

Google Developers



> _

Data science Track



intro to Big Data

Tawadros Gamal

@ DSC Assiut



Content

- Introduction
- What is Big Data
- Characteristic of Big Data
- Structure of Big Data
- Why Big Data
- Big Data sources
- Application of Big Data
- Risks of Big Data
- Benefits of Big Data
- Difference between Big Data and Data Science
- Tools to use in Big Data
- Hierarchy of Data
- BigQuery

Introduction

- Big Data may well be the Next Big Thing in the IT world.
- Big data burst upon the scene in the first decade of the 21st century.
- Like many new information technologies, big data can bring about dramatic cost reductions, substantial improvements in the time required to perform a computing task, or new product and service offerings

Let's have a look at the data generated per minute
on the internet



"2.1Million"



"3.8Million"



"1.0Million"



"4.5Million"



"188Million"

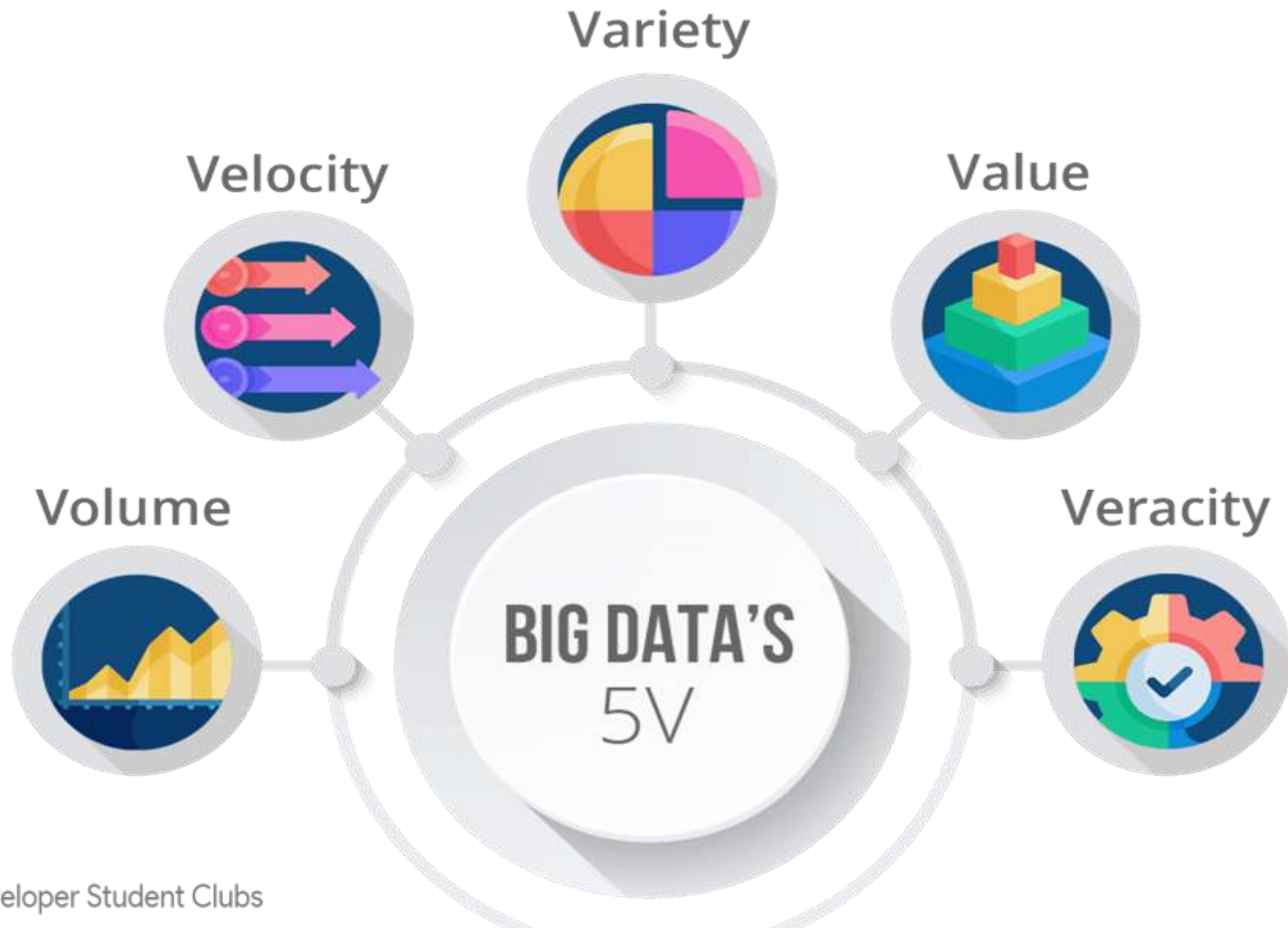
That's a lot of data



What is BIG DATA?

