Testing and Result

When placing data in one place, there is no limit to what we can do to benefit from it and analyze it efficiently. The result of this use will benefit all stakeholders, including farmers, scientists, and others, and enable them to make good decisions and provide creative solutions, which will definitely increase the efficiency of the agricultural process itself.

For example, we can see the total productivity in a given locality relative to a specific crop. We can also study the productivity of a specific type of soil in a rainy season. And other analyzes that open up endless possibilities to benefit from the data

In this chapter, we will present a number of these analyzes that we carried out on the data that we obtained after doing the disposal and storage operations in order to test the success of the platform's work mechanism. We'll use HQL in Hive View to do the visualization with Zeppelin.

HiveQL Sample Queries

1. Show the production, harvested and yield for crop

Quering the dataset to get data about production, yield and harvested areas for specific type of crops. We use the four following queries to accomplish this:

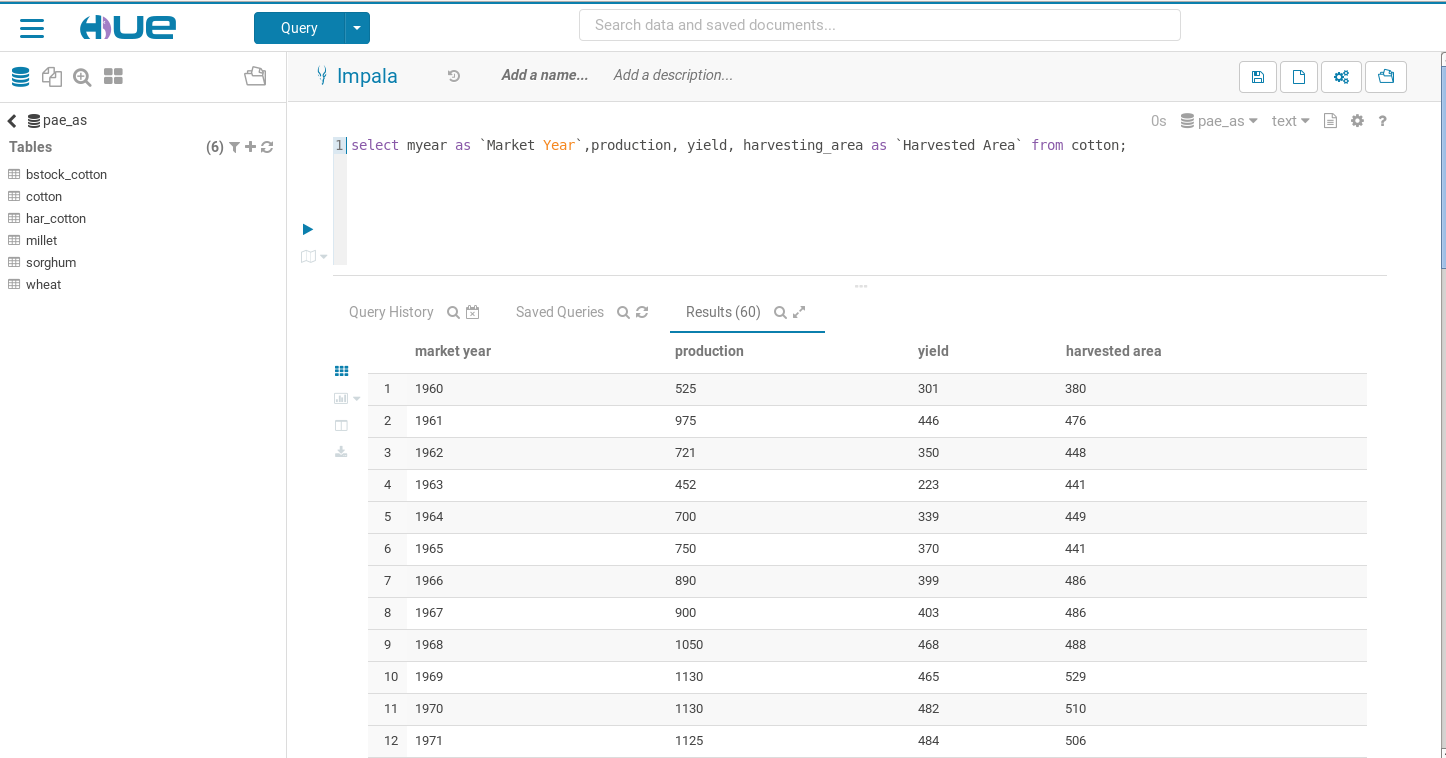
select myear as `Market Year`,production, yield, harvesting\_area as `Harvested Area` from cotton;

select myear as `Market Year`,production, yield, harvesting\_area as `Harvested Area` from wheat;

select myear as `Market Year`,production, yield, harvesting\_area as `Harvested Area` from sorghum;

select myear as `Market Year`,production, yield, harvesting\_area as `Harvested Area` from millet;

the result of the first query about cotton shown in figure ??



