**Data Accessing:**

* Data : file://10.201.71.125/shared/COMMON/COPY&DELETE/CI\_Data/
* Open the above link in run command so that it will open in folder mode
* Check that you have access to the shared path
* Extract the Ticket num and Summary(Ticket Description) from the file name starts with CI (Cognitive Insight)

**Data Pre-processing**

* NLP on the ticket summary like remove stop words and cleaning
* Ticket ID might be same for different accounts
* Based on the ticket keywords mapping sheet, create N-grams with count.

**Others:**

Data load (.Net) plan to replace java/Node/angular JS

We are handling excel files

MySQL data base SQL storage

Analytics tool-Qlikview

Aggregation function is used in the database tables.

Keyword mapping will give by pre sales team

We are taking the data from the table, they have added 3 new columns also

Table Name- All grams

Record size= 50K, we have to increase to 2 lakhs

Data load takes 3 to 7 days since it is mapping linearly

Excel to csv so that 50 K to 2 lakhs records can be handles since large account have high number of tickets.

In application, key words will be different.

500+ classification sets

We have to use domain keyword check

Based on the vertical, we can drill down the search.

We will go for PG (Postgre)

ADHOC reporting, user--> report.