driver for MySQL (Connector/J): Allows us to connect to the DB from JAVA.
    - [ADO.NET Driver for MySQL (Connector/NET)]
- You might also need to install Visual C++
  Redistributable for Visual Studio 2015 from:
  <a href="https://www.microsoft.com/en-us/download/details.aspx?id=48145">https://www.microsoft.com/en-us/download/details.aspx?id=48145</a>