L3HARRIS

FAST. FORWARD.

ICEAA 2024

HOW TO CHOOSE A DATABASE STORAGE MODEL

TTRISTAN JUDD | Systems Engineer Data Analyst

May 16, 2024

Introduction

Tristan Judd

SUNY New Paltz class of 2022

B.S. in Computer Engineering

L3Harris employee since January 2023

Last year I was tasked with setting up the data pipeline for our team.

I did an analysis of our data needs and concluded that dimensional databases are the best solution for us.

I'm here to tell you how I came to that conclusion and what benefits come with dimensional databases.

Our data needs

Requirement Analysis

- Large amounts of structured data

- Querying in PowerBI

- Ability to Time phase data

- Scalable structure

Database Options

Excel, storing documents in file explorer

- Optimized for ease of use

Traditional Relational databases

- Optimized for normalization

Dimensional databases

- Optimized for ease of Querying

What is relational

- Type of Database, typically managed using Structured Querying Language (SQL)

- Data is organized into tables, where each row represents a record and each column represents a data field

- Each row has a unique primary key used to identify the row, and information can be linked using these keys
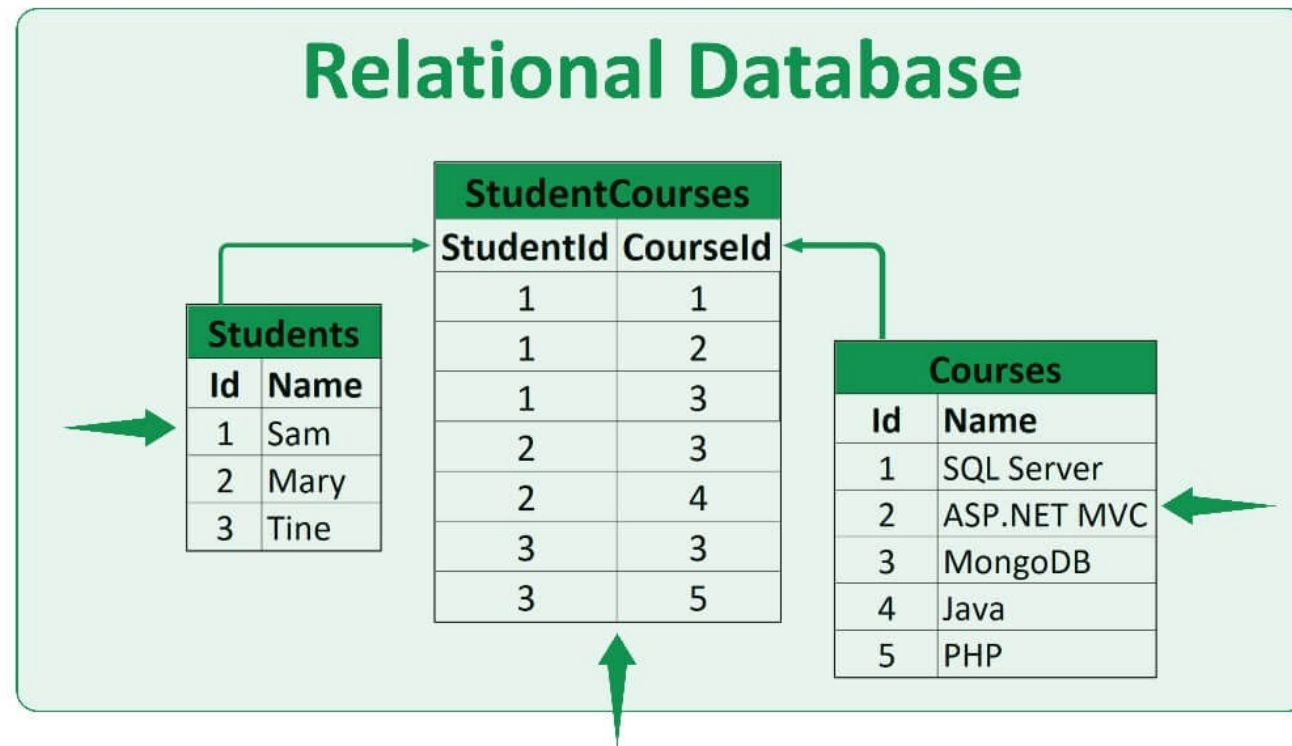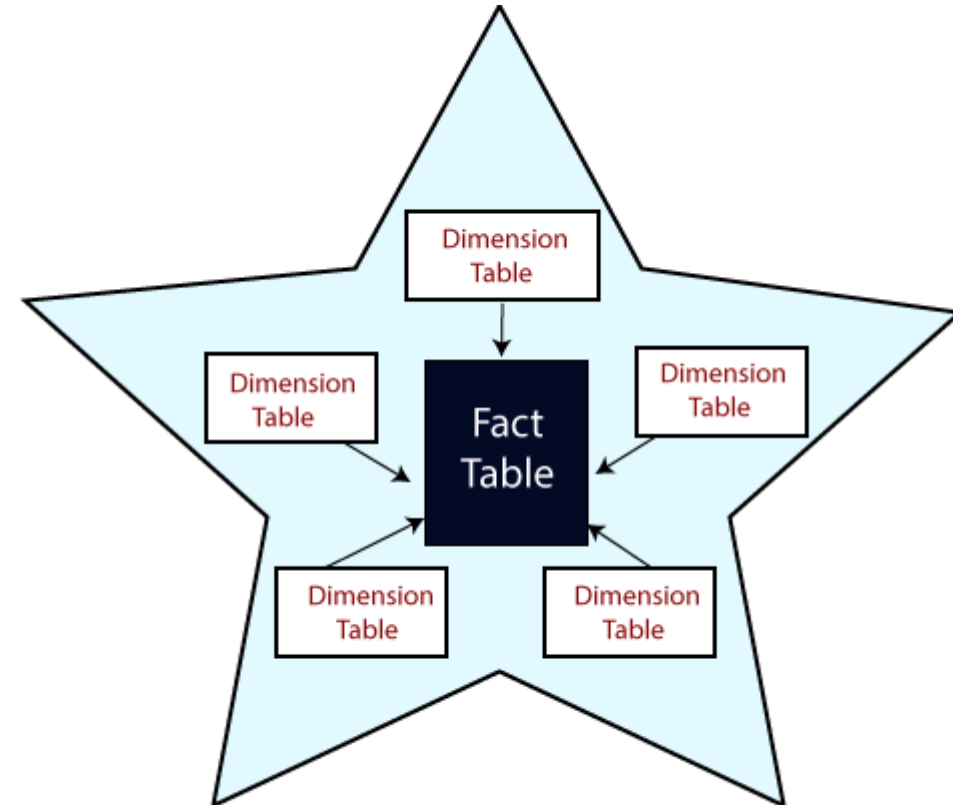
What is dimensional?

