

Project Requirement:

Data for this task can be found below;

CSV Format – video_data.csv (1,048,575 rows)

Use this raw data and construct a star schema Data Warehouse which will be used to track VideoStarts over time.

Show the SQL queries you would use to populate the Data Warehouse Dimensions and Fact table.

- VideoStart metric;
 - Determined from the “events” column containing “206”
 - All rows without 206 should be discarded
- DimDate
 - should go to the minute grain
- Dimensions
 - DimDate
 - DimPlatform
 - DimSite
 - DimVideo
- DimPlatform
 - Split VideoTitle by pipe ‘|’
 - If VideoTitle.split('|')[0] contains something that looks like a platform (iPhone, Android Phone etc) then use that as the platform
 - If VideoTitle.split('|')[0] doesn't contain a platform but looks like a site, assume the platform is Desktop
 - If VideoTitle.split('|').count = 1, discard the row.
- DimSite
 - Split VideoTitle by pipe |
 - If VideoTitle.split('|').count = 1, discard the row.
 - If VideoTitle.split('|')[0] looks like a site name, save the site name
- DimVideo
 - Last piece of VideoTitle.split('|') contains the video title
 - You can ignore any middle pieces