

# Databases - Introduction

Amos Azaria

# Administration

- Lecturers: Amos Azaria and Merav Chkroun
- TAs: Avigail Stekel Chaya Liberman, and Keren Nivasch, Gil Levi
- Classes (4 groups):
  - Tuesdays: 9:00-12:00 and 14:00-15:00 (Amos)
  - Sundays: 9:00-12:00 and Mondays: 11:00-14:00 (Merav)
- Office hour: Tuesdays at 13:00 (11.2.11)

# Home Assignments

- Grading policy:
  - 4 Home assignments: Each assignment can add up to 2 bonus points to the final grade.
  - Final Grade: Test + bonus points
- Submission in pairs.
- First home assignment should be published on Monday (you will have 2 weeks to submit it).
- 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.
- ב. הורשע הסטודנט בהכנסת שינוי כלשהו בבחינה – עונשו של המורשע יכלול הרחקה מהלימודים למשך סמסטר אחד לכל הפחות.
- ג. הורשע הסטודנט הונאה בעבודות סמינריות 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

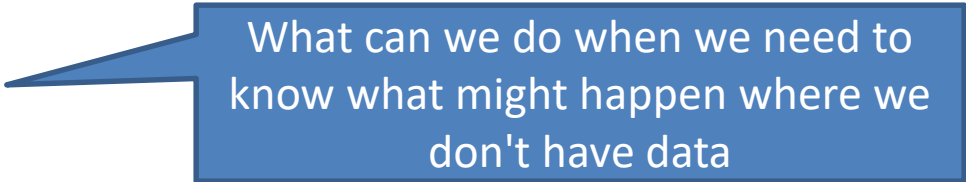
- You may not post your assignments to a public repository until after the deadline.
- You may want to consider the usage of Bitbucket instead of Github.

# Course Structure

- Relational Databases Management Systems
  - SQL
  - Building databases
  - ERD Entity-Relational-Diagrams (next semester)
- 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
  - Spark
- Naïve Bayes and Regression
  - Linear regression
  - Logistic regression



Handling huge amounts of data



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

# Relational Databases

# Relational Databases

- Relational databases are built from tables (relations), with attributes (columns) and entries (rows).
- 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.



# Database management system usage

(<http://db-engines.com/en/ranking>)

343 systems in ranking, February 2019

Rank			DBMS	Database Model	Score		
Feb 2019	Jan 2019	Feb 2018			Feb 2019	Jan 2019	Feb 2018
1.	1.	1.	Oracle	Relational, Multi-model	1264.02	-4.82	-39.26
2.	2.	2.	MySQL	Relational, Multi-model	1167.29	+13.02	-85.18
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1040.05	-0.21	-81.98
4.	4.	4.	PostgreSQL	Relational, Multi-model	473.56	+7.45	+85.18
5.	5.	5.	MongoDB	Document	395.09	+7.91	+58.67
6.	6.	6.	IBM Db2	Relational, Multi-model	179.42	-0.43	-10.55
7.	7.	8.	Redis	Key-value, Multi-model	149.45	+0.43	+22.43
8.	8.	9.	Elasticsearch	Search engine, Multi-model	145.25	+1.81	+19.93
9.	9.	7.	Microsoft Access	Relational	144.02	+2.41	+13.95
10.	10.	11.	SQLite	Relational	126.17	-0.63	+8.89
11.	11.	10.	Cassandra	Wide column	123.37	+0.39	+0.59
12.	13.	17.	MariaDB	Relational, Multi-model	83.42	+4.60	+21.77
13.	12.	13.	Splunk	Search engine	82.81	+1.39	+15.55
14.	14.	12.	Teradata	Relational	75.97	-0.22	+2.98
15.	15.	18.	Hive	Relational	72.29	+2.38	+17.23
16.	16.	14.	Solr	Search engine	60.96	-0.52	-2.91
17.	17.	16.	HBase	Wide column	60.28	-0.12	-1.43



