Data Analytics Training – SQL

Introduction

- Bachelor of Engineering
- Data analyst at Tesco
- Rotman MBA at UofT, and graduated in 2015
- Data scientist at Capital One Canada, a bank specializing in credit cards and known for its data-driven approach
- Extensive experience in job hunting, and would like to share with everyone

One of the most popular jobs in the era of Big Data

Data scientists, data engineers and data analysts are among the most sought-after positions worldwide

Machine Learning Engineers, Data Scientists and Big Data Engineers rank among the top emerging jobs

LinkedIn

Data Job growth in the next decade is expected to outstrip growth during the previous decade, creating 11.5M jobs by 2026

U.S. Bureau of Labor Statistics

The average salary for a Data Scientist is **$95,173** per year in Canada

Indeed

- Demand For Data Scientists Will Soar 28% By 2020
- 59% of all Data Science and Analytics (DSA) job demand is in Finance and Insurance, Professional Services, and IT

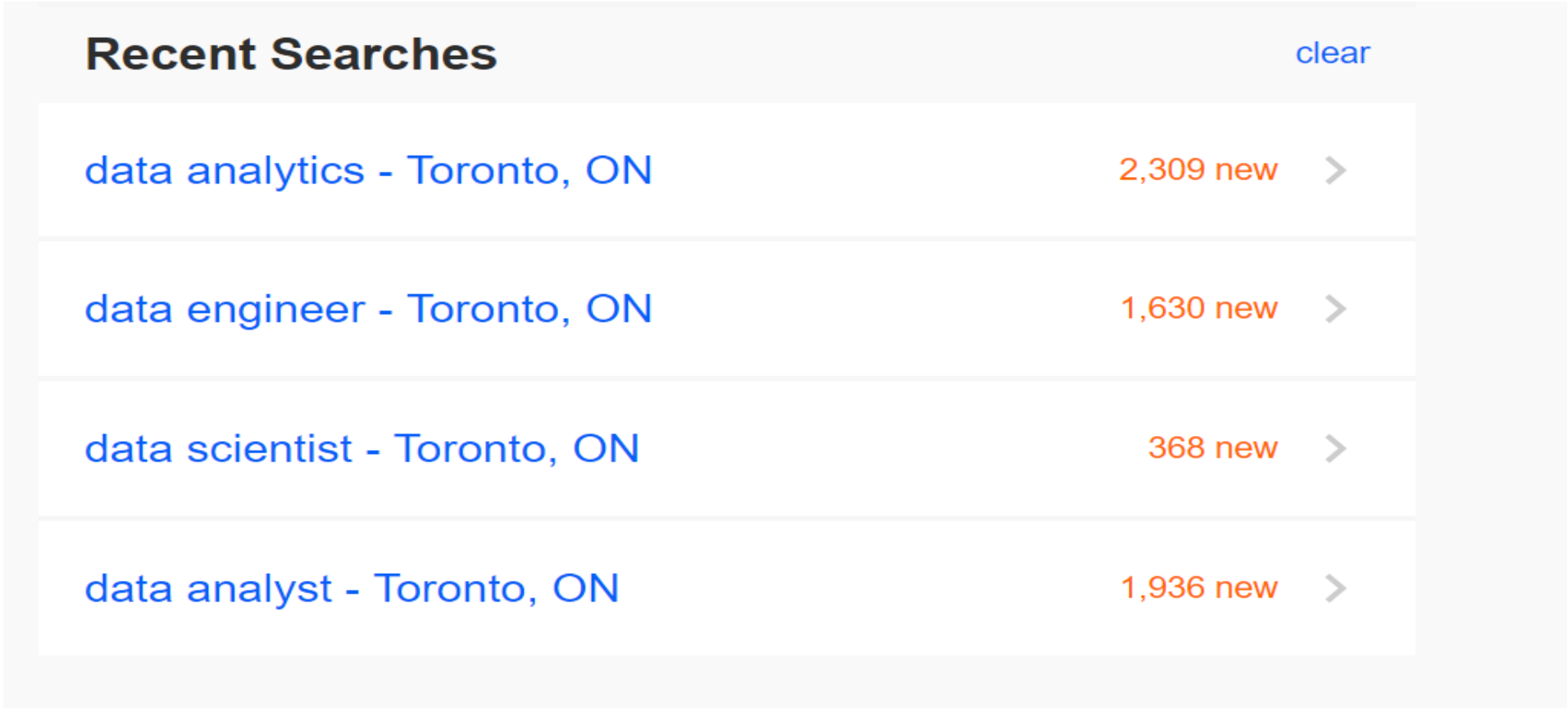
IBM

Over the next three years, there will be an additional 2.7m new jobs created in data science and analytics

PwC

One of the most popular jobs in the era of Big Data

Many openings on Canada Indeed and grow faster than other jobs



One of the most popular jobs in the era of Big Data

Salary is competitive

Machine Learning Engineer
99 salaries reported
Machine Learning Engineer Jobs
