DATA ANALYTIC SERVICES - TRAINING

FUNDAMENTALS OF DATAFLOW AND SQL

Fundamentals of Data Flow

SELECT
FROM
WHERE
ORDERBY
LIMIT

Matthew Morris

Git: Morrisdata
Msmorris@costco.com

PREVIOUSLY IN DATA ANALYTICS

Presentations
Installing ORACLE SQL Developer
Basic Navigation
Simple Select statement

Matthew Morris
Git: Morrisdata
Msmorris@costco.com

Prompt for Project

You have Identified a problem with an actionable solution, gathered requirements, created some metadata.

You are now ready to; Obtain your data, Understand your data and Prepare it using SQL.

Prompt for Project

Initial Draft of Findings:

Describe the primary dimensions represented in the data.

Describe any initial findings related to your research questions.

Describe what you still need to do for your presentation in order to make a sound recommendation

SQL Queries + Questions:

List the appropriate business questions you were trying to answer
Use correct, efficient syntax for the functions and operators needed
Include corresponding output listed in the .sql file
Transfer simple Queries to Webi
Create an alteryx workflow for complex queries

Prompt for Project

10-minute Presentation:

Clarify the problem statement to be solved

Summarize the data in your presentation

Summarize your exploration of your assumptions and exploration of data

Using filters, aggregations, and relationships in data provide insights

Compare the two features and create a coherent recommendation.

Identify follow-up problems and questions for future analysis

GATHERING REQUIREMENTS

GATHERING REQUIREMENTS

GATHERING REQUIREMENTS

First Name	Last Name	Address	City	Age
Mickey	Mouse	123 Fantasy Way	Anaheim	73
Bat	Man	321 Cavern Ave	Gotham	54
Wonder	Woman	987 Truth Way	Paradise	39
Donald	Duck	555 Quack Street	Mallard	65
Bugs	Bunny	567 Carrot Street	Rascal	58
Wiley	Coyote	999 Acme Way	Canyon	61
Cat	Woman	234 Purrfect Street	Hairball	32
Tweety	Bird	543	Itottaw	28

GATHERING REQUIREMENTS

GATHERING REQUIREMENTS

WHAT FIELDS DO YOU WANT?
(dimensions, measures, calculations)

WHERE CAN YOU CURRENTLY FIND THIS INFORMATION?
(spreadsheets,SAP, Tableau, AS400)

HOW WOULD YOU LIKE THIS INFORMATION FILTERED?

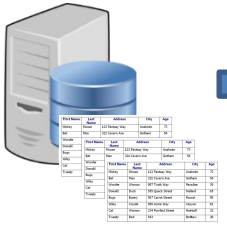
HOW WOULD YOU LIKE IT GROUPED?

