



Problem Statement

Input Dataset: Input is a MySQL database table which is getting data from some old traditional system which can't be changed.

Problem Statement: Assume that MySQL table has huge data and this cannot be processed using MySQL. We will need to put this data in HDFS and then process the data using Hive. Once the data is processed by creating Hive external table and running HiveQL, processed data need to be put into MySQL back for reporting purpose.

We will use Sqoop for import and export of data from and to MySQL and HIVE for doing the processing.

Output: To find out total volume for each stock_symbol in each quarter. Output table would look like this:

```
mysql> desc quarter_stock_volumes;
```

Field	Type	Null	Key	Default	Extra
quarter	int(11)	YES		NULL	
stock	varchar(35)	YES		NULL	
volume	bigint(20)	YES		NULL	

3 rows in set (0.00 sec)