

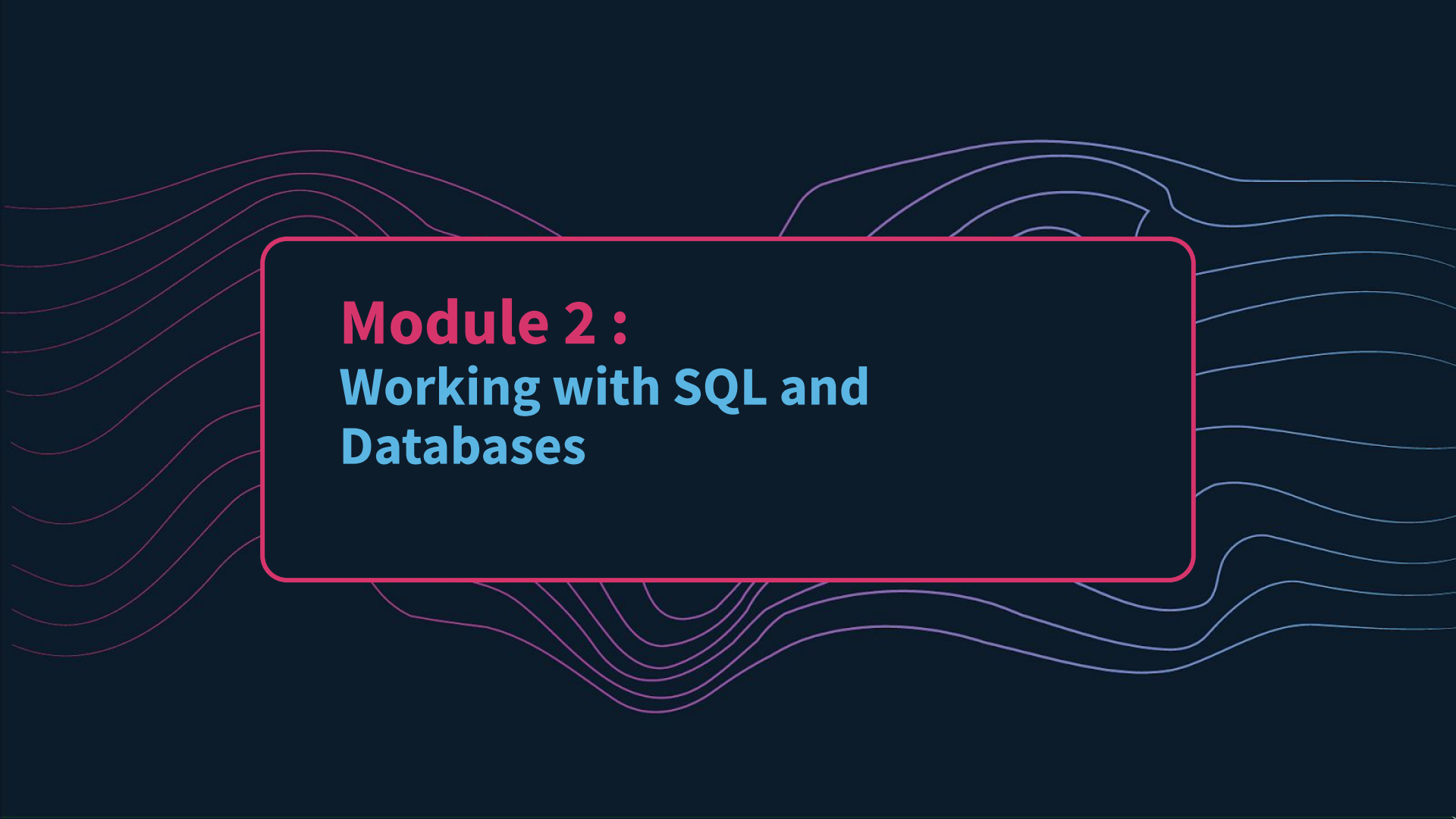


WeCloudData

SQL Fundamentals

Kick off your career in data science & analytics





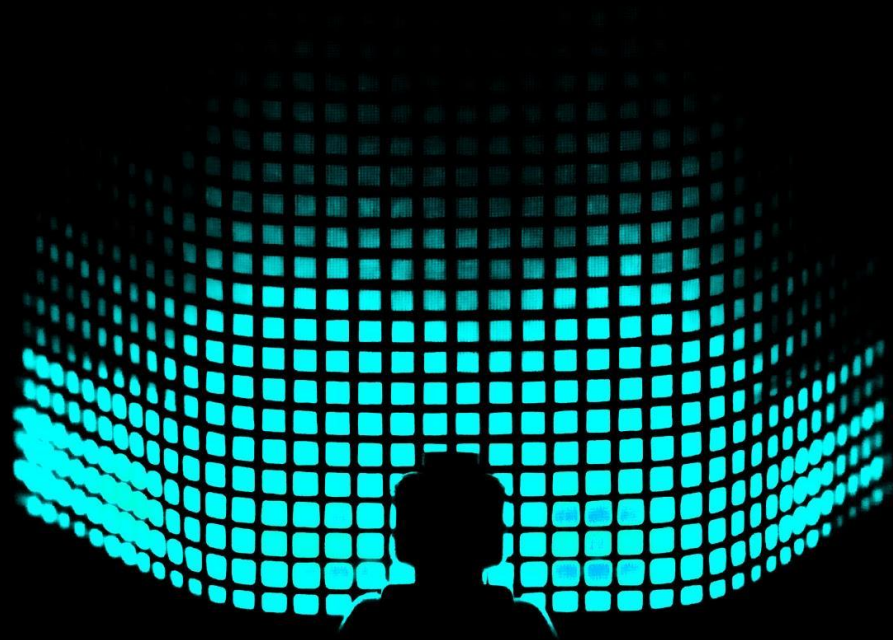
Module 2 : **Working with SQL and Databases**



Learning Objectives

In this module, we will introduce how to write queries in SQL. Specifically, we will share with you:

- What we can achieve with SQL
- SQL data types
- Working with databases and tables



RDBMS & Operations

What is SQL?

Why learn SQL?

What can SQL do?

SQL Data Types

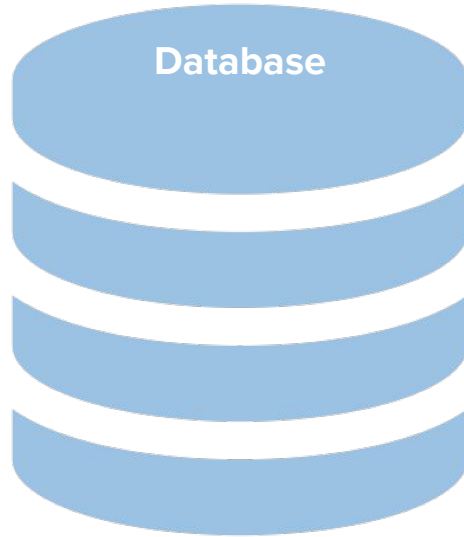
Working with Databases and Tables

Agenda.



What are the Components of a RDBMS?

Introduction to SQL



RDBMS a.k.a Relational
DataBase Management System
has the following major components:

- Table
- View
- Index
- Schema





What are the Components of a RDBMS?: Table

Introduction to SQL

A **table** is a collection of data represented in rows and columns.
The column's data type is explicitly defined.

Column Header

ProductID	ProductName	ProductCategory
5889	White GlueTop Scratch Pads	Office Supplies
15497	Fellowes 8 Outlet Superior Workstation Surge Protector	Office Supplies
16735	"Belkin 325VA UPS Surge Protector, 6"	Office Supplies
23721	Gyration Ultra Cordless Optical Suite	Technology
27473	Epson FX-980 Dot Matrix Printer	Technology
30902	"Adams Phone Message Book, Professional, 400 Message C...	Office Supplies
34354	"O'Sullivan Elevations Bookcase, Cherry Finish"	Furniture
37176	"SAFCO PlanMaster Heigh-Adjustable Drafting Table Base, 4...	Furniture
37996	Belkin 8 Outlet Surge Protector	Office Supplies
41129	"Letter/Legal File Tote with Clear Snap-On Lid, Black Granite"	Office Supplies
46361	Letter or Legal Size Expandable Poly String Tie Envelopes	Office Supplies
47324	Xerox 1980	Office Supplies
48396	"It's Hot Message Books with Stickers, 2 3/4"" x 5"""	Office Supplies
49746	"Hewlett-Packard Business Color Inkjet 3000 [N, DTN] Series...	Technology

product	ProductID	int(11)
product	ProductName	varchar(200)
product	ProductCategory	varchar(20)
product	ProductSubCateg...	varchar(50)
product	ProductContainer	varchar(20)
product	ProductBaseMargin	decimal(4,2)

Data Type

Column




What are the Components of a RDBMS?: Schema

Introduction to SQL

A **schema** is a collection of database objects including tables, views, constraints, indexes, sequences, etc.

Non-schema objects include users and roles.



Info Tables Columns Indexes Triggers Views Stored Procedures Functions Grants Events										
Name	Engine	Version	Row Format	Rows	Avg Row Length	Data Length	Max Data Length	Index Length	Data Free	Au
customer	InnoDB	10	Dynamic	1832	98	176.0 KiB	0.0 bytes	0.0 bytes	0.0 bytes	
customer_multi_order	InnoDB	10	Dynamic	44	372	16.0 KiB	0.0 bytes	0.0 bytes	0.0 bytes	
order_multi_prod	InnoDB	10	Dynamic	2032	56	112.0 KiB	0.0 bytes	0.0 bytes	0.0 bytes	
orders	InnoDB	10	Dynamic	7909	200	1.5 MiB	0.0 bytes	0.0 bytes	4.0 MiB	
product	InnoDB	10	Dynamic	1234	185	224.0 KiB	0.0 bytes	0.0 bytes	0.0 bytes	
product_new	InnoDB	10	Dynamic	1234	146	176.0 KiB	0.0 bytes	0.0 bytes	0.0 bytes	
returns	InnoDB	10	Dynamic	572	85	48.0 KiB	0.0 bytes	0.0 bytes	0.0 bytes	





What are the Components of a RDBMS?: Index

Introduction to SQL

An **index** is a copy of selected columns of data from a table. It can be used to speed up searches/queries when retrieving data from the database.

Indexes include a low-level disk block address/direct link to the complete row of data it was copied from.

[wikipedia](https://en.wikipedia.org/wiki/Index_(database))

INDEX	
ABC, 164, 321 ⁿ	Anello, Douglas, 60
academic journals, 262, 280–82	animated cartoons, 21–24
Adobe eBook Reader, 148–53	antiretroviral drugs, 257–61
advertising, 36, 45–46, 127, 145–46, 167–68, 321 ⁿ	Apple Corporation, 203, 264, 302
Africa, medications for HIV patients in, 257–61	architecture, constraint effected through, 122, 123, 124, 318 ⁿ
Agee, Michael, 223–24, 225	archive.org, 112
agricultural patents, 313 ⁿ	see also Internet Archive
Aibo robotic dog, 153–55, 156, 157, 160	archives, digital, 108–15, 173, 222, 226–27
AIDS medications, 257–60	Aristotle, 150
air traffic, land ownership vs., 1–3	Armstrong, Edwin Howard, 3–6, 184, 196
Akerlof, George, 232	Arrow, Kenneth, 232
Alben, Alex, 100–104, 105, 198–99, 295, 317 ⁿ	art, underground, 186
alcohol prohibition, 200	artists:
<i>Alice's Adventures in Wonderland</i> (Carroll), 152–53	publicity rights on images of, 317 ⁿ
	recording industry payments to, 52, 58–59, 74, 195, 196–97, 199, 301, 329 ⁿ –30 ⁿ

Example:

A book without an index would make it very difficult to find subjects of interest!



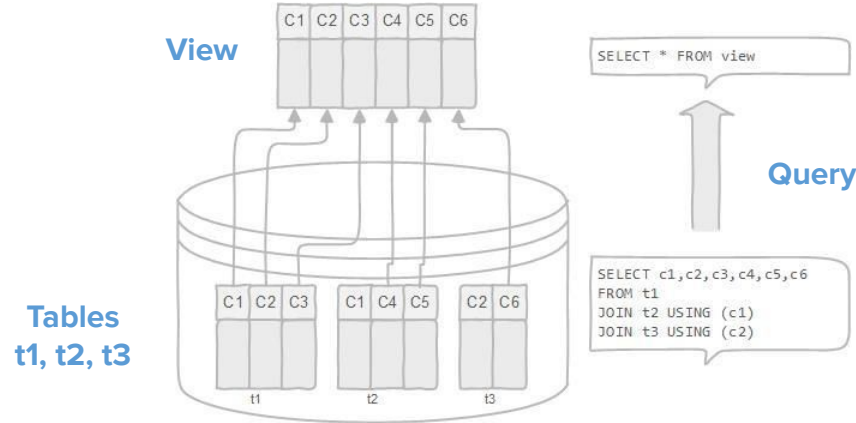


What are the Components of a RDBMS?: View

Introduction to SQL

A **view** is the resulting set of a stored query on data, which users can query just as they would for a persistent database collection object.

