**Apache Pig Use Case - Electrical Bulb Testing**

In this part of the e-book, we will work on a case involving electric bulbs and work with the date and time concepts in Pig.

In this example, Pig is used in the local mode to load the local data. We can use Pig in HDFS mode as per our convenience.

**DATASET DESCRIPTION**

The dataset used in this case is a sample from a light bulb production house where bulbs are tested at random intervals of time. The first column is **StartDate** which is the date and time when the testing of the bulb started, and the second column is **EndDate** which is the date when the testing ended.

A few rows may be empty that indicates that data is not available, this may be due to various reasons. But as developers, we need not worry about the missing data. With the help of Data Filtering, we can remove the unnecessary data.

**Loading Data into the Pig Environment and Processing**

**Step 1** -> Since Pig uses default as tab(\t) delimited data, it is not mandatory to state USING PigStorage('\t’) in the code while loading, nevertheless, it is good practice to write it.

**Step 2** -> Now that we have data inside Pig, the first step is to filter data in the column we are working on.Here, we remove all the rows with null data.

**Step 3 ->** Now that we have data inside Pig, the first step is to filter data in the column we are working on.Here, we remove all the rows with null data.

Now that we have data inside Pig, the first step is to filter data in the column we are working on. Here, we remove all the rows with null data.

**Here, we use two predefined functions.**

ToDate()

MinutesBetween()

**Where:**

ToDate()- Converts **the character array to date**-**time** readable **structure** that can be interpreted **by** Pig MinutesBetween() - Takes **the difference between any** two **date**-**time parameters** provided

The ToDate function can be used in different formats of a year, month, and date. Example

YYYY-MM-DD

DD/MM/YYYY

DD-YY-MM

**Sample Output:**

**Case1** -> Result of lightbulb testing that displays the time interval for which a lightbulb stays “switched on”



**Case2** -> find the maximum or minimum time a bulb can stay“switched on” among other things.