



Bulk Loading of Data into HBase Tables

Until now we must have got use to the idea of loading data into HBase tables cell-by-cell that is we first use a row key and then we consider a column family under a column family we choose a particular column and then we try to insert the cell value or simply the column data into that particular column family in a corresponding column.

However in reality we're not going to be using this method because HBase is going to be mainly used in cases where table size is extremely large that is we have millions of columns literally it not practically and billions of rows which most of the cases on big data would be the case.

Now how do we load data into HBase tables when we are dealing with such large volumes of data? We have to use the options available for bulk loading of data. There are some proprietary tools to actually move data into HBase tables however from the most commonly used method to load data into HBase tables

We're going to use MapReduce that is write a Java Program to load the data which is present either in the local file system or in the HDFS into your HBase tables. The much



Simpler and easier way and the most sort after way of loading the data when compared to the Java MapReduce is going to be PIG. So how do we use PIG to actually load data into HBase tables. If the data is going to be there in local file system or if it's going to be there in your HDFS we can use PIG to load data into your HBase tables. We shall actually see a demo of that.

Another very interesting and most commonly used way of loading data into HBase tables would be using Sqoop. Now this would be the case when we already have data in our RDBMS tables. So when we have the data in our RDBMS data base servers, we can directly load the data from the tables into HBase using Sqoop.