

Final Project(PowerBI)

My Topic is PowerBI, and when I searched on internet I found PowerBI desktop software to create beautiful drill down reports and publish them to PowerBi portal, then I realized this cannot be that simple and found that you can create report with live stream data using **Event Hub** and **Analytical Stream Job**. So I created two demos, the main demo is obviously using the greatest and latest technology using **event hub** and **stream analytic job** to report live stream data and other short demo just to show how to make beautiful graphs, and drill down ,slice/dice the data using desktop version of BI on static data and publish them in **PowerBI portal**.

Demo1 (Main demo)

Event hub live stream & PowerBI Analysis

Problem Statement:

Data is being generated in every form today, there are several tools available in the market to analyze the data but Microsoft PowerBI provides interface with Azure and provide ability to analyze the live stream data. In power BI you can almost connect to all the data sources like relational databases, live Facebook feed, tweeter feed, text/CSV and as well as data created by event hubs and analyze them in one central places in real time.

We will present a use case of real-time fraud detection based on phone-call data and. We will Bring streaming events into an instance of Azure Event Hubs and then send the real-time data to power BI engine to examine the data in real time.

In this example, we will consider fraudulent usage to be calls that originate from the same user but in different locations within 5 seconds of one another. For example, the same user can't legitimately make a call from the US and Australia at the same time.

Overview of the Technology:

To Solve the problem, I have created an even hub in Azure portal that will read the data in real time by using a client application that will generate a phone-call data including fraud calls. Some of the records that the client app produces look like fraudulent calls. That data will then send to PowerBI engine to analyze in real time to detect the fraudulent calls.

High Level Steps:

- 1) Download the client program to generate the sample data
- 2) Configure even hubs in Azure portal
- 3) Create a Stream Analytics job to manage streaming data
- 4) Run a client application to generate the live telephone fraud call data
- 5) Create a query in Analytics job in Azure to transform the data ingested by event hub
- 6) Configure the output in analytic job in Azure to send the transform data to powerBI
- 7) In Powerbi portal create a dashboard to process the output of stream analytic query

Data source

Test data generator application

<http://download.microsoft.com/download/8/B/D/8BD50991-8D54-4F59-AB83-3354B69C8A7E/TelcoGenerator.zip>

Hardware used

Windows 7 64 bit

Software used:

Azure portal (tool)

PowerBI portal (tool)

Code :**Software Code For Analytic stream job query**

```
SELECT System.Timestamp AS WindowEnd, COUNT(*) AS FraudulentCalls
      INTO "fraudstreamoutput"
      FROM "fraudstream" CS1 TIMESTAMP BY CallRecTime
      JOIN "fraudstream" CS2 TIMESTAMP BY CallRecTime
      ON CS1.CallingIMSI = CS2.CallingIMSI
      AND DATEDIFF(ss, CS1, CS2) BETWEEN 1 AND 5
      WHERE CS1.SwitchNum != CS2.SwitchNum
      GROUP BY TumblingWindow(Duration(second, 1))
```

