HospitalQ

[Input](#h.2ipkcms9zpq8)

[Processing](#h.thbti8fpfzc0)

[Description](#h.c6s3fcciac4r)

[Uploading to HDFS](#h.mckwhqw093lj)

[To verify in HDFS](#h.5glq0wbdvwmp)

[Uploading to Hive](#h.3n9vnguhhv5t)

[To verify in Hive](#h.eqcgf17m1590)

[Output](#h.3r2b6mfropbb)

[Description](#h.yp27snvbztzu)

# **Input**

* Database input (Hospital information)
  + Access (or CSV)
* Stream (Twitter)
  + Flume

# **Processing**

## Description

* Hive database
  + hospitalmaster
  + hospitalcompare
  + hospitalopvolume
  + hospitalreadmit
  + hospitaltimelyeffectivecare

## Uploading to HDFS

hadoop fs -put /home/Power2Power/DataSets/TimelyEffectiveCareHospital.csv Power2Power/datafiles

hadoop fs -put /home/Power2Power/DataSets/ValueCareNational.csv Power2Power/datafiles

## To verify in HDFS

hadoop fs -ls Power2Power/datafiles

## Uploading to Hive

create database power2power;

use power2power;

create external table Hospital\_master(Provider\_ID string,HospitalName string,city string,state string,Zipcode string,countyName string,PhoneNumber string,Hospital\_type string,EmergencyServices string) row form row format delimited fields terminated by ",";

create external table Hospital\_TimelyEffectiveCare(Provider\_ID string,HospitalName string,condition string,measure\_id string,measure\_name string,score string,sample string,footnote string,mesaureStartDate string, measureEndDate string) row format delimited fields terminated by ",";

LOAD DATA INPATH '/user/hadoop/Power2Power/datafiles/HospitalGeneralInformation.csv' OVERWRITE INTO TABLE Hospital\_master;

LOAD DATA INPATH '/user/hadoop/Power2Power/datafiles/TimelyEffectiveCareHospital.csv' OVERWRITE INTO TABLE Hospital\_TimelyEffectiveCare;

create table HQI\_HOSP\_MSPB(Provider\_ID string,measure\_id string,score string,footnote string,mesaureStartDate string, measureEndDate string)

LOAD DATA INPATH '/user/hadoop/Power2Power/datafiles/MedicareHospitalSpendingPatientHospital.csv' OVERWRITE INTO TABLE HQI\_HOSP\_MSPB;

## To verify in Hive

Select \* from Hospital\_master;

select \* from Hospital\_TimelyEffectiveCare;

select \* from HQI\_HOSP\_MSPB;

# **Output**

## Description

* Query based on the following criteria
  + Zipcode
* Result output
  + Hospital Phone Number
  + Emergency (Yes / No)
  + Outpatient care (Yes / No)
  + Outpatient care programs

hive;

use healthdatabase;

Select \* from Hospital\_master;

select \* from Hospital\_TimelyEffectiveCare;

select \* from HQI\_HOSP\_MSPB;

**Query format**

select PhoneNumber, Address, Emergency, Outpatient from

========

hadoop fs -put /home/Power2Power/DataSets/HospitalGeneralInformation.csv Power2Power/datafiles

hadoop fs -put /home/Power2Power/DataSets/FootnoteCrosswalk.csv Power2Power/datafiles

hadoop fs -put /home/Power2Power/DataSets/HCAHPSHospital.csv Power2Power/datafiles

hadoop fs -put /home/Power2Power/DataSets/HealthcareAssociatedInfectionsHospital.csv Power2Power/datafiles

hadoop fs -put /home/Power2Power/DataSets/MeasureDates.csv Power2Power/datafiles

hadoop fs -put /home/Power2Power/DataSets/MedicareHospitalSpendingPatientHospital.csv Power2Power/datafiles

hadoop fs -put /home/Power2Power/DataSets/PaymentValueCareHospital.csv Power2Power/datafiles

======

**Oct 24- Latest**

drwxr-xr-x - cloudera cloudera 0 2015-10-24 08:32 /user/cloudera/hospitalq/datafiles

[cloudera@quickstart datafiles]$ hadoop fs -put /home/cloudera/hospitalq/datafiles/\*.\* /user/cloudera/hospitalq/datafiles/

[cloudera@quickstart datafiles]$ hadoop fs -ls /user/cloudera/hospitalq/datafiles

Found 5 items

-rw-r--r-- 1 cloudera cloudera 13782271 2015-10-24 08:33 /user/cloudera/hospitalq/datafiles/hospitalReadmit.csv

-rw-r--r-- 1 cloudera cloudera 21658463 2015-10-24 08:33 /user/cloudera/hospitalq/datafiles/hospitalcompare.csv

-rw-r--r-- 1 cloudera cloudera 823170 2015-10-24 08:33 /user/cloudera/hospitalq/datafiles/hospitalinfo.csv

-rw-r--r-- 1 cloudera cloudera 526750 2015-10-24 08:33 /user/cloudera/hospitalq/datafiles/hospitalopvolume.csv

-rw-r--r-- 1 cloudera cloudera 36754001 2015-10-24 08:33 /user/cloudera/hospitalq/datafiles/hospitaltimelyeffective.csv

[cloudera@quickstart datafiles]$ hive

Logging initialized using configuration in jar:file:/usr/jars/hive-common-1.1.0-cdh5.4.2.jar!/hive-log4j.properties

WARNING: Hive CLI is deprecated and migration to Beeline is recommended.

hive> create database healthq;

OK

Time taken: 2.29 seconds

hive>

create table hospitalmaster(providerID string,hospitalName string,city string,state string,zipcode string,countyName string,phoneNumber string,hospitalType string,emergencyServices string) row format delimited fields terminated by ",";

create table hospitaltimelyeffectivecare(providerID string,hospitalName string,condition string,measureId string,measureName string,score string,sample string,footnote string,mesaureStartDate string, measureEndDate string) row format delimited fields terminated by ",";

create table hospitalopvolume (providerID string,hospitalName string,measureId string,gastrointestinal string,eye string,nervousSystem string,musculoskeletal string,skin string,genitourinary string,cardiovascular string, startDate string, end\_dDate string) row format delimited fields terminated by ",";

create table hospitalcompare (providerID string,measureName string,measureId string,comparedToNational string,score string, footnote string,measureStartDate string, measureEndDate string) row format delimited fields terminated by ",";

create table hospitalreadmit (providerID string,measureName string,measureId string,comparedToNational string,denominator string, score string,lowerEstimate string, higherEstimate string) row format delimited fields terminated by ",";

--drop table hospitalcompare drop table hospitalreadmit;

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitalinfo.csv' OVERWRITE INTO TABLE hospitalmaster;

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitaltimelyeffective.csv' OVERWRITE INTO TABLE hospitaltimelyeffectivecare;

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitalopvolume.csv' OVERWRITE INTO TABLE hospitalopvolume;

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitalcompare.csv' OVERWRITE INTO TABLE hospitalcompare;

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitalReadmit.csv' OVERWRITE INTO TABLE hospitalreadmit;

PIG:

pig >

grunt>

HMaster = LOAD '/user/hive/warehouse/healthq.db/hospitalmaster' USING PigStorage() AS (providerID:chararray, hospitalName:chararray,city:chararray,state:chararray,zipcode:chararray,countyName:chararray,phoneNumber:chararray,hospitalType:chararray,emergencyServices:chararray);

dump HMaster;

HTimely = LOAD '/user/hive/warehouse/healthq.db/hospitaltimelyeffectivecare' USING PigStorage() AS (providerID:chararray,hospitalName:chararray,condition:chararray,measure\_id:chararray,measureName:chararray,score:chararray,sample:chararray,footnote:chararray,mesaureStartDate:chararray, measureEndDate:chararray)

HOPVolume= LOAD '/user/hive/warehouse/healthq.db/hospitalopvolume' USING PigStorage() AS (providerID:chararray,hospitalName:chararray,measureId:chararray,gastrointestinal:chararray,eye:chararray,nervousSystem:chararray,musculoskeletal:chararray,skin:chararray,genitourinary:chararray,cardiovascular:chararray, startDate:chararray, endDate:chararray);

HCompare= LOAD '/user/hive/warehouse/healthq.db/hospitalcompare' USING PigStorage() AS (providerID:chararray,measureName:chararray,measureId:chararray,comparedToNational:chararray,score:chararray, footnote:chararray,measureStartDate:chararray, measureEndDate:chararray);

HReadmit = LOAD '/user/hive/warehouse/healthq.db/hospitalreadmit' USING PigStorage() AS (provider\_ID:chararray,measureName:chararray,measureId:chararray,comparedToNational:chararray,denominator:chararray, score:chararray,lowerEstimate:chararray, higherEstimate:chararray);

==================

Oct 24- Latest

cd /user/cloudera/hospitalq/datafiles

$ hadoop fs -put /home/cloudera/hospitalq/datafiles/\*.\* /user/cloudera/hospitalq/datafiles/

hadoop fs -ls /user/cloudera/hospitalq/datafiles

$ hive

hive> create database healthq;

create table hospitalmaster (providerID string,hospitalName string,streetName string,city string,state string,zipcode string,countyName string,phoneNumber string,hospitalType string,emergencyServices string) row format delimited fields terminated by "\t";

create table hospitaltimelyeffectivecare(providerID string,hospitalName string,condition string,measureId string,measureName string,score string,sample string,footnote string,mesaureStartDate string, measureEndDate string) row format delimited fields terminated by ",";

create external table hospitalopvolume (providerID string,hospitalName string,measureId string,gastrointestinal string,eye string,nervousSystem string,musculoskeletal string,skin string,genitourinary string,cardiovascular string, startDate string, endDate string) row format delimited fields terminated by ",";

create table hospitalcompare (providerID string,measureName string,measureId string,comparedToNational string,score string, footnote string,measureStartDate string, measureEndDate string) row format delimited fields terminated by ",";

create table hospitalreadmit (providerID string,measureName string,measureId string,comparedToNational string,denominator string, score string,lowerEstimate string, higherEstimate string) row format delimited fields terminated by ",";

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitalinfo.csv' OVERWRITE INTO TABLE hospitalmaster;

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitaltimelyeffective.csv' OVERWRITE INTO TABLE hospitaltimelyeffectivecare;

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitalopvolume.csv' OVERWRITE INTO TABLE hospitalopvolume;

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitalcompare.csv' OVERWRITE INTO TABLE hospitalcompare;

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitalReadmit.csv' OVERWRITE INTO TABLE hospitalreadmit;

PIG:

pig >

grunt>

HMaster = LOAD '/user/hive/warehouse/healthq.db/hospitalmaster' USING PigStorage(',') AS (providerID:chararray, hospitalName:chararray,streetName:chararray,city:chararray,state:chararray,zipcode:chararray,countyName:chararray,phoneNumber:chararray,hospitalType:chararray,emergencyServices:chararray);

dump HMaster;

HTimely = LOAD '/user/hive/warehouse/healthq.db/hospitaltimelyeffectivecare' USING PigStorage(',') AS (providerID:chararray,hospitalName:chararray,condition:chararray,measureId:chararray,measureName:chararray,score:chararray,sample:chararray,footnote:chararray,mesaureStartDate:chararray, measureEndDate:chararray)

HOPVolume= LOAD '/user/hive/warehouse/healthq.db/hospitalopvolume' USING PigStorage(',') AS (providerID:chararray,hospitalName:chararray,measureId:chararray,gastrointestinal:chararray,eye:chararray,nervousSystem:chararray,musculoskeletal:chararray,skin:chararray,genitourinary:chararray,cardiovascular:chararray, startDate:chararray, endDate:chararray);

HCompare= LOAD '/user/hive/warehouse/healthq.db/hospitalcompare' USING PigStorage(',') AS (providerID:chararray,measureName:chararray,measureId:chararray,comparedToNational:chararray,score:chararray, footnote:chararray,measureStartDate:chararray, measureEndDate:chararray);

HReadmit = LOAD '/user/hive/warehouse/healthq.db/hospitalreadmit' USING PigStorage(',') AS (provider\_ID:chararray,measureName:chararray,measureId:chararray,comparedToNational:chararray,denominator:chararray, score:chararray,lowerEstimate:chararray, higherEstimate:chararray);

ZipCodeResult = FILTER HMaster by zipcode == '35631';

\*\*\*\*\*\*\*\*\*\*\*\*\*

**DEBUGGING WORKING VERSION**

=====

**hive**

create database BookDB;

use BookDB;

create table hospitalmaster(providerID string,hospitalName string,streetName string, city string,state string,zipcode string,countyName string,phoneNumber string,hospitalType string,emergencyServices string) row format delimited fields terminated by ",";

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/Book1.csv' OVERWRITE INTO TABLE hospitalmaster;

**PIG:**

HMaster = LOAD '/user/hive/warehouse/bookdb.db/hospitalmaster' USING PigStorage(‘,’) AS (providerID:chararray,hospitalName:chararray,streetName:chararray,city:chararray,state:chararray,zipcode:chararray,countyName:chararray,phoneNumber:chararray,hospitalType:chararray,emergencyServices:chararray);

dump HMaster;

ZipCodeResult = FILTER HMaster by zipcode == '35631';

dump ZipCodeResult;

B = FOREACH A generate hospitalName;

dump B

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Oct 25 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

create table hospitalmaster (providerID string,hospitalName string,streetName string,city string,state string,zipcode string,countyName string,phoneNumber string,hospitalType string,hospitalownership string, emergencyServices string) row format delimited fields terminated by ",";

create table hospitaltimelyeffectivecare(providerID string,hospitalName string,condition string,measureId string,measureName string,score string,sample string,footnote string,mesaureStartDate string, measureEndDate string) row format delimited fields terminated by ",";

create table hospitalopvolume (providerID string,hospitalName string,measureId string,gastrointestinal string,eye string,nervousSystem string,musculoskeletal string,skin string,genitourinary string,cardiovascular string, startDate string, endDate string) row format delimited fields terminated by ",";

create table hospitalcompare (providerID string,measureName string,measureId string,comparedToNational string,score string, footnote string,measureStartDate string, measureEndDate string) row format delimited fields terminated by ",";

create table hospitalreadmit (providerID string,measureName string,measureId string,comparedToNational string,denominator string, score string,lowerEstimate string, higherEstimate string) row format delimited fields terminated by ",";

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitalinfo.csv' OVERWRITE INTO TABLE hospitalmaster;

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitaltimelyeffective.csv' OVERWRITE INTO TABLE hospitaltimelyeffectivecare;

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitalopvolume.csv' OVERWRITE INTO TABLE hospitalopvolume;

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitalcompare.csv' OVERWRITE INTO TABLE hospitalcompare;

LOAD DATA INPATH '/user/cloudera/hospitalq/datafiles/hospitalreadmit.csv' OVERWRITE INTO TABLE hospitalreadmit;

PIG:

pig >

grunt>

HMaster = LOAD '/user/hive/warehouse/healthq.db/hospitalmaster' USING PigStorage(',') AS (providerID:chararray, hospitalName:chararray,streetName:chararray,city:chararray,state:chararray,zipcode:chararray,countyName:chararray,phoneNumber:chararray,hospitalType:chararray,hospitalownership:chararray,emergencyServices:chararray);

dump HMaster;

HTimely = LOAD '/user/hive/warehouse/healthq.db/hospitaltimelyeffectivecare' USING PigStorage(',') AS (providerID:chararray,hospitalName:chararray,condition:chararray,measureId:chararray,measureName:chararray,score:chararray,sample:chararray,footnote:chararray,mesaureStartDate:chararray, measureEndDate:chararray);

HOPVolume= LOAD '/user/hive/warehouse/healthq.db/hospitalopvolume' USING PigStorage(',') AS (providerID:chararray,hospitalName:chararray,measureId:chararray,gastrointestinal:chararray,eye:chararray,nervousSystem:chararray,musculoskeletal:chararray,skin:chararray,genitourinary:chararray,cardiovascular:chararray, startDate:chararray, endDate:chararray);

HCompare= LOAD '/user/hive/warehouse/healthq.db/hospitalcompare' USING PigStorage(',') AS (providerID:chararray,measureName:chararray,measureId:chararray,comparedToNational:chararray,score:chararray, footnote:chararray,measureStartDate:chararray, measureEndDate:chararray);

HReadmit = LOAD '/user/hive/warehouse/healthq.db/hospitalreadmit' USING PigStorage(',') AS (provider\_ID:chararray,measureName:chararray,measureId:chararray,comparedToNational:chararray,denominator:chararray, score:chararray,lowerEstimate:chararray, higherEstimate:chararray);

compareMasterResult = JOIN HCompare by providerID, HMaster by providerID

ZipCodeResult = FILTER HMaster by zipcode == '35631';

VT4Result= FILTER HTimely by measureId == 'VTE\_4';

A = FOREACH HMaster generate hospitalName;

procedureResult = FILTER HOPVolume by providerID == '010012';

--Dee Notes

create table hospitalmaster (providerID string,hospitalName string,streetName string,city string,state string,zipcode string,countyName string,phoneNumber string,hospitalType string,hospitalownership string, emergencyServices string) row format delimited fields terminated by ",";

create table hospitaltimelyeffectivecare(providerID string,hospitalName string,condition string,measureId string,measureName string,score string,sample string,footnote string,mesaureStartDate string, measureEndDate string) row format delimited fields terminated by ",";

create table hospitalopvolume (providerID string,hospitalName string,measureId string,gastrointestinal tinyint,eye tinyint,nervousSystem tinyint,musculoskeletal tinyint,skin tinyint,genitourinary tinyint,cardiovascular tinyint, startDate string, endDate string) row format delimited fields terminated by ",";

create table hospitalcompare (providerID string,measureName string,measureId string,comparedToNational string,score tinyint, footnote string,measureStartDate string, measureEndDate string) row format delimited fields terminated by ",";

--Dee : This file wasnt matching with the datasets, few columns were missing, added 3 columns

create table hospitalreadmit (providerID string,hospitalname string,measureName string,measureId string,comparedToNational string,denominator string, score string,lowerEstimate string, higherEstimate string,footnote string,measurestart string,measureend string) row format delimited fields terminated by ",";