Basically we are dealing with mostly with the sales data. With this sales data, we are generating the data based on weekly sales, monthly sales, half yearly sales and yearly sales.

In ikea we are dealing with 550 stores across the world. They baically deals with the furnitures, home decors kitchen untensils and many more.

All the stores will send the data in the form of the flat files. Each stores will be having access to one shared path where they will keep the sales files beofre they are going to close and also they have fixed cut off time (2.30 AM CET==6 IST) before that each store has to place the file in the shared path. There is one designed oozie workflow which runs at 3.30 AM UTC( 7 IST), in that there are 4 actions present.

1. Reading the files:- Since all the stores has placed the file in the shared path, a job will run which reads the files seqeuntly and inserts the data in the Oracle DB.
2. SQOOP:- Once reading is completed from the files and all the data has been inserted in the oracle DB, immediately SQOOP incremental jobs will be triggered, which takes the data from Oracle DB and copy the data in HDFS will the replication Factor 3
3. HiveQL:- Once the data is moved into HDFS, which we will insert those records in the HIVE managed/temporary table.
4. Deletion from Oracle DB:- Once the data is copied into Hive, we will delete the records from Oracle DB because that is space occuiped and also we have the replication factor as 3 which means the data is already distributed to multiple nodes.

Once the process the completes, we will use the managed tables to insert those data into Hive partioned table which is partitioned on year, month, week number, day wise for each stores.

Partioning will be easy to get the data particularly store wise.

Once the data is stored permanently in hive partitioned tables, we will start processing using SPARK SQL, because for processing SPARK-SQL is very faster compare to HIVE.

Now on the hive partioned table, we will perform the aggregation based on week number of the particular store

* Select sto\_no, sum(ART\_VAL\_SALES) from TABLE\_NAME group by(sto\_no, week\_no)

Once we generate the report for weekly sales, we will compare the current week data and the previous week data, so that we can we can get to know histogram of the sales for the particular stores.

Once the months end, we will be generating the report on the month wise on the first day of the next month for the sales and we will send the data to BI team to generate the report and pass the report to client.