SQL 3

The Data Bootcamp | February 16, 2018

© 2017 Trilogy Education Services

Subqueries

- A query within a query
- A subquery can almost always be replaced by a join
  - Joins may be faster.
  - SQL Server, for example, will often optimize these and rewrite them as joins under the hood. Other database implementations vary.
- Which should we do? Focus on designing a readable query first.

```
select *
from inventory
where film_id IN
(
  select film_id
  from film
  where title = 'Early Home'
);
```

© 2017 Trilogy Education Services

Subqueries

SELECT LINE

- Access information from tables in SELECT LINE
  - Print out orders and compare to average product price
- Query must return a single value

Food Truck Example:

```
sql_query = """
select avg(Price)
from Products
"""
products = pd.read_sql_query(sql_query, engine)
products.head()
```

	avg(Price)
0	6.25

```
sql_query = """
select Name, Price - (select avg(Price) from Products)
from Products;
"""
products = pd.read_sql_query(sql_query, engine)
products.head()
```

	Name	Price - (select avg(Price) from Products)
0	carne asada	5.75
1	diet coke	-3.75
2	pollo burrito	1.75
3	sprite	-3.75

© 2017 Trilogy Education Services

Subqueries

Where Line

- Useful for comparing values from other tables
  - Find customers who have placed orders
  - Find orders containing a particular product category

We did a similar example to this one Thursday with the Sakila database using joins.

```
sql_query = """
select *
from inventory
where film_id IN
(
  select film_id
  from film
  where title = 'Early Home'
);
"""
joined = pd.read_sql_query(sql_query, engine)
joined.head()
```

	inventory_id	film_id	store_id	last_update
0	1206	268	2	2006-02-15 05:09:17
1	1207	268	2	2006-02-15 05:09:17

© 2017 Trilogy Education Services

OLTP, OLAP, ETL, ELT, Oh my…

Sr Data Engineer

American National Bank ★★★★☆ 58 reviews - Omaha, NE

Apply Now

Experience, Education & Training Requirements

Experience Needed:

- Requires 7+ years of related experience in Enterprise Data Engineering/Management
- Experience developing and maintaining an enterprise wide data governance framework
- Experience with big data technologies
- Experience using data management software and BI Reporting tools like Cognos
- Experience writing and maintaining SQL
- Experience with master data management, data modeling and database design, OLTP and OLAP, data warehousing ETL/ELT workflows
- Financial services/banking industry experience is desired

Data Warehouse Architect

ClickBank ★★★★½ 11 reviews - Broomfield, CO

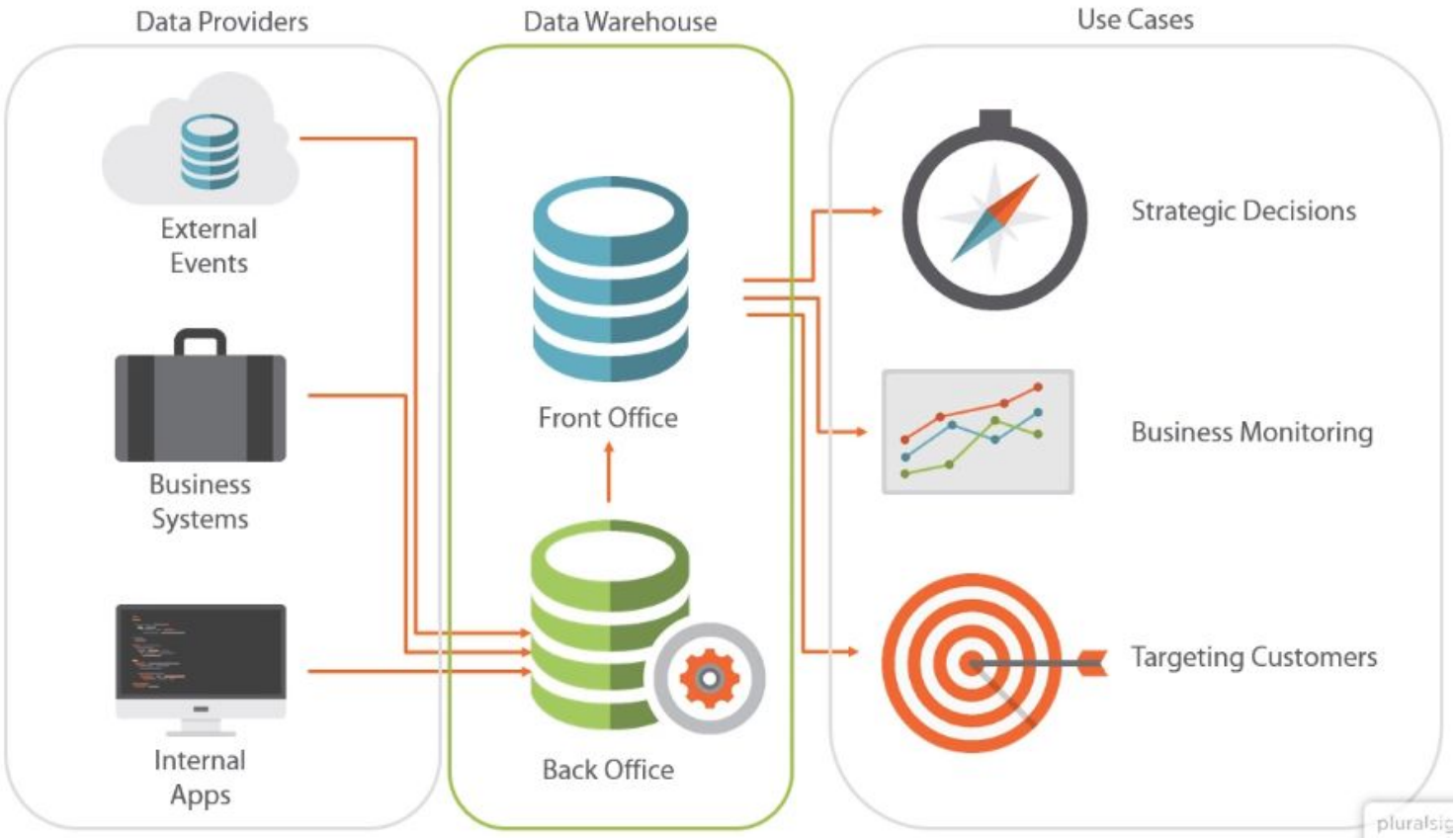
Apply Now

Key Responsibilities include, but are not limited to:

- Works with the business intelligence and the ETL development team to extract data from the OLTP system and present it in a format that supports data mining, analysis and reporting.
- Ability to maintain existing and develop new ETL jobs which support extracting source systems, transforming data based on business rules and requirements and loading of data into warehouse structures.
- Troubleshoot data issues and implement fixes to ensure data accuracy.
- Good communication skills with both technical and non-technical staff members.
- Able to translate business requirements and design ETL jobs to support them.
- Ability to re-design data models to support data mining and enhanced analytics efforts.

© 2017 Trilogy Education Services

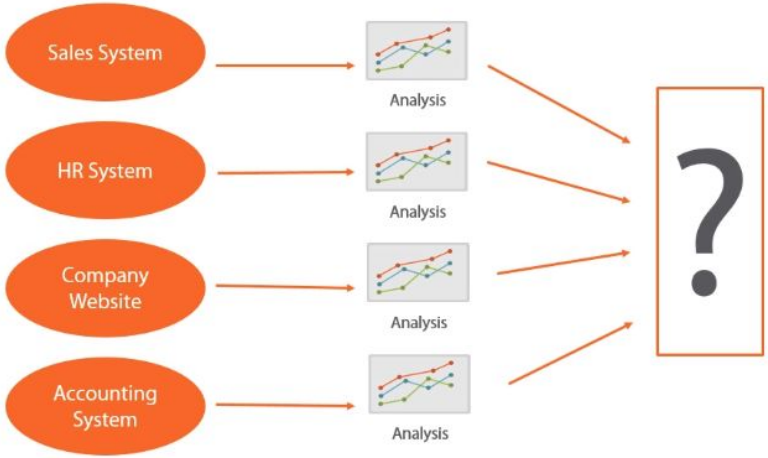
The Big Picture

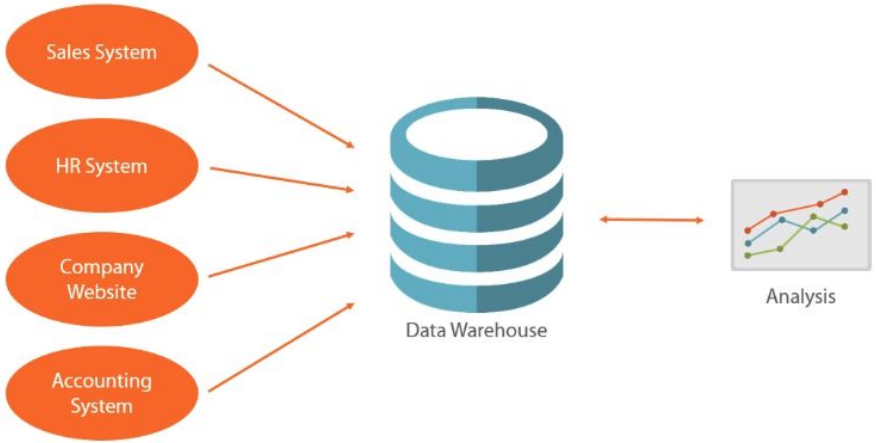


© 2017 Trilogy Education Services

Data Warehousing

Data warehousing opens up opportunities to gain value from your raw data through the process of analysis.



