Analyzing machine & sensor data

Sensor Data: A sensor is a device that measures a physical quantity and transforms it into a digital signal. Sensors are always on, capturing data at a low cost, and powering the “Internet of Things.”

Potential Uses of Sensor Data: Sensors can be used to collect data from many sources, such as: ✓ To monitor machines or infrastructure such as ventilation equipment, bridges, energy meters, or airplane engines. This data can be used for predictive analytics, to repair or replace these items before they break. ✓ To monitor natural phenomena such as meteorological patterns, underground pressure during oil extraction, or patient vital statistics during recovery from a medical procedure.

This case study is about how to refine data from heating, ventilation, and air conditioning (HVAC) systems using the Cloudera Data Platform, and how to analyze the refined sensor data to maintain optimal building temperatures.

Input Data

In this case study, we will focus on sensor data from building operations. Specifically, we will refine and analyze the data from Heating, Ventilation, Air Conditioning (HVAC) systems in 20 large buildings around the world

In order to perform analysis, we will use the below data as input data.

✓ HVAC.csv – contains the targeted building temperatures, along with the actual (measured) building temperatures. The building temperature data was obtained using Apache Flume. Flume can be used as a log aggregator, collecting log data from many diverse sources and moving it to a centralized data store. In this case, Flume was used to capture the sensor log data, which we can now load into the Hadoop Distributed File System (HFDS).

✓ building.csv – contains the “building” database table. Apache Sqoop can be used to transfer this type of data from a structured database into HFDS.

High Level Steps:

1. Download and extract the sensor data files(HVAC.csv & building.csv).

2. Load the sensor data into the HDFS.

3. Create HVAC & Building tables in Hive or Pig

4. Using Hive/pig/impala to refine the sensor data.

5. Calculate three variables (temp\_diff, temprange, extremetemp) in Hvac table

✓ Temp\_didff = actual temperature – target temperature

✓ temprange column indicates whether the actual temperature was:

✓ NORMAL – within 5 degrees of the target temperature.

✓ COLD – more than five degrees colder than the target temperature.

✓ HOT – more than 5 degrees warmer than the target temperature.

✓ Extrememetemp: If the temperature is outside of the normal range, extremetemp is assigned a value of 1; otherwise its value is 0. 6. Create final file by joining the two tables and perform the analysis

Expected Output: We would like to accomplish three goals with this data: ✓ Reduce heating and cooling expenses. ✓ Keep indoor temperatures in a comfortable range between 65-70 degrees. ✓ Identify which HVAC products are reliable, and replace unreliable equipment with those models.

These analysis will be very helpful for facilities department to initiate data-driven strategies to reduce energy expenditures and improve employee comfort.

Analysis need to be performed: 1. Data visualization/analysis by mapping the buildings that are most frequently outside of the optimal temperature range. Calculate count of extremetemp (i.e where the temperature was more than five degrees higher or lower than the target temperature) by each country and temprange 2. Which country offices run hot (Hot offices can lead to employee complaints and reduced productivity) and which offices run cold (Cold offices cause elevated energy expenditures and employee discomfort). Calculate count of offices run in hot and count of office run in cold by country. 3. Our data set includes information about the performance of five brands of HVAC equipment, distributed across many types of buildings in a wide variety of climates. We can use this data to assess the relative reliability of the different HVAC models(i.e We can see that the which model seems to regulate temperature most reliably and maintain the appropriate temperature range). Calculate count of extreamtemp by hvacproduct.

Sample outputs:-