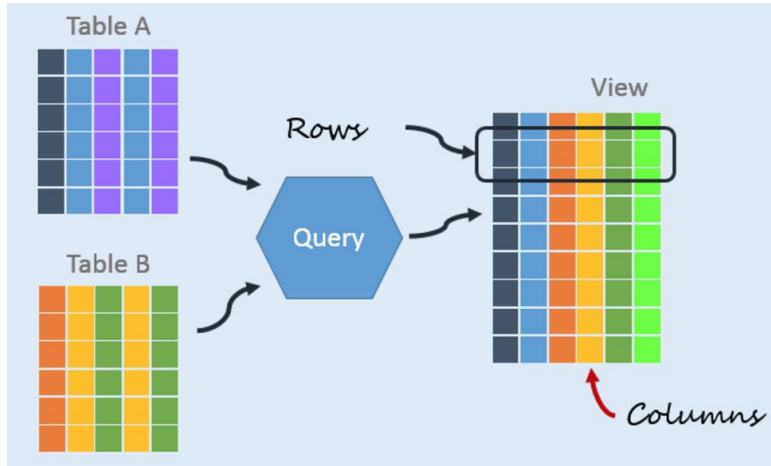
It allows you to simplify complex queries as well as provide extra security by limiting data access to specific users.





What are the Components of a RDBMS?: View

Introduction to SQL



Anatomy of a **View**

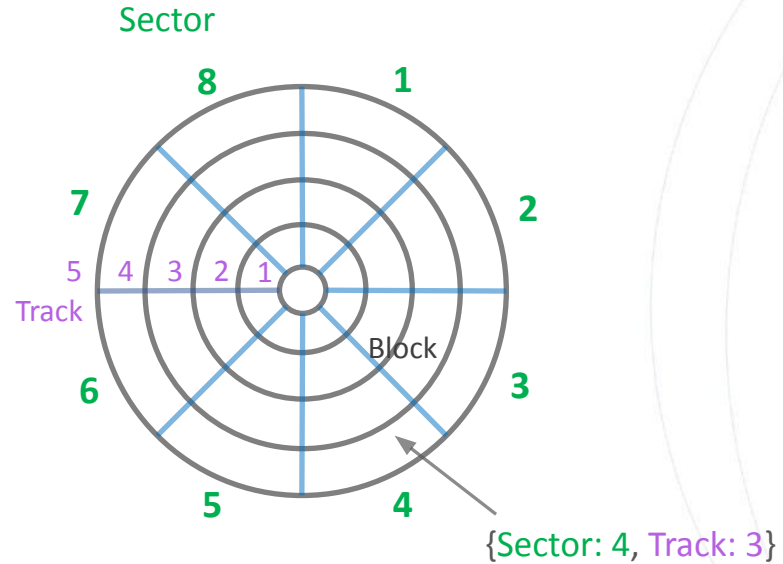
- One or more tables make up a view
- Query follows “SELECT” statement format
- Views are generally read-only
- Views don’t require additional storage





How Does RDBMS Store Data?

Introduction to SQL



1 Block = 512 Bytes





How Does RDBMS Organize Data?

Introduction to SQL

Column Size in **Bytes**

10 _{byte}	20 _{byte}
ProductID	Painter
15	□
35	□
37	□
236	□
751	□

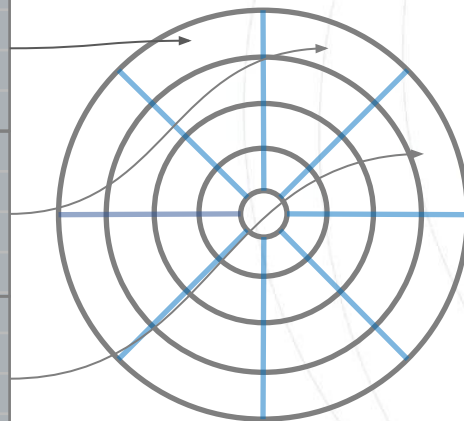
Index

Index is stored on disk as well.

In the above example, it takes
30 Bytes to index one row 128 Bytes

Column Size in **Bytes**

10 _{byte}	50 _{byte}	25 _{byte}	25 _{byte}	18 _{bytes}
ProductID	P_Name	sP_Cat	S_Subcat	P_Price
15	python	book	tech	\$64.5
35	7 habits	book	business	\$34.2
37	macbook	computer	mac	\$3,500.0
236	converse	sport	basketball	\$45.0
751	soccer	shoe	nike	\$213.0
Block 1 (4 rows)				
Block 2 (4 rows)				
Block 3 (4 rows)				



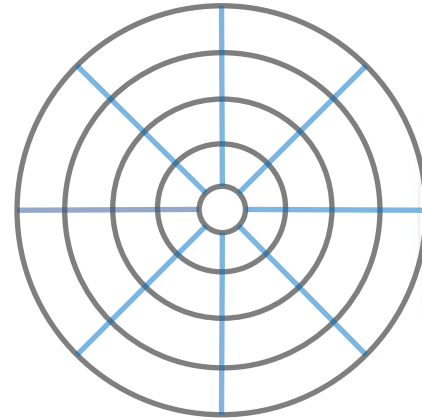


How Does a Hard Disk Drive Store Data?

Introduction to SQL

Column Size in **Bytes**

10 _{bytes}	50 _{bytes}	25 _{bytes}	25 _{bytes}	18 _{bytes}
ProductID	P_Name	P_Cat	P_Subcat	P_Price
15	python	book	tech	\$64.5
35	7 habits	book	business	\$34.2
37	macbook	computer	mac	\$3,500.0
236	converse	sport	basketball	\$45.0
751	soccer	shoe	nike	\$213.0



Row size = 10 + 50 + 25 + 25 + 18 = **128 Bytes**

Block size = **512 Bytes**

Rows per block = **512 / 128 = 4**

Row 1	Row 2	Row 3	Row4
-------	-------	-------	------

Block = 512 Bytes



RDBMS & **Operations**

What is SQL?

Why learn SQL?

What can SQL do?

SQL Data Types

Working with Databases and Tables

Agenda.



