**Certificate in Big Data and Hadoop Class Project**

This project is to look at US efforts to reduce Heathcare-Associated Infections (HAI) in Year 2013.

4 measures are included:

hospital-acquired central line-associated bloodstream infections (CLABSI)

Methicillin-resistant Staphylococcus aureus (MRSA)

catheter-associated urinary tract infections (CAUTI)

Clostridium difficile (C.diff.)

**Data source**: Medicare.gov, Official Hospital Compare Data

These are the steps taken to process the data

**Step 1**: Download data source as csv files

HAI\_Hospital.csv

HAI\_State.csv

HAI\_State\_Prior.csv

**Step 2**: mkdir directory medicare. Copy files into HDFS

**Step 3**: Use Hive to create database medicare

Run LoadStagingTables.hsql

This is to create and load staging tables:

hai\_hospital\_stg

hai\_state\_stg

Hai\_state\_prior\_yr\_stg

**Step 4**: Run Create\_Load\_tables.hsql

This is to:

* + Create tables hai\_hosptial and populate from hai\_hospital\_stg table

filtering on measureid like '%SIR' (only include standardized infection ratio) and score <> 'Not Available', partitioned by measureid

* + Create table hai\_hospital\_cnt\_by\_state and populate from hai\_hospital\_stg table group by state to get the number of participating hospitals
  + Create table hai\_state and populate from hai\_state\_stg filtering on measureid like '%SIR' (only include standardized infection ratio) and score <> 'Not Available'
  + Create table hai\_state\_prior\_yr and populate from hai\_state\_prior\_yr\_stg filtering on score <> 'Not Available'

**Step 5**: Run Hai\_Analysis

This is to create table hai\_analysis and populate by joining hai\_state, hai\_hospital\_cnt\_by\_state and hai\_state\_prior\_yr to create one dataset for data visualization.

**Step 6**: Use sqoop to export hai\_analysis to mysql:

sqoop export -m 1 --connect jdbc:mysql://192.165.155.1/medicare --username root --password root --table HAI\_Analysis --export-dir /user/cloudera/medicare/hai\_analysis --input-fields-terminated-by '\t'

Example of data visualization:

This chart is comparing different HAI – SIR score for50 states.

