--Data Loading into BigInsights:

wget -O Building\_and\_Safety\_Permit\_Information.csv https://www.dropbox.com/s/m5j0yqkvtqskim7/Building\_and\_Safety\_Permit\_Information.csv

hdfs dfs -mkdir Building

hdfs dfs -put Building\_and\_Safety\_Permit\_Information.csv Building

--Creating Hive table to Query Building & Safety Permit Data:

hive

CREATE EXTERNAL TABLE building\_permit (

`Unique\_ID` string,

`Assessor Book` string,

`Assessor Page` string,

`Assessor Parcel` string,

`Tract` string,

`Lot` string,

`PCIS Permit #` string,

`Permit` string,

`Permit Type` string,

`Status` string,

`Permit Sub-Type` string,

`Status Date` string,

`Permit Category` string,

`Initiating Office` string,

`Issue Date` string,

`Address Start` string,

`Address End` string,

`Street Direction` string,

`Street Name` string,

`Street Suffix` string,

`city` string,

`Zip Code` string,

`Work Description` string,

`Valuation` string,

`Contractor's Business Name` string,

`Contractor Address` string,

`Contractor City` string,

`Contractor State` string,

`License Type` string,

`License #` string,

`License Expiration Date` string,

`Census Tract` string,

`Latitude` string,

`Longitude` string,

`State` string,

`Country` string)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION '/user/mvaghela/Building/' TBLPROPERTIES ('skip.header.line.count'='1');

--Describe all fields:

DESCRIBE building\_permit ;

--View all rows:

SELECT \* FROM building\_permit;

--Query 2 – Permit Sub-Type - Query to Find Count of Permits by Permit Sub-Type

DROP TABLE IF EXISTS permit\_count;

CREATE TABLE permit\_count AS

SELECT `Permit`, `Permit Sub-Type`, COUNT(`unique\_ID`) AS ID

FROM building\_permit

GROUP BY `Permit Sub-Type`, `Permit`;

INSERT OVERWRITE DIRECTORY '/user/mvaghela/Output/permit\_count/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM permit\_count;

--Query 3 – Permit Category - Query to find number of permits by Plan Check or No Plan Check

DROP TABLE IF EXISTS permit\_category;

CREATE TABLE permit\_category

AS SELECT COUNT(`Permit Category`),`Permit`, `Permit Category`

FROM building\_permit

GROUP BY `Permit`, `Permit Category`;

INSERT OVERWRITE DIRECTORY '/user/mvaghela/Output/permit\_category/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM permit\_category;

--Query 4 – Display Permit - Query to find number of permits by City and Permit Subtype

DROP TABLE IF EXISTS map\_permit;

CREATE TABLE map\_permit AS

SELECT `Permit`, `Permit Type`, `City`, `State`

FROM building\_permit;

INSERT OVERWRITE DIRECTORY '/user/mvaghela/Output/map\_permit/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM map\_permit;

--Query 5 – Valuation of Permit - Query to find average valuation of different types of Permits

DROP TABLE IF EXISTS valuation;

CREATE TABLE valuation AS

SELECT Permit, AVG(Valuation) as average

FROM building\_permit

GROUP BY Permit;

INSERT OVERWRITE DIRECTORY '/user/bjadhav/Output/valuation/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM valuation;

--Query 6 – Contractor's Business Name - Query to find Contractor's Business Name by number of permits and their location

DROP TABLE IF EXISTS contractor;

CREATE TABLE contractor AS

SELECT `Contractor's Business Name`, COUNT(`unique\_ID`) AS ID, `Permit`, `Contractor City`, `Contractor State`

FROM building\_permit

GROUP BY `Contractor's Business Name`, `Permit`, `Contractor City`, `Contractor State`;

INSERT OVERWRITE DIRECTORY '/user/ bjadhav /Output/contractor/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM contractor;

--Query 7 – Status - Query to find number of permits issued and in other status

DROP TABLE IF EXISTS permit\_status;

CREATE TABLE permit\_status AS

SELECT COUNT(`Unique\_ID`) ,`Permit` ,`Status`

FROM building\_permit

GROUP BY `Status`,`Permit`;

INSERT OVERWRITE DIRECTORY '/user/ bjadhav /Output/permit\_status/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM permit\_status;

--Query 8 – Permit Issue Date - Query to find number of permits issued in particular year

DROP TABLE IF EXISTS permit\_year;

CREATE TABLE permit\_year AS

SELECT `Permit`, count(`Issue Date`)

FROM building\_permit

WHERE `Issue Date`rlike '.\*(2013).\*'

GROUP BY `Permit`;

INSERT OVERWRITE DIRECTORY '/user/bhjadhav/Output/permit\_year1/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

select \* from permit\_year;

DROP TABLE IF EXISTS permit\_year2;

CREATE TABLE permit\_year2 AS

SELECT `Permit`, count(`Issue Date`)

FROM building\_permit WHERE `Issue Date`rlike '.\*(2014).\*'

GROUP BY `Permit`;

INSERT OVERWRITE DIRECTORY '/user/bhjadhav/Output/permit\_year2/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

select \* from permit\_year2;

DROP TABLE IF EXISTS permit\_year3;

CREATE TABLE permit\_year3 AS

SELECT `Permit`, count(`Issue Date`)

FROM building\_permit

WHERE `Issue Date`rlike '.\*(2015).\*'

GROUP BY `Permit`;

INSERT OVERWRITE DIRECTORY '/user/bhjadhav/Output/permit\_year3/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

select \* from permit\_year3;

DROP TABLE IF EXISTS permit\_year4;

CREATE TABLE permit\_year4 AS

SELECT `Permit`, count(`Issue Date`)

FROM building\_permit

WHERE `Issue Date`rlike '.\*(2016).\*'

GROUP BY `Permit`;

INSERT OVERWRITE DIRECTORY '/user/bhjadhav/Output/permit\_year4/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

select \* from permit\_year4;

--Query 9 – License Type - Query to find count of License Type by it’s average valuation

DROP TABLE IF EXISTS license;

CREATE TABLE license AS

SELECT `License Type`, count(`Unique\_ID`) , Avg(`Valuation`)

FROM building\_permit

GROUP BY `License Type`;

INSERT OVERWRITE DIRECTORY '/user/nmali/Output/license/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM license;

--Query 10 – License Expiration Date - Query to find number of license types by it’s expiration date

DROP TABLE IF EXISTS license\_exp;

CREATE TABLE license\_exp AS

SELECT `License Type`, count(`Unique\_ID`)

FROM building\_permit

WHERE `License Expiration Date` rlike '.\*(2017).\*' OR `License Expiration Date` rlike '.\*(2018).\*'

GROUP BY `License Type`;

INSERT OVERWRITE DIRECTORY '/user/nmali/Output/license\_exp/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM license\_exp;