GATHERING REQUIREMENTS

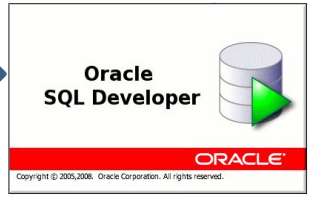
DATA ECO SYSTEM

Basic Ecosystem

Real time



Near Real time



Google Sheets

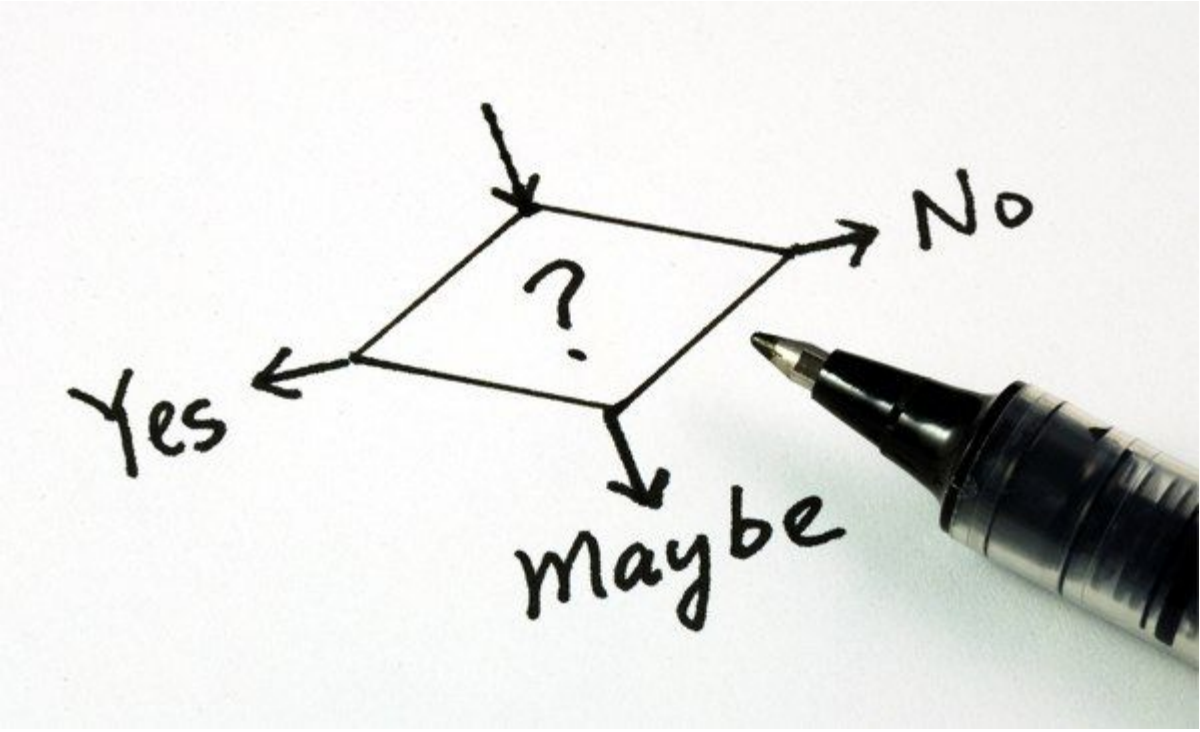
Basic Ecosystem

DATA INACCURATE

Timing data not in sync

Inaccurate programming

Report needs review



Basic Ecosystem

Timing/data not in sync.

There are numerous programs running to validate the complex programs running that are constantly moving and loading data. We have a team of people always reviewing these tables and programs.

Solution Open a Service Desk ticket or try again. Most issues are resolved quickly. Bigger issues sustain usually past 10AM PST

INACCURATE PROGRAMMING

This scenario entails SAP or iSeries400 that has a report with inaccurate logic for a calculation. You may see this with rolled up summary like YTD from detailed information. This is rare if at all the case. Be ready to show extensive proof that you have found a bug.