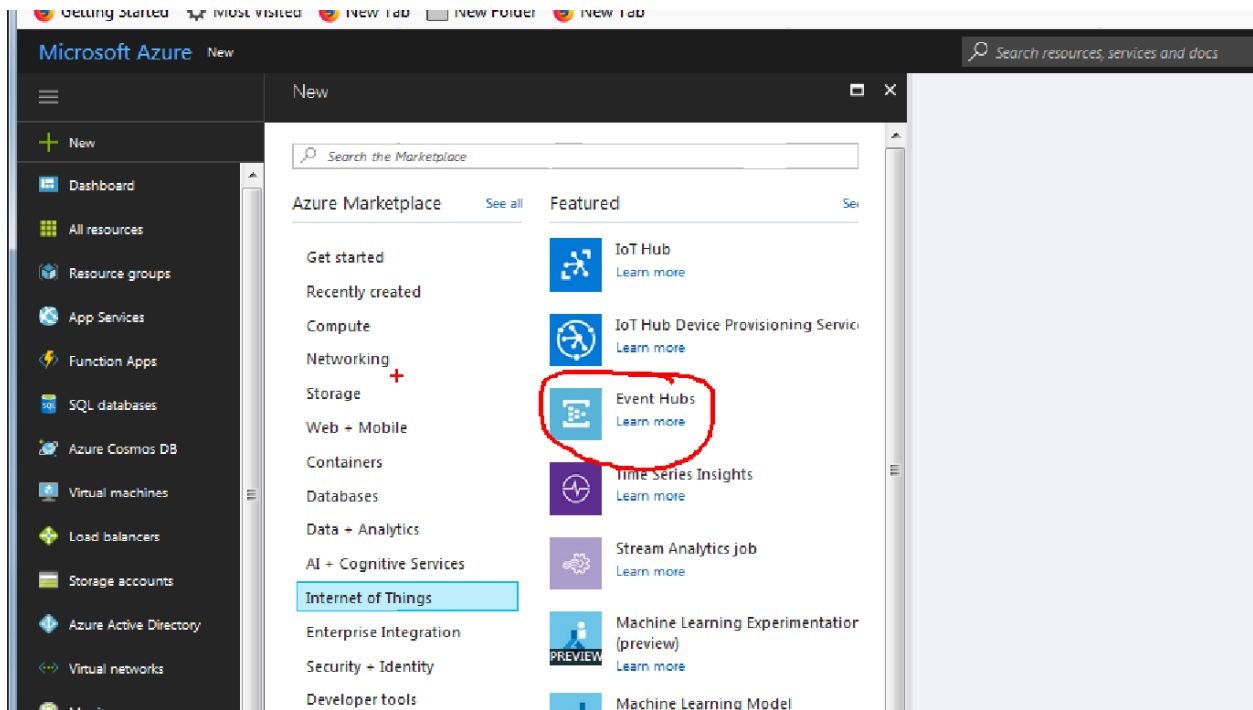
Software Code date dim table in PowerBI desktop

```
Date_dim = ADDCOLUMNS ( CALENDAR (DATE(2000,1,1), DATE(2025,12,31)), "DateAsInteger",
      FORMAT ( [Date], "YYYYMMDD" ), "Year", YEAR ( [Date] ), "MonthNo", FORMAT ( [Date], "MM"
      ), "YearMonthNo", FORMAT ( [Date], "YYYY/MM" ),
      "YearMonth", FORMAT ( [Date], "YYYY/mmm" ),
      "MonthShort", FORMAT ( [Date], "mmm" ), "MonthLong", FORMAT ( [Date], "mmmm" ),
      "WeekNo", WEEKDAY ( [Date] ), "WeekDay",
      FORMAT ( [Date], "dddd" ), "WeekDayShort", FORMAT ( [Date], "dddd" ), "Quarter", "Q" &
      FORMAT ( [Date], "Q" ), "YearQuarter",
      FORMAT ( [Date], "YYYY" ) & "/" & "Q" & FORMAT ( [Date], "Q" ))
```

Here is step by step Process

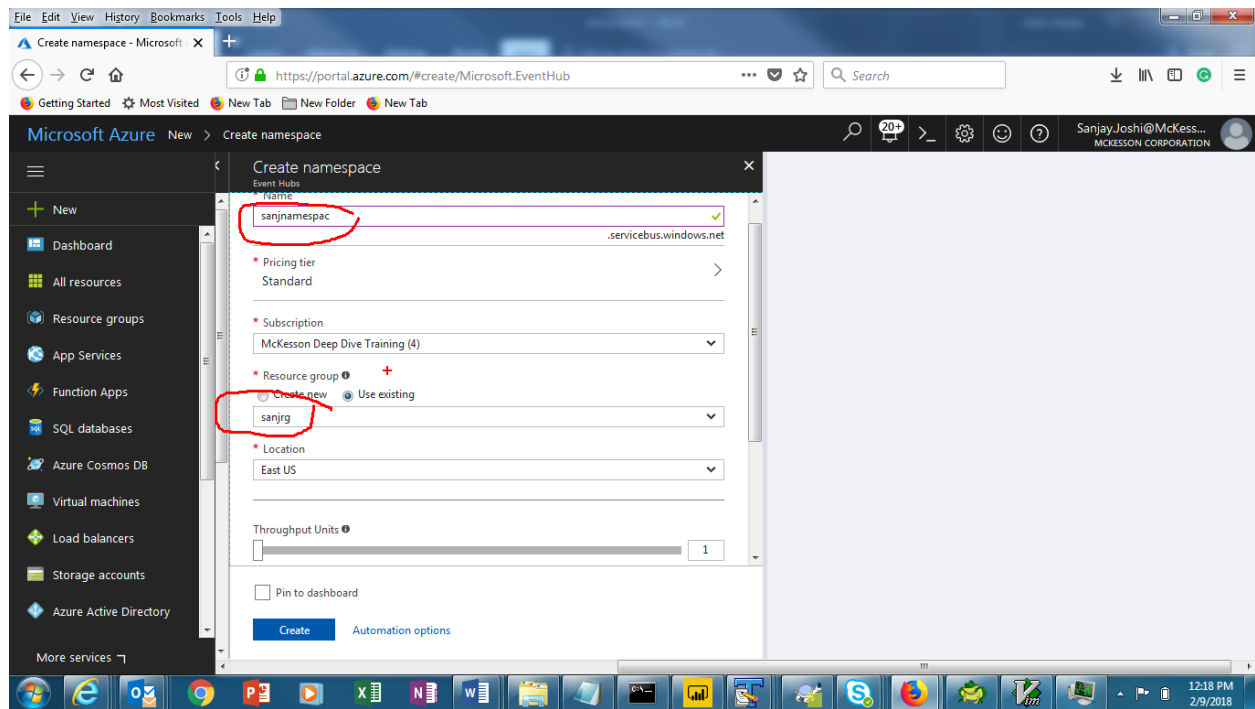
1) Create event hub

Login to Azure portal and click on the new and select the 'Internet of thing' blade and select even hub to create a event hub

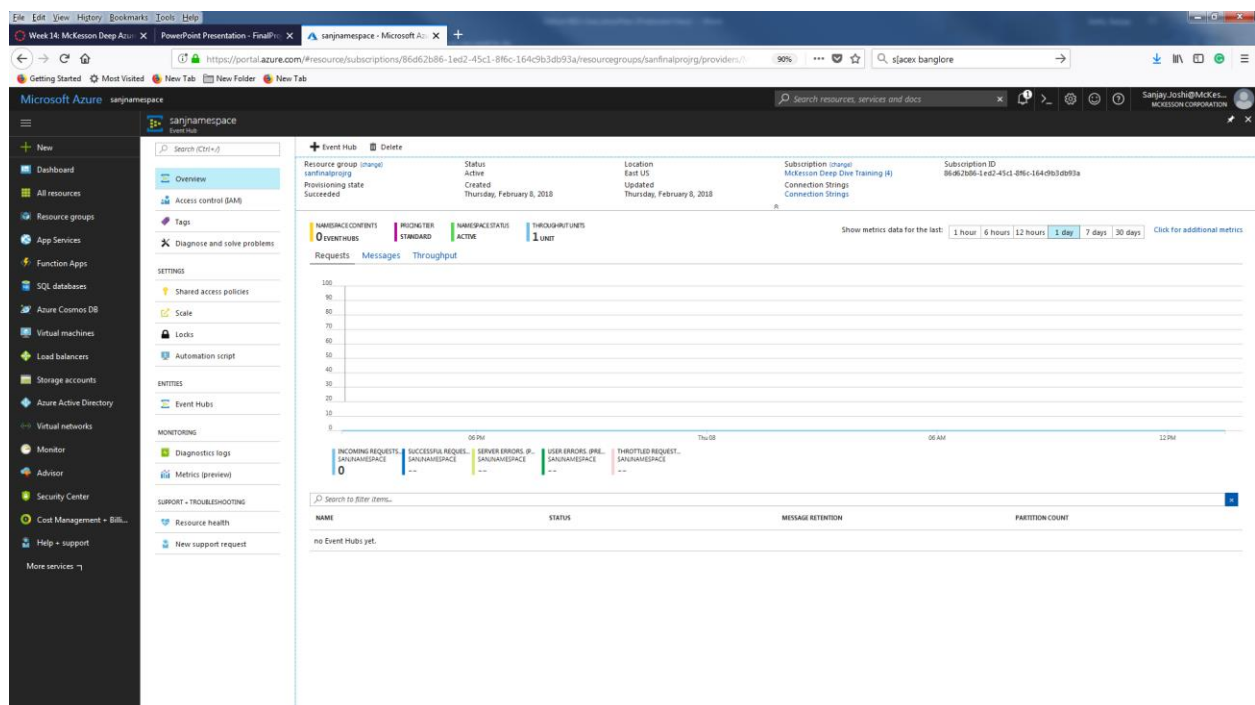


Create namespace

The following screen will appear to create a namespace , select name , provide resource group name and select location and hit create button as seen below



It will take few seconds to create namespace, select the namespace you created (You may find it by searching under all resources or you could have pinned it dashboard for easy access). The following screen will appear.