RDBMS Database Operations: Create

Introduction to SQL

CRUD

Create | Read | Update | Delete

Index

ProductID	Pointer
15	□
35	□
37	□
236	□

Insert ③

751	□
-----	---

Table

ProductID	P_Name	P_Cat	P_Subcat	P_Price
15	python	book	tech	\$64.5
35	7 habits	book	business	\$34.2
37	macbook	computer	mac	\$3,500.0
236	converse	sport	basketball	\$45.0
751	soccer	shoe	nike	\$213.0

Block 1 (4 rows)

Insert ①

Insert ②





RDBMS Database Operations: Read

Introduction to SQL

CRUD

Create | Read | Update | Delete

Index

ProductID	Pointer
15	□
35	□
37	□
236	□
751	□

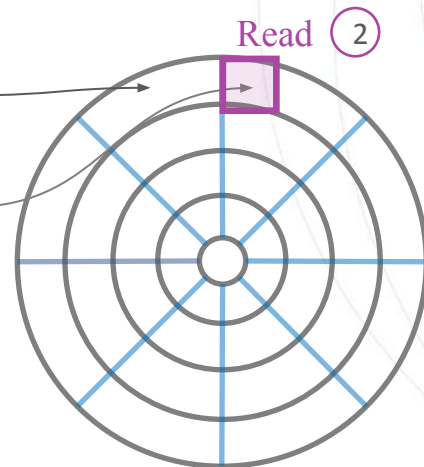
Read ①

Table

ProductID	P_Name	P_Cat	P_Subcat	P_Price
15	python	book	tech	\$64.5
35	7 habits	book	business	\$34.2
37	macbook	computer	mac	\$3,500.0
236	converse	sport	basketball	\$45.0
751	soccer	shoe	nike	\$213.0

Read ③

Block 1 (4 rows)





RDBMS Database Operations: Update

Introduction to SQL

CRUD

Create | Read | Update | Delete

Index

ProductID	Pointer
15	□
35	□
37	□
236	□
751	□

Update ①

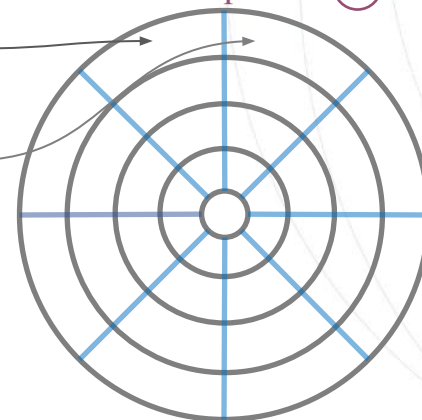
Table

ProductID	P_Name	P_Cat	P_Subcat	P_Price
15	python	book	tech	\$64.5
35	7 habits	book	business	\$34.2
37	macbook	computer	mac	\$3,500.0
236	converse	sport	basketball	\$45.0
751	soccer	shoe	nike	\$258.0

Block 1 (4 rows)

Update ②

Update ③





RDBMS Database Operations: Delete

Introduction to SQL

CRUD

Create | Read | Update | Delete

Index

ProductID	Pointer
15	<input type="checkbox"/>
35	<input type="checkbox"/>
37	<input type="checkbox"/>
236	<input type="checkbox"/>
751	<input type="checkbox"/>

Delete

①

Table

ProductID	P_Name	P_Cat	P_Subcat	P_Price
15	python	book	tech	\$64.5
35	7 habits	book	business	\$34.2
37	macbook	computer	mac	\$3,500.0
236	converse	sport	basketball	\$45.0
751	soccer	shoe	nike	\$213.0

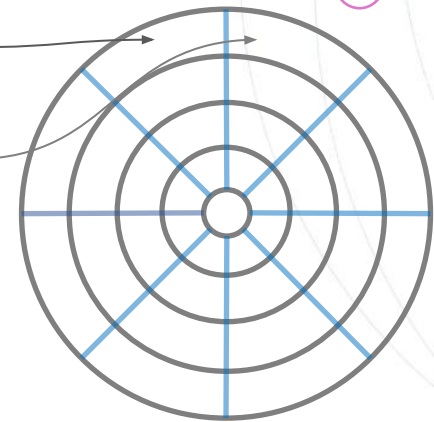
Block 1 (4 rows)

Delete

②

Delete

③



RDBMS & Operations

What is SQL?

Why learn SQL?

What can SQL do?

SQL Data Types

Working with Databases and Tables

Agenda.



What is SQL?

Introduction to SQL



designed by  freepik

Image: [Technology vector created by pikisuperstar - www.freepik.com](https://www.freepik.com)

SQL a.k.a Structured Query Language allows users to:

- Create databases and objects within them
- Store data in databases
- Change and analyze data
- Produce reports, web pages, and/or other visualizations



RDBMS & Operations

What is SQL?

Why learn SQL?

What can SQL do?

SQL Data Types

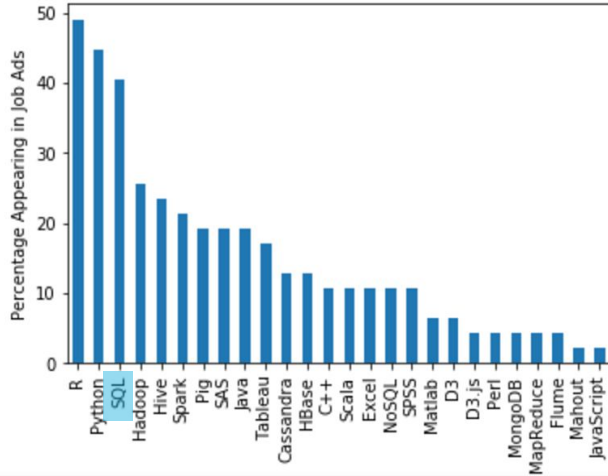
Working with Databases and Tables

Agenda.

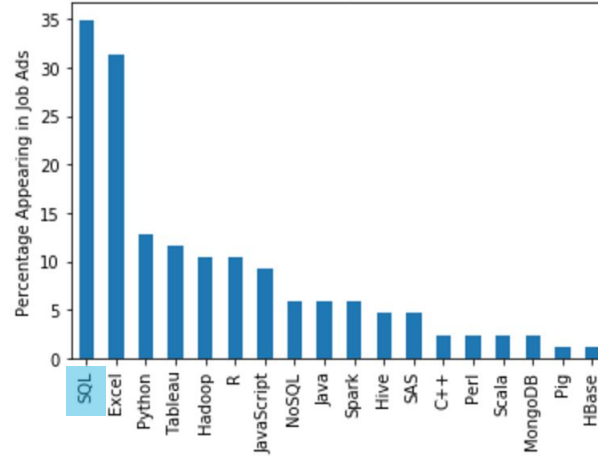


Why learn SQL?: SQL is In-Demand

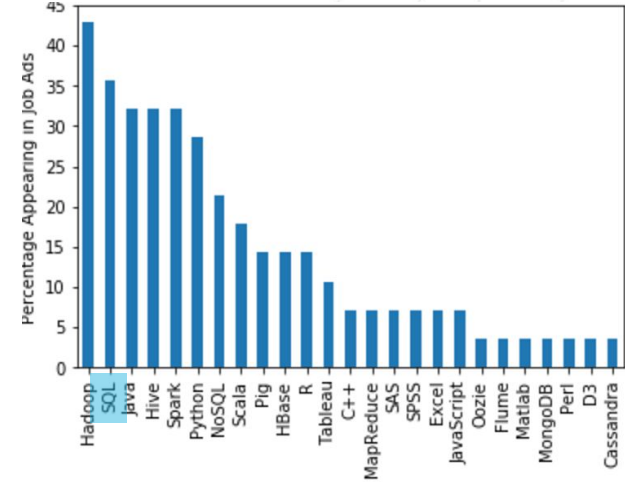
Introduction to SQL



Data
Scientist



Data
Analyst



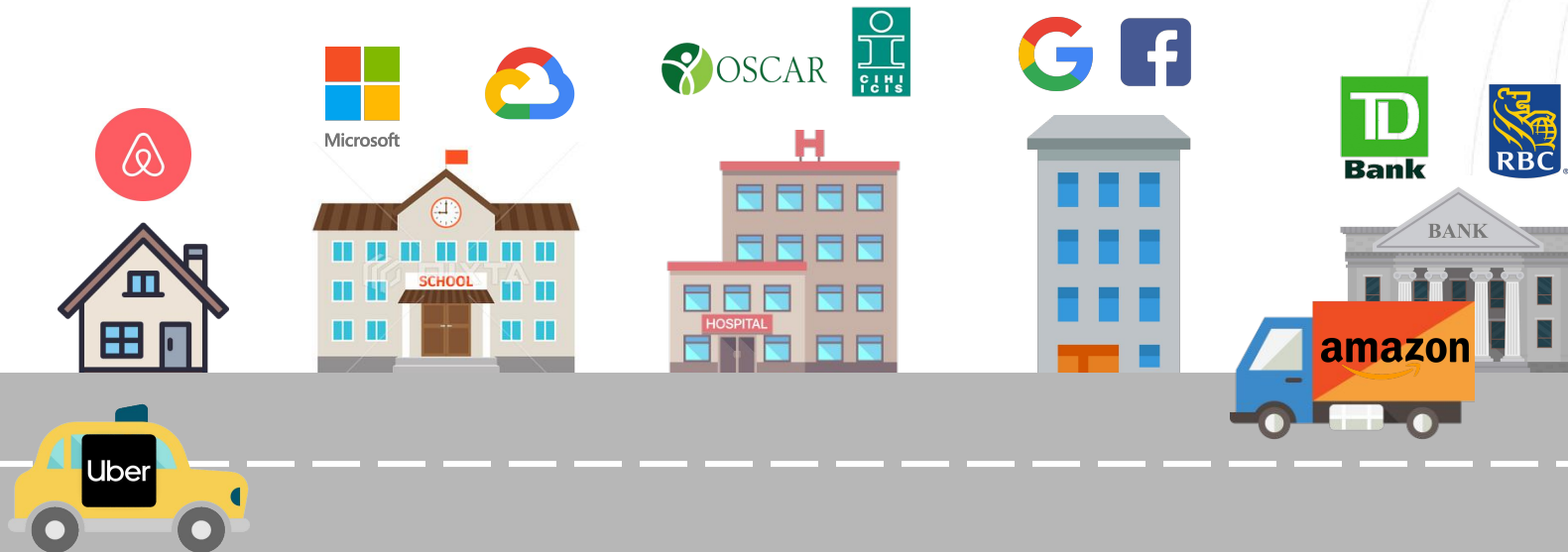
Data
Engineer





Why learn SQL?: SQL is Everywhere

Introduction to SQL



RDBMS & Operations

What is SQL?

Why learn SQL?

What can SQL do?

SQL Data Types

Working with Databases and Tables

Agenda.



Relational Databases: Operational Databases

Introduction to SQL

Enterprise Resource Planning (ERP)

- Manages data about employees and productivity of a company
- Focuses on the processes and reducing cost.

Customer Relationship Management (CRM)

- Manages data about existing and potential customers of a company
- Focuses on the customer
- Helps companies stay connected to customers, streamline processes, and improve sales/profitability





Relational Databases: CRM System and Databases

Introduction to SQL

The screenshot displays the Salesforce CRM interface for an Opportunity record titled "Presidio Technology + Anypoint Connectors". The interface includes a sidebar with navigation icons, a top search bar, and a main content area. The Opportunity details show the Account as "Presidio Technology", the Close Date as "Jun 31, 2015", the Amount as "\$385,000", and the Owner as "Jason Dewar". The Opportunity is in the "Price Quote" stage of a sales process, with a progress bar showing stages: Log Call, Event, Email, Price Quote, Negotiation, and Closed. The "Next Steps" section lists two tasks: "Review proposals for EBC deck with larger team and have Marketing review this" assigned to Lei Chan, and "Presidio Technology - EBC Meeting" assigned to Pete Schaffer. The "Past Activity" section shows a "Mobile conversation on Monday" from yesterday. The right sidebar displays a list of contacts associated with the Opportunity, including Lei Chan (Primary), Gwen Jones, and Pete Schaffer.

OPPORTUNITY
Presidio Technology + Anypoint Connectors

ACCOUNT Presidio Technology **CLOSE DATE** Jun 31, 2015 **AMOUNT** \$385,000 **OWNER** Jason Dewar

Price Quote Negotiation Closed

ACTIVITY **COLLABORATE** **DETAILS**

Task Log Call Event Email

Subject Add Task

Next Steps

- ☐ Review proposals for EBC deck with larger team and have Marketing review this Today
Name Lei Chan Assigned to Betty Mason Priority Medium
- ☐ Presidio Technology - EBC Meeting Today 2:00 PM
Lets get together to review the theater's layout and facilities. We'll also discuss potential...
Location 300 Pike St, San Francisco CA Name Pete Schaffer

Past Activity

- ☐ Mobile conversation on Monday Yesterday

Contacts (3)

- Lei Chan** PRIMARY
Role Executive Buyer
Title President, Operations
- Gwen Jones**
Role Executive Buyer
Title President, Operations
- Pete Schaffer**
Role Marketing Contact
Title Director, Marketing

Notes (3)

