

Building & Safety Permits Data Analysis

By

Spriha Batar (CIN: 305053299)

Sumani Pidikiti (CIN: 305065103)

Hasitha Reddy Pittu (CIN: 305084421)

Introduction and Prerequisites

In this tutorial, we will analyze the building and safety permits dataset using different types of Hive queries and import the output into the excel for displaying it in graphical view. Through this we can learn and analyze the problems/complaints received and provide a business solution.

The data includes types of permits, contractors name and address, applicants name and address, issued and expiry dates etc. Using this data, we can find out number of permits issued every year, which type of permit was issued the most etc.

Prerequisites

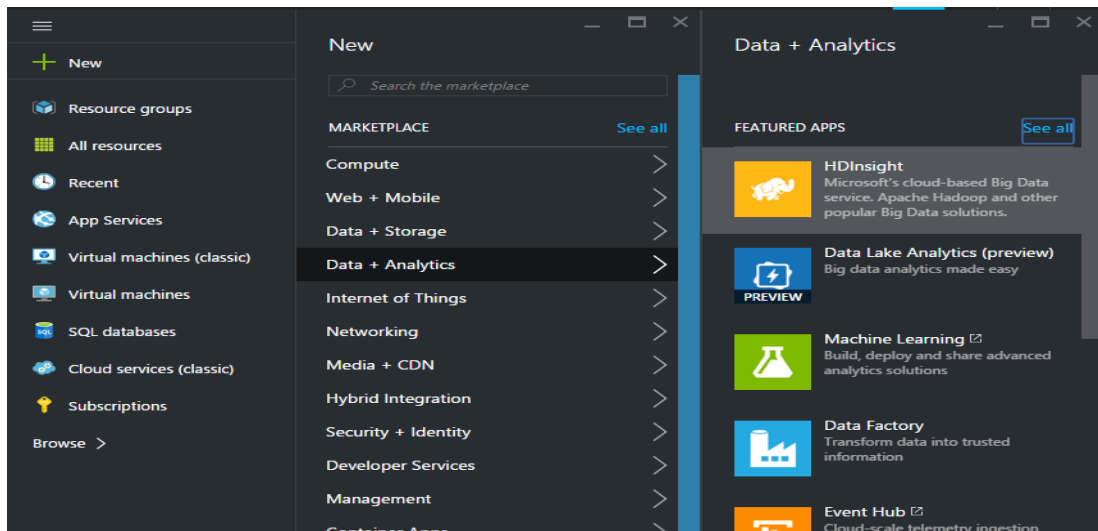
Microsoft Azure account [Click here for trial account](#)

Cloud Berry for Azure [To download click here](#)

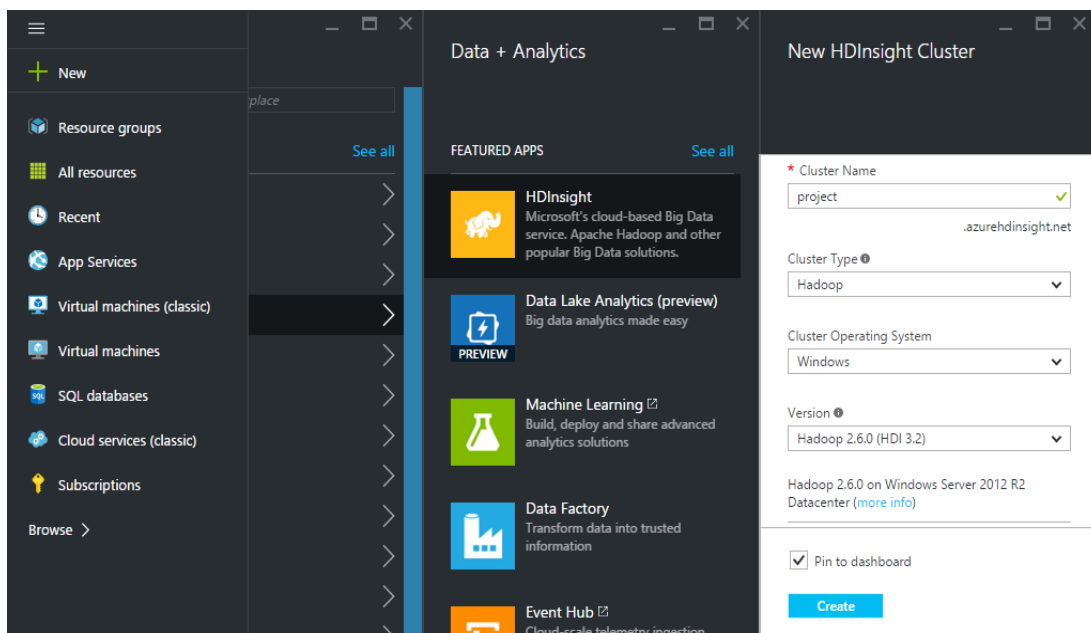
Steps

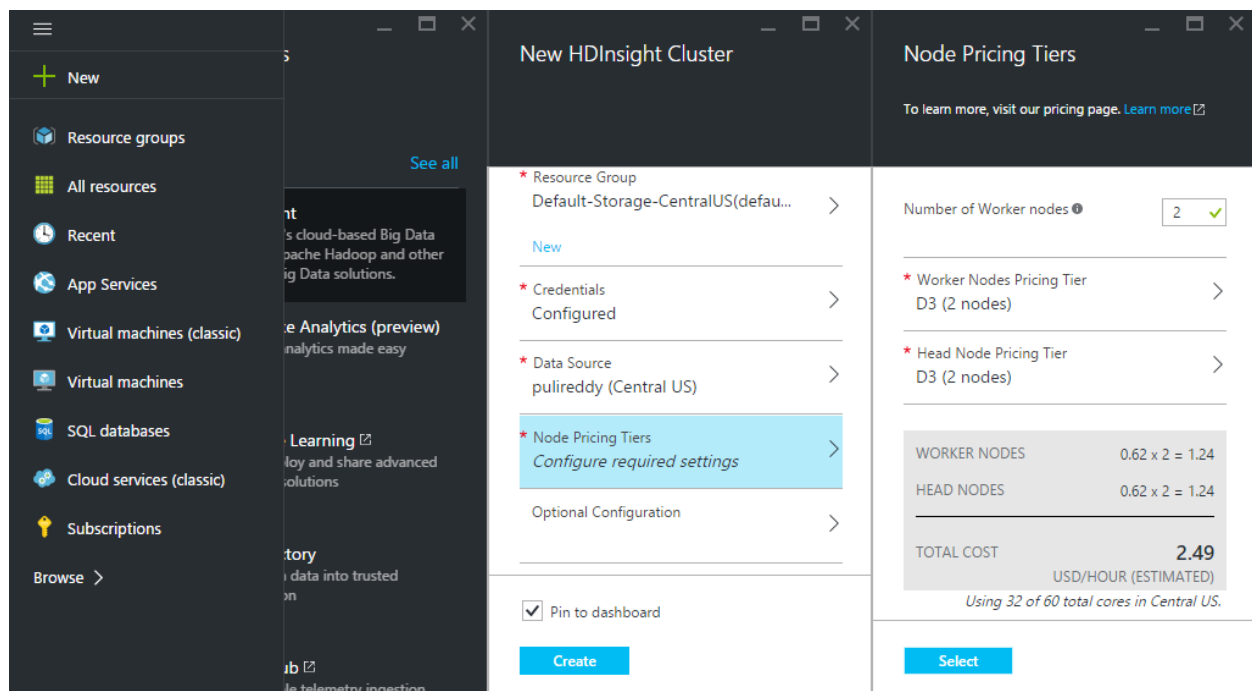
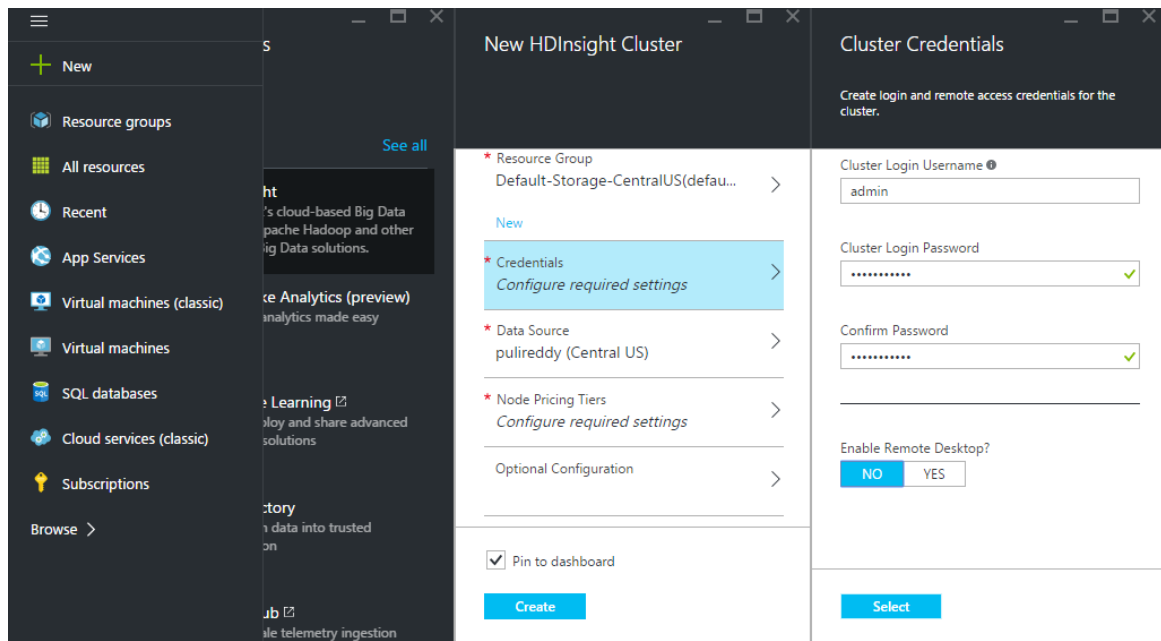
1. After creating an account on Microsoft's Azure, create an HDInsight cluster.

Go to Azure portal after creating an account in Azure, Click on New, In Data + Analytics, Select HDInsight

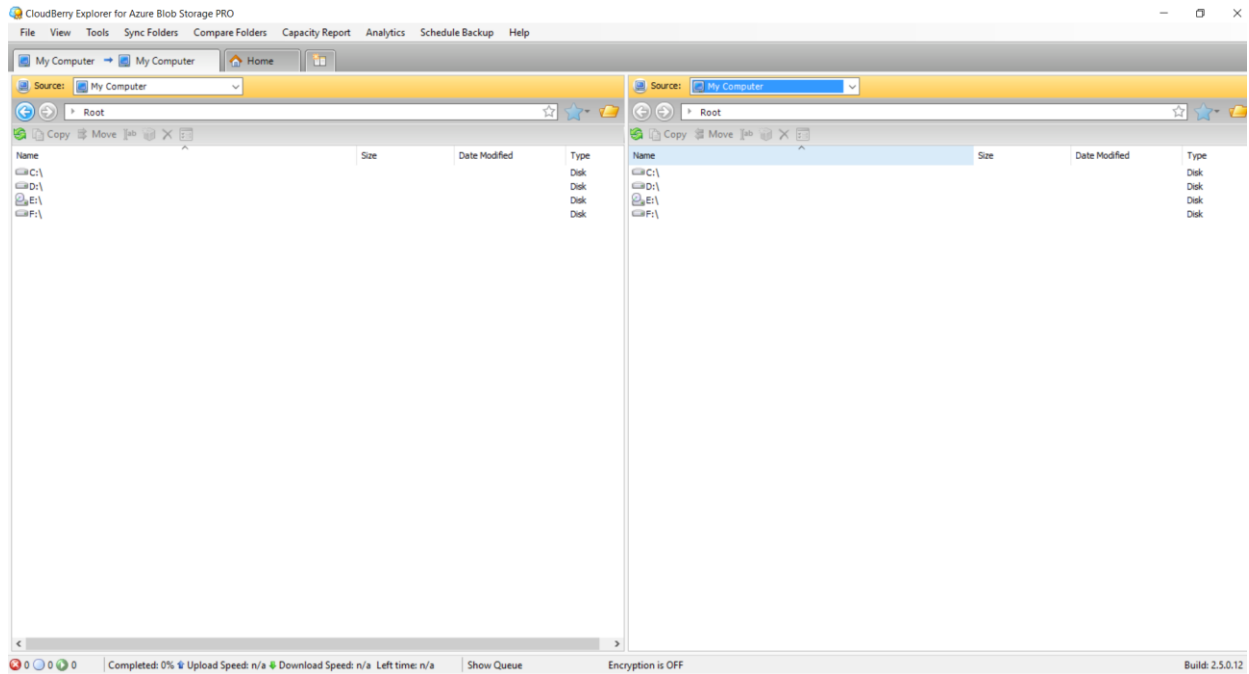


A new pop-up will appear, Type the cluster name. Select type as "Hadoop", Operating System as "Windows", Select Total node as 2, Put Login name and password and click on *create*.

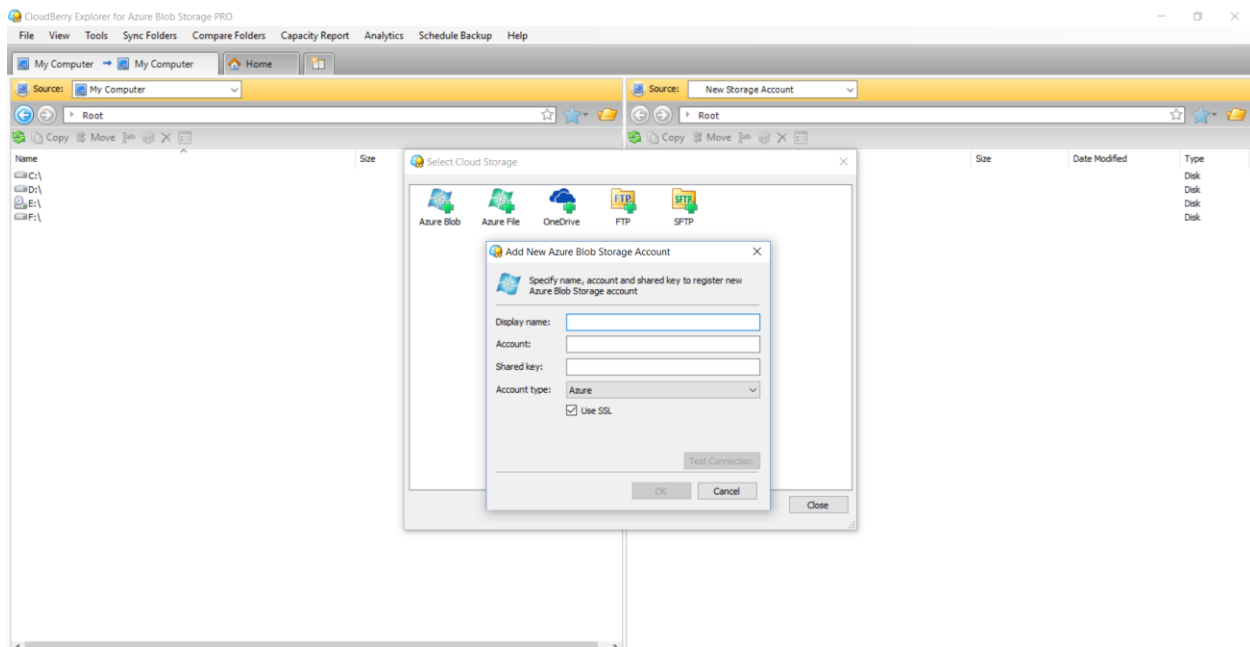




- Azure will start deploying a cluster. It will take few minutes to complete the process.
2. After Creating the cluster, we will upload data into the cluster.
 3. To upload data in the cluster we will use Cloudberry for Azure.
 - Open the Cloudberry, once it is installed on local machine.

