Tables Design:

VIDEOSTART_RAW: data landing table

COLUMN_NAME	DATA_TYPE	PK	NULLAB LE	DATA_DEFAULT	COLUMN_ID	COMMENTS
DATETIME	VARCHAR(30 BYTE)	N	Yes	null	1	Data from raw file
VIDEOTITLE	VARCHAR(200 BYTE)	N	Yes	null	2	Data from raw file
EVENTS	VARCHAR(150 BYTE)	N	Yes	null	3	Data from raw file

VIDEOSTART_DLT: after data transformation from VIDEOSTART_RAW

COLUMN_NAME	DATA_TYPE	PK	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
DATETIME	TIMESTAMP (6)	N	No	null	1	Data reformatted from VIDEOSTART_RAW. DATETIME
PLATFORM	VARCHAR(200 BYTE)	N	No	null	2	Data derived from VIDEOSTART_RAW. VIDEOTITLE
SITE	VARCHAR(200 BYTE)	N	No	null	3	Data derived from VIDEOSTART_RAW. VIDEOTITLE
VIDEO	VARCHAR(200 BYTE)	N	No	null	4	Data derived from VIDEOSTART_RAW. VIDEOTITLE

DIMDATE_DLT: delta table is used to record the date info

COLUMN_NAME	DATA_TYPE	PK	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
DATETIME	VARCHAR(12 BYTE)	N	No	null	1	Data reformatted from VIDEOSTART_DLT. DATETIME

DIMPLATFORM_DLT: delta table is used to record the platform info

COLUMN_NAME	DATA_TYPE	PK	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
PLATFORM	VARCHAR2(200 BYTE)	N	No	null	1	Data derived from VIDEOSTART_DLT. PLATFORM

DIMSITE_DLT: delta table is used to record the site info

COLUMN_NAME	DATA_TYPE	PK	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
SITE	VARCHAR2(200 BYTE)	N	No	null	1	Data derived from VIDEOSTART_DLT. SITE

DIMVIDEO_DLT: delta table is used to record the video info

COLUMN_NAME	DATA_TYPE	PK	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
VIDEO	VARCHAR2(200 BYTE)	N	No	null	1	Data derived from VIDEOSTART_DLT. VIDEO

MIDDATE: date dimension table without surrogate key since date time is already unique which can be used as unique key.

COLUMN_NAME	DATA_TYPE	PK	NULLAB LE	DATA_DEFAULT	COLUMN_ID	COMMENTS
DATETIME_SKEY	NUMBER(38,0)	Y	No		1	Data derived from DIMDATE_DTL. DATETIME

MIDPLATFORM: platform dimension table with surrogate key

COLUMN_NAME	DATA_TYPE	PK	NULLAB LE	DATA_DEFAULT	COLUMN_ID	COMMENTS
PLATFORM_SKEY	NUMBER(38,0)	Y	No		1	
PLATFORM	VARCHAR2(20 0 BYTE)	N	No	null	2	Data derived from DIMPLATFORM_D LT. PLATFORM

DIMSITE: site dimension table with unique surrogate key

COLUMN_NAME	DATA_TYPE	PK	NULLAB LE	DATA_DEFAULT	COLUMN_ID	COMMENTS
SITE_SKEY	NUMBER(38,0)	Y	No		1	
SITE	VARCHAR2(20 0 BYTE)	N	No	null	2	Data derived from DIMSITE_DLT. SITE

DIMVIDEO: video dimension table with unique surrogate key

COLUMN_NAME	DATA_TYPE	PK	NULLAB LE	DATA_DEFAULT	COLUMN_ID	COMMENTS
VIDEO_SKEY	NUMBER(38,0)	Y	No		1	
VIDEO	VARCHAR2(20 0 BYTE)	N	No	null	2	Data derived from DIMVIDEO_DLT. VIDEO

There are 6 stages designed in sequence to do the ETL process on mysql workbench with multiple tools:

1. Data Landing: load the raw videostarts file into videostart_raw table

Load the video data into RStudio and have a quick review of the data in video_data.csv file. There are 2686002 observations with 3 columns for each and the content of each observation is showing as below snapshot:

```
> dim(videodata)
[1] 2686002        3
>
```

```
> head(videodata)
           DateTime                                                    VideoTitle                                        events
1 2017-01-11T00:00:31.000Z          App Web|Clips|a-current-affair;2016|William Tyrrell twist     157,120,160,104,162,161,163,164,165,166,171,229
2 2017-01-11T00:00:53.000Z                                    news| Shark attacks spearfisherman         127,157,120,160,104,162,161,171,206
3 2017-01-11T00:00:21.000Z                                    news| Shark attacks spearfisherman 127,157,120,160,104,162,161,163,164,165,166,171,229
4 2017-01-11T00:01:27.000Z                                    news| Chilean navy films UFO         157,120,160,104,162,161,170,171,237
5 2017-01-11T00:00:33.000Z          news| Video shows alleged axe attack at Sydney service station     157,120,160,104,162,161,163,164,165,166,171,229
6 2017-01-11T00:00:18.000Z App Web|Clips|today;2017|Pint-sized Aussie surfer hits the global stage 127,157,120,160,104,162,161,163,164,165,166,171,229
```

The landing stage is only designed to swallow the raw data information as original and quick as possible. Therefore, a table videostart_raw table is designed for data landing with 3 attributes : date_time, video_title and events.

Using MySQL Workbench tool to do table data import from video_data.csv into new created table videostart_raw: