**HP Inc . project.**

**Requirement Analysis:**

This project is about identification and reduction of causes for defects in the printers by analyzing data from returned printer units. ( HP Inc. has several manufacturing locations under its belt which manufacture thousands of printers of various models and types with adherence to specific industry standards.)

**Purpose:**

To move/migrate the whole system from RDBMS to HDFS System for faster analysis of data and to detect the exact cause and region where the defect is being originated.

**Purpose of data migration:**

1. The whole Printer defective software tracking system needs to be moved from RDBMS system to HDFS System as the whole data was becoming huge and difficult to manage live data coming from the units such as Manufacturing location to Warehouse, Warehouse to reseller and reseller to the customer from where the printer’s defective reports are received.
2. The data that contains the damage details of the printer usually comes from various sources which needs to be updated live into HDFS System.
3. Analyzing patterns like increase in defective units from one manufacturing location or defective units being reported from lots shipped by particular shipping company will help find probable root cause of defects.
4. **The project has three important Big Data Components which can be used for data migration**:

* Moving Data from Traditional RDMS to HDFS (SQOOP)
* Storing Live data from resellers to HDFS (FLUME with Spark streaming)
* Querying the data from HDFS using HIVE to provide useful analytical information for the decision makers. (HIVE finally connected to ClickView)

**How to reduce defect or damage reduction rate:**

1. Manufacturing locations send details of the units manufactured, the lot codes, number of units, the shipping companies used to ship the units from Manufacturing locations to Warehouses. Warehouses in turn send data related to the units with respect to where the units were shipped from, how many days they are stocked in inventory, which specific reseller they were shipped to and the shipping companies used for the to and fro shipment.
2. So this vast amount of data from Manufacturing locations, Warehouses and Shipping companies that have been stored in the RDMS is moved to HDFS which will intern provide better performance for querying and analyzing in detecting the root cause of the defect’s origin and for the remediation of the same.