production, harvested and yield for cotton by yeas

1. Total production

Planning a country's food and financial security requires consideration of a number of factors, including the total production of food and cash crops. The following query is for the total production of the four selected crops. we select the total production and average for harvested area by year for all crops. And here we connect between three table

SELECT crop\_name,loc\_name, cc.mdate as years, sum(production) as total\_production, avg(harvested) as avg\_harvested

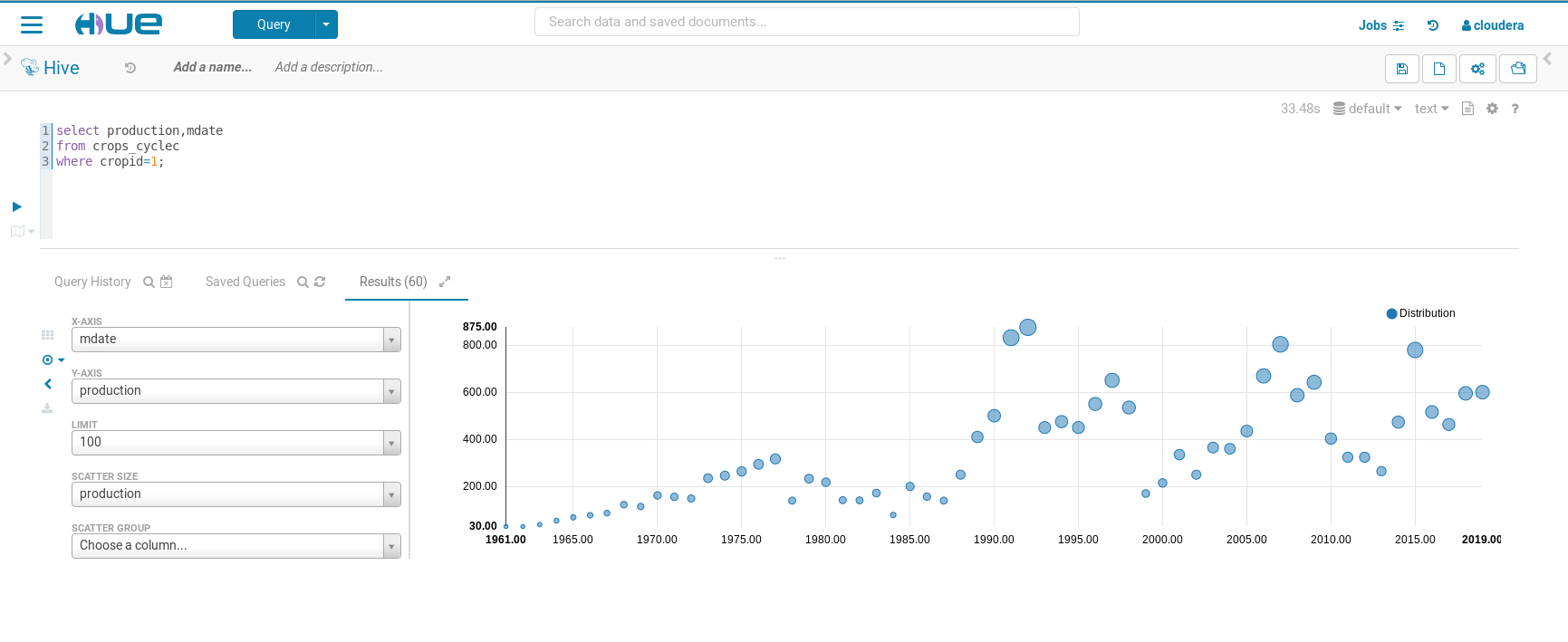
FROM `pae\_db`.`crops\_cyclec` cc

join crops c on c.cropid= cc.cropid

join locality l on l.localityid=cc.localityid

GROUP BY c.crop\_name,l.loc\_name,cc.mdate;

Total production and average of harvested for Khartom locality and wheat crop are shown in figure ??.