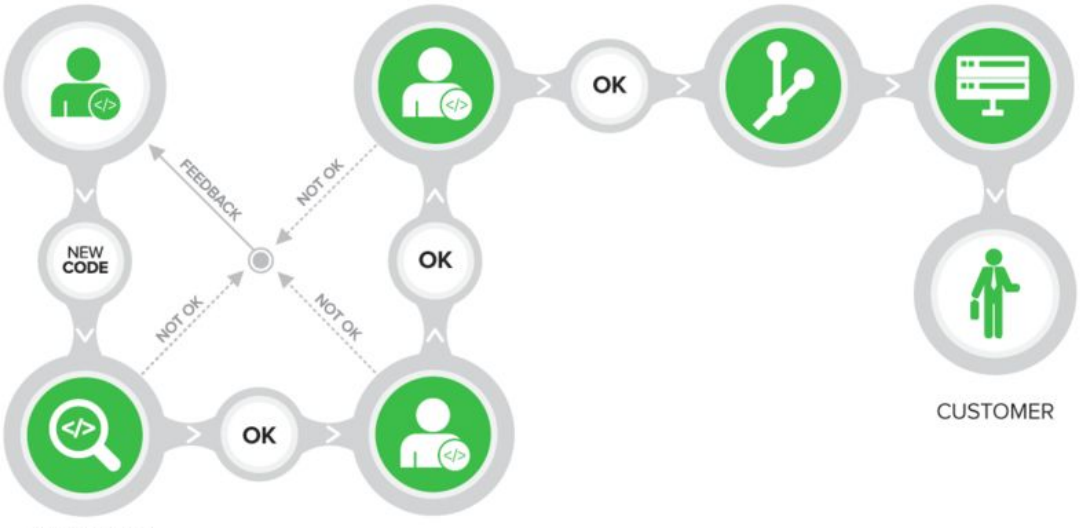
Basic Ecosystem

REPORT NEEDS REVIEW

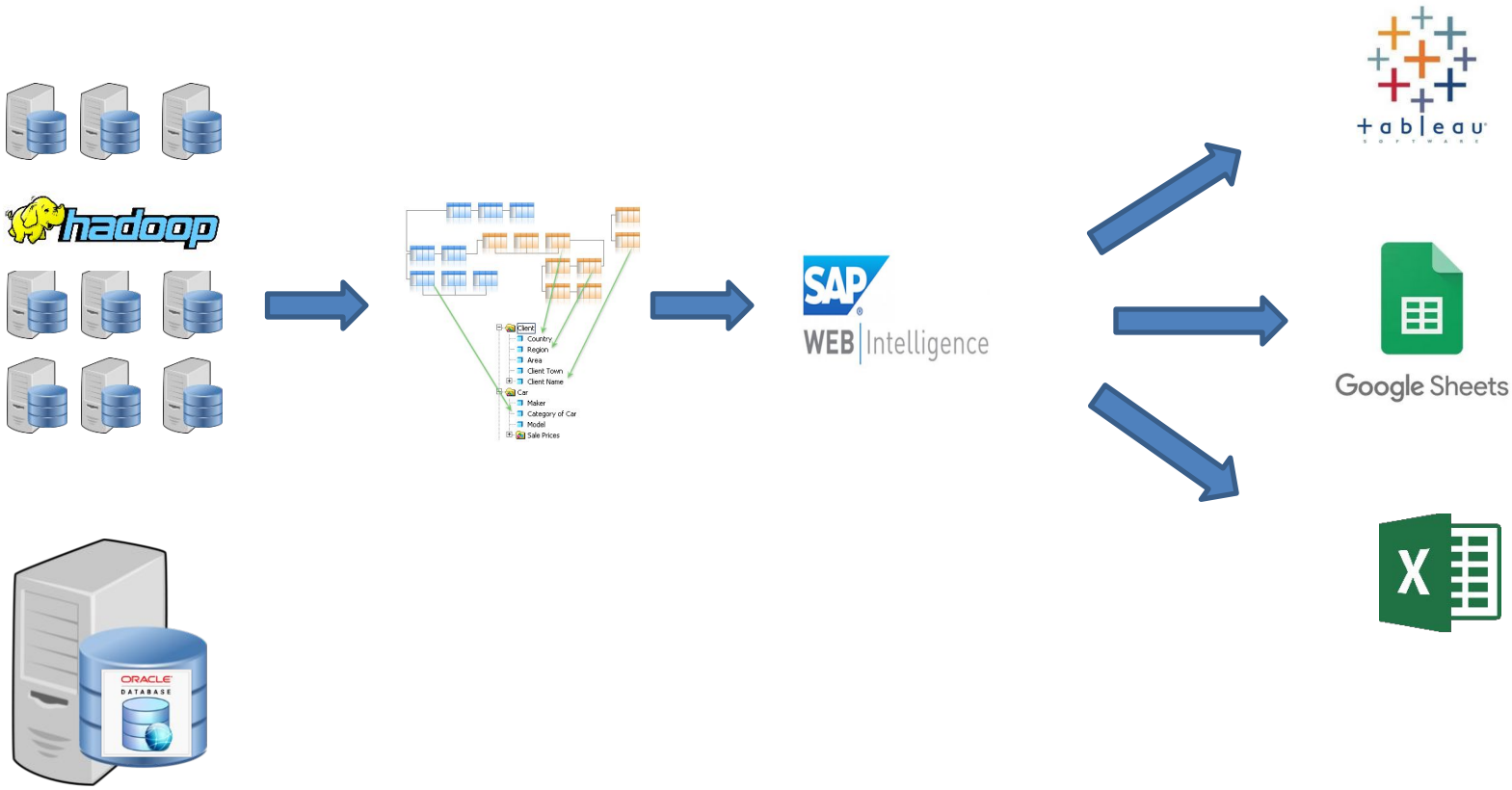
Over 90% of the time this is the culprit. Are you using the tables from the source you are comparing to? Are the fields correct? Are the calcs correct? Are the filters correct? Are the functions correct? Are the Joins correct? Is your 3 Value logic Correct?

Basic Ecosystem

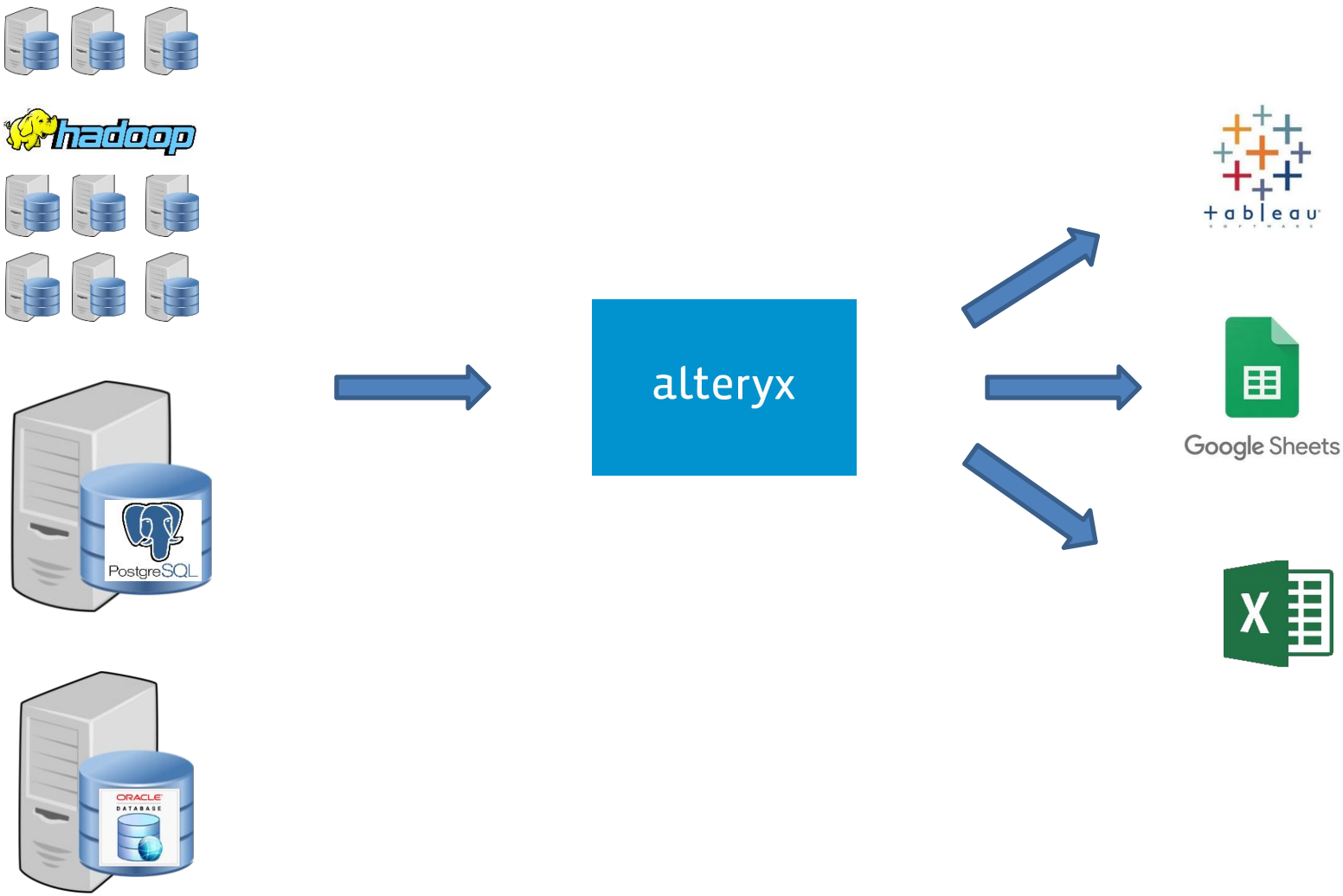
CODE REVIEW



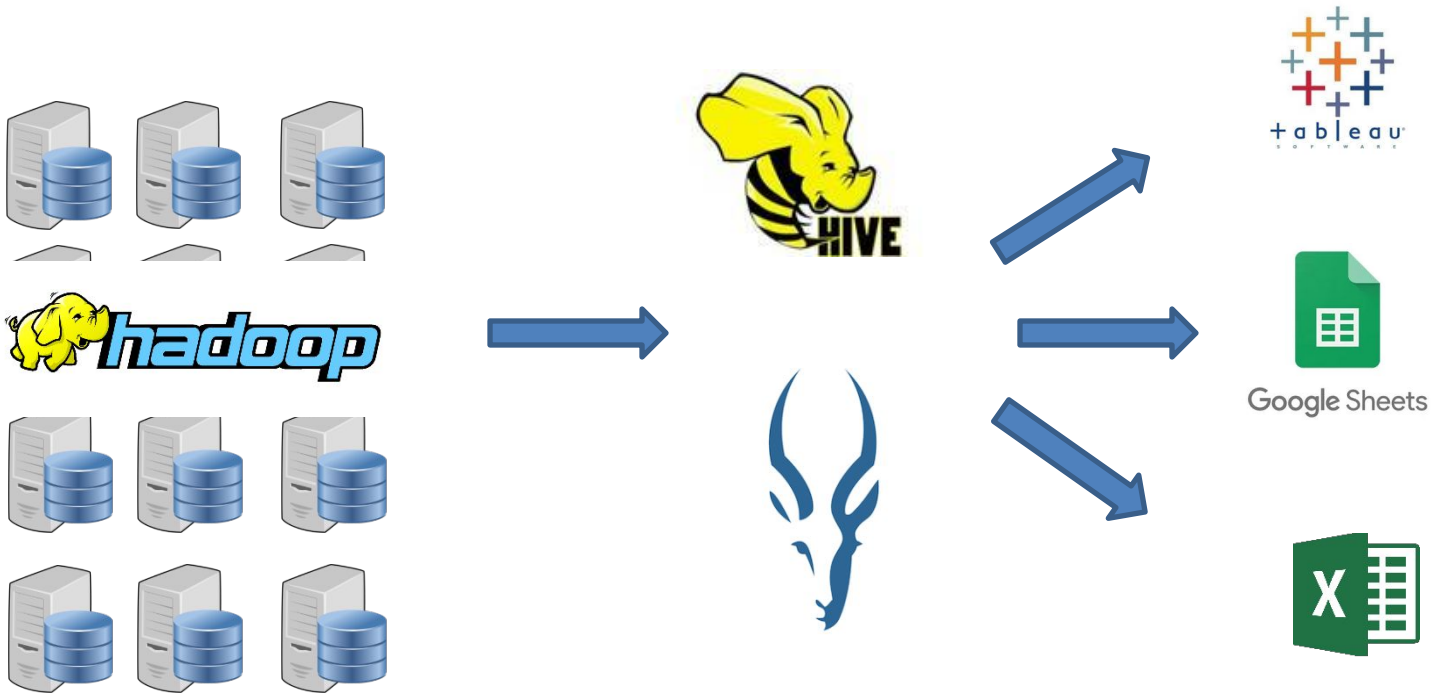
Basic Ecosystem



Basic Ecosystem



Basic Ecosystem



Relational DB

SCHEMAS/LIBRARIES/COLLECTIONS

TABLES/FILES/OBJECTS

MEMBERS/PARTITION IN FILES

Relational DB

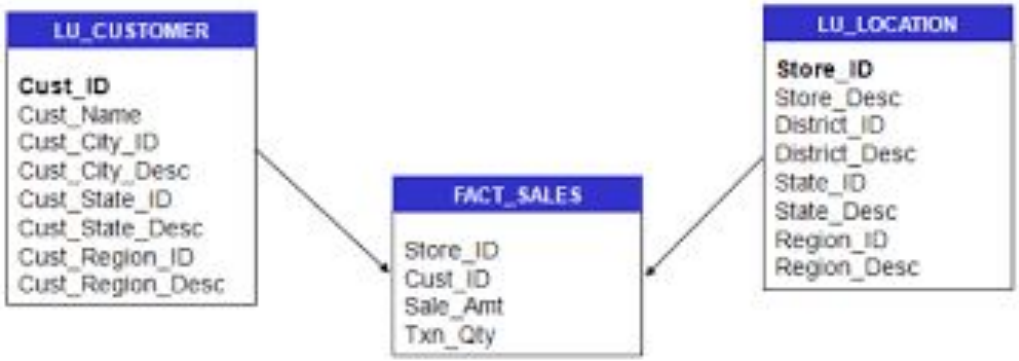
Id	Release date	Record Label	Artist	Song	Album	sales
1						
2						
3						
4						
5						

Id	Artist	Song	Album	sales
1				
2				
3				
4				
5				

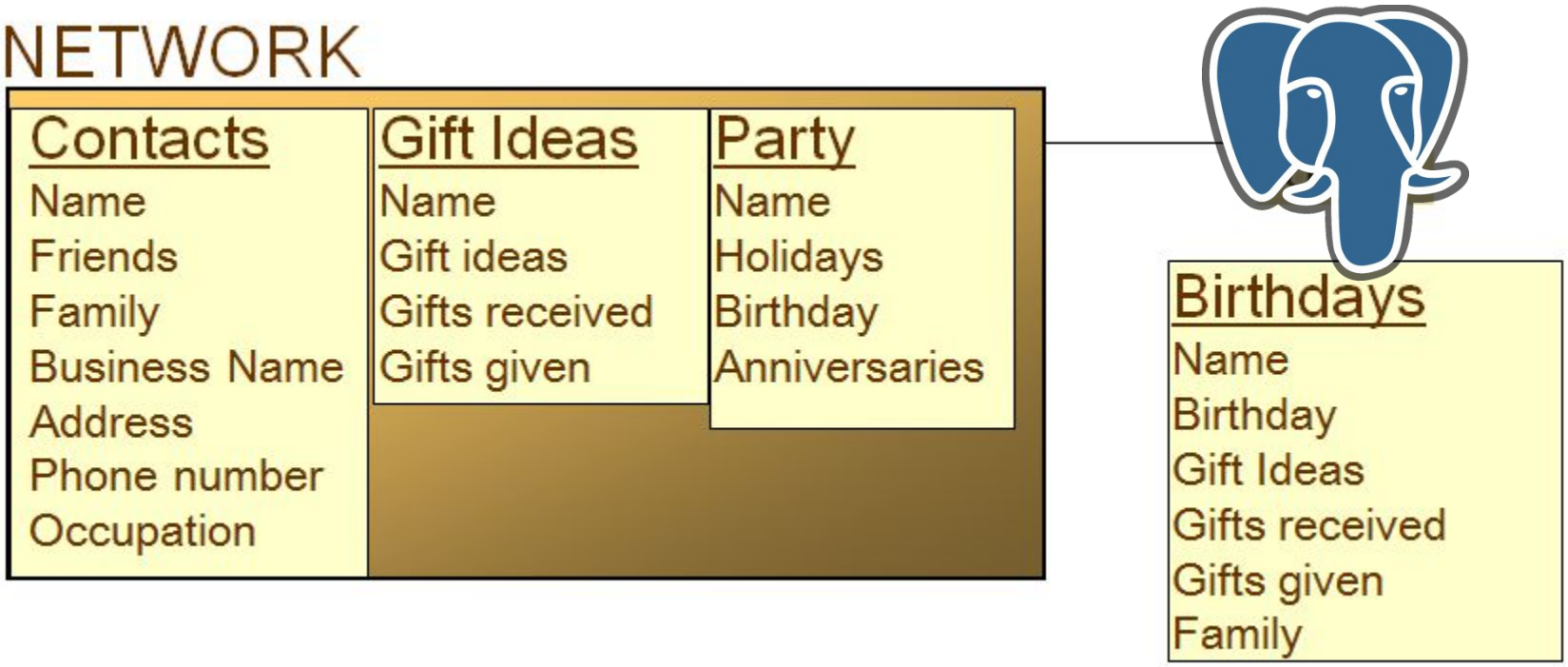
Relational DB

SALES

FIELD	TYPE	LENGTH
ID	PK	1
ARTIST	Char	25
SONG	Char	225
ALBUM	Char	225



Relational DB



Relational DB

Understanding your data.

What are the tables?

What are the fields?

How might you link the tables?

Basic SQL

SELECT
FROM
WHERE
ORDER BY
LIMIT

Basic SQL

SELECT *

SELECT FIELD1, FIELD2 …

SELECT (FIELD1+FIELD2), FIELD 3…

SELECT SUM(FIELD1), FIELD2

```
SELECT DISTINCT  Location, NumberOfSales
```

Location	NumberOfSales
Seattle	101
Seattle	40
Tacoma	72

SELECT DISTINCT Location, NumberOfSales, Date

Location	NumberOfSales	Date
Seattle	101	10/28/17
Seattle	101	10/27/17
Seattle	40	10/26/17
Tacoma	72	10/28/17
Tacoma	72	10/27/17

Basic SQL

WHERE COUNTRY = US

WHERE COUNTRY = US
AND STATE = WA

WHERE COUNTRY = US
AND STATE = WA
AND SALES > 100

ORDER BY 1

ORDER BY 1,2 DESC

LIMIT 1000

ROWNUM <= 1000

Basic SQL

CAST(field AS type)

A helpful String function

Oracle uses TO_
(TO_DATE, TO_CHAR ETC)

Basic SQL

1. Select various fields from the INID table that interest you.

   *BE sure to use LIMIT 1000

2. Practice using filters.
3. Use AND to apply multiple filters

Change the sort.

4. Save your Query

Use your new skills to review Tables from 1 or more of your options

Q & A

Better a diamond with a flaw than a pebble without.
- confucius

Conclusion

SQL is the language of questions and retrieving answers.

Developing good questions is the first step in developing a good SQL Statement.

The next step is knowing where your data resides. In order to find that out you need to find your source behind many reports

FEEDBACK

CLASS : FUNDAMENTALS OF DATA FLOW AND SQL

QUESTION:

Which Clause would you use to create your Filters?