$102,765 per year

Data Analyst
339 salaries reported
Data Analyst Jobs
$59,877 per year

Data Scientist
161 salaries reported
Data Scientist Jobs
$96,808 per year

Database Analyst
55 salaries reported
Database Analyst Jobs
$66,914 per year

Reporting Analyst
77 salaries reported
Reporting Analyst Jobs
$66,749 per year

Salary estimated from 1,394 employees, users, and past and present job advertisements on Indeed in the past 36 months. Last updated: November 10, 2018

Lots of data analytics jobs

Demand in Canada is high across different industries

- Banking, especially from retail banking, e.g. credit card, mortgage business
- Insurance, e.g. SunLife, Manulife
- Retail, e.g. Walmart, Sobeys, Loblaw
- Technology, e.g. Amazon, Google, Facebook
- Government/Healthcare

Also, across different job functions

- Data analytics/Data scientist/Data engineer
- Machine Learning
- Reporting
- Risk management
- Marketing
- Trading
- Financial analyst

Finding a data-related job is not difficult

In a short-term:

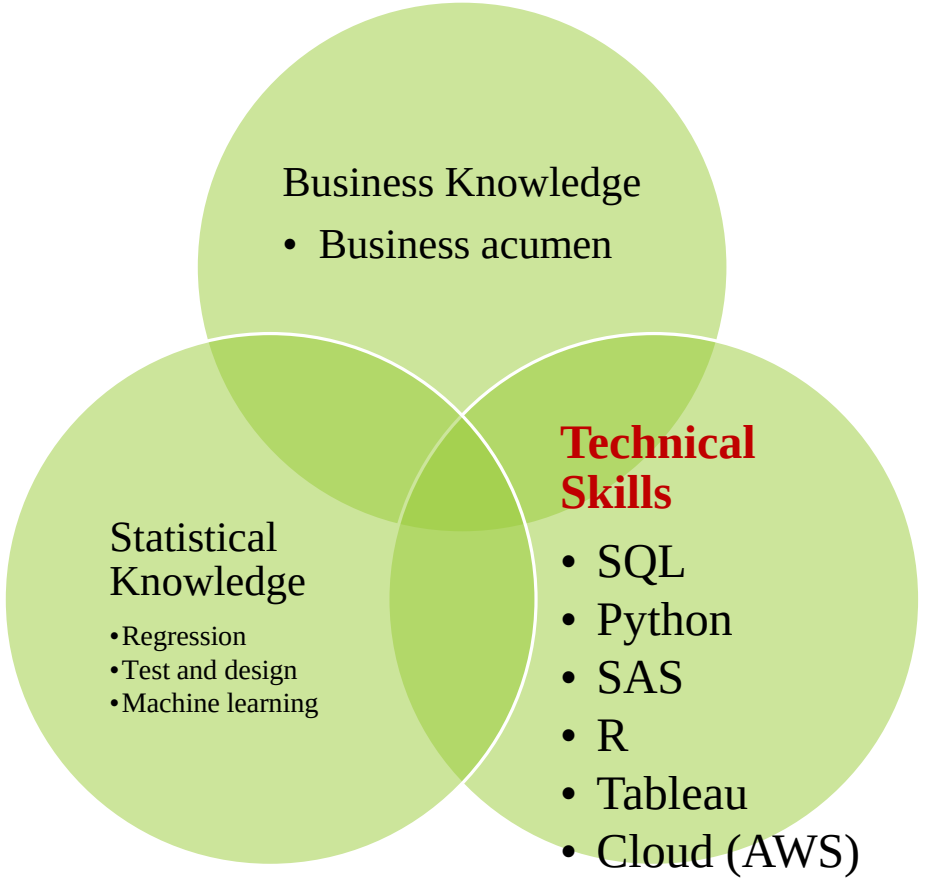
- Your education background doesn’t matter too much
- Very advanced IT or statistical knowledge is not necessary
- The relevant skill set is more important than previous work experience
- Never too late to switch to data area if you have a passion
- **You can always put relevant project experience on your resume to prove you are capable for the job!**