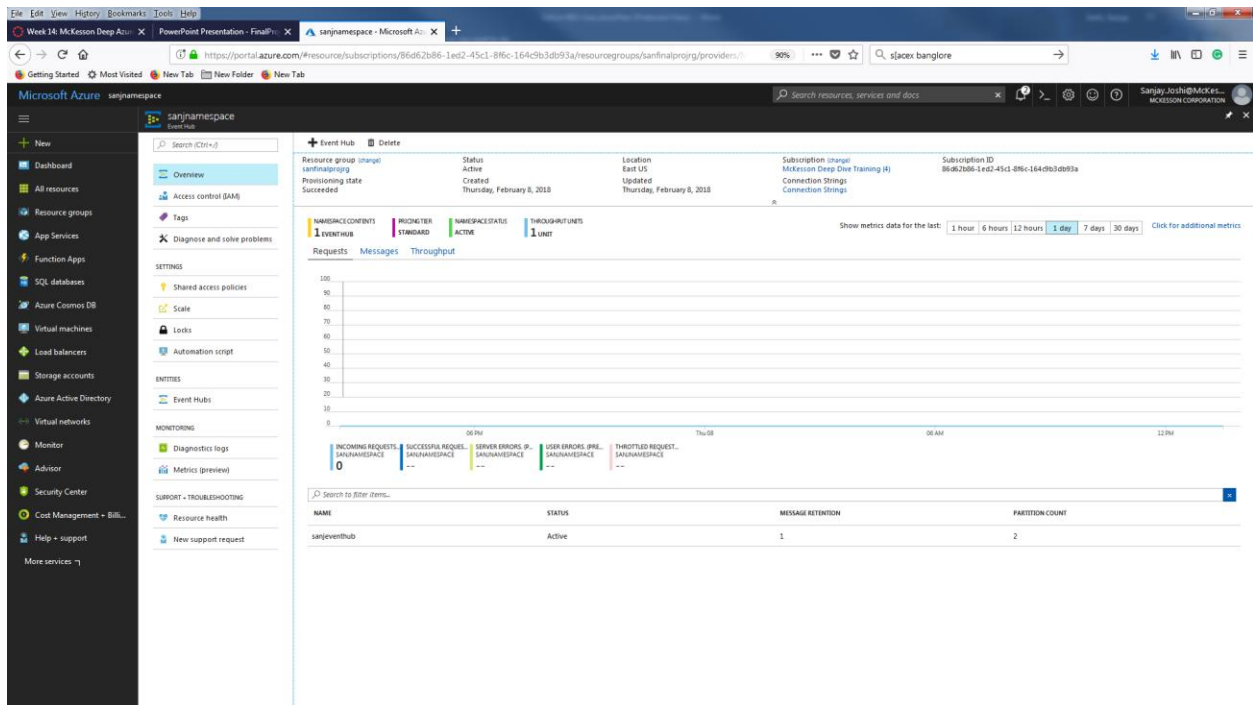


Click on +Event hub icon to create the even hub following screen will appear, provide the name and hit the create button

The screenshot shows the 'Create Event Hub' dialog in the Azure portal. The dialog has a dark header with the title 'Create Event Hub' and a close button. Below the header, there's a sidebar on the left with a list of Azure services. The main area contains the following fields:

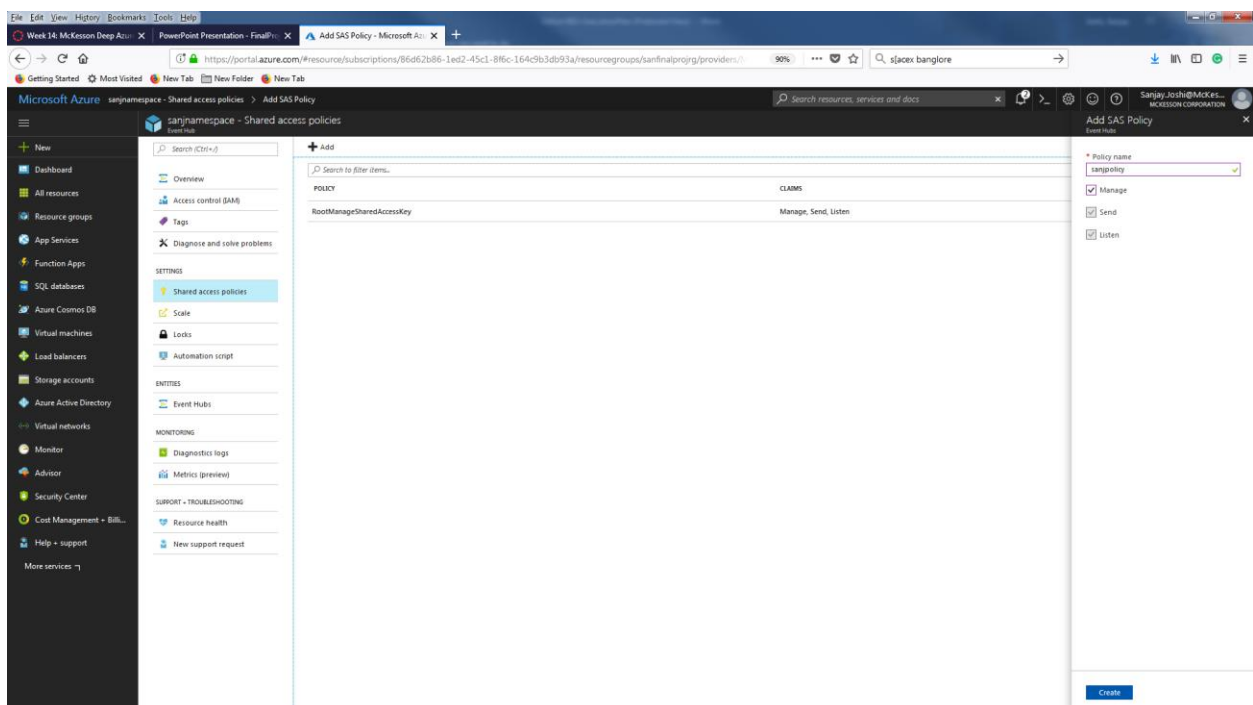
- Name:** A text input field with the value 'sanjeventhub'. It is highlighted with a red circle.
- Partition Count:** A slider control set to 2.
- Message Retention:** A slider control set to 1.
- Capture:** A toggle switch set to 'Off'. A red plus sign is visible next to this section.

