**PROBLEM STATEMENT**

With the number of active cases on a rise in India, a predictive model can help predict the trend of cases going on and help find out when the rise will peak and number of cases will drop down. The social effects of COVID-19 are also seen. By taking this dataset we can analyse by using the pig and hive commands. Hive allow users to read , write , and manage petabytes of data using SQL . Hive is built on top of Apache Hadoop , which is an open-source framework used to efficiently store and process large datasets. As a result , Hive is closely integrated with Hadoop , and is designed to work quickly on petabytes of data. Hive provides an SQL dialect ,called Hive query Language. it’s an effective , reasonably intuitive model for organizing and using data. Pig is a high -level platform or tool which is used to process the large datasets. It provides a High-Level abstraction for processing over the MapReduce . It provides a high-level scripting language ,known as pig-Latin which is used to develop the data analysis codes.

HIVE COMMANDS

hive

create table corona(confirmed int,recovered int,active int newcase int,newdeaths int);

insert into corona values(10,6,5,2,1);

insert into corona values(45,8,6,22,1);

insert into corona values(45,6,8,3,2);

insert into corona values(51,5,9,3,5);

insert into corona values(60,4,6,2,8);

select \* from corona;

select recovered,newcase from corona;

select count(\*) from corona;

select sum(active) from corona;

select floor(recovered) from corona;

select sqrt(confirmed) from corona;

select max(newdeaths) from corona;

PIG COMMANDS

corona = LOAD '/home/hadoop/corona.txt' USING PigStorage(',') as (confirmed:int,recovered:int,active:int,newcase:int,newdeaths:int);

dump corona;

group\_data = GROUP corona by recovered;

dump group\_data;

order\_by\_data = ORDER corona BY confirmed DESC;

dump order\_by\_data;

corona2 = JOIN corona BY active, corona1 BY active;

dump corona2;

limit\_data = LIMIT corona 2;

dump limit\_data;