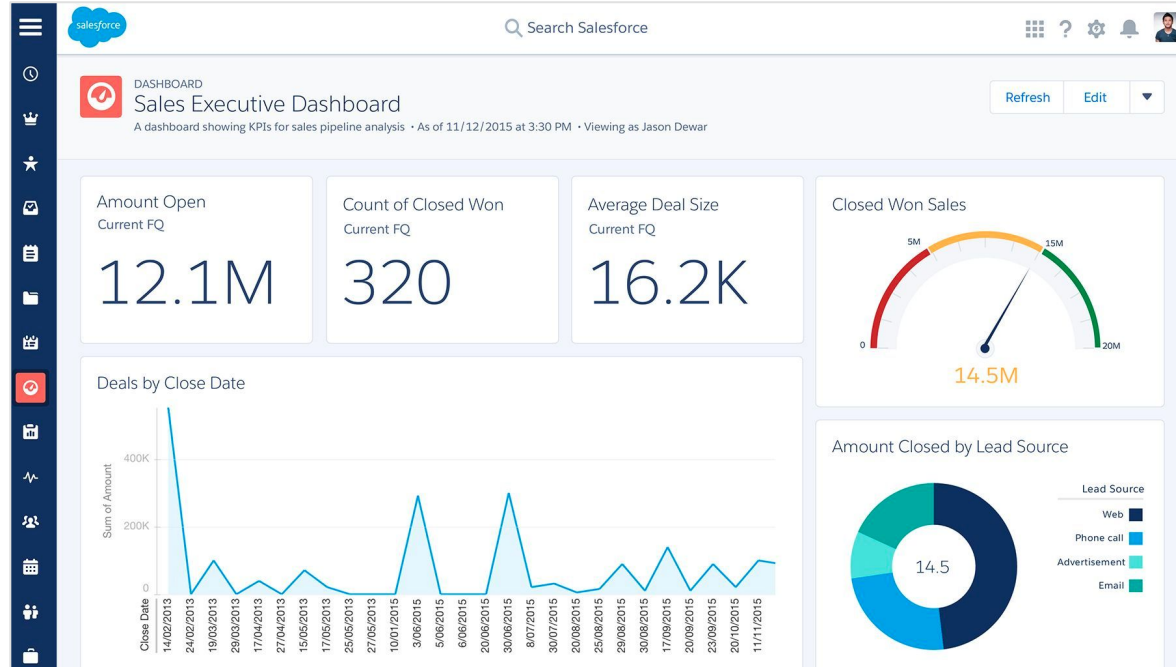
- Demo prep with CloudHub...**
5/27/15 by Jason Dewar





Relational Databases: BI and Reporting

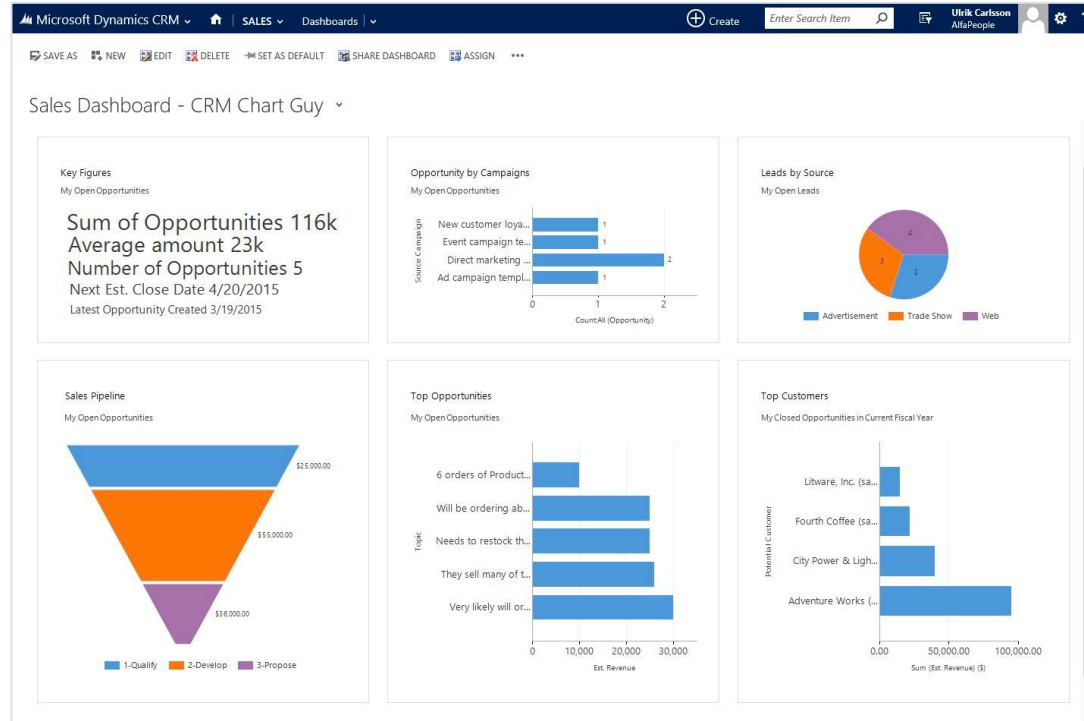
Introduction to SQL





Relational Databases: BI and Reporting (Cont'd)

Introduction to SQL



RDBMS & Operations

What is SQL?

Why learn SQL?

What can SQL do?

SQL Data Types

Working with Databases and Tables

Agenda.



SQL Data Types

Introduction to Database Queries



Image: [Data vector created by stories - www.freepik.com](https://www.freepik.com)

SQL supports many **data types** which can be grouped into three categories:

- Numeric
- String
- Date and Time



Data Types: Numeric

Introduction to Database Queries

Data Type Syntax	Maximum Size	Description
INT(<i>m</i>)	2, 147, 483, 647	<ul style="list-style-type: none">Integer types
BIGINT(<i>m</i>)	9, 223, 372, 036, 854, 775, 807	<ul style="list-style-type: none">Integer: -9, 223, 372, 036, 854, 775, 808 to 9, 223, 372, 036, 854, 775, 807
DECIMAL(<i>m</i> , <i>d</i>)	<i>m</i> → precision <i>d</i> → scale	<ul style="list-style-type: none">Any values with <i>m</i> digits and <i>d</i> decimals
FLOAT(<i>m</i> , <i>d</i>)	<i>m</i> → precision <i>d</i> → scale	<ul style="list-style-type: none">The FLOAT types represent approximate numeric data values
DOUBLE(<i>m</i> , <i>d</i>)	<i>m</i> → precision <i>d</i> → scale	<ul style="list-style-type: none">The DOUBLE types represent approximate numeric data values





Data Types: String

Introduction to Database Queries

Data Type Syntax	Maximum Size	Description
CHAR(size)	255	<ul style="list-style-type: none">• Fixed length
VARCHAR(size)	255	<ul style="list-style-type: none">• Variable length• VARCHAR is store inline with the table
TEXT(size)	65, 535	<ul style="list-style-type: none">• Used for large text blobs• TEXT is stored off the table• Table just has a pointer to the location of the actual storage





Data Types: Date and Time

Introduction to Database Queries

Data Type Syntax	Format	Description
DATE	YYYY-MM-DD	<ul style="list-style-type: none">Used when you need values that contain only the date information
DATETIME	YYYY-MM-DD HH:MM:SS	<ul style="list-style-type: none">Used when you need values that contain both date and time information
TIMESTAMP	YYYY-MM-DD HH:MM:SS	<ul style="list-style-type: none">TIMESTAMP values are converted from the current timezone to UTC for storage, and converted back from UTC to the current timezone for retrieval



RDBMS & Operations

What is SQL?