In a long-term:

- You can become an expert in a specific field
- Skills are easily transferable - You can switch to different job functions/industries with your skill set

Skills you need to develop from now on

- You should have a focus based on which specific job category you want to do in the future, but technical skills are the most important!



What is SQL?

Structured Query Language (SQL) is a special-purpose programming language

SQL's purpose:

To manipulate sets of data; typically from a relational database ANSI and ISO standards

What is a Relational Database?

Database:

A container to help organize data

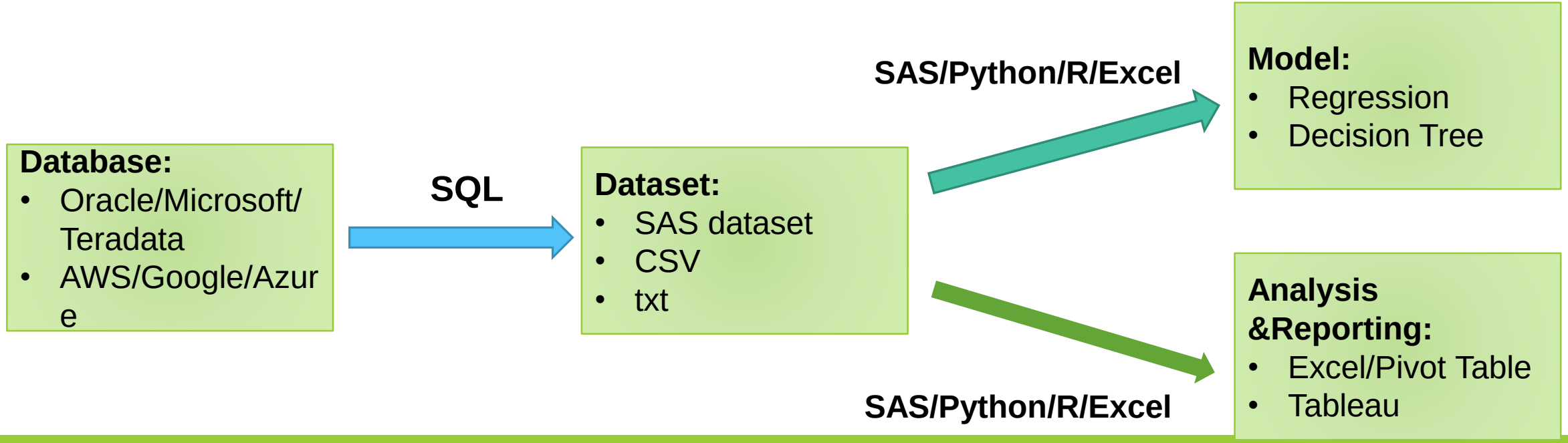
A way to efficiently store and retrieve data