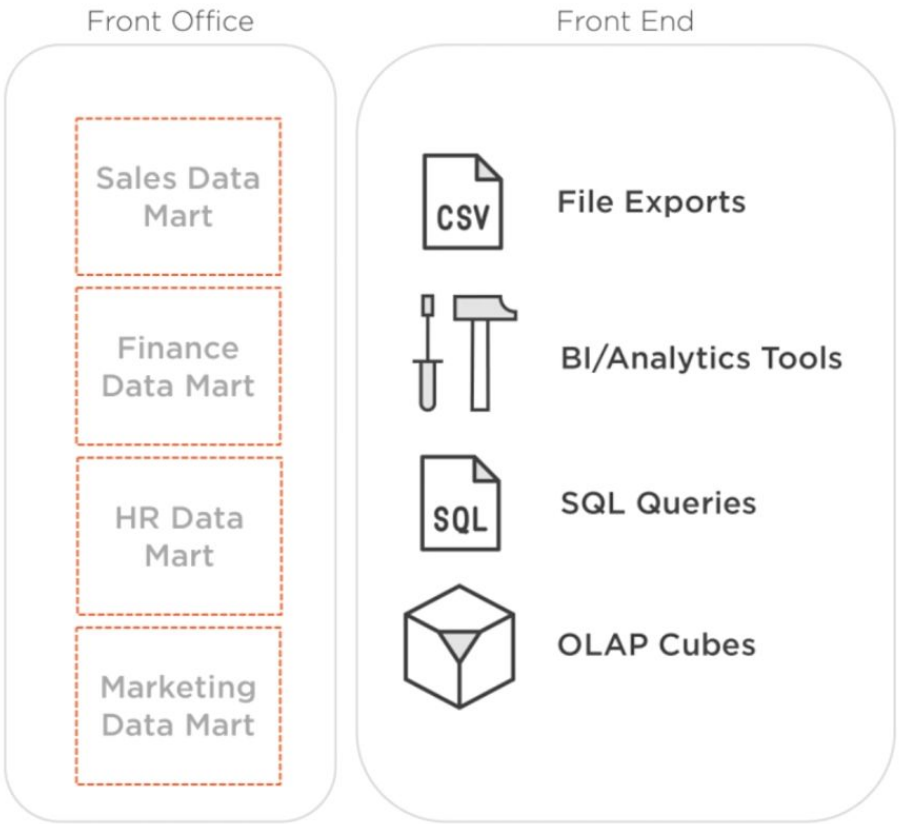


© 2017 Trilogy Education Services

The Big Picture

© 2017 Trilogy Education Services

The Big Picture



- Strategic Analysis
- Data Science
- Operational Efficiency
- Research & Development
- Sales Optimization
- Etc...

© 2017 Trilogy Education Services

The Big Picture

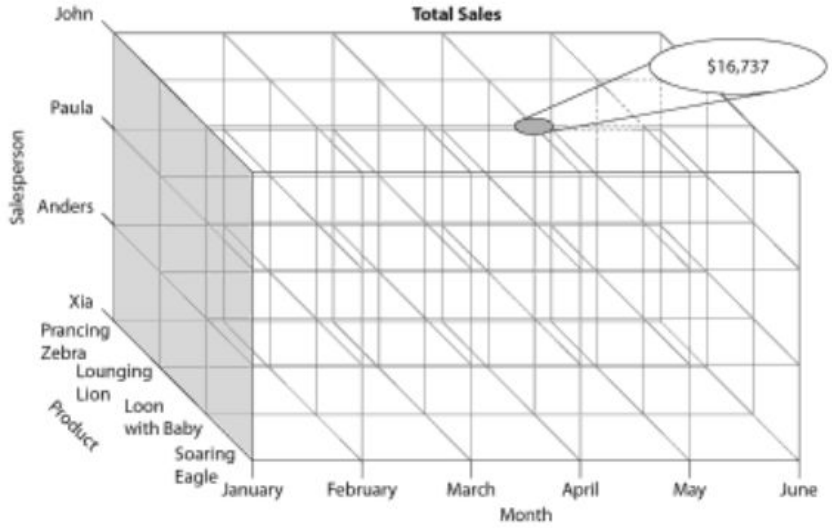
OLAP Cubes

OLAP We will go over in the next slide.