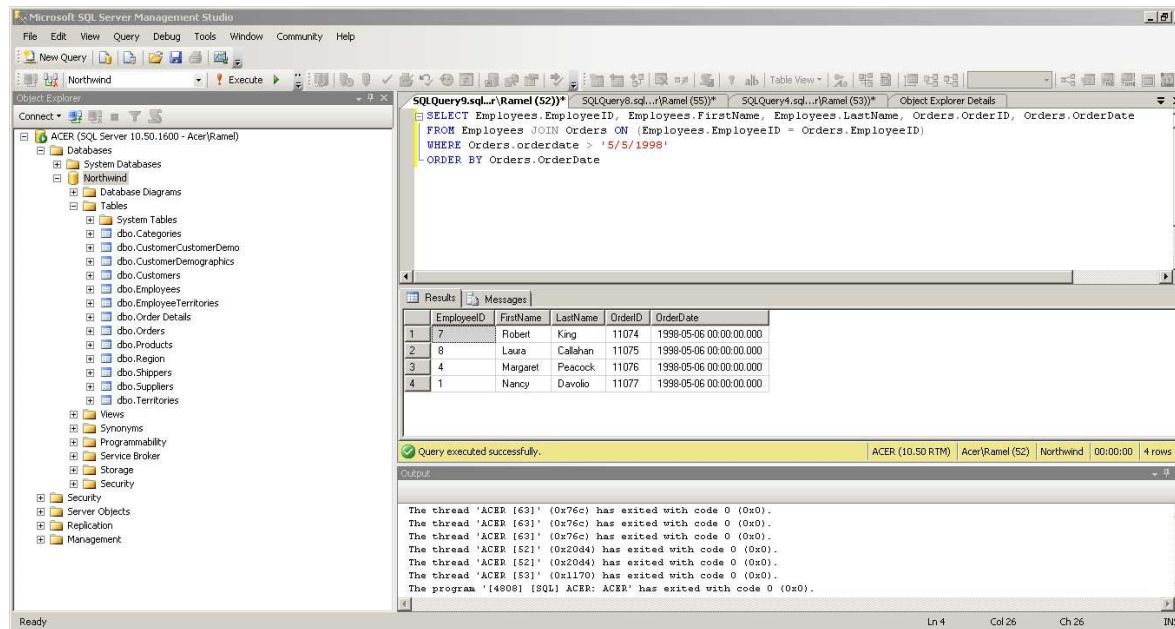
- 
- The screenshot displays the MySQL Workbench application window. The title bar reads 'MySQL Workbench'. The main menu bar includes 'File', 'Edit', 'View', 'Query', 'Database', 'Server', 'Tools', 'Scripting', and 'Help'. Below the menu is a toolbar with various icons for file operations, editing, and database management.
- The interface is divided into several panes:
- Left Sidebar (Navigator):** Contains a tree view of the database structure. Under 'MANAGEMENT', there are links for Server Status, Current Connections, Users and Privileges, Status and System Variables, Data Export, and Data Import/Restore. Under 'INSTANCE', there are links for Startup / Shutdown, Server Logs, and Options File. Under 'PERFORMANCE', there are links for Dashboard, Performance Reports, and Performance Schema Setup. Under 'SCHEMAS', there is a search bar and a list of databases: 'test' (selected), 'student', 'Views', 'Stored Procedures', 'Functions', and 'world'.
  - Top Panel:** Shows the current context: 'Query 1', 'SQL File 3\*', 'queries', and 'student - Table'. It includes fields for 'Table Name: student', 'Schema: test', 'Collation: utf8 - default collation', and 'Engine: InnoDB'. There is also a 'Comments' field.
  - Table Structure View:** A table showing the columns of the 'student' table:
 

Column Name	Datatype	P.	N.	U.	B.	U.	Z.	A.	Default
id	INT(11)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
age	DOUBLE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
gender	BIT(1)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
degree	INT(11)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
firstName	VARCHAR(20)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
lastName	VARCHAR(20)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
  - Column Properties View:** Below the table structure, there is a section for 'Column Name: firstName'. It shows 'Data Type: VARCHAR(20)' and 'Default:'. There are checkboxes for 'Primary Key', 'Not Null', 'Unique', 'Binary', 'Unsigned', 'Zero Fill', and 'Auto Increment'.
  - Bottom Panel:** Contains the 'Columns' tab (selected), 'Indexes', 'Foreign Keys', 'Triggers', 'Partitioning', and 'Options'. Below this is the 'Output' pane, which shows a list of actions:
 

	Time	Action	Message	Duration / Fetch
1	17:16:04	Apply changes to Student		
2	17:18:53	Apply changes to student	Changes applied	

# SQL Server

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



# MySQL Installation

- Go to <https://dev.mysql.com/downloads/installer/>, 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: <https://www.microsoft.com/en-us/download/details.aspx?id=48145>

# ACID

**Actions on a relational database are batched in transactions.**

- **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:** the DBMS can execute many transactions in parallel, as long as the result is identical to a sequential execution.
- **Durability:** a committed transaction always remains in database even in case of sudden power shortage or hardware/software failure etc.