Final Project
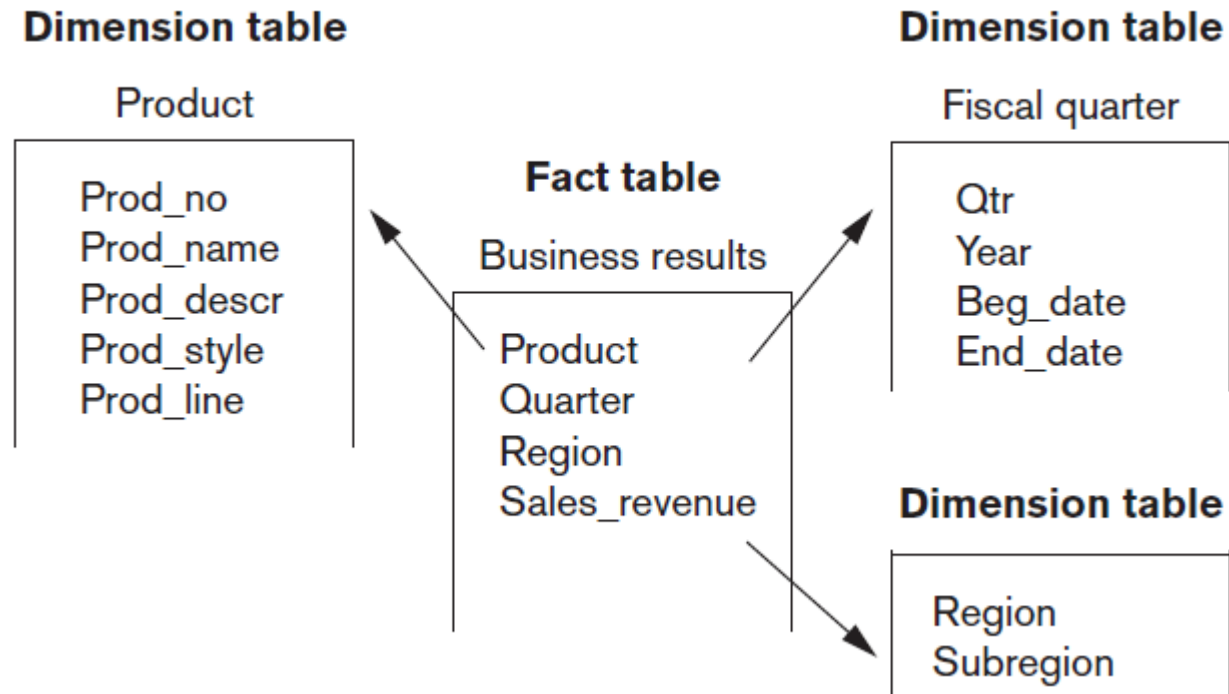
CS 327E
March 26, 2018

Final Project Guidelines

http://www.cs.utexas.edu/~scohen/project/fp_guidelines.pdf
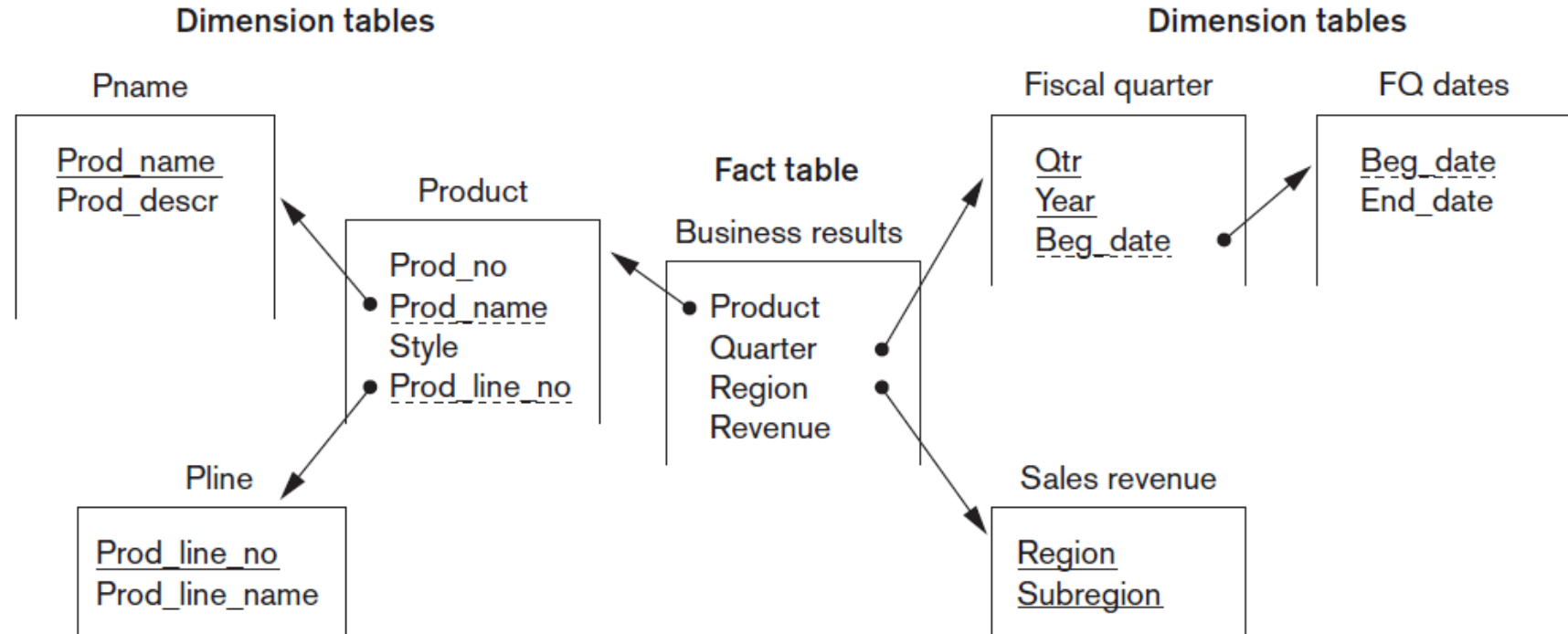
1) Which is **not** a characteristic of a data warehouse?

A) Transaction processing
B) Massive data volumes
C) Dimensional model
D) Extract-Transform-Load process

2) The text discusses an instance of a dimensional model as shown below. Which query can be answered by this schema?

A) Monthly sales revenues.
B) Weekly sales revenues.
C) Daily sales revenues.
D) None of the above.

Dimension table

Product

Prod_no
Prod_name
Prod_descr
Prod_style
Prod_line

Fact table

Business results

Product
Quarter
Region
Sales_revenue

Dimension table

Fiscal quarter

Qtr
Year
Beg_date
End_date

Dimension table

Region
Subregion

3) What makes this schema a snowflake (as opposed to a star schema)?

Dimension tables

Pname

Prod_name
Prod_descr

Product

Prod_no
Prod_name
Style
Prod_line_no

Fact table

Business results

Product
Quarter
Region
Revenue

Dimension tables

Fiscal quarter

Qtr
Year
Beg_date

FQ dates

Beg_date
End_date

Pline

Prod_line_no
Prod_line_name

Sales revenue

Region
Subregion

A) There are multiple tables for each Fact

B) There are multiple tables for each Dimension

C) Both A and B

4) What implicit assumption(s) does the dimensional model make?