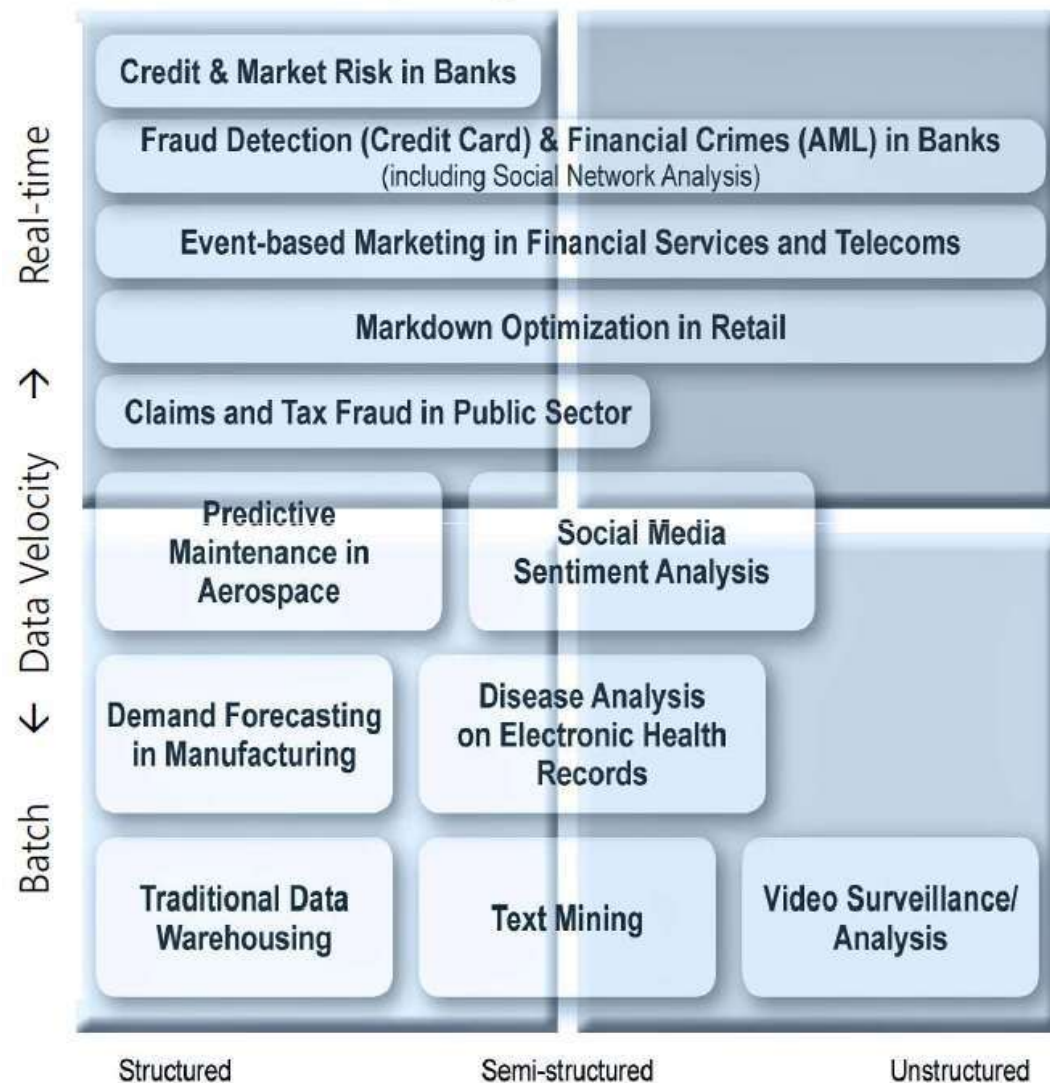
- 'Big Data' is similar to 'small data', but bigger in size
- but having data bigger it requires different approaches:
Techniques, tools and architecture
- an aim to solve new problems or old problems in a better way.
- Big Data generates value from the storage and processing of very large quantities of digital information that cannot be analyzed with traditional computing techniques.

