-- make tables for each file

-- EpisodeV

CREATE TABLE episodeV (name STRING , line STRING)

ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'

TBLPROPERTIES("skip.header.line.count"="2");

--load data into tables

LOAD DATA LOCAL INPATH 'root/input/episodeV\_dialouges.txt' INTO TABLE episodeV;

--Run Query on Hive

SELECT name, COUNT(name) AS no\_of\_lines FROM episodeIV GROUP BY name ORDER BY no\_of\_lines DESC;

SELECT name, COUNT(name) AS no\_of\_lines FROM episodeV GROUP BY name ORDER BY no\_of\_lines DESC;

SELECT name, COUNT(name) AS no\_of\_lines FROM episodeVI GROUP BY name ORDER BY no\_of\_lines DESC;

SELECT COUNT(\*) FROM episodeIV where INSTR(line,'Luke')>=1;