APAN PS5400: Managing Data

Week 7: Interacting with data

Lecturer: Francois Scharffe

COLUMBIA UNIVERSITY
School of Professional Studies

Recap of last week

- Data Quality
- Data Governance
- Data Policies

This week

Interacting with data using

- GUI interfaces
- Web forms
- Programs
- APIs
- Intelligent Agents

We will also cover JSON objects as a data model

Interacting vs. Using

Interaction involves

- Retrieving
- Updating
- Deleting

Using involves downstream utilization of data

- Creating reports
- Feeding to analytical pipelines, etc

DBMS provided Query Interface

It is possible to interact with MySQL using command line (i.e., without any GUI), but why would you want to do that?

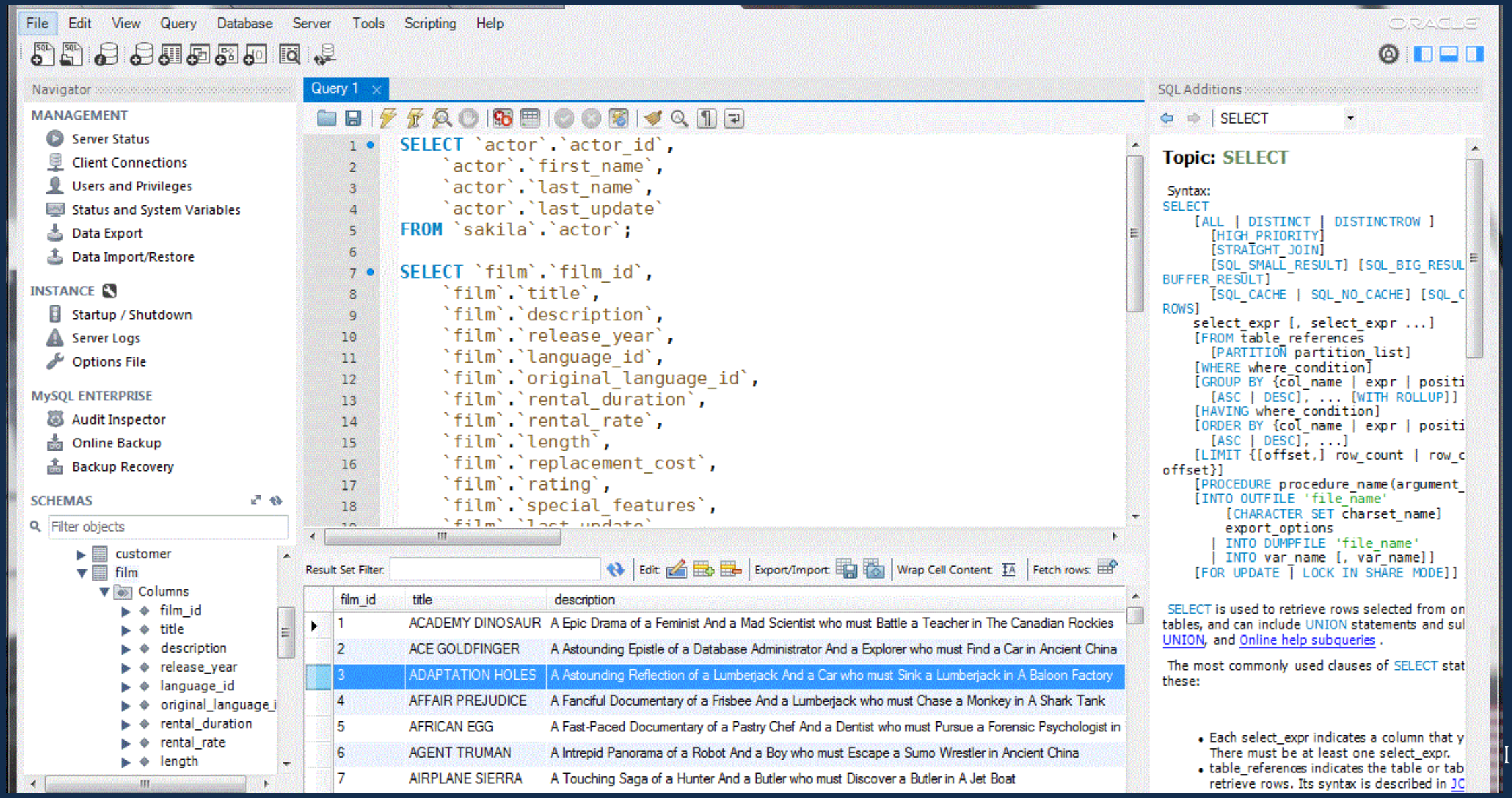
MySQL provides an interface (which you used for your project 1.1 and 1.2)

