**BI Lab Exp 1**

K Suryanarayan Dora Vu4f2122034 Div: A Batch: B

**Star Schema**

What is a star schema?

A **Star Schema** is a multi-dimensional data model used to organize data in a database so that it is easy to understand and analyze. Star schemas can be applied to data warehouses, databases, data marts, and other tools. The star schema design is optimized for querying large data sets.

Introduced by Ralph Kimball in the 1990s, star schemas are efficient at storing data, maintaining history, and updating data by reducing the duplication of repetitive business definitions, making it fast to aggregate and filter data in the data warehouse.



Fact tables and dimension tables

A star schema is used to denormalize business data into dimensions (like time and product) and facts (like transactions in amounts and quantities).

A star schema has a single fact table in the center, containing business "facts" (like transaction amounts and quantities). The fact table connects to multiple other dimension tables along "dimensions" like time, or product. Star schemas enable users to slice and dice the data however they see fit, typically by joining two or more fact tables and dimension tables together.

Denormalized data

Star schemas **denormalize** the data, which means adding redundant columns to some dimension tables to make querying and working with the data faster and easier. The purpose is to trade some redundancy (duplication of data) in the data model for increased query speed, by avoiding computationally expensive join operations.

In this model, the fact table is normalized but the dimensions tables are not. That is, data from the fact table exists only on the fact table, but dimensional tables may hold redundant data.

Benefits of star schemas

* Fact/dimensional models like star schemas are **simple** to understand and implement, and make it easy for end users to find the data they need. They can be applied to data marts and other data resources.
* **Great for simple queries** because of their reduced dependency on joins when accessing the data, as compared to normalized models like snowflake schemas.
* **adapt well to fit OLAP models**
* **Improved query performance** as compared to normalized data, because star schemas attempt to avoid computationally expensive joins.

How does a star schema differ from 3NF (Third Normal Form)?

3NF,or Third Normal Form is a method of reducing data-redundancy through normalization. It is a common standard for databases that are considered fully normalized. It typically has more tables than a star schema due to data normalization. On the flip-side, queries tend to be more complex due to the increased number of joins between large tables.

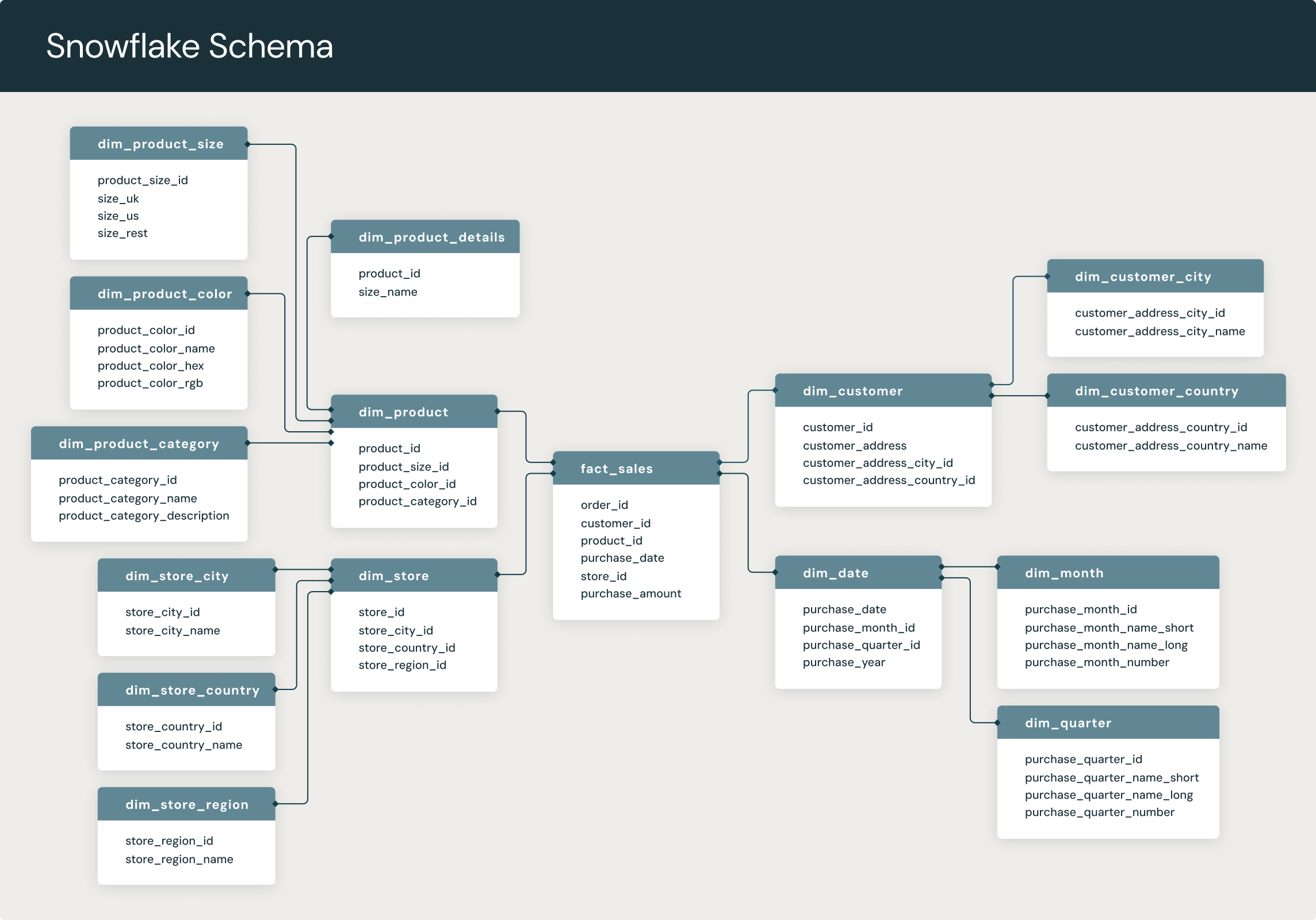
**Snowflake Schema**

What is a snowflake schema?

A **snowflake schema** is a multi-dimensional data model that is an extension of a star schema, where dimension tables are broken down into subdimensions. Snowflake schemas are commonly used for business intelligence and reporting in OLAP data warehouses, data marts, and relational databases.

In a snowflake schema, engineers break down individual dimension tables into logical subdimensions. This makes the data model more complex, but it can be easier for analysts to work with, especially for certain data types.

It's called a snowflake schema because its entity-relationship diagram (ERD) looks like a snowflake, as seen below.



Snowflake schemas vs. star schemas

Like star schemas, snowflake schemas have a central fact table which is connected to multiple dimension tables via foreign keys. However, the main difference is that they are more normalized than star schemas.

Snowflake schemas offer more storage efficiency, due to their tighter adherence to high normalization standards, but query performance is not as good as with more denormalized data models. Denormalized data models like star schemas have more data redundancy (duplication of data), which makes query performance faster at the cost of duplicated data.

Benefits of snowflake schemas

* Fast data retrieval
* Enforces data quality
* Simple, common data model for data warehousing

Drawbacks of snowflake schemas

* Lots of overhead upon initial setup
* Rigid data model
* High maintenance costs