Facts – The lowest grain you are storing

Dimensions – Tables that helps define the fact

Star Schema – Dimensional layout with a single
 fact table

Star Schema

Relational vs Dimensioal

The primary difference between relational and dimensional is their design and intended purpose.

Relational are optimized for **data integrity** and **normalization**.

Dimensional databases are optimized for **data reading** and **querying speed**.

Database option chart

	Excel	Relational	Dimensional
Large amounts of Storage	Yes	Yes	Yes
Quarriable in PowerBI	No	No	Yes
Time Phasing	No	No	Yes
Scalable Structure	No	No	Yes

Dimensional Examples

Football

- Fact: Yardage lost or gained

- Dimensions: Players, teams, time, date

Weather

- Fact: Temperature, humidity, wind speed
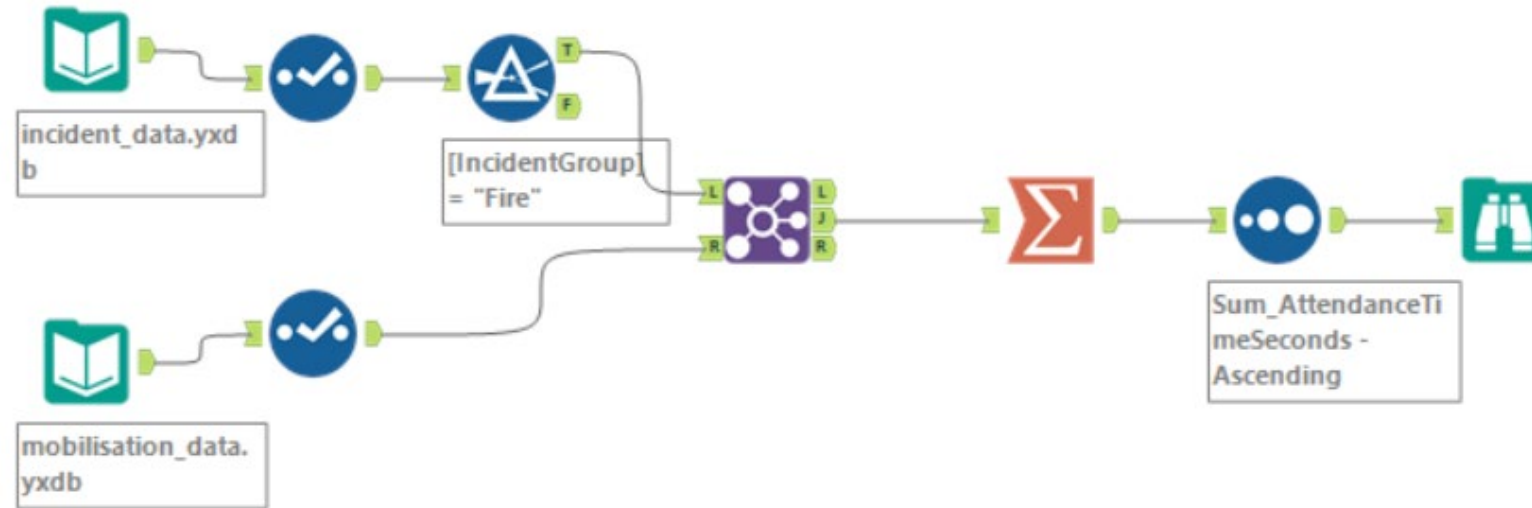
- Dimension: Date, time, units, equipment type

Car Sales

- Fact: Sales record

- Dimension: Customer, Salesperson, Date, Time

ETL

Splitting up the data comes with a little extra work, but we have tools to make it easy

Can also write python scripts to split up this data automatically. Comprises of comparison operations, splits, and merges.

Time phasing

Type 0 – No changes allowed

Type 1 – When a change occurs, the existing data is overwritten

Type 2 – Multiple records are kept to track changes over time. When a change occurs, a new row is added, with a start and end date to signify when the data is up to date

Type 3 – Instead of keeping separate records, changes are tracked by storing previous values in extra columns of the same record. Simply put a new column is added with the new values in the same row.

Conclusion

Dimensional Databases are a great option for teams with large amounts of structured data they would like to easily query

Star Schema gives a format for how to structure the database, and makes visualizing data in PowerBI easy

With a bit of time invested into learning SQL and Dimensional Design, a team can see many benefits when making the switch to a Dimensional Database.

QUESTIONS?