Characteristics of Big Data



The Structure of Big Data

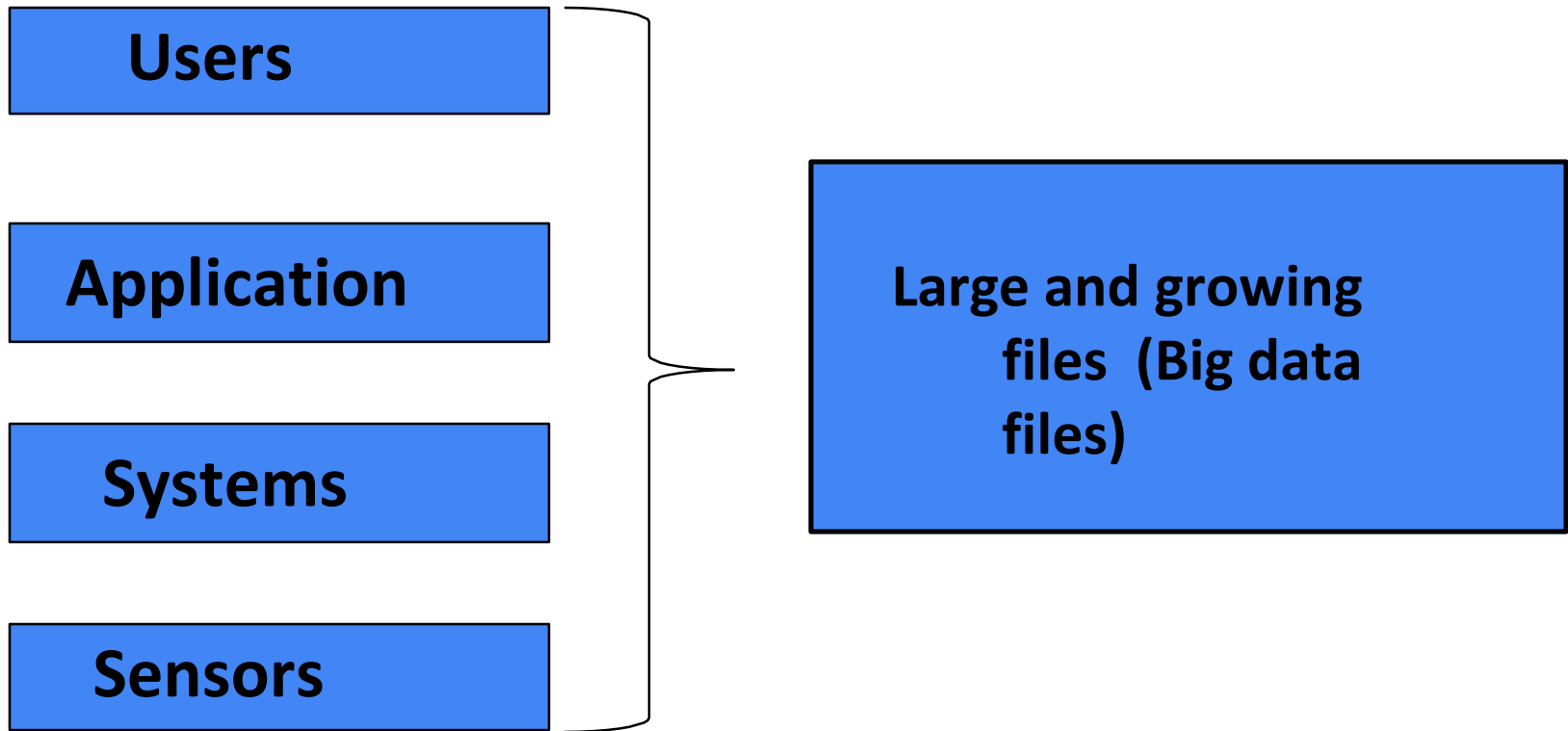
- Structured
 - Most traditional data sources
- Semi-structured
 - Many sources of big data
- Unstructured
 - Video data, audio data