If we only want to to retrive the data about production and yield among the years we can use the following query:

select c.crop\_name,production,mdate,yield

from crops\_cyclec cc

join crops c on c.cropid=cc.cropid

group by c.crop\_name,production,yield,mdate;

production for one crop

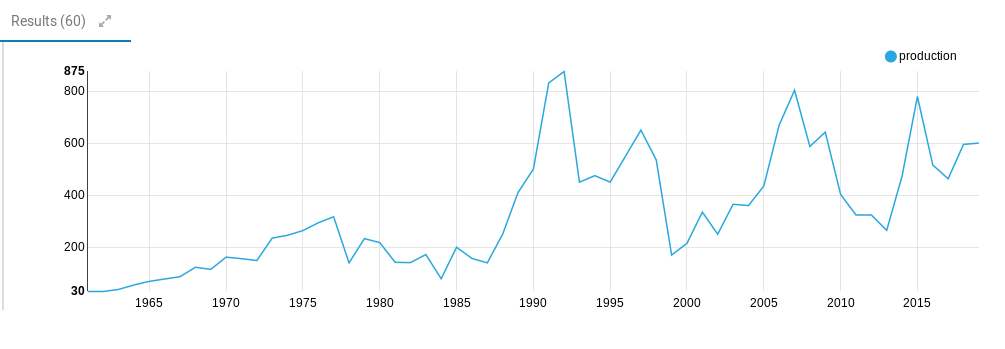
quering crop cycle table to get the production for specific crop can be done by the following query

select mdate,production

from crops\_cyclec

where cropid=1;

the result will be as shown in fig 0603 which show the production for wheat for year between 1960 to 2019. The fluctuation in the quantities produced over the years is evident



Change in production of wheat crop in the years 1960-2019

For the rest of the crops: sorghum, millet and cotton, the queries are as follows in same order

select mdate,production

from crops\_cyclec

where cropid=2;

select mdate,production

from crops\_cyclec

where cropid=3;

select mdate,production

from crops\_cyclec

where cropid=4;

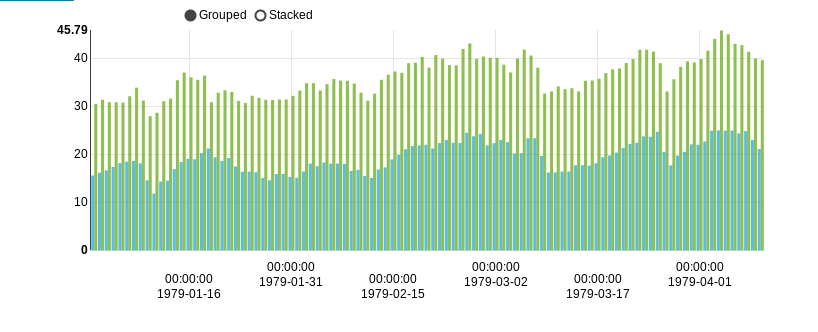
Temperature and Precipitation

The importance of knowing the climatic conditions comes in that it is considered a key to the success of the agricultural season. As temperatures affect the cultivation of wheat that are preferred to grow in low temperatures. As for the amount of rain, its importance is clearer for agriculture in the rainy sector, while this importance decreases somewhat in the irrigated sector. to show min and max temperature degree during the period from the first of January 1979 to the end of July 2014

select cast(rec\_date as date), min, max from temperature;

We circumvented Hive's restrictions on date fields by storing them as string fields and then using the cast function to convert them to date again at the query stage.

The fig 0604 shows us a partial scene of the upper and lower temperatures in the year 1979

High and low temperatures for some months in 1979

group temperature by year

Since most of the data used are using the years to explain the output and harvest, so we have used the following query to find the average temperatures for the specified year using the average function