User has to explicitly write the SQL statements

Minimal GUI support

Example on next slide

MySQL query interface



ITY

Third-party GUI interfaces

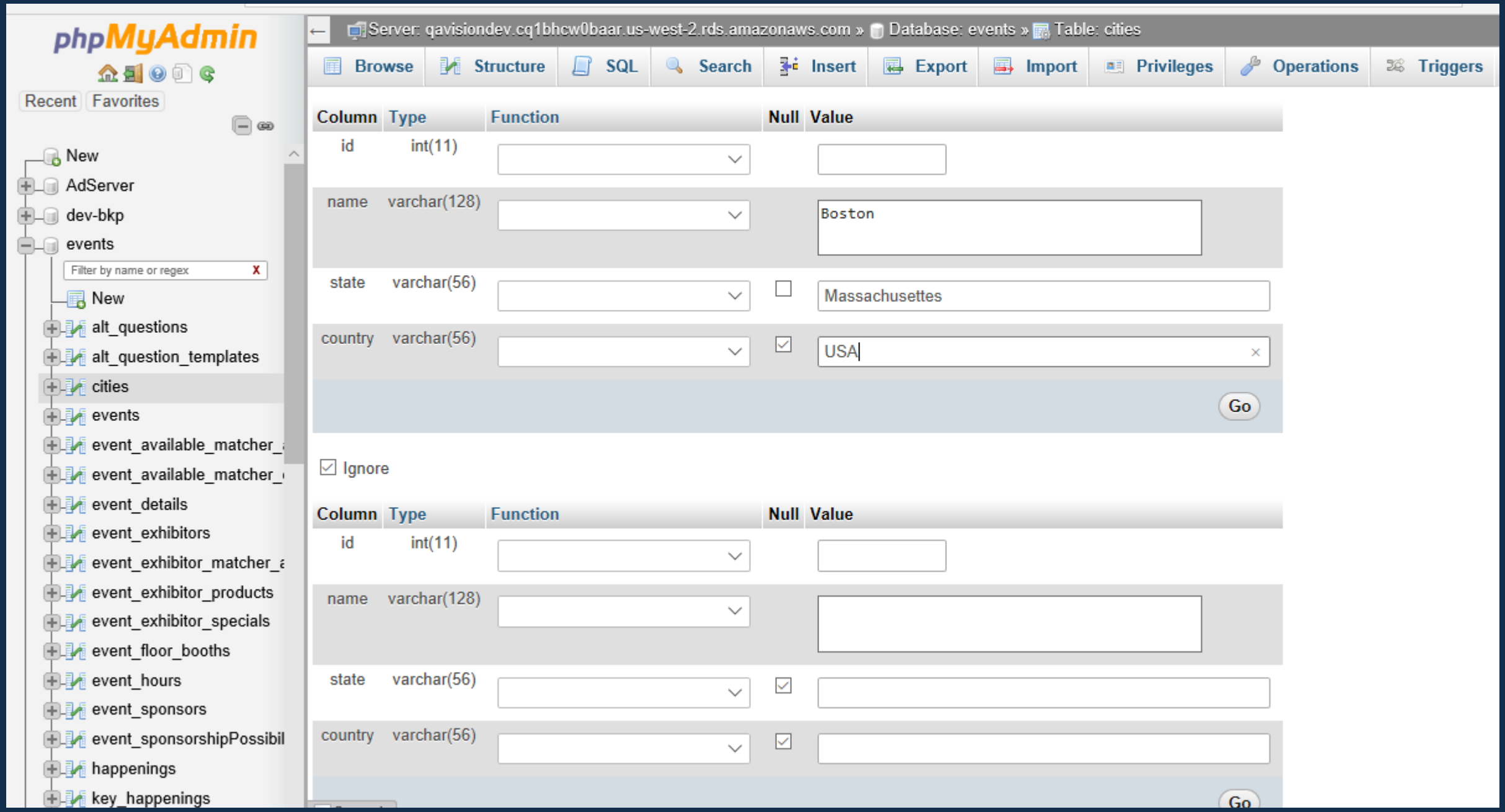
Example: PhpMyAdmin for MySQL databases

Can create queries by filling out a form

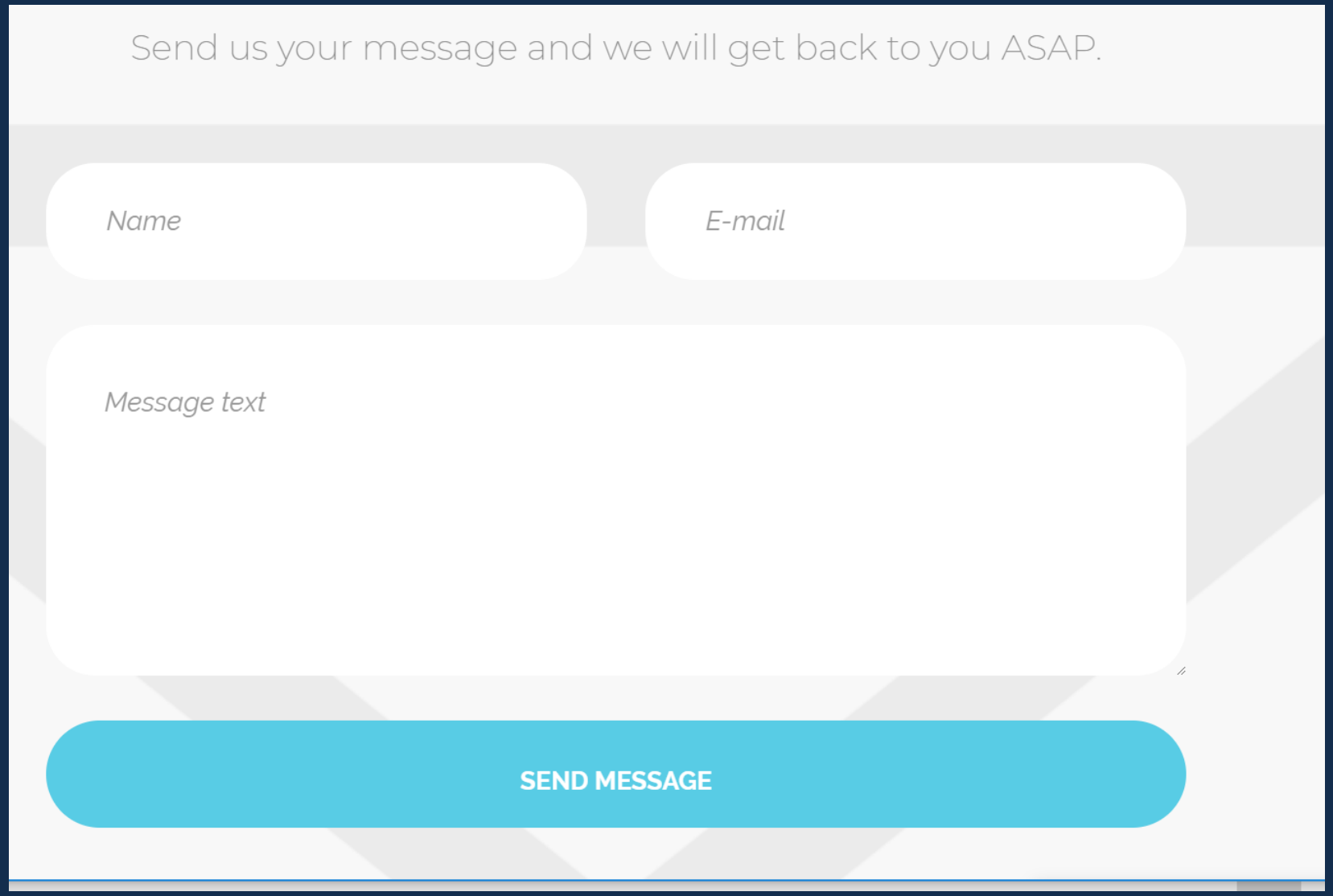
- Automatically creates SQL code based on the fields filled out in the form
- Displays the SQL code generated
- Provides the ability to modify the generated SQL code

Also provides users the means to write their own SQL code

Example next slide


phpMyAdmin
Recent
Favorites
New
AdServer
dev-bkp
events
Filter by name or regex
New
alt_questions
alt_question_templates
cities
events
event_available_matcher_
event_available_matcher_
event_details
event_exhibitors
event_exhibitor_matcher_a
event_exhibitor_products
event_exhibitor_specials
event_floor_booths
event_hours
event_sponsors
event_sponsorshipPossibil
happenings
key_happenings
Server: qavisiondev.cq1bhcw0baar.us-west-2.rds.amazonaws.com » Database: events » Table: cities
Browse
Structure
SQL
Search
Insert
Export
Import
Privileges
Operations
Triggers
Column
Type
Function
Null
Value
id
int(11)
name
varchar(128)
Boston
state
varchar(56)
Massachusettes
country
varchar(56)
USA
Go
Ignore
Column
Type
Function
Null
Value
id
int(11)
name
varchar(128)
state
varchar(56)
country
varchar(56)
Go

Example of a web form



Under the hood

- A web form may be used to perform these actions on a database:
  - enter data into a database or update existing data
  - retrieve data from a database, which is presented on the web page in which the web form resides
- These actions are triggered by pressing an action button (Send Message, etc.)
- The action button is associated with a script in some programming language such as PHP or JavaScript, which does the following when the button is clicked:
  - Establishes a connection to the database and sends an SQL command to be executed on the database
  - Retrieves the results and displays them in the appropriate place on the web form
- A web form provides very limited flexibility for interacting with data
  - Basically, the underlying queries are hard-coded into a program, which gets executed when the Action button is pressed.

