[Build an application for reporting and analysis on Hadoop framework](http://stackoverflow.com/questions/27889368/build-an-application-for-reporting-and-analysis-on-hadoop-framework)

**Problem Statement:**

I have an application with SAS where I pull the data from Oracle and produce report to excel using Base SAS and SAS macros. Now the problem is day by day my database is getting huge and fetching data from Oracle is taking more time, as a result my jobs are running slow.

So I want my application to be built on Hadoop for Reporting and analysis purpose. Can someone please suggest me any approach and what are the tools I need to use for this.

The short answer is: it depends.

For unloading data from Oracle I would recommend you to use Sqoop (<http://sqoop.apache.org/>), it is designed for this specific use case and can even do incremental loads and can create Hive table for unloaded data

When the data is unloaded, you can use Impala to build the report you need. Impala can natively work with Hive tables, so the sings are really simple. Of course, you would have to rewrite your SAS code to a set of SQL statements that would run on top of Impala.

Next, if you need visualization tool to run on top of it, you can either try something like Tableau or any other tool that is capable of utilizing ODBC/JDBC to connect to Impala

Finally, I think Hadoop + Sqoop + Impala would cover your needs. But I'd recommend you also to take a look at the MPP databases, because using SAS means you have pretty structured data and MPP database would be a better fit for this case

My understanding:

Project 1:

1. Import table data form db using Sqoop
2. Process a data and build hive tables on top of data
3. Use any reporting tool to analyze data(here I took care of first 2 steps- for step 3 we had different team)
4. Some time we also use to analyze weblog data

(https://www.rittmanmead.com/blog/2014/04/simple-data-manipulation-and-reporting-using-hive-impala-and-cdh5/)

To understand step 4 use above link

-it is using hive command with Regx to process data.

Project 2:

We use to get weblogs and we were suppose to analyze

* We used spark->create hive table and run analytics on that.