After few second eventhub will be created and will look like this.

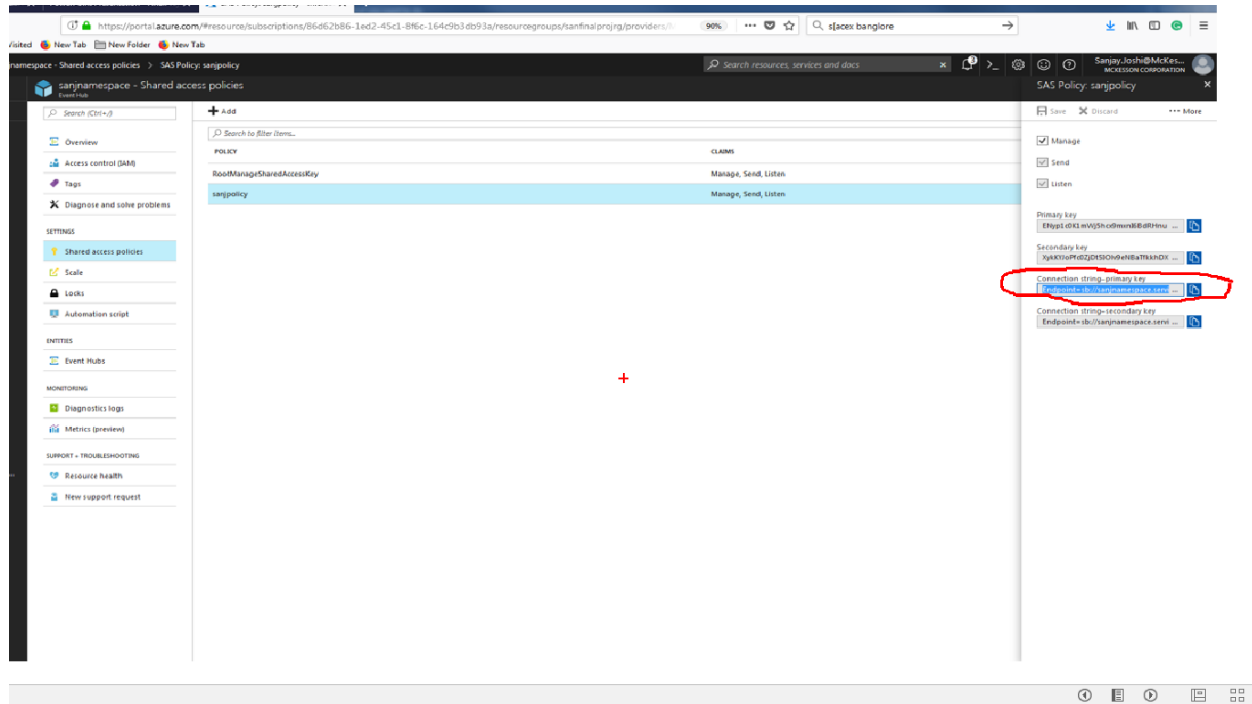


Click on the Shared access policy blade under setting and hit +Add button to create new policy , don't touch the default RootMangedSharedAccesskey .