Relational:

- A way to describe data and the relationships between data entities

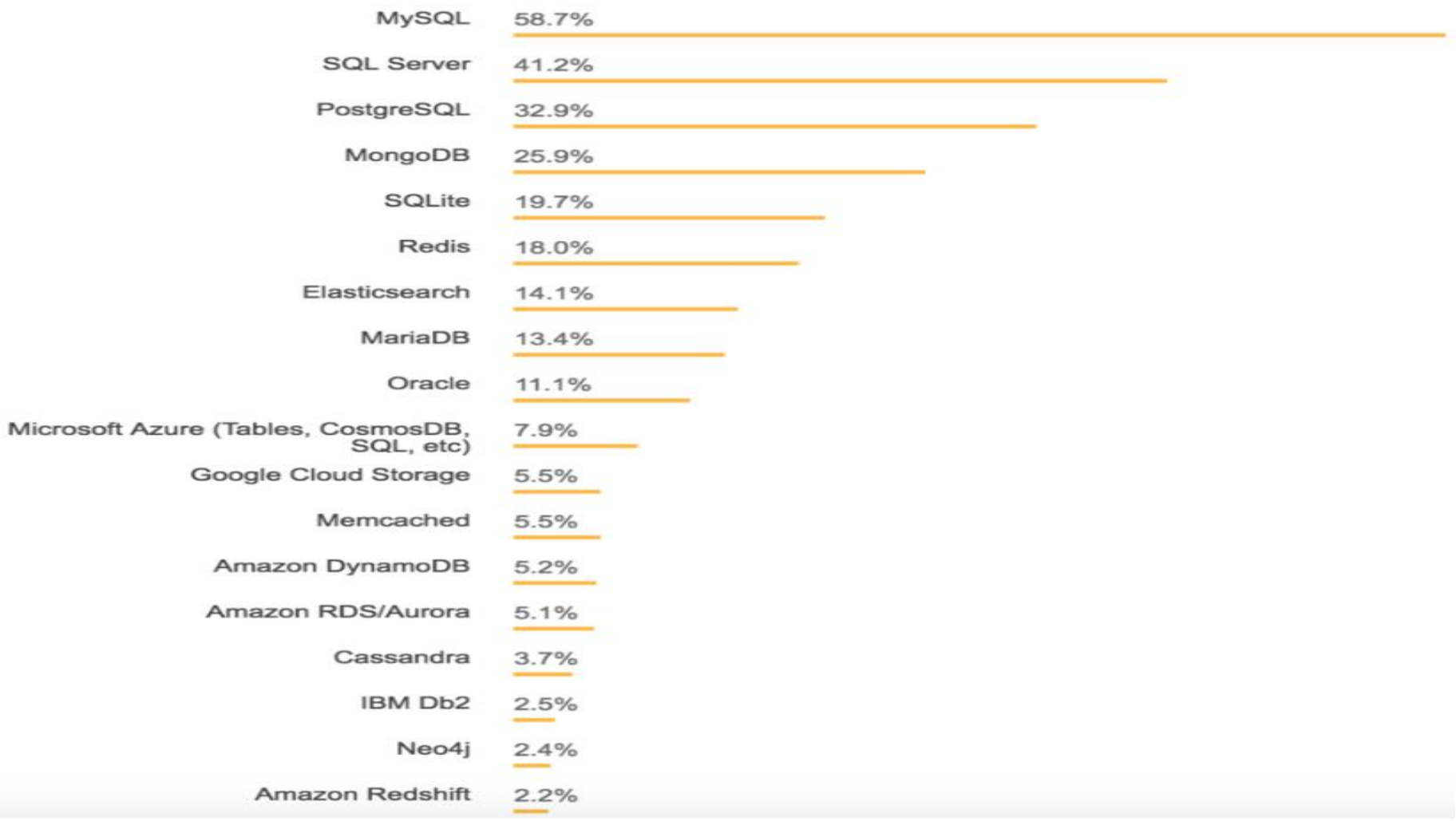
SQL is the most important skill for any Data related job!