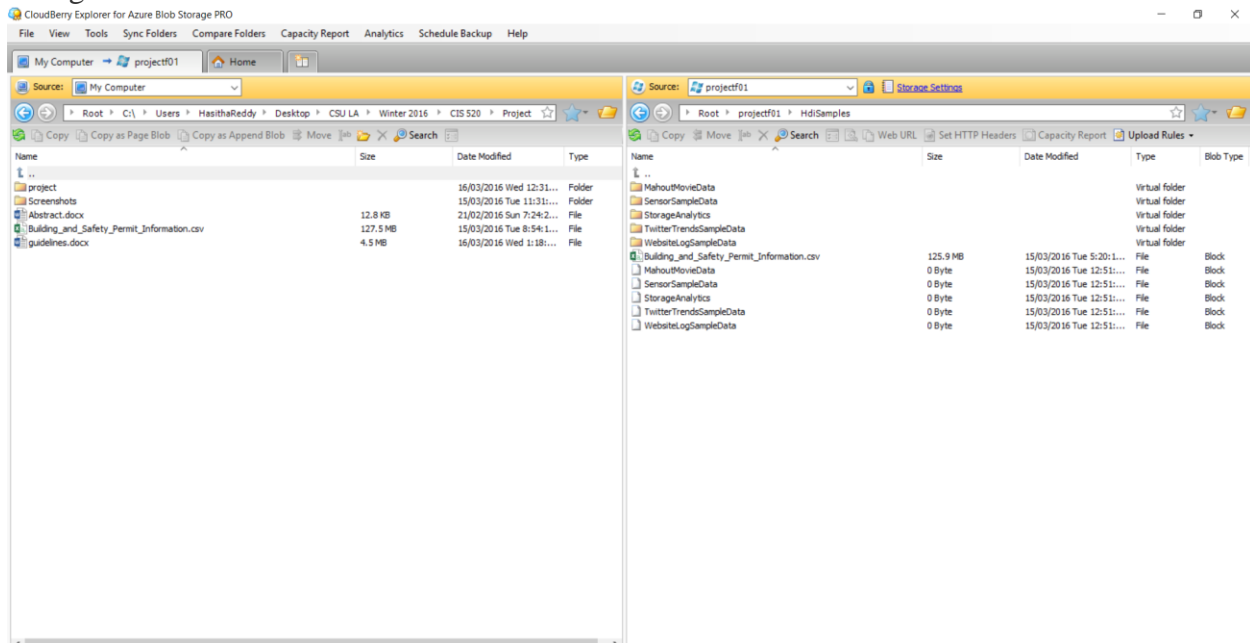


- Select the source on the right and create a new storage account. Select Azure Blob In the dialog box, add the details as mentioned while creating cluster and upload file to the cluster.

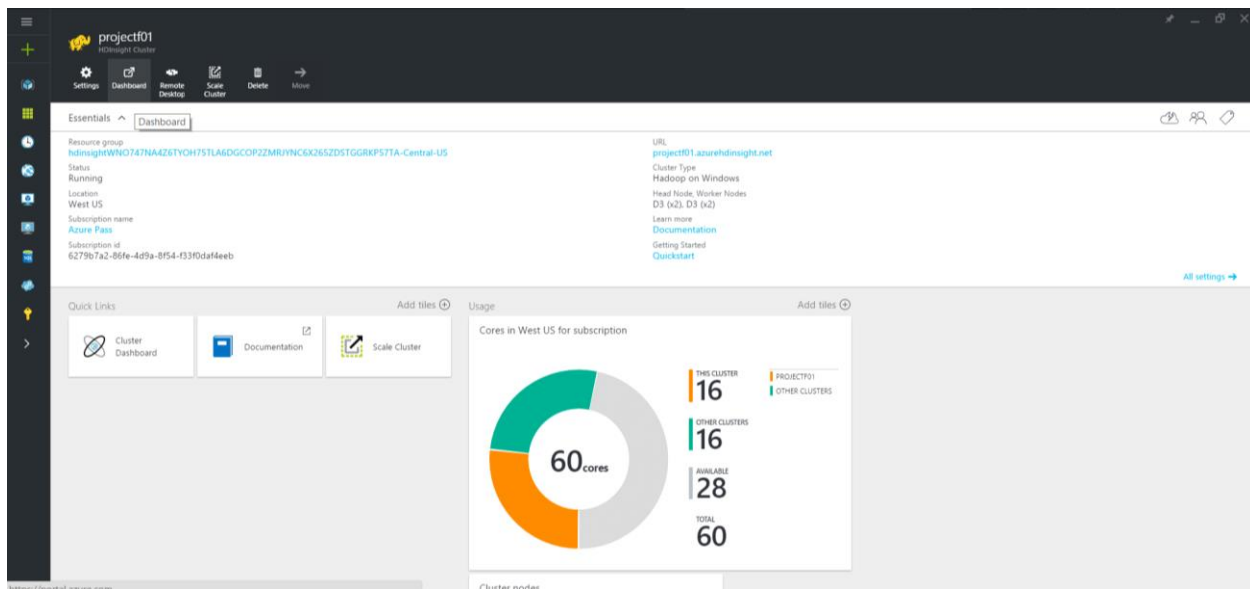


4. Go to the path where your data set is saved on the left hand side. On the right hand side select the storage account you just created. Copy the data set from your system to the cloudberry

storage.



5. After uploading the data, Go to the Azure where the cluster is created and select the Dashboard. In the Dashboard select the Hive Editor.



Microsoft Azure
HDInsight Query Console

[Getting Started Gallery](#)
[Hive Editor](#)
[Job History](#)
[File Browser](#)
[Hadoop UI](#)
[Yarn UI](#)
[Help + Feedback](#)

Solutions With Your Azure Data

We've created jobs for common data analysis scenarios, so that you have a starting point for creating your own solutions. You can use your own data with the job to see how it works, then when you are ready, create your own solution based on the pre-built job.

Microsoft Azure Website Log Analysis Solution

Microsoft Azure Storage Analytics Solution

Coming Soon

Solutions With Sample Data

Learn how to work with HDInsight by walking through basic scenarios such as analyzing web logs and sensor data. You will learn not only how to use HDInsight to analyze data, but how you can connect other applications and services to your data, such as visualizing data in Microsoft Excel.

Sensor Data Analysis

Twitter Trend

Website Log Analysis

Mahout Movie Recommendation

Coming Soon

5. In Hive, First thing is to create table.

- For creating the table run this query in a query editor.