Why learn SQL?

What can SQL do?

SQL Data Types

Working with Databases and Tables

Agenda.



Dashboard

Introduction to Database Queries

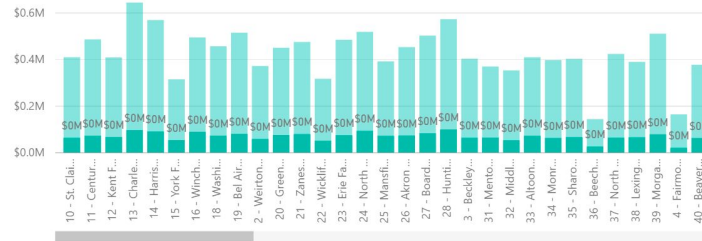
Example of a dashboard built for sales reporting

District Monthly Sales

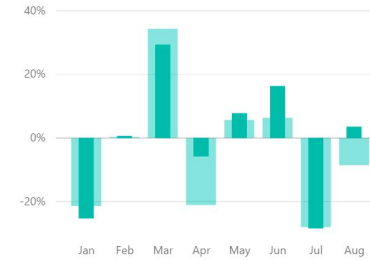
District Manager

- ☐ Allan Guinot
- ☐ Andrew Ma
- ☐ Annelie Zubar
- ☐ Brad Sutton
- ☐ Carlos Grilo
- ☐ Chris Gray
- ☐ Chris McGurk
- ☐ Tina Lassila
- ☐ Valery Ushakov

This Year Sales by StoreNumberName

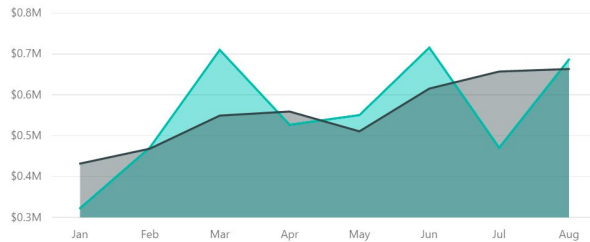


Total Sales Variance % by FiscalMonth

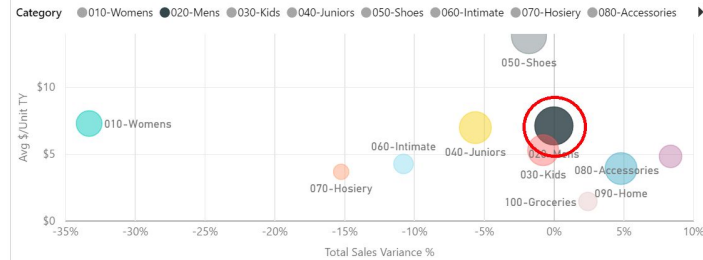


This Year Sales and Last Year Sales by FiscalMonth

● This Year Sales ● Last Year Sales



Total Sales Variance %, Avg \$/Unit TY and This Year Sales by Category



obviEnce llc ©



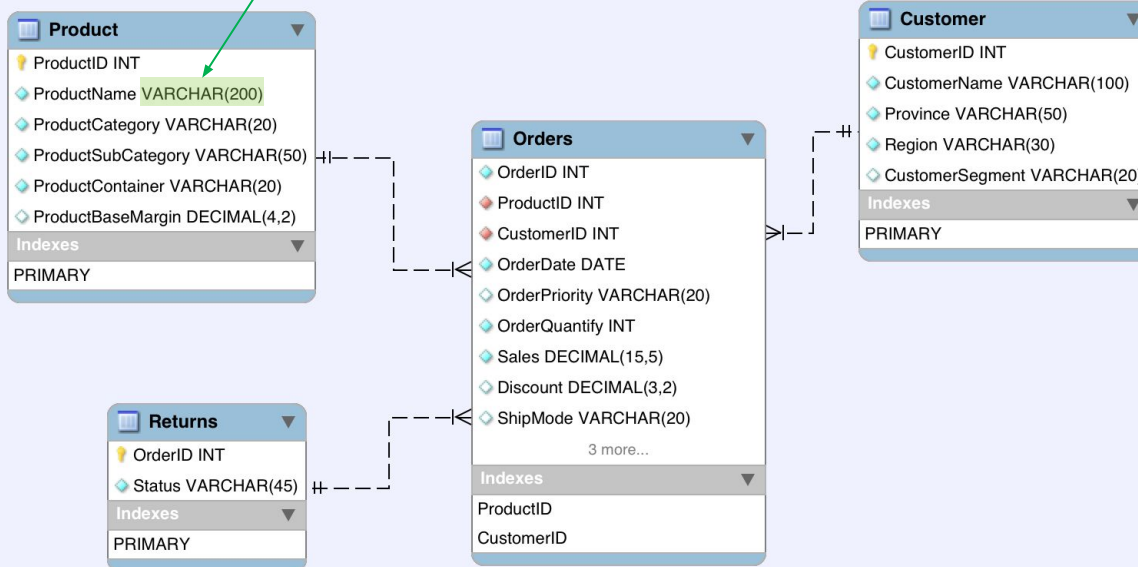


Superstore Database: Entity Relationship Model

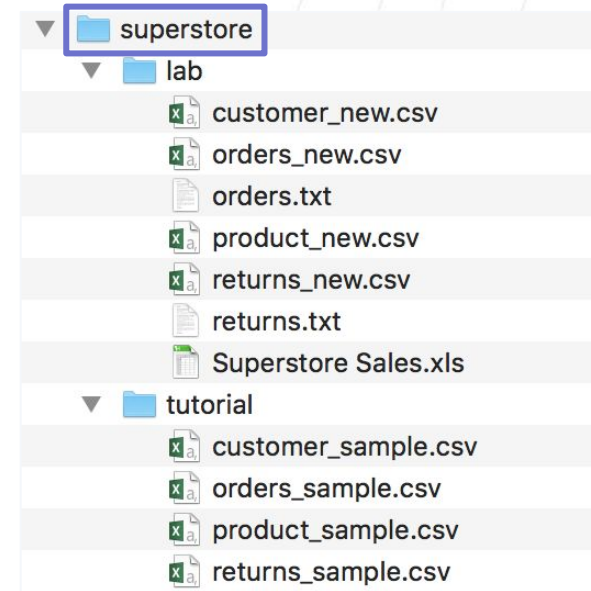
Introduction to Database Queries

Superstore Database

Notice the data types are shown!