- SQL – Structured Query Language
- At least 90% of your daily work is to write SQL query and pull data from database



Different types of Relational Databases?

Why MySQL?



StackOverflow_2018_survey

Who use MySQL?

Features of this SQL Trainings

Provide practices in each class based on real business daily jobs – please code by yourselves and improve hands-on skills!

2 quiz (online testing) to test your general SQL knowledge

2 mock technical interviews – solve real business issue within 30 mins

2 hands-on projects to put on resume

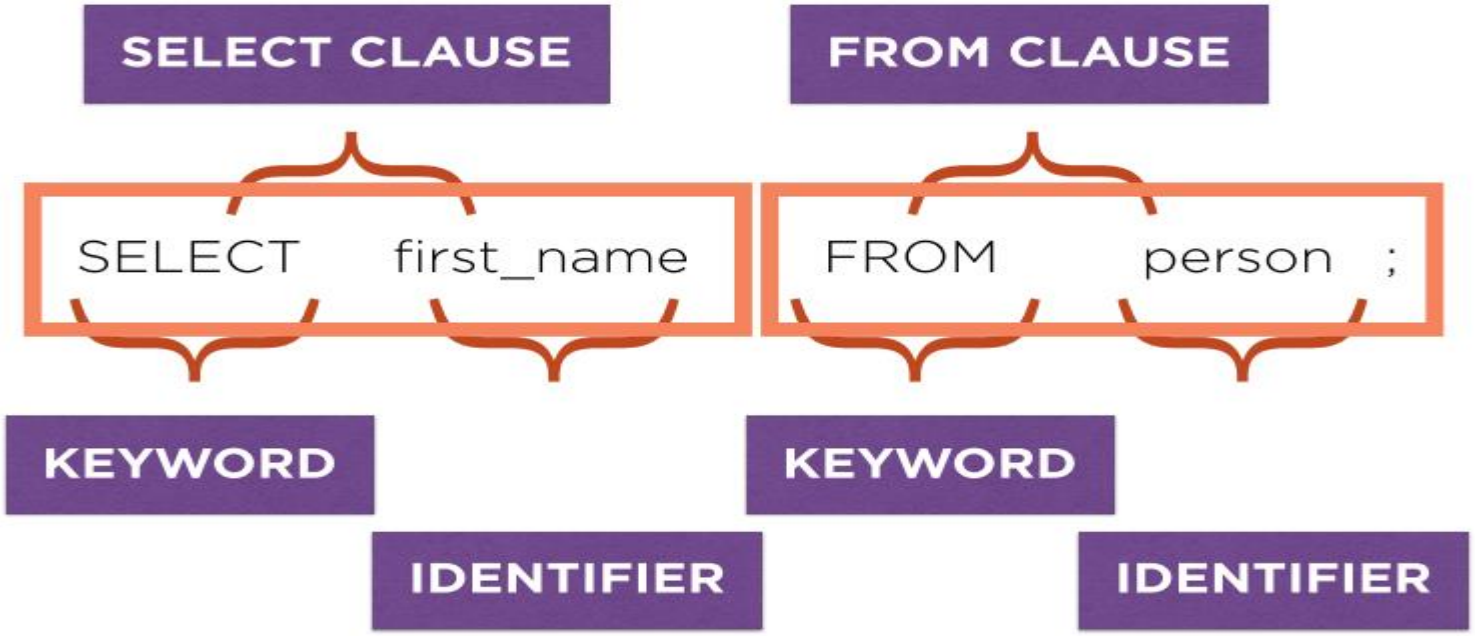
The purpose of my SQL trainings is:

- Put '**Advanced SQL skills**' on your resume
- Help you pass any SQL technical interview
- **Most importantly, build your confidence for your job search**

Basic SQL Syntax

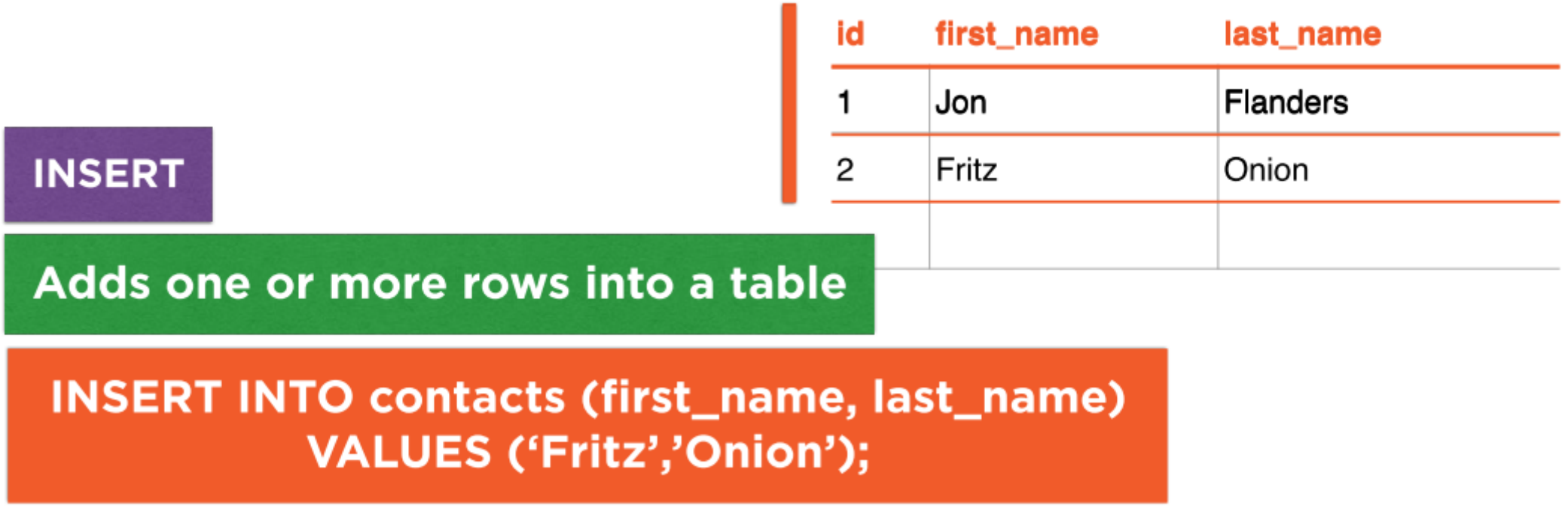
SQL Statement

A SQL Statement is an expression that tells a database what you want it to do:

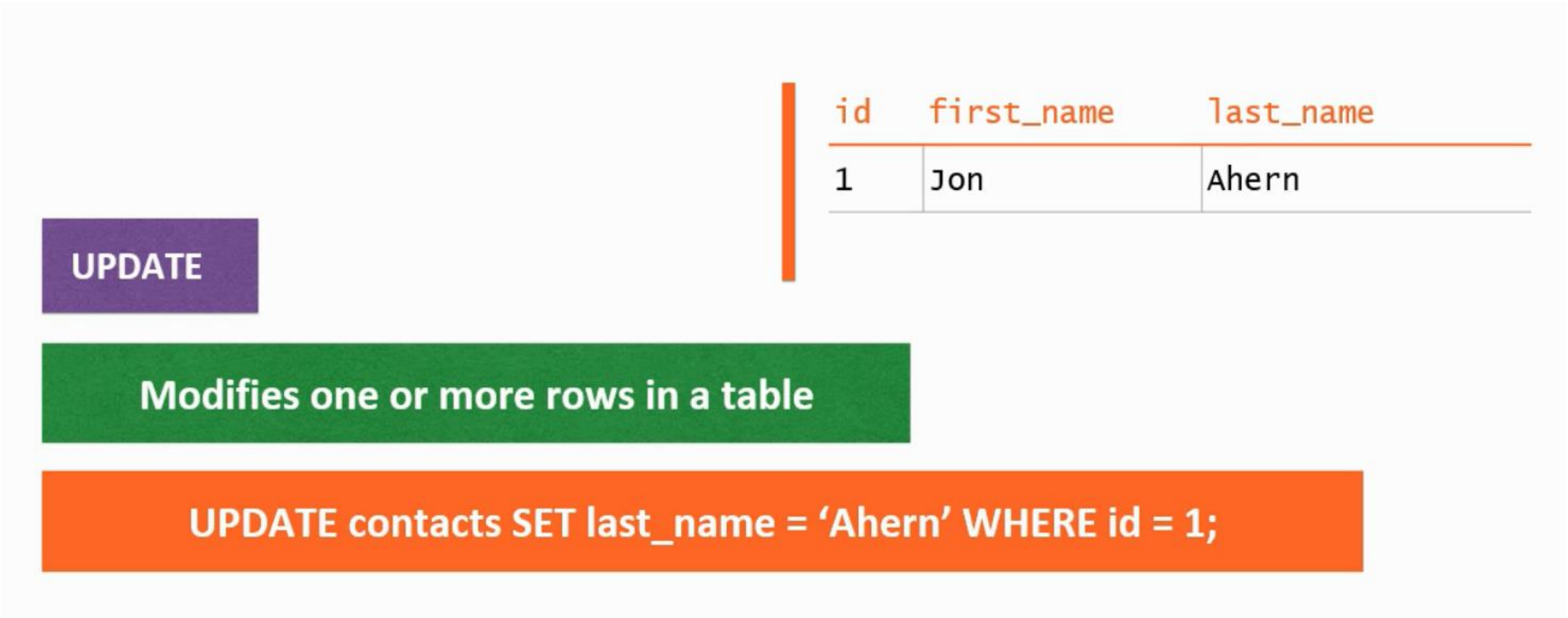


Basic SQL Commands

Basic SQL Commands



Basic SQL Commands



Basic SQL Commands

id	first_name	last_name
1	Jon	Flanders
2	Fritz	Onion

DELETE

Removes one or more rows from one table

```
DELETE FROM contacts where id = 2;
```

SELECT Statement

The SELECT Statement

- Most of the time it contains a list of columns from a table you want to query
- Then, a FROM clause is required
- After every column comes a comma
- *Except: no comma after the last column*

```
SELECT <COLUMN_NAME>, <COLUMN_NAME>  FROM  <TABLE_NAME> ;
```

```
SELECT prod_id, prod_name, prod_price FROM Products;
```

The SELECT Statement

- Use **Select *** to pulls all the columns from a table

```
SELECT * FROM Products;
```

- It is a bad practice.

Use Limit to constrain the display of records

SELECT * FROM Products **Limit 5**;

Ways to Constrain the Number of Results

- **DISTINCT Qualifier**
- **WHERE Clause**

Distinct

- Without DISTINCT:

  SELECT vend_id FROM Products;

- **With Distinct:**

  SELECT DISTINCT vend_id FROM Products;

  SELECT DISTINCT vend_id, prod_price FROM Products;

You always put distinct before any column and it will apply to all the following columns

Use Comments

1. Use two hyphens

```
SELECT prod_name   -- this is a comment
FROM Products;
```

2. A # at the start of a line makes the entire line a comment

```
# This is a comment
SELECT prod_name
FROM Products;
```

3. /* starts a comments, and */ ends it. Anything between /* and */ is comment text.

```
/* SELECT prod_name, vend_id
FROM Products; */
SELECT prod_name
FROM Products;
```

Practice 1.1:

Use SCHEMAS(Database) **world** and table **country** to practice:

- Select all columns from table country and only display 5 records
- Only want to check code, name, region, population columns from table country;
- Want to check different values of region in table country
- Comment out one of the query you just wrote down

Sort Retrieved Data by Using Order By

Order By clause:

- ORDER BY takes the name of one or more columns by which to sort the output
- Position of ORDER BY Clause When specifying an ORDER BY clause, be sure that it is the **last clause** in your SELECT statement

If it is not the last clause, an error will be generated.

```
SELECT prod_name FROM Products
ORDER BY prod_name;
```

Sort by Multiple Columns

- To sort by multiple columns, simply specify the column names **separated by commas** (just as you do when you are selecting multiple columns):

```
SELECT prod_id, prod_price, prod_name
FROM Products
ORDER BY prod_price, prod_name;
```

- It is important to understand that when you are sorting by multiple columns, the sort sequence is exactly as specified

Sort by Column Position

- ORDER BY also supports ordering specified by **relative column position**

```
SELECT prod_id, prod_price, prod_name
FROM Products
ORDER BY 2, 3;
```

- The relative positions of selected columns in the SELECT list are specified. ORDER BY 2 means sort by the second column in the SELECT list, the prod_price column

- Obviously you cannot use this technique when sorting by columns that are not in the SELECT list

But you can mix and match actual column names and relative column positions in a single statement

Sort by Column Position

- ORDER BY also supports ordering specified by **relative column position**

```
SELECT prod_id, prod_price, prod_name
FROM Products
ORDER BY 2, 3;
```

- The relative positions of selected columns in the SELECT list are specified. ORDER BY 2 means sort by the second column in the SELECT list, the prod_price column

- Obviously you cannot use this technique when sorting by columns that are not in the SELECT list

But you can mix and match actual column names and relative column positions in a single statement

Specifying Sort Direction

- By default, it is **ascending sort order**
- To sort by descending order, the keyword DESC must be specified and put after the column you want to order by descending

```
SELECT prod_id, prod_price, prod_name
FROM Products
ORDER BY prod_price DESC;
```

- If you want to sort descending on multiple columns, be sure each column has its **own** DESC keyword

Filter Data by using Where Clause

- Retrieving just the data you want involves specifying search criteria, also known as a filter condition
- The WHERE clause is specified right **after the table name** (the FROM clause) as follows:

```
SELECT prod_name, prod_price
FROM Products
WHERE prod_price = 3.49;
```

- When using both ORDER BY and WHERE clauses, make sure that **ORDER BY comes after the WHERE**, otherwise an error will be generated

Where Clause Operators

Operator	Description
=	Equality
<>	Non-equality
!=	Non-equality
<	Less than
<=	Less than or equal to
!<	Not less than
>	Greater than
>=	Greater than or equal to
!>	Not greater than
BETWEEN	Between two specified values
IS NULL	Is a NULL value

- Some of the operators listed above are redundant

```
SELECT prod_name, prod_price
FROM Products
WHERE prod_price <= 10;
```

Where Clause Operators

- Check for Nonmatches:

```
SELECT vend_id, prod_name
FROM Products
WHERE vend_id <> 'DLL01';
```

- **When to use Quotes**: the single quotes are used to delimit a string
- **Be careful - the value is case-sensitive!**

- Check for a Range of Values:

```
SELECT prod_name, prod_price
FROM Products
WHERE prod_price BETWEEN 5 AND 10;
```

Where Clause Operators

- Check for No Value - When a column contains no value, it is said to contain a NULL value:

```
SELECT prod_name
FROM Products
WHERE prod_price IS NULL;
```

Practice 1.2:

Use SCHEMAS(Database) **world** and table **country** to practice:

- Display the TOP 5 countries with the largest population
- Rank the country by descending region, and ascending surfaceArea
- List the countries with lifeExpectancy => 75 and rank by ascending population
- List countries became independent between 1980 and 1990;
- List countries in region Eastern Asia and indepYear is null;
- Select countries in Western Europe, with population less than 80000000 and surfacearea larger than 3000, and rank these countries by descending Code column