Assignment 1 – ETL Process

Data for this task can be found at these links;
CSV Format – https://s3-ap-southeast-2.amazonaws.com/jiangren-de-bucket/assignments/video_data.csv
Parquet Format – https://s3-ap-southeast-2.amazonaws.com/jiangren-de-bucket/assignments/video_data.gz.parquet

Use whichever format is easier for you to work with.
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

Assignment 2 – Reporting

Data for this task can be found at these links;
If you happen to use MSSQL, everything has been setup in this backup file - https://s3-ap-southeast-2.amazonaws.com/jiangren-de-bucket/assignments/TechTest2_DB_MSSQL.bak
Want to use a different system? No troubles! The CSV data and CREATE TABLE scripts can be found here - https://s3-ap-southeast-2.amazonaws.com/jiangren-de-bucket/assignments/TechTest.zip

Use the database provided and create a report (or set of reports) that can be used to analyse site performance by day.
Report can be created in any platform of your choice (Tableau, SSRS, Excel etc).

- User should be able to select multiple sites, and devices to compare performance
- User should be able to select a date range OR prior day range (eg last 30 days)
- Site metrics to use are PageViews, Visits, Unique Visitors
- Should use only LedgerName='Actual'
- Should use only PlatformType='Web'