A **Cube** is a structure that contains a value for one or more measures for each unique combination of the members of all its dimensions.

OLAP Cubes are fading as Tableau / PowerBI / Pandas are moving into this space.

That said, they are still out there.. everywhere...



More Information. **Delivering Business Intelligence with Microsoft SQL Server 2016**
https://www.amazon.com/exec/obidos/ASIN/1259641481/alvinashcraft-20

© 2017 Trilogy Education Services

OLTP Versus OLAP

OLAP
Online
Analytical
Processing

OLTP
Online
Transaction
Processing

© 2017 Trilogy Education Services

OLTP and OLAP

OLTP - Transactions	OLAP - Analytics
Daily operations. Always on	Periodically updated
All kinds of users	Managers, Data Analysts
Directly supports business processes. Orders, shipping, customer info	For long-term planning, not tracking orders.
Current Status. Real Time Data.	Maintains History
Entity-Relationship Model	Dimensional Model
Normalized to 3NF	Denormalized dimensions

© 2017 Trilogy Education Services

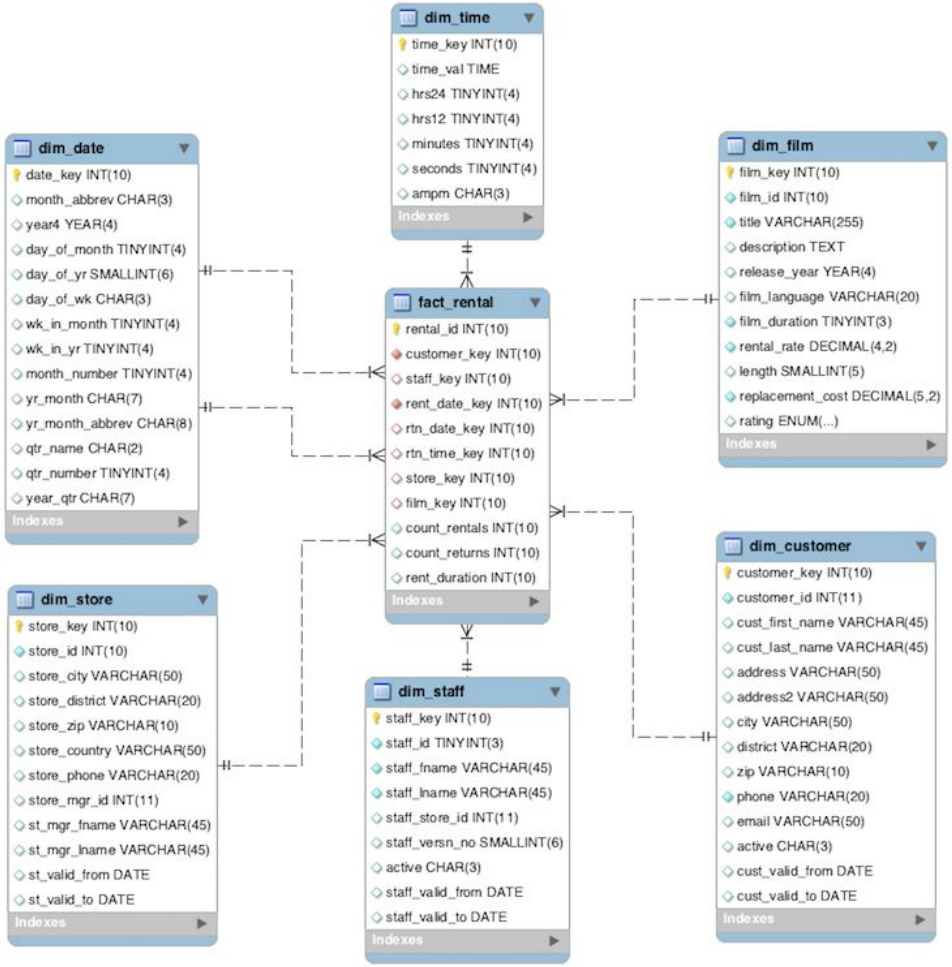
Why have two systems?

Normalized (OLTP)	Denormalized (OLAP)
Machine oriented. Quick writes.	Human Oriented
Complex Joins	Simple Joins
Eliminates redundancy. Product name not repeated so you don't mess it up.	Use redundancy to simplify

© 2017 Trilogy Education Services

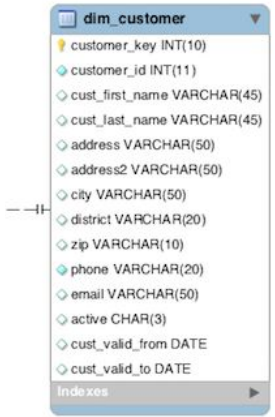
Star Schema

A data mart design pattern consisting of fact and dimension tables resembling a star like shape.

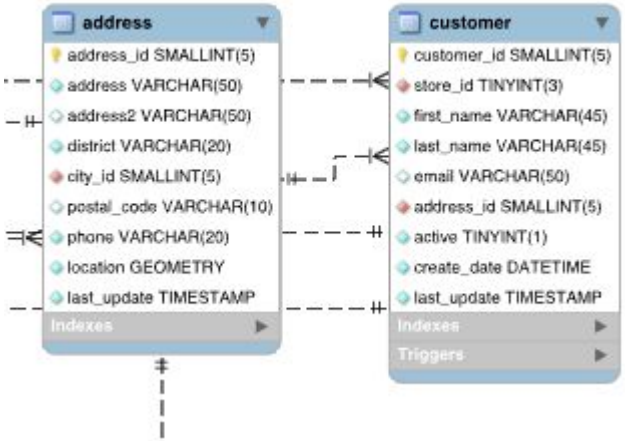

dim_time
time_key INT(10)
time_val TIME
hrs24 TINYINT(4)
hrs12 TINYINT(4)
minutes TINYINT(4)
seconds TINYINT(4)
ampm CHAR(3)
Indexes
dim_date
date_key INT(10)
month_abbrev CHAR(3)
year4 YEAR(4)
day_of_month TINYINT(4)
day_of_yr SMALLINT(6)
day_of_wk CHAR(3)
wk_in_month TINYINT(4)
wk_in_yr TINYINT(4)
month_number TINYINT(4)
yr_month CHAR(7)
yr_month_abbrev CHAR(8)
qtr_name CHAR(2)
qtr_number TINYINT(4)
year_qtr CHAR(7)
Indexes
fact_rental
rental_id INT(10)
customer_key INT(10)
staff_key INT(10)
rent_date_key INT(10)
rtn_date_key INT(10)
rtn_time_key INT(10)
store_key INT(10)
film_key INT(10)
count_rentals INT(10)
count_returns INT(10)
rent_duration INT(10)
Indexes
dim_film
film_key INT(10)
film_id INT(10)
title VARCHAR(255)
description TEXT
release_year YEAR(4)
film_language VARCHAR(20)
film_duration TINYINT(3)
rental_rate DECIMAL(4,2)
length SMALLINT(5)
replacement_cost DECIMAL(5,2)
rating ENUM(...)
Indexes
dim_store
store_key INT(10)
store_id INT(10)
store_city VARCHAR(50)
store_district VARCHAR(20)
store_zip VARCHAR(10)
store_country VARCHAR(50)
store_phone VARCHAR(20)
store_mgr_id INT(11)
st_mgr_fname VARCHAR(45)
st_mgr_lname VARCHAR(45)
st_valid_from DATE
st_valid_to DATE
Indexes
dim_staff
staff_key INT(10)
staff_id TINYINT(3)
staff_fname VARCHAR(45)
staff_lname VARCHAR(45)
staff_store_id INT(11)
staff_versn_no SMALLINT(6)
active CHAR(3)
staff_valid_from DATE
staff_valid_to DATE
Indexes
dim_customer
customer_key INT(10)
customer_id INT(11)
cust_first_name VARCHAR(45)
cust_last_name VARCHAR(45)
address VARCHAR(50)
address2 VARCHAR(50)
city VARCHAR(50)
district VARCHAR(20)
zip VARCHAR(10)
phone VARCHAR(20)
email VARCHAR(50)
active CHAR(3)
cust_valid_from DATE
cust_valid_to DATE
Indexes

Star Schema / Normalized Schema Comparison

Star Schema - OLAP



3NF - OLTP



ETL Software

In a startup, you might do this with Python and SQLAlchemy / MongoDB.

A medium to large company at scale will almost undoubtedly use one of these ETL Software Packages

© 2017 Trilogy Education Services

ETL Stages

Stage

Cleanse

Deliver

Extract: Bring everything in from everywhere.

Transform: Cleanup names, move things around. Most of the work.

Load: Deliver to front office database

© 2017 Trilogy Education Services

To Learn More

Star Schema Foundations:
https://app.pluralsight.com/library/courses/star-schema-foundations/table-of-contents

The Kimball Group Reader
https://www.amazon.com/gp/product/1119216311

© 2017 Trilogy Education Services