Interacting with a database programmatically

- A program written in Java, Python, C#, etc., may need to perform some operations on a database
  - Example: It may need to do some computation on parameters passed to it using information stored in the database
  - The program may be passed information about the closing price of a list of stocks, and it may be asked to compute the percentage of change in price from the previous day's closing (which is stored in a database).
  - To do this computation, the program needs to retrieve this information from the database

Marrying programs to databases

- Like any user of the database, the program needs to establish a connection to the database
- To do that it needs a special routine (method) called a ‘connector’
- The data types of the programming language and the data types of the data manipulation language of the database (i.e., SQL) may not match
  - This is called ‘impedance mismatch’
  - The programming language needs a routine (often, called a ‘wrapper’) for mapping its data types to the data types of SQL.

Example in Java

```
public static void classDemo() throws Exception
{
Class.forName("com.mysql.jdbc.Driver"); //jdbc is the connector

Connection con = DriverManager.getConnection(
Config.DEMO_DB_PATH, Config.DEMO_USER_NAME,
Config.DEMO_PASSWORD);

String sql = “INSERT INTO cities(name, state, country) VALUES
(?,?,?) “;

PreparedStatement query = con.prepareStatement(sql);
query.setString(1,  "Kansas City");
query.setString(2,  "Missouri");
query.setString(3,  "USA");
query.executeUpdate(sql);

con.close();
}
```

Looking ahead

For Project 2 you will have to interact with either a Cassandra database or a MongoDB database through a Python program.

You may want to look at examples of how to interact with your MySQL database using a Python program (just to feel comfortable using Python to connect with a database).

Through APIs

Application Program Interface

- An application (say, Yelp) gives a program an interface to retrieve information from its database
- The program must make a request to the application using some protocol like http
- The answers are returned in the form of a JSON object or an XML document
- Essentially, an API is a web service plus an authentication protocol like Oauth
  - A web service is a service that can be requested/invoked via the web
  - You ask something, you get something. You don't need to know how the service does its job.

Interacting using Intelligent Assistants

- Intelligent agents, such as Siri, Google Now, etc., can be asked to do simple transactions
  - “Siri, what time is my appointment with Jen on Tuesday”
  - Siri: At 3 pm
  - When an Intelligent agent makes an appointment or retrieves information about an appointment it performs an action on a database.
  - There are intelligent agents that can perform operations on a CRM system like SalesForce or on an accounting system like QuickBooks
- It is also possible to build a natural language interface to databases which takes instructions in English and turns it into a SQL query to be executed on a database.

What is JSON?

- JavaScript Object Notation
- It is a format for representing data for storing and exchanging data
  - It is a data model
- In MongoDB (which we will study later) the data (documents) are stored as JSON objects
- APIs return data in JSON format
  - Also, optionally, as XML documents

JSON Syntax

- Data is in key/value pairs
- Data is separated by commas
- Curly braces hold objects
  - In this context an object is a set of key/value pairs separated by commas
- Square brackets hold arrays
  - An array is a sequence of values which can be accessed randomly, i.e., without having to traverse the previous values
  - E.g., You can get to the 5th position in an array without having to go through the previous values in the sequence

JSON Key/Value pairs

- Consists of a key in double quotes (i.e., a string) followed by a colon followed by a value
  - E.g., “city” : “New York”
- Values must be one of
  - a string
  - a number
  - a JSON object
  - an array
  - a boolean
  - null

JSON Example

```
{
“name” : “Swami”,
“position” : “Vice President”,
“current” : “Goldman Sachs”,
“previous” : [“PwC”, “McKinsey”, “Chase”],
“address” : {“apt” : 66, “street” : “235 W. Broadway”, “city”: “New York” }
}
```

an array

a JSON object

JSON exercise

Do this in class—10 minutes

Represent the following information in a well-formed JSON object. Represent every fact about her in separate key/value pairs. So don't put 'Palm Beach, FL' as one value—break it up into a city and a state value.

Vera BigShot is the President of BigDeal, Inc. She has two pets: a dog named Woof who is three years old and a cat named Meow who is ten years old. She has three houses, one in New York, NY; one in Palm Beach, FL; and one in Flagstaff, AZ. She is not married. Her hobbies are golfing and tormenting student interns. She was born on June 10, 1970 in Oakland, California.

Looking ahead to next week

In the remaining time we will discuss how to install and configure Cassandra database programs on your computers.

Recap of week

- Interacting with data through
  - Web forms
  - Programs
  - APIs
  - Other channels (chat bots?)
- JSON objects as a data model

Next Week

- Introduction to NoSQL databases
- The CAP theorem
- Cassandra database
- How to install Cassandra on your computer