A) The grain of the Fact table is at the appropriate level-of-detail.
B) The actual facts / measures are key performance indicators for the business.
C) The Dimension tables contain all the relevant attributes about the dimension they are representing (Product, Customer, etc.).
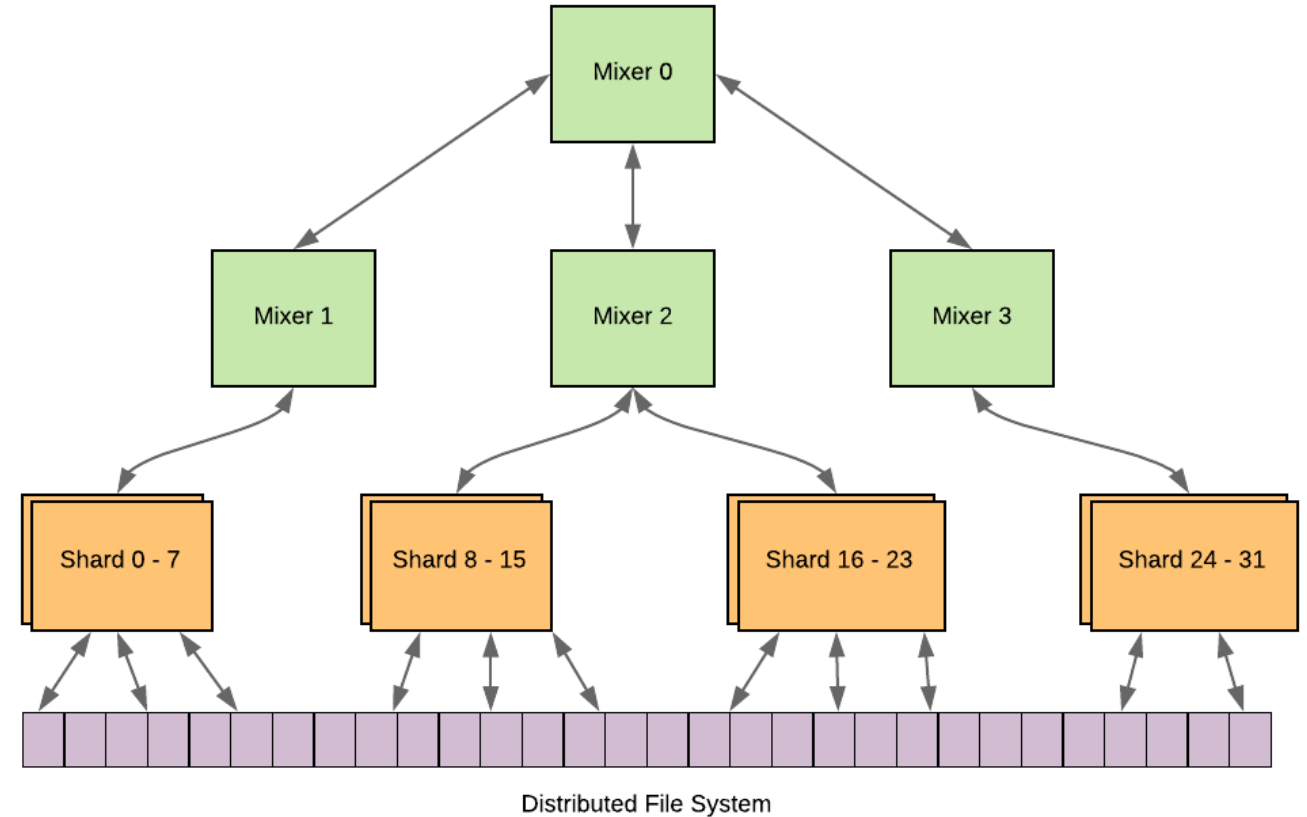D) All of the above.

5) What can go wrong in a data warehouse?

A) The reporting data goes stale.
B) The ETL process runs overtime.
C) The source data contains bad records.
D) All of the above.

BigQuery / Dremel

Key Concepts:

- Scale-out architecture
- Structured data storage
- Columnar storage format
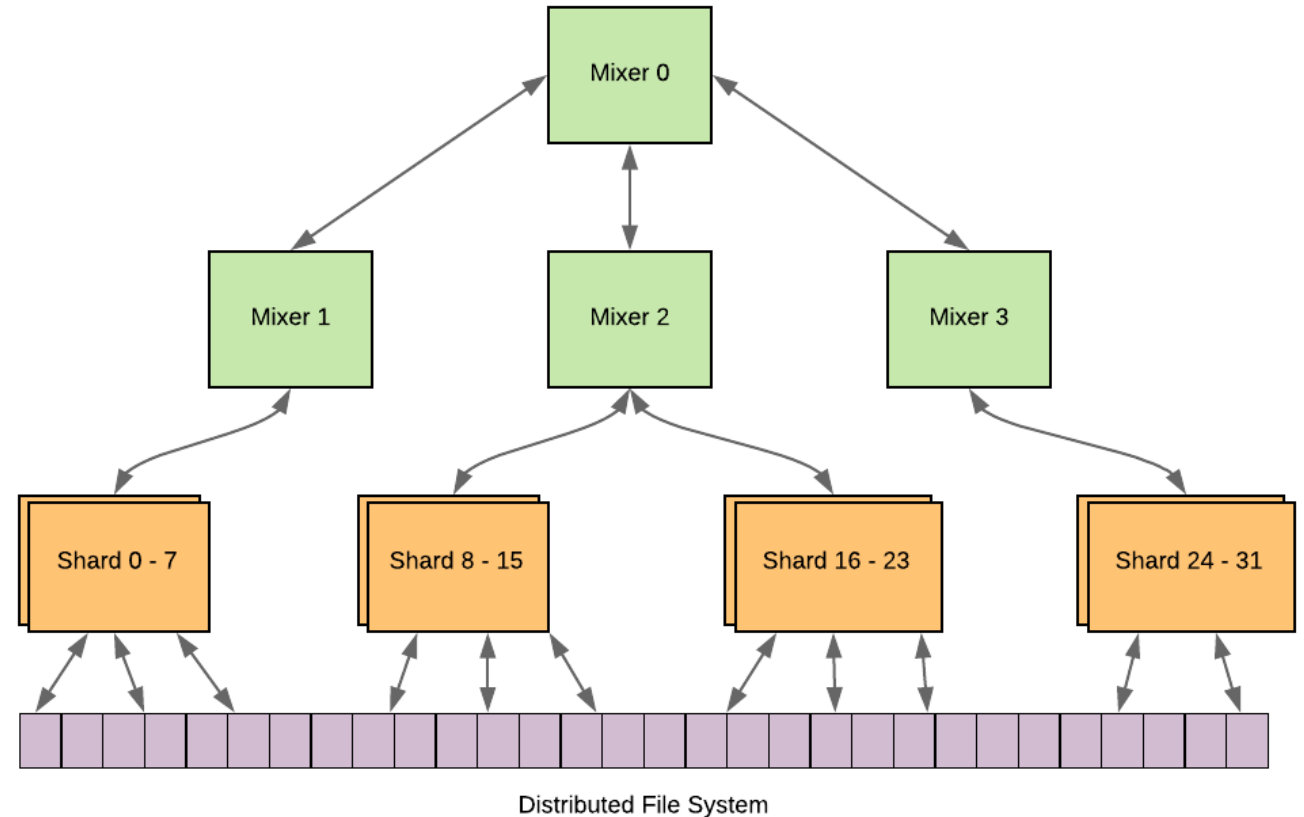- Standard SQL queries
- Jobs

BigQuery Architecture

Key Concepts:

- Scale-out architecture
- Structured data storage
- Columnar storage format
- Standard SQL queries
- Jobs

Tradeoffs:

- Limited DML
- Limited concurrency
- No B-trees

BigQuery Architecture

Demo

BigQuery References

Documentation Home Page: https://cloud.google.com/bigquery/docs/

Data Loading Tutorial: https://cloud.google.com/bigquery/docs/loading-data-cloud-storage-csv

SQL Reference Guide:
https://cloud.google.com/bigquery/docs/reference/standard-sql/

Public Datasets: https://cloud.google.com/bigquery/public-data/