Why Big Data

- Growth of Big Data is needed
 - Increase of storage capacities
 - Increase of processing power
 - Availability of data(different data types)

Big Data sources



Data generation points Examples

Mobile Devices

Readers/Scanners

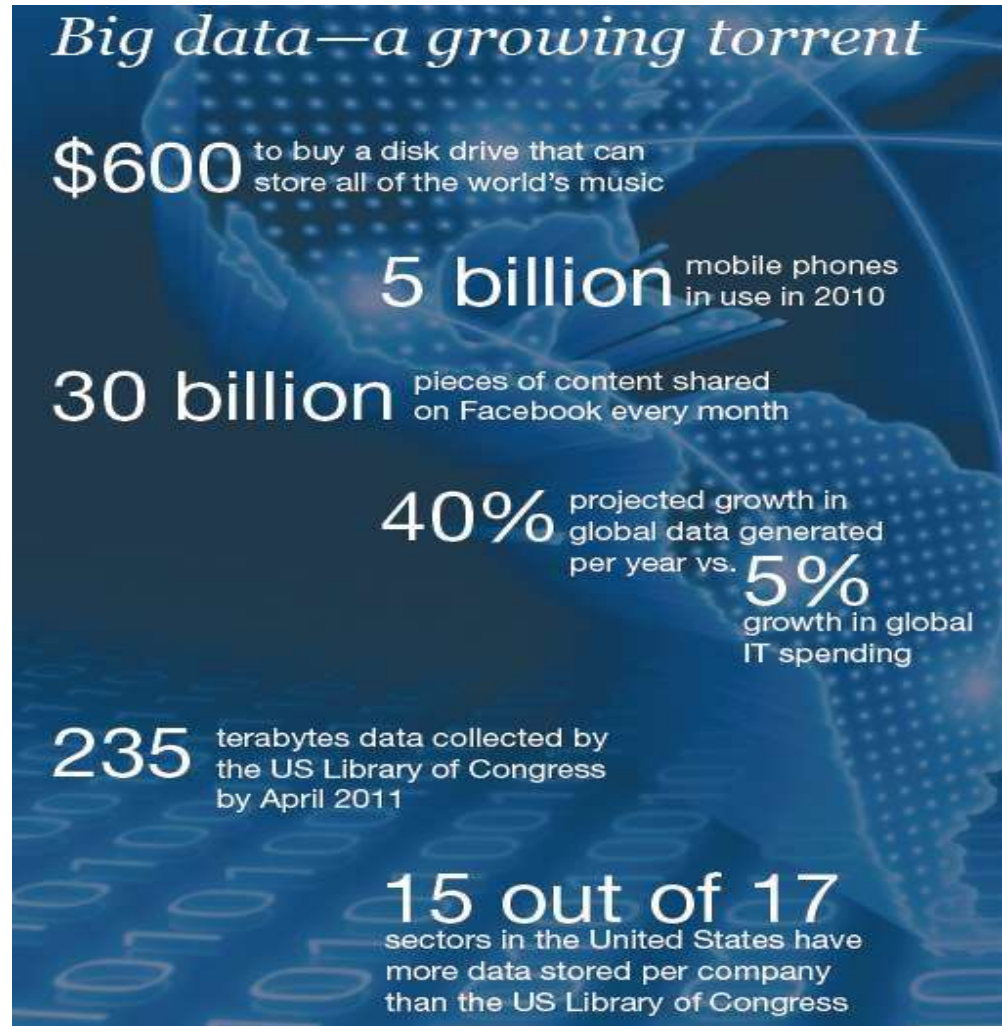
Science facilities

Programs/Software

Web Social Media

Cameras

Microphones



Big Data Analytics

- Examining large amount of data
- Appropriate information
- Identification of hidden patterns, unknown correlations
- Competitive advantage
- Better business decisions: strategic and operational
- Effective marketing, customer satisfaction, increased revenue