Here are the **datasets** we'll use for the lecture and lab:





Superstore Database: Entity Relationship Model

Introduction to Database Queries

**Product
Dataset**

ProductID	ProductName	ProductCat...	ProductSubCateg...	ProductC...	Pr...
5889	White GlueTop Scratch Pads	Office Supplies	Paper	Wrap Bag	0.39
15497	Fellowes 8 Outlet Superior Workstation Surge Protector	Office Supplies	Appliances	Small Box	0.56
16735	"Belkin 325VA UPS Surge Protector, 6"	Office Supplies	Appliances	Small Box	0.60
23721	Gyration Ultra Cordless Optical Suite	Technology	Computer Peripherals	Small Box	0.46
27473	Epson FX-980 Dot Matrix Printer	Technology	Office Machines	Jumbo Drum	0.59

**Orders
Dataset**

OrderID	ProductID	OrderDate	OrderPrior...	OrderQuantity	Sales	Discount	ShipMode	Profit	UnitPrice	ShippingCost
8710	657768	2009-01-04	Critical	42	151.35000	0.07	Express Air	8.33	3.71	1.93
16326	657768	2010-05-10	High	39	147.46000	0.06	Regular Air	14.13	3.71	1.93
59815	657768	2010-12-15	Not Specified	14	51.56000	0.09	Regular Air	-1.06	3.71	1.93
58470	657768	2011-06-08	High	13	49.08000	0.06	Regular Air	-0.31	3.71	1.93
50657	657768	2011-12-18	High	16	60.02000	0.10	Express Air	0.94	3.71	1.93

**Customer
Dataset**

CustomerID	CustomerName	Province	Region	CustomerSegme...
40732966	Tamara Dahlen	Ontario	Ontario	Corporate
68464052	Bill Donatelli	Ontario	Ontario	Corporate
63834266	Christy Brittain	British Columbia	West	Consumer
38512011	Barry Blumstein	British Columbia	West	Small Business
82335880	Aleksandra Gannaway	New Brunswick	Atlantic	Corporate

**Returns
Dataset**

OrderID	Status
65	Returned
69	Returned
134	Returned
135	Returned
230	Returned

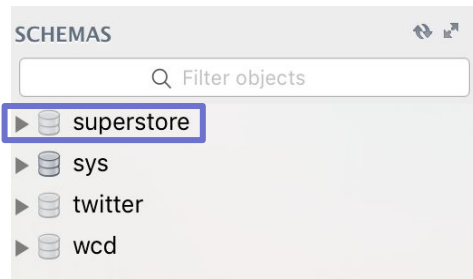




Create a Database

Introduction to Database Queries

```
-- create a database  
create database superstore;
```



```
-- list databases  
show databases;
```

Database
information_schema
mysql
performance_schema
superstore
sys
twitter
wcd

Syntax

SHOW DATABASES;

CREATE DATABASE **table_name**;

NOTE:

- The first step is always to create a database
- In a real work environment, the database admin/developer should have done that for you
- You rarely will need to create your own database as an analyst

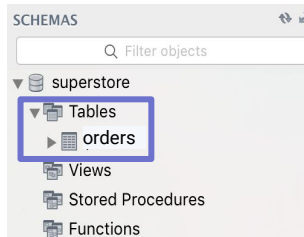




Create a Table

Introduction to Database Queries

```
/* create and load orders table */
drop table if exists superstore.orders;
create table superstore.orders (
  OrderID int,
  ProductID int,
  OrderDate date,
  OrderPriority varchar(20),
  OrderQuantity int,
  Sales decimal(15,5),
  Discount decimal(3,2),
  ShipMode varchar(20),
  Profit decimal(15,2),
  UnitPrice decimal(15,2),
  ShippingCost decimal(15,2)
);
```



Syntax

DROP TABLE [IF EXISTS] table_name;

CREATE TABLE table_name [AS] (
 column1 datatype [options],
 ...,
 columnN datatype [options]
);

NOTE:

- In real work environments, you may need to create tables occasionally
- However, most of the production databases and tables have been created by the database experts and data has been ingested into the tables
- As a data analyst/scientist, you need to focus on writing queries to extract information



Load Data into a Table

Introduction to Database Queries

```
-- clear the data in this table while keeping the table in database
-- to avoid repeated insertion of same data
truncate superstore.product;
```

```
-- load data into the product table (please change the file path accordingly)
load data local infile 'data/superstore/tutorial/product_sample.csv'
into table superstore.product character set 'latin1'
fields terminated by ','
lines terminated by '\n'
;
```

Remember to change
the path accordingly

Syntax

TRUNCATE `table_name`;

LOAD DATA LOCAL INFILE `'/PATH/file.txt'`
INTO TABLE `table_name`
FIELDS TERMINATED BY `','`;
LINES TERMINATED BY `'\n'`;

ProductID	ProductName	ProductCateg...	ProductSubCategory	ProductContain...	ProductBaseMar...
5889	White GlueTop Scratch Pads	Office Supplies	Paper	Wrap Bag	0.39
23721	Gyration Ultra Cordless Optical Suite	Technology	Computer Peripherals	Small Box	0.46
115501	Newell 308	Office Supplies	Pens & Art Supplies	Wrap Bag	0.60
213268	"Advantus Employee of the Month Certificate Frame, 11..."	Furniture	Office Furnishings	Small Pack	0.44
284312	Belkin 105-Key Black Keyboard	Technology	Computer Peripherals	Small Box	0.68





Describe and List Tables

Introduction to Database Queries

```
-- describe a table  
describe superstore.product;
```

Field	Type	Null	Key	Default	Extra
ProductID	int(11)	NO	PRI	NULL	
ProductName	varchar(200)	YES		NULL	
ProductCategory	varchar(20)	YES		NULL	
ProductSubCategory	varchar(50)	YES		NULL	
ProductContainer	varchar(20)	YES		NULL	
ProductBaseMargin	decimal(4,2)	YES		NULL	

Syntax

SHOW TABLES;

DESCRIBE db.table_name;

NOTE:

- Knowing the data types of columns in your table is always helpful, especially when you apply SQL functions
- Describing table schema is a good habit when you work with new data sources

