DIMENSIONAL MODELLING

By GROUP 8

ANKUR SRIVASTAVA

GOUTHAM BEERAM

PRASAD ACHARYA

RISHABH MITTAL

RITIK GUPTA

Introduction

To gain experience and better understanding of dimensional modelling in data warehouse design, we have developed a basic ERD diagram by taking the available stock data.

We have also made some assumptions about the design.

Finally, we have written few basic queries which can be further developed to analytic queries

Dimensions and Fact Tables

- SP500_STOCKS is a dimension table that has all the information of companies. Change_ID is added to handle the changes to the companies like stock splits, merges and acquisitions etc.

- SP500_EOD_STOCK_FACTS is a fact table that holds all intraday transactions. Congress_ID is added to identify the political relation to stock market analysis.

- POLITICAL_PARTIES is a fact table that has the information of the ruling government for each year

- JULIAN_DAYS is a time dimension table that holds all time dimensions. It does not hold any transaction data. Any changes would update the current records.

- SECTORS is a dimension table which records sector information whenever there is a change in company business activity such as stock splits, merges etc.

- STOCK_CHANGES_FACT is a fact table where all changes at the granular level to be recorded in the fact table for preserving history.

- SPLIT_TYPES is a dimension table that holds types of splits if any. This design will help the process more data driven.

ERD DIAGRAM
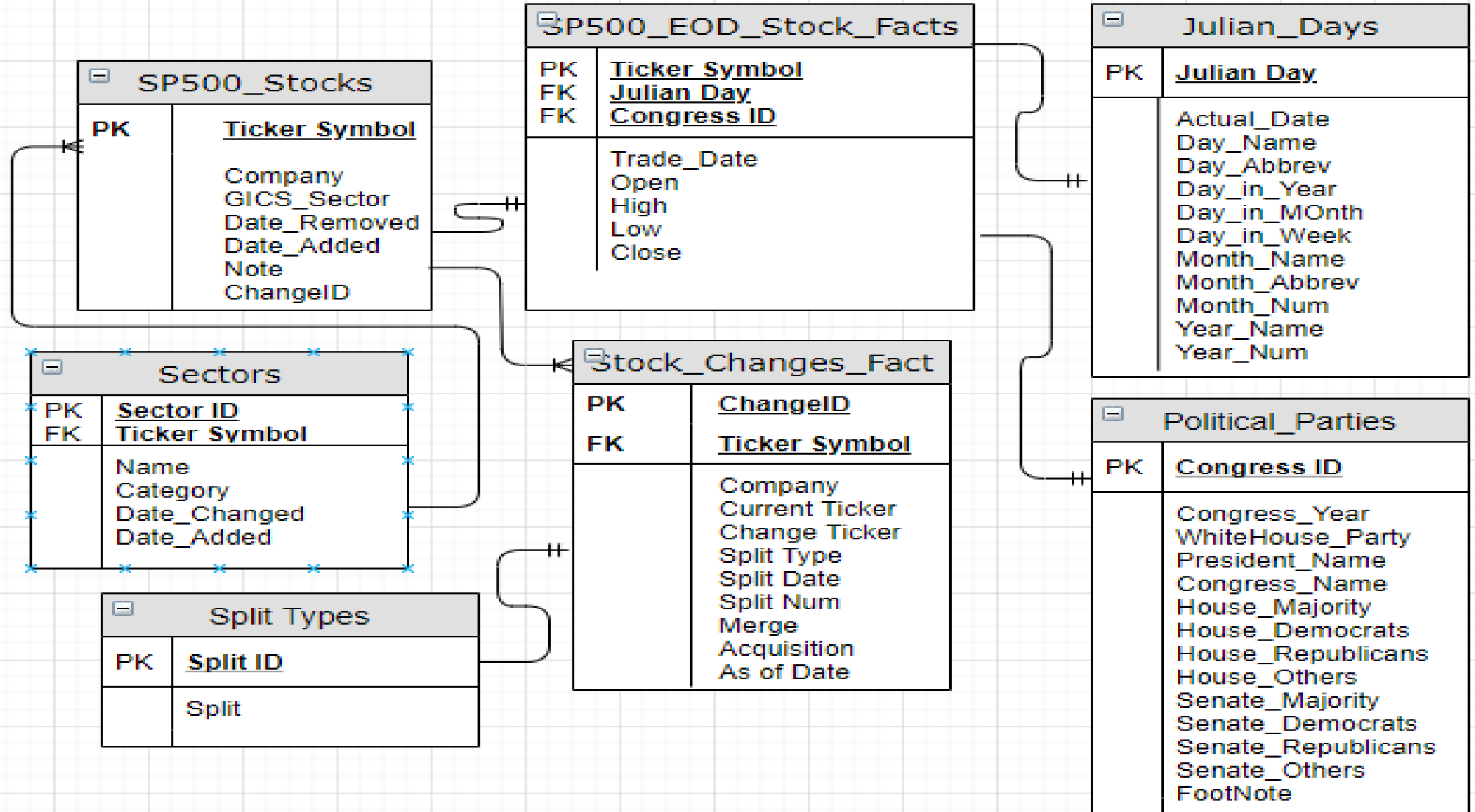
Assumptions

- The type of government and the political parties in action can have an influence on the stock prices

- SP500_STOCKS will contain the sub-details of the stock model focusing on a particular section

- The degree of relationship between the SP500_EOD_STOCK_FACTS and POLITICAL_PARTIES can be analyzed and determined with analytical queries

SOME QUERIES

1. The top 10 company ticker and how they performed over different political parties in power between time period of 1990 through 2020.

SELECT top 10 percent

sf.ticker_symbol, president_name, whitehouse_party, house_majority, senate_majority

FROM SP500_EOD_STOCK_FACTS sf

inner join SP500_STOCKS ss

on

sf.ticker_symbol = ss.ticker_symbol

inner join STOCK_CHANGES_FACT scf

on

```sql
sf.ticker_symbol = scf.ticker_symbol

inner join political_parties pp

on

sf.congress_id=pp.congress_id

where trade_date_str between 1/1/1990 and 1/1/2020

group by sf.ticker_symbol

having

sf.high > (select avg(high)

from SP500_EOD_STOCK_FACTS

where trade_date_str

between 1/1/1990 and 1/1/2020)

order by sf.ticker_symbol
```

2. The company that were undergone atleast a change in business like merger, acquisition or even stock splits.

SELECT ss.company, scf.high, scf.low

from SP500_STOCKS ss

inner join SP500_EOD_STOCK_FACTS scf

on

ss.ticker_symbol = scf.ticker_symbol

inner join julian_days jd

on

scf.julian_day = jd.julian_day

where

ss.ticker_symbol

in

(select ticker_symbol

from sectors

where changeid is not null)

group by

ss.company, scf.high, scf.low

order by

ss.Company

STREAMING SERVICE

The Star Schema Below follows Business Process of a streaming service like Netflix and how can it be leveraged to find Business Intelligence which could be used to derive Recommendation Systems.

FACT AND DIMENSIONS TABLE

■ The fact table contains details about user watching history. The granularity covers user_id, show_id, Date_watched_id, Registration_id, Payment_id, Time_spent, Pause_time, max_genre_watch.

■ User dimension : This deals with basic user details including location of the user so to have demographic intelligence

■ Registration Dimension : This deals with registration details like registration type(Premium, standard, etc.).

FACT AND DIMENSIONS TABLE

- Transaction_Dimension: This dimension covers basic transaction details like payment_id, date of payment, Account type, Payment info

- Show Dimension: This are the details about shows, like show type(movie, tv, documentary) and all the show details like genre, description, production cast, release date, etc

- User_members: This is the extension from user dimension. As an account would be registered by an user_id but one user can have 3 to 4 members.

FACT AND DIMENSIONS TABLE

- Date Dimension: The Date Dimension is the most crucial element of the design assuming each date_watched_id is like every other day. While the fact table corresponds to this date identifier as the user who watched on this date. For instance if the date identifier is 1 it denotes it is January 1 of the year streaming service has started and then 2 means the second day and it goes on for every single day.

- It has details like week month day and binary variables like if it is weekend or holiday on that specified date. This accounts to a good business intelligence use case.

User_Dimension

PK	User_id
	User_name
	User_email
	User_Location
	user_members_number
	Latitute
	Longitude
	user_age
	user_preference

Transaction Dimension

PK	Payment_id
	Payment_type
	Account_type
	Payment
	Payment_info

Registration_dimension

PK	Registration_id
	Registration_type
	Registration_rate
	Registration_date

Fact_table

FK	user_id
FK	show_id
FK	Date_watched_id
FK	Registration_id
FK	Payment_id
	Time_spent
	Pause_time
	max_genre_watched
	Cumulative time spent

Show Dimensions

PK	show_id
	show_type
	show_name
	show_genre
	show_description
	show production
	show cast
	show Director
	show_country
	show_episodes
	show_time
	show_rating
	show_liked_ratio
	show_release_date

User_Members

FK	User_id
	Member_name
	Member_age
	Member_location
	Member Preference
	Latitude
	Longitude

Date Dimension

PK	Date_watched_id
	Date_week
	Date_month
	Date_day
	is_weekend
	is_Holiday