Application Of Big Data analytics

Smarter
Healthcare



Homeland
Security



Traffic
Control



Manufacturing



Multi-channel
sales



Teleco
m



Trading
Analytics

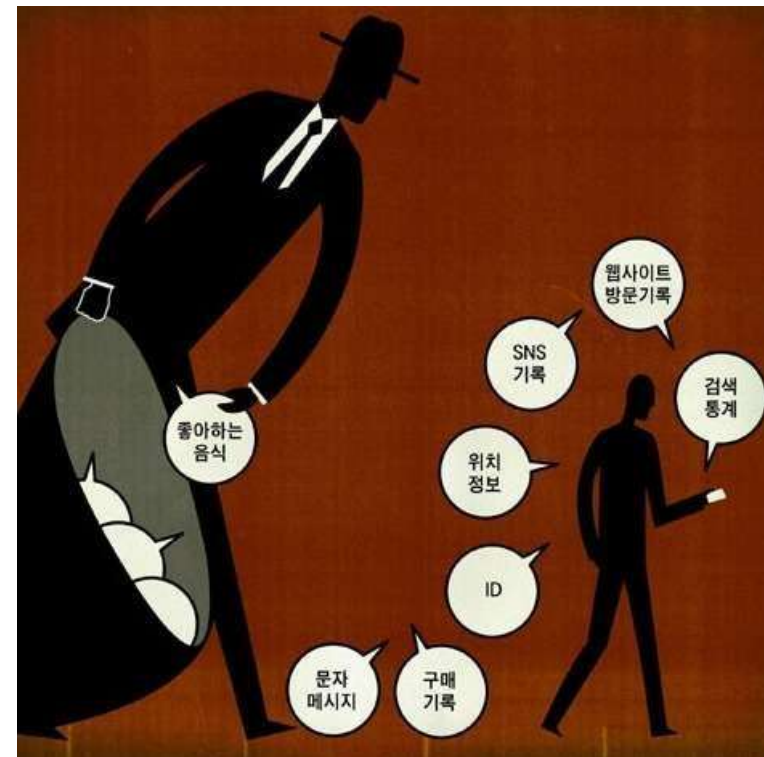


Search
Quality



Risks of Big Data

- Will be so overwhelmed
 - Need the right people and solve the right problems
- Costs escalate too fast
 - Isn't necessary to capture 100%
- Many sources of big data is privacy



Benefits of Big Data

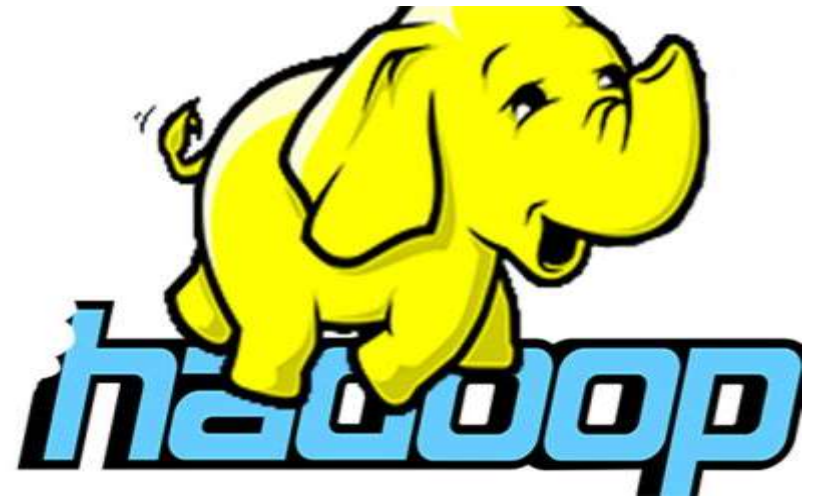
- Make better decisions and take meaningful actions at the right time.
- Store Huge data with Hadoop and Scala.
- Run queries without changing the data structures underneath.

Tools to use in Big Data

- Bigquery
- Scala
- Hadoop



BigQuery



Difference between Data Science and Big Data

- **Meaning**
- **Applications Area**
- **Concept**
- **Approche**
- **Basis of formation**

#1.MEANING

BIGDATA



- Huge volumes of data which cannot be handled using traditional database programming
- Characterized by volume, variety, and velocity

DATA SCIENCE



- A data focused scientific activity
- Approaches to process big data
- Harnesses the potential of big data for business decisions
- Similar to data mining

#2. Concept

BIG DATA



- Diverse data types generated from multiple data sources
- Includes all types and formats of data