Microsoft Azure
HDInsight Query Console

[Getting Started Gallery](#)
[Hive Editor](#)
[Job History](#)
[File Browser](#)
[Hadoop UI](#)
[Yarn UI](#)
[Help + Feedback](#)

Use this editor and submit a hive query to your cluster

Query Name

```

1 create table if not exists permits (Permit_No string, Status string, Status_Date string, Permit_Type string,
Permit_Sub_Type string, Permit_Category string, Initiating_Office string, Issue_Date string, Address_Start string,
Address_End string, Street_Direction string, Street_Name string, Street_Suffix string, Unit_Range_Start string,
Unit_Range_End string, Zip_Code int, Work_Description string, Valuation string, Contractor_Business_Name string,
Contractor_Address string, Contractor_City string, Contractor_State string, License_Type string, License_No string,
Principal_First_Name string, Principal_Middle_Name string, Principal_Last_Name string, License_Expiration_Date string,
Zone string, Occupancy string, Census_Tract string, Latitude_and_Longitude string)
2 row format delimited fields terminated by ','
3 TBLPROPERTIES("skip.header.line.count"="1");

```

Job Session

Query Name	Date	Id	Action	Status
No data available in table				

Submit

```

create table if not exists permits (Permit_No string, Status string, Status_Date string, Permit_Type string,
Permit_Sub_Type string, Permit_Category string, Initiating_Office string, Issue_Date string, Address_Start
string, Address_End string, Street_Direction string, Street_Name string, Street_Suffix string,
Unit_Range_Start string, Unit_Range_End string, Zip_Code int, Work_Description string,
Valuation string, Contractor_Business_Name string, Contractor_Address string, Contractor_City string,
Contractor_State string, License_Type string, License_No string, Principal_First_Name string,
Principal_Middle_Name string, Principal_Last_Name string, License_Expiration_Date string, Zone string,
Occupancy string, Census_Tract string, Latitude_and_Longitude string)

```

```

row format delimited fields terminated by ','

```

```
TBLPROPERTIES("skip.header.line.count"="1");
```

- After the table is created, add data to the table by the following query.

[Use this editor and submit a hive query to your cluster](#)

Query Name

```
1 LOAD DATA INPATH '\HdiSamples\Building_and_Safety_Permit_Information.csv' OVERWRITE INTO TABLE permits;
```

Job Session

Submit

Query Name	Date	Id	Action	Status
No data available in table				

```
LOAD DATA INPATH
'\HdiSamples\Building_and_Safety_Permit_Information.csv' OVERWRITE INTO TABLE
permits;
```

- Now data is ready to analyze.

6. Now, Analyse the data using following queries.

- Types of Permits Issued

```
select Permit_Types, count(*) from permits
```

```
Group By Permit_Types;
```

- Number of Initiating Offices in LA.

```
select initiating_office, count(*) from permits Group
```

```
By initiating_office;
```

- Permits status over the years.

```
select Issue_Date, Status, count(Status) from permits group
```

```
by Issue_Date, Status;
```


- Different contractors all over the country

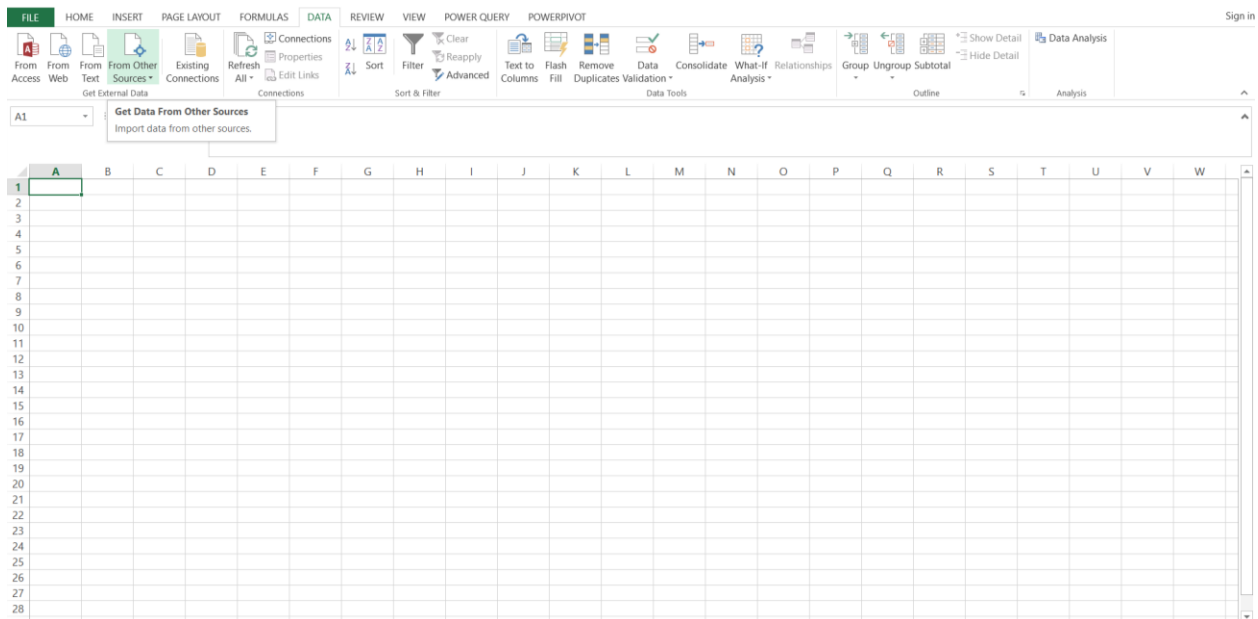
```
select DISTINCT Contractor_State from permits;
```