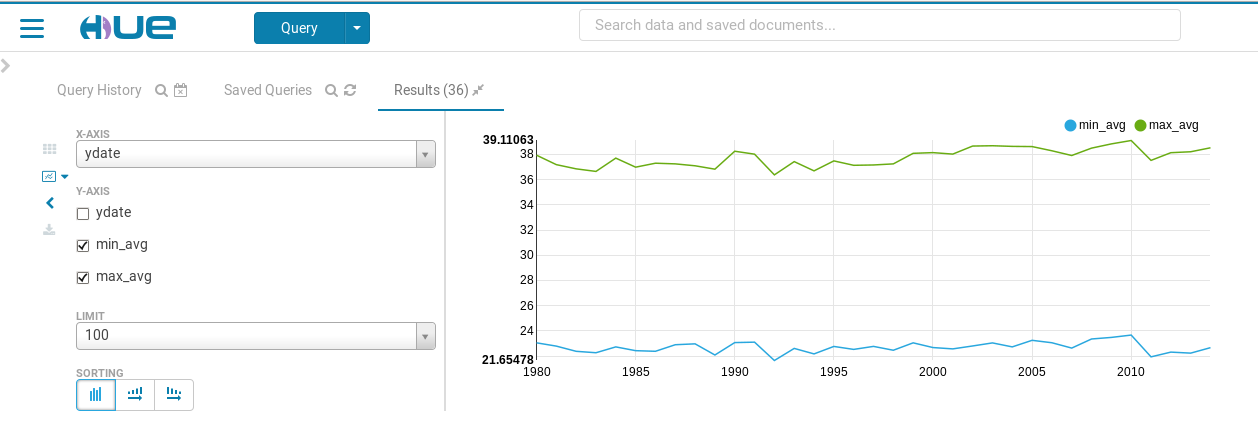
select year(cast(rec\_date as date)) as ydate, avg(min) as min\_avg, avg(max) as max\_avg

from temperature

group by year(cast(rec\_date as date))

limit 100;

The fig 0605 shows us the average of the upper and lower limits of the temperature for a number of years



Average high and low temperatures in the years 1980-2014

group rainfall by year

For rain, we also used the AVG function to calculate the amount of precipitation per year because our weather data give us values per day and we want them per year.

select year(cast(rec\_date as date)) as ydate, avg(precipitation)as precipitation\_avg

from rainfalls

group by year(cast(rec\_date as date));

We created the precipitation\_avg view for our need to use it later when comparing with production to show the impact of production on the amount of rain

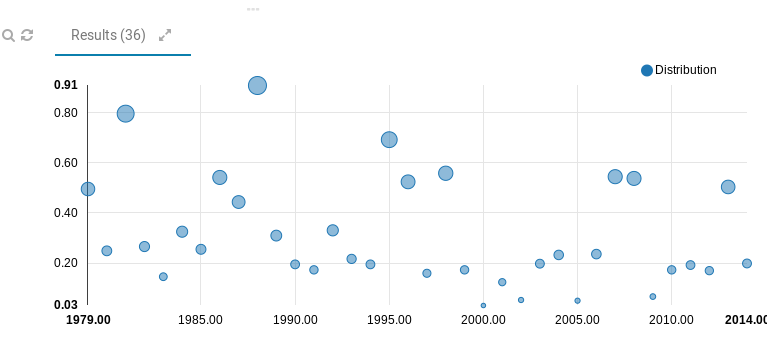
create view precipitation\_avg as

select year(cast(rec\_date as date)) as ydate, avg(precipitation)as precipitation\_avg

from rainfalls

group by year(cast(rec\_date as date));

The fig 0606 shows the average precipitation in the years shown



Average precipitation in years from 1980 to 2014

Also with regard to temperature, we have created a display to show average high and low grades throughout the study period so that we can use it later for comparison operations

create view tem\_year as

select year(cast(rec\_date as date)) as ydate, avg(min) min\_avg,avg(max) as max\_avg

from temperature

group by year(cast(rec\_date as date))

limit 100;

The effect of weather changes on production

The study of weather change helps to predict production. It can be used among several factors to help avoid the resulting gaps in the impact of weather on agriculture.

Production and temerature

In the following query, we use the mean and minimum average maximum temperatures per year compared to the production of millet in the same year:

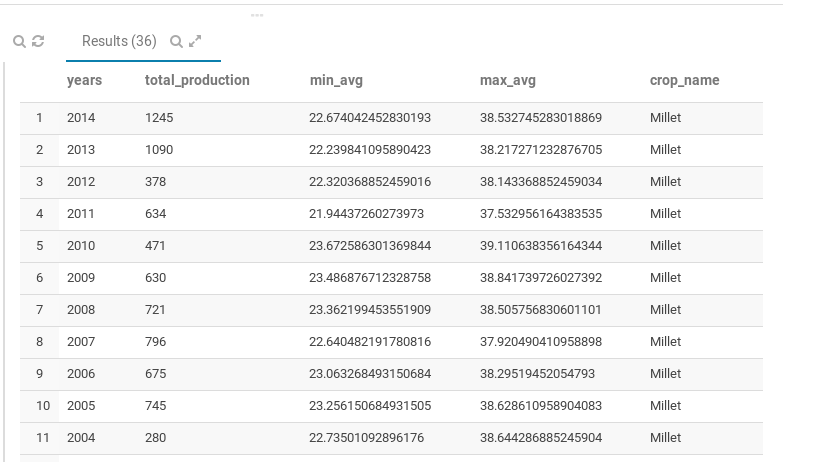
select years, total\_production,min\_avg,max\_avg, crop\_name

from totalproduction

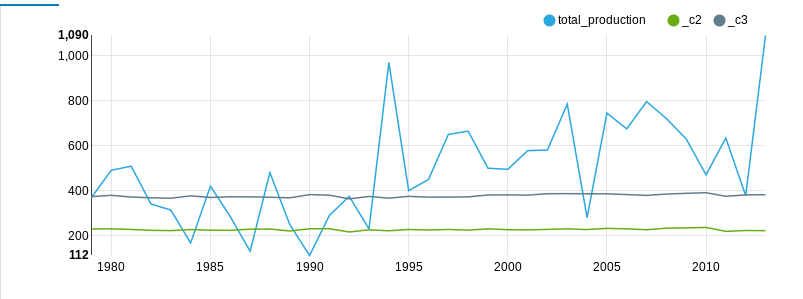
right join tem\_year on years=ydate

where crop\_name=”Millet”;

The previous query produces data that appears in the figure 0607 , which shows the year, production, minimum degree and maximum degree of millet yield. We even show if there is any indication of a change in temperature to produce millet



We used the data in the fig 0607 to produce the chart in the fig 0608. In order for the comparison to become clearer, we measured the temperatures by multiplying them by 10 only for the purpose of illustration to be closer to the production line in the chart.



Comparing millet production with temperatures in the period 1980-2014

Production and precipitation

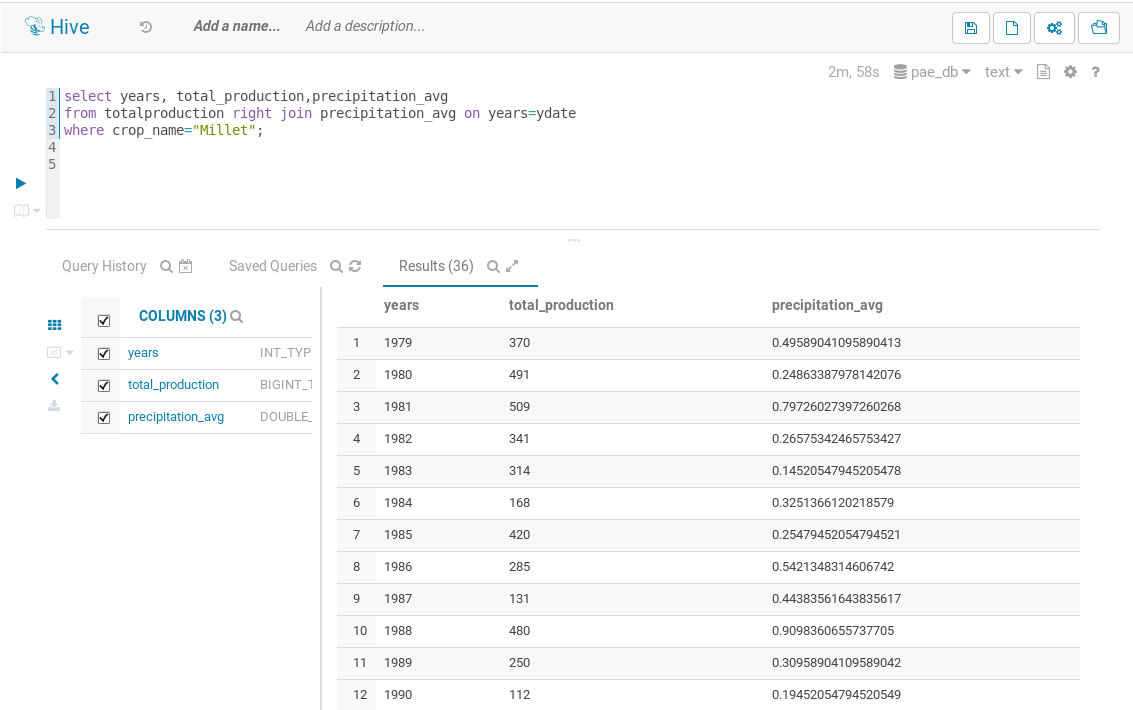
In the same way that we compare the relationship between millet production and temperatures in the following query, we compare the production of millet and the amounts of rain. And we have used the average for precipitation amount in year comparing to production of millet

select years, total\_production,precipitation\_avg

from totalproduction right join precipitation\_avg on years=ydate

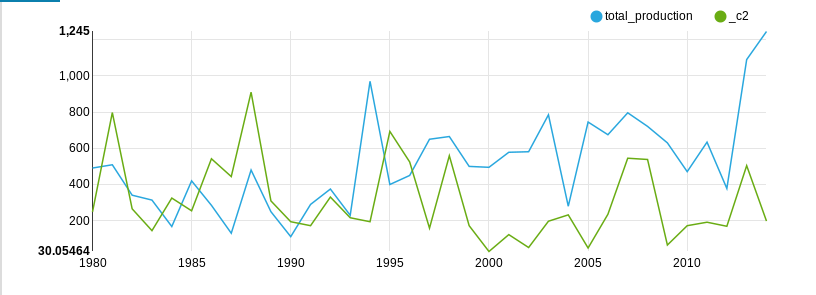
where crop\_name=”Millet”;

The fig0609 shows the result of the previous query



Total production and precipitation amount according to years

In the previous query, we obtained the total production and the average amount of precipitation according to the years from 1980 to 2014 for Millet in order to compare the amount of rain and total production for each year. It is shown in the fig 0610. We also measured the average amount of rain per year multiplied by 100 until the precipitation line came close to the production line in the chart to facilitate comparison.



A comparison between millet production and average precipitation for the years 1980-2014

Parametric Queries

Parametric queries provide us with the ability to customize queries to answer more than one in the same query. To write a parametric query in Hue, we use the dollar sign with curly braces. Then the query user must enter criteria to execute the query. The following query shows the total production and average harvest for a crop in a given locality. The parameters to be entered are the local name and the name of the crop to be queried.

SELECT crop\_name,

loc\_name,

cc.mdate AS years,

sum(production) AS total\_production,

avg(harvested) AS avg\_harvested

FROM `pae\_db`.`crops\_cyclec` cc

JOIN crops c ON c.cropid= cc.cropid

JOIN locality l ON l.loccalityid=cc.localityid

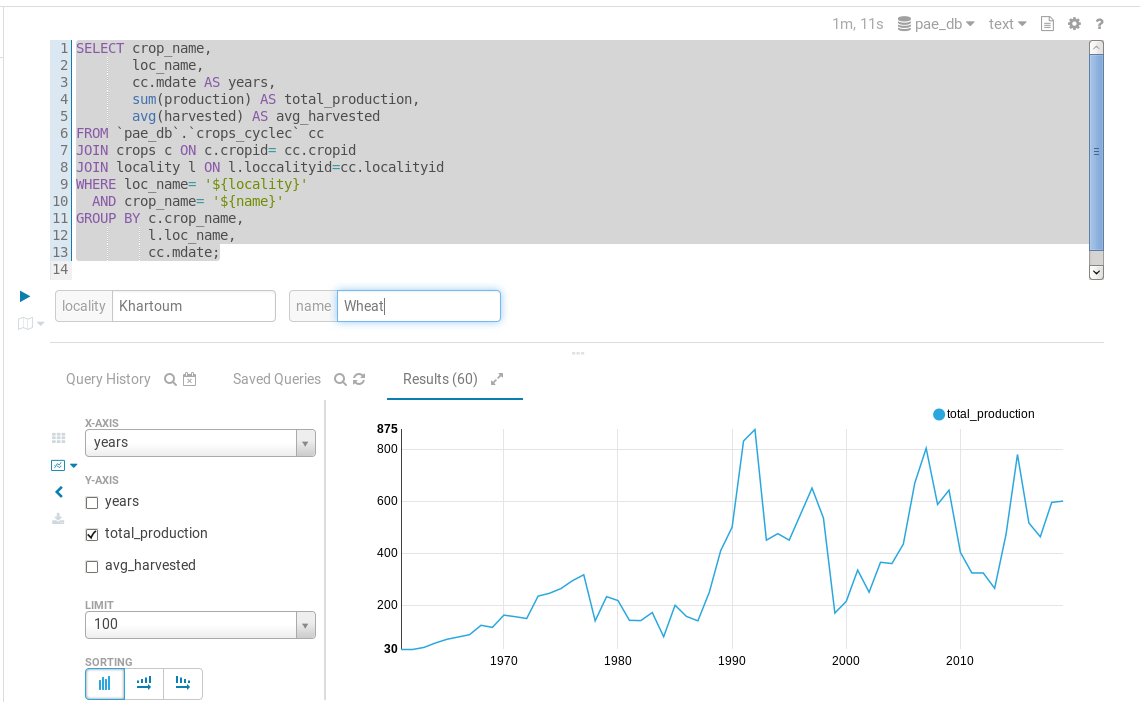
WHERE loc\_name= '${locality}'

AND crop\_name= '${name}'

GROUP BY c.crop\_name,

l.loc\_name,

cc.mdate;



Parametric query to display production for a crop in a specific locality

In the previous query, we used three tables: the crop cycle, the crop, and the locality. Which give us the figure fig 0611 that show the query result after enter Khartoum as locality and Wheat as Crop name.

The previous queries and charts were used as an example showing what this platform can offer us in terms of analysis and presentation. We have just presented some queries for millet, but of course the same thing can be repeated for the rest of the crops under study.

Conclusion

All stakeholders in the agricultural field are searching for the best technologies to ensure the efficiency of the agricultural process and maximize the benefit. With the world’s food shortage and the growing population and waste in supply chains, the effectiveness of the agricultural process becomes one of the important things to meet this growing demand. Stakeholders, scientists, farmers, and decision makers are looking for accurate information that enables them to make the right decisions.

Agricultural data analysis techniques and devices provided by the Internet of Things with the platforms provided by big data have become very useful in this aspect. On this platform, we provided a solution to the Department of Planning and Agricultural Economy in the Republic of Sudan, which can be applied anywhere. This solution helps to ingest data from several sources and store it in a central repository, what makes providing it for analysis and visualization easy and efficient.

This study showed the importance of providing such a platform to collect data and store it in one warehouse, which saves a lot of time and effort that were spent on collecting and analyzing this data. As this platform was able to provide some working days that were wasted to collect this data and analyze it without mentioning the resources used for that.

Challenges

Despite the completion of the research achieving the goals that he intended, which is building an integrated platform for collecting, storing and analyzing data for the Department of Planning and Agricultural Economy, in order to complete this research, we passed many challenges and limitations including:

1. Accessing raw data in an acceptable digital form is a major challenge. In addition to the bureaucratic restrictions on making public data available for research.
2. The availability of accurate integrated data is another difficulty with the weak information infrastructure in Sudan, forcing the researcher to search for sources other than the official government agencies.
3. Big data needs high hardware. Especially we suggested a block of five nodes. With the huge records provided by agricultural activity, these resources are used extensively. For example, in the initial experiments on a virtual machine, some queries took seven minutes. These high requirements are justified by the benefit the platform brings in terms of efficiency and speed of results.
4. Big data requires the efforts of work teams that include data engineers and data scientists to branch out its molecules and the technologies that provide them, which took a lot of time with the limited time allocated for research.

Future Work

during working on building and developing this platform, we have seen a number of capabilities that increase its efficiency if it is included in the future. Therefore, the researcher suggests working on including these features in any future work to improve this platform:

1. Expanding the data under study to include more aspects of the agricultural process that are important for decision-making in the Department of Agricultural Planning and Economics. Such as the risks of agricultural financing, prices of crops and livestock, and the contributions of sectors such as rain and irrigation.
2. Ensure to provide detailed data instead of dealing with data for a full year, which helps in the accuracy of the analysis.
3. Providing an API to relevant parties so that they can take advantage of the platform either by providing it with data or by obtaining information from it.
4. Building an interactive dashboard interface to facilitate the inquiry by the beneficiaries.
5. Add predicting module to make prediction for price, temperature degree and rain precipitation.