DATA SCIENCE



- A specialized area involving scientific programming tools, models and techniques to process big data
- Provides techniques to extract insights and information from large data sets
- Supports organizations in decision making

#3. Basis of formation

BIG DATA



- Internet users/traffic
- Electronic devices (sensors, RFID, etc.)
- Audio/video streams including live feeds
- Online discussion forums
- Data generated in organizations (transactions, DB, spreadsheets, emails, etc.)
- Data generated from system logs

DATA SCIENCE



- Applies scientific methods to extract knowledge from big data
- Related to data filtering, preparation, and analysis
- Capture complex patterns from big data and develop models
- Working apps are created by programming developed models

#4. Application areas

BIG DATA



- Financial services
- Telecommunications
- Optimizing business processes
- Performance optimization
- Health and sports
- Improving commerce
- Research and development
- Security and law enforcement

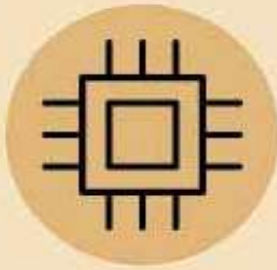
DATA SCIENCE



- Internet search
- Digital advertisements
- Search recommenders
- Image/Speech recognition
- Fraud, risk detection
- Web development
- Other miscellaneous areas/utilities

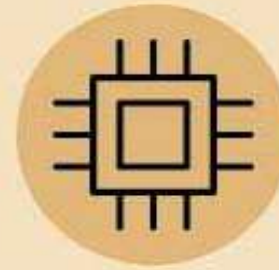
#5. Approach

BIG DATA



- To develop business agility
- To gain competitiveness
- Leverage datasets for business advantage
- Establish realistic metrics and ROI
- To achieve sustainability
- To understand markets and gain new customers

DATA SCIENCE



- Involves extensive use of mathematics, statistics, and other tools
- State-of-the-art techniques/ algorithms for data mining
- Programming skills (SQL, NoSQL), Hadoop platforms
- Data acquisition, preparation, processing, publishing, preserve or destroy
- Data visualization, prediction

Hierarchy of Data

- Column
- Row
- Record
- Table
- Dataset
- Warehouse



Google Cloud



BigQuery

- BigQuery is an enterprise data warehouse that solves this problem by enabling super-fast SQL queries using the processing power of Google's infrastructure



Some of MYSQL

- **SELECT**
- **INSERT**
- **DELETE**
- **UPDATE**
- **Where**
- **AS**
- **ORDER BY - GROUP BY**

Cont.

```
SELECT USER FROM example_table
```

```
SELECT USER, SHIPPED FROM example_table
```

```
SELECT USER FROM example_table WHERE  
SHIPPED='YES'
```

	A	B	C
1	USER	PRICE	SHIPPED
2	SEAN	\$35	YES
3	ROCKY	\$50	NO
4	AMANDA	\$20	YES
5	EMMA	\$65	YES
6	ANDRES	\$10	NO
7	CASEY	\$55	YES
8	HANNAH	\$15	NO
9	JOCELYN	\$30	NO

Cont.

```
SELECT USER FROM example_table WHERE  
SHIPPED='YES' GROUP BY USER
```

```
SELECT USER FROM example_table WHERE  
SHIPPED='YES' GROUP BY USER ORDER BY USER  
DESC
```

```
SELECT USER AS MyUser FROM example_table  
WHERE SHIPPED='YES' GROUP BY USER ORDER  
BY USER ASC
```

Cont.

```
CREATE TABLE t(  
  Id INT PRIMARY KEY  
  Name VARCHAR NOT NULL  
  Price INT DEFAULT 0  
)
```

```
INSERT INTO t (Id,Name,Price)  
VALUES ("id5","Orange",5)
```

```
UPDATE t  
SET Price == 10  
WHERE Id == "id5"
```

```
DELETE FROM t WHERE Id == "id5"
```



Qwiklabs labs

Qwiklabs labs are hands-on labs.

A hands-on lab is an online learning environment along with a set of instructions to walk you through a live, real world, and scenario-based use case. In a lab you have access to the actual environment you want to learn about, not a simulation or demo environment. You can access the lab environment from anywhere on the Internet using a standard browser.



Any Questions



Remember That #U_are_Our_mission

Thank You!