- Types of permits issued over the years

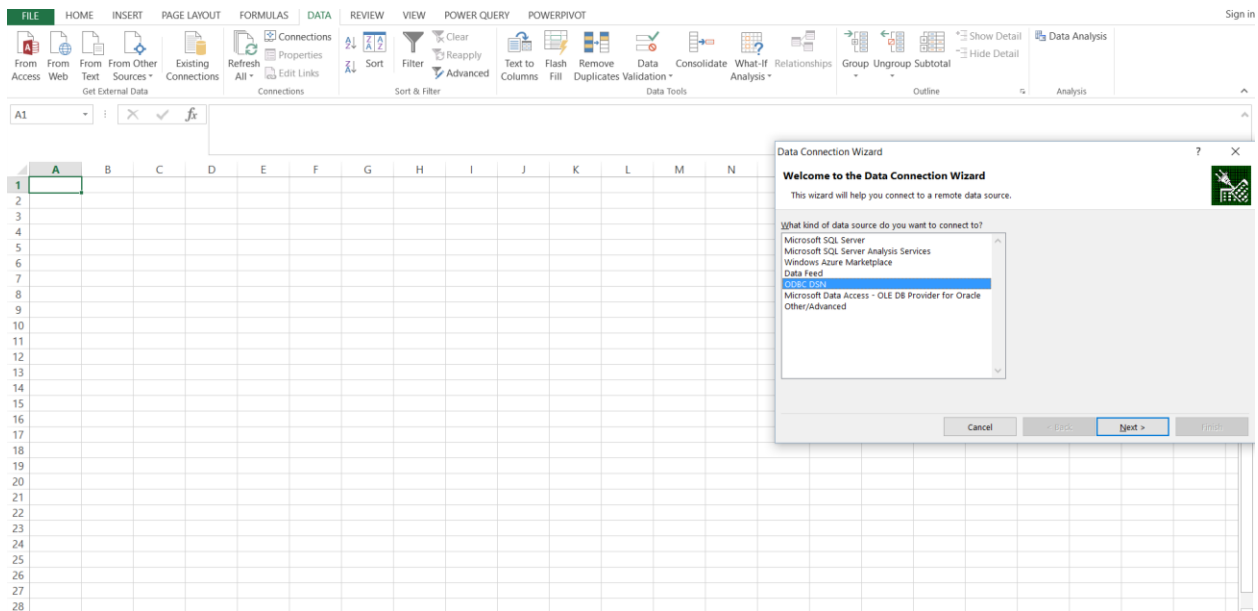
```
select Issue_Date, Permit_Type, count(Permit_Type) from
permits group by Issue_Date, Permit_Type;
```

7. After data is analysed in Hive, go to Microsoft excel.

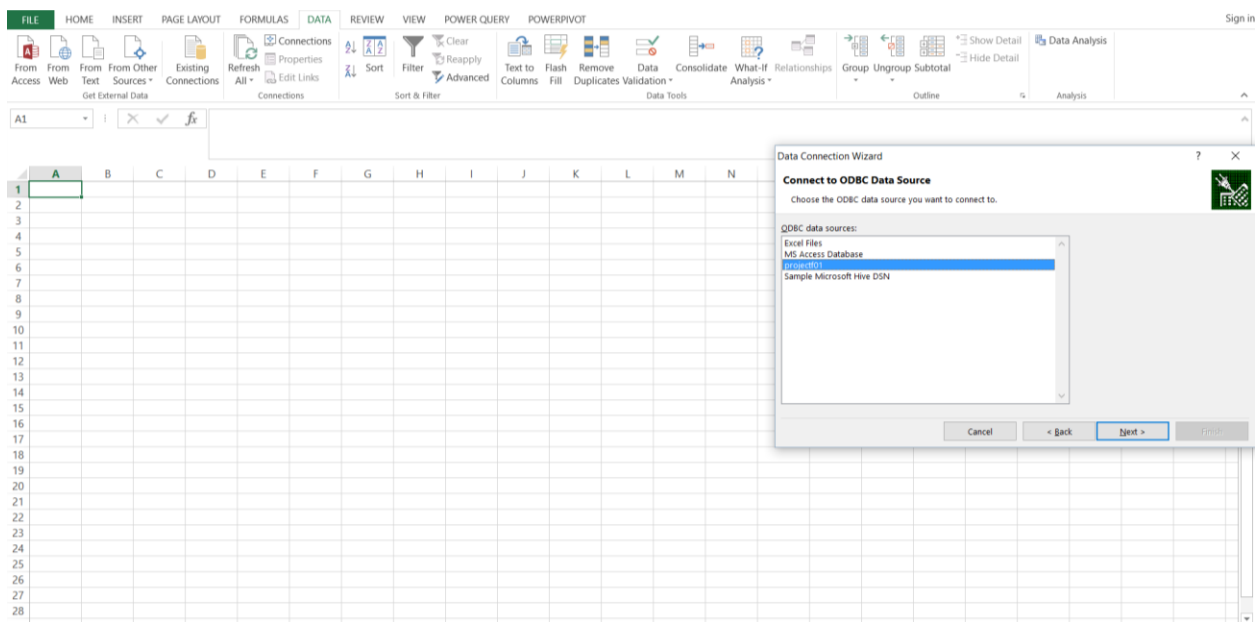
- Select Data tab on top and navigate to From Other Sources



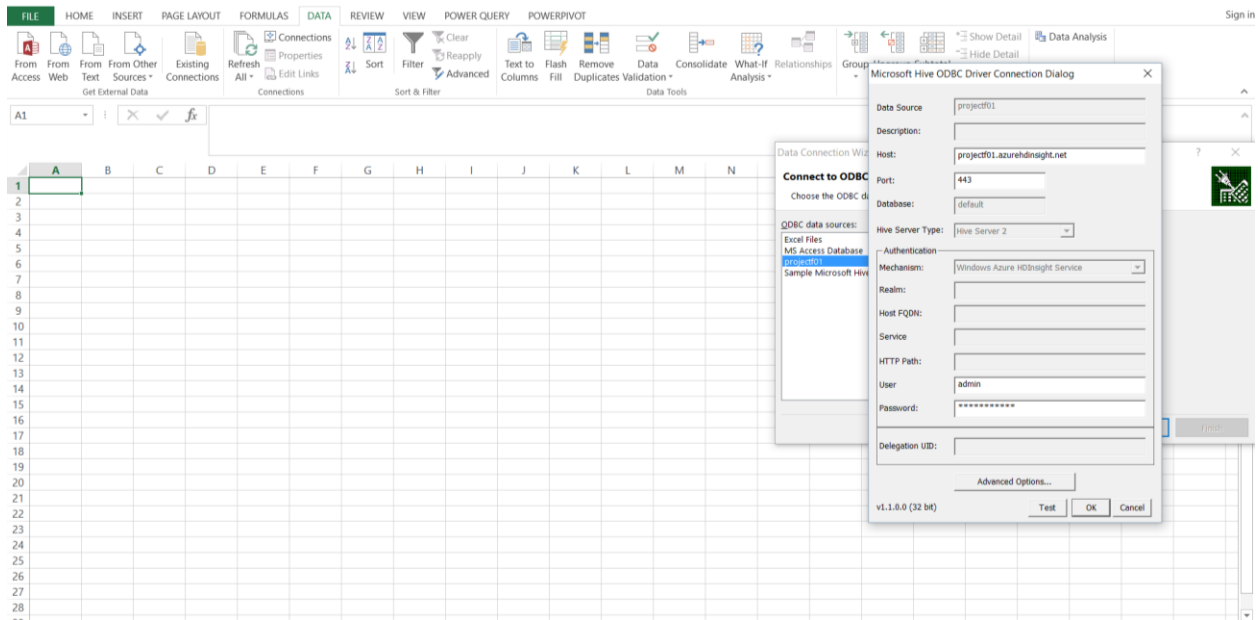
- Select From Data Connection Wizard (Have your ODBC driver previously installed before this. Go to ODBC settings and configure) and select ODBC DSN.



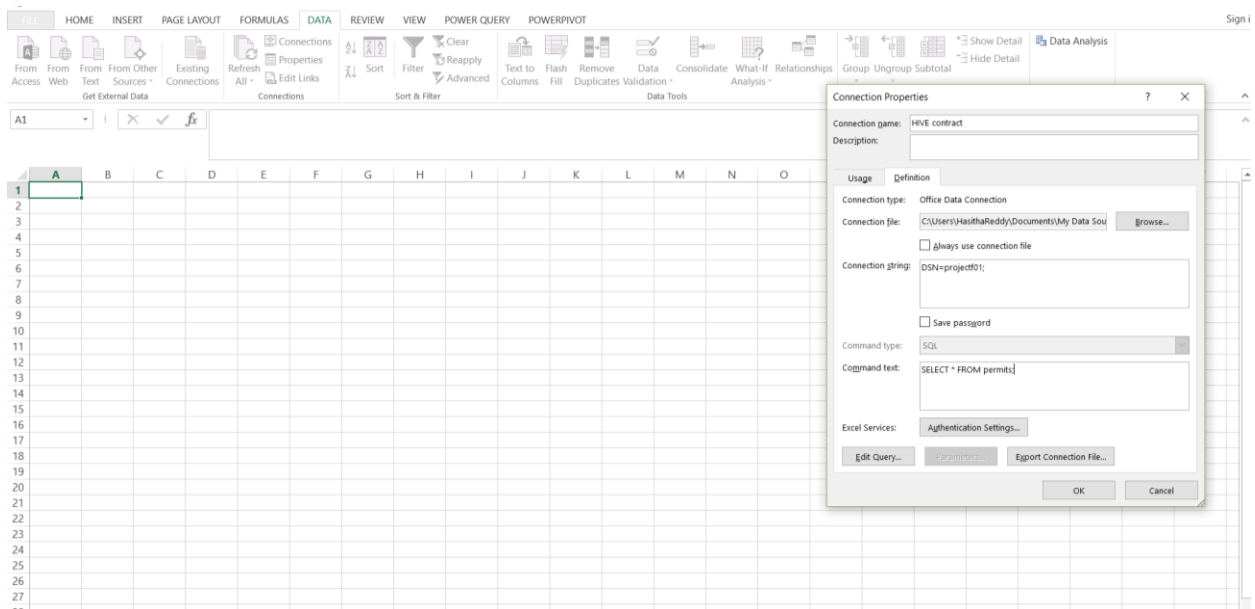
- Select your ODBC data source



- Give the password and select OK. Select your table and click next and then finish.



- Change the properties and give the hive the queries in the under the definitions tab in properties



- The table will be loaded by excel and will look as follows:

<div> <div> <div>FILE</div> <div>HOME</div> <div>INSERT</div> <div>PAGE LAYOUT</div> <div>FORMULAS</div> <div>DATA</div> <div>REVIEW</div> <div>VIEW</div> <div>POWER QUERY</div> <div>POWERPIVOT</div> <div>DESIGN</div> </div> <div> <div>Table Name: Table_HVAC_permit</div> <div> <div>Summarize with PivotTable</div> <div>Remove Duplicates</div> <div>Resize Table</div> <div>Convert to Range</div> </div> <div> <div>Insert Slicer</div> <div>Export Refresh</div> <div>Open in Browser</div> <div>Unlink</div> </div> <div> <div>Properties</div> <div>External Table Data</div> </div> </div> <div> <div> <div>Header Row</div> <div>Total Row</div> <div>Banded Rows</div> </div> <div> <div>First Column</div> <div>Last Column</div> <div>Banded Columns</div> </div> <div> <div>Filter Button</div> <div>Table Style Options</div> </div> </div> <div> <div>Table Styles</div> </div> </div>											
1	2	3	4	5	6	7	8	9	10	11	12
13041-90000-00002	Permit Finaled	01/04/13 0:00	Electrical	1 or 2 Family Dwelling	No Plan Check	INTERNET	01/01/13	5334	5334	S	CHESLEY
13042-90000-00001	Permit Finaled	01/04/13 0:00	Plumbing	1 or 2 Family Dwelling	No Plan Check	INTERNET	01/01/13	2537	2537	N	PANORAMA
13042-90000-00006	Permit Finaled	01/08/13 0:00	Plumbing	1 or 2 Family Dwelling	No Plan Check	INTERNET	01/01/13	4929	4929	W	MELROSE
13042-90000-00005	Permit Finaled	04/30/2015 12:00:00 AM	Plumbing	1 or 2 Family Dwelling	No Plan Check	INTERNET	01/01/13	3651	3651	N	ALTA MESA
13042-90000-00002	Permit Finaled	01/07/13 0:00	Plumbing	1 or 2 Family Dwelling	No Plan Check	INTERNET	01/01/13	619	619	N	COMMONWEALTH
13044-90000-00001	Permit Finaled	01/11/13 0:00	HVAC	Commercial	No Plan Check	INTERNET	01/01/13	7321	7321	N	RESEDA
13041-90000-00001	Permit Finaled	01/18/2013 12:00:00 AM	Electrical	1 or 2 Family Dwelling	No Plan Check	INTERNET	01/01/13	3401	3401	W	FERNCROFT
13042-90000-00007	Permit Finaled	07/17/2013 12:00:00 AM	Plumbing	Commercial	No Plan Check	INTERNET	01/01/13	790	790	S	CERES
13042-90000-00004	Permit Expired	01/27/2015 12:00:00 AM	Plumbing	1 or 2 Family Dwelling	No Plan Check	INTERNET	01/01/13	744	744	N	SIERRA BONITA
13042-90000-00003	Permit Finaled	01/07/13 0:00	Plumbing	1 or 2 Family Dwelling	No Plan Check	INTERNET	01/01/13	617	617	N	COMMONWEALTH
13042-90000-00004	Permit Finaled	01/09/13 0:00	Electrical	1 or 2 Family Dwelling	No Plan Check	INTERNET	01/01/13	15906	15906	W	VALERIO
13041-10000-00037	Permit Finaled	05/06/13 0:00	Electrical	1 or 2 Family Dwelling	No Plan Check	METRO	01/02/13	22050	22050	W	MAYALL
13044-20000-00015	Permit Finaled	04/03/13 0:00	HVAC	1 or 2 Family Dwelling	No Plan Check	VAN NUYS	01/02/13	1325	1325	E	WARREN
13016-10000-00028	Permit Finaled	01/03/13 0:00	Bldg Alter/Repair	1 or 2 Family Dwelling	No Plan Check	METRO	01/02/13	607	607	W	115TH
13016-10000-00049	Permit Finaled	02/11/13 0:00	Bldg Alter/Repair	Apartment	No Plan Check	METRO	01/02/13	9800	9800	S	GRAPE
13042-90000-00038	Permit Expired	01/27/2015 12:00:00 AM	Plumbing	1 or 2 Family Dwelling	No Plan Check	INTERNET	01/02/13	8119	8119	N	CANTALOUPE
13041-20000-00021	Permit Closed	12/20/2013 12:00:00 AM	Electrical	Commercial	No Plan Check	VAN NUYS	01/02/13	3171	3171	S	BUNDY
13042-90000-00086	Permit Finaled	01/07/13 0:00	Plumbing	1 or 2 Family Dwelling	No Plan Check	INTERNET	01/02/13	3910	3910	S	COOLIDGE
13041-20000-00098	Permit Finaled	01/24/2013 12:00:00 AM	Electrical	1 or 2 Family Dwelling	No Plan Check	VAN NUYS	01/02/13	11600	11600	N	RUFFNER
13042-20000-00027	Permit Expired	01/27/2015 12:00:00 AM	Plumbing	Apartment	No Plan Check	VAN NUYS	01/02/13	14454	14454	W	DICKENS
13016-90000-00078	Permit Finaled	03/25/2014 12:00:00 AM	Bldg Alter/Repair	1 or 2 Family Dwelling	No Plan Check	INTERNET	01/02/13	1001	1001	S	MAIN
13016-90000-00007	Permit Finaled	01/09/13 0:00	Bldg Alter/Repair	Apartment	No Plan Check	INTERNET	01/02/13	6600	6600	N	VANALDEN
12044-20000-01724	Permit Finaled	07/25/2013 12:00:00 AM	HVAC	Apartment	No Plan Check	VAN NUYS	01/02/13	4440	4440	N	VANTAGE
12016-10000-15864	Issued	01/02/13 0:00	Bldg Alter/Repair	Commercial	Plan Check	METRO	01/02/13	315	315	W	5TH
13042-20000-00054	Permit Finaled	01/07/13 0:00	Plumbing	1 or 2 Family Dwelling	No Plan Check	VAN NUYS	01/02/13	11547	11547	W	KLING
13042-70000-00018	Permit Expired	01/27/2015 12:00:00 AM	Plumbing	Apartment	No Plan Check	SOUTH LA	01/02/13	4325	4325	W	CLINTON
13044-10000-00005	Permit Finaled	05/07/13 0:00	HVAC	Commercial	No Plan Check	METRO	01/02/13	8500	8500	W	BEVERLY

- Different types of visualization can be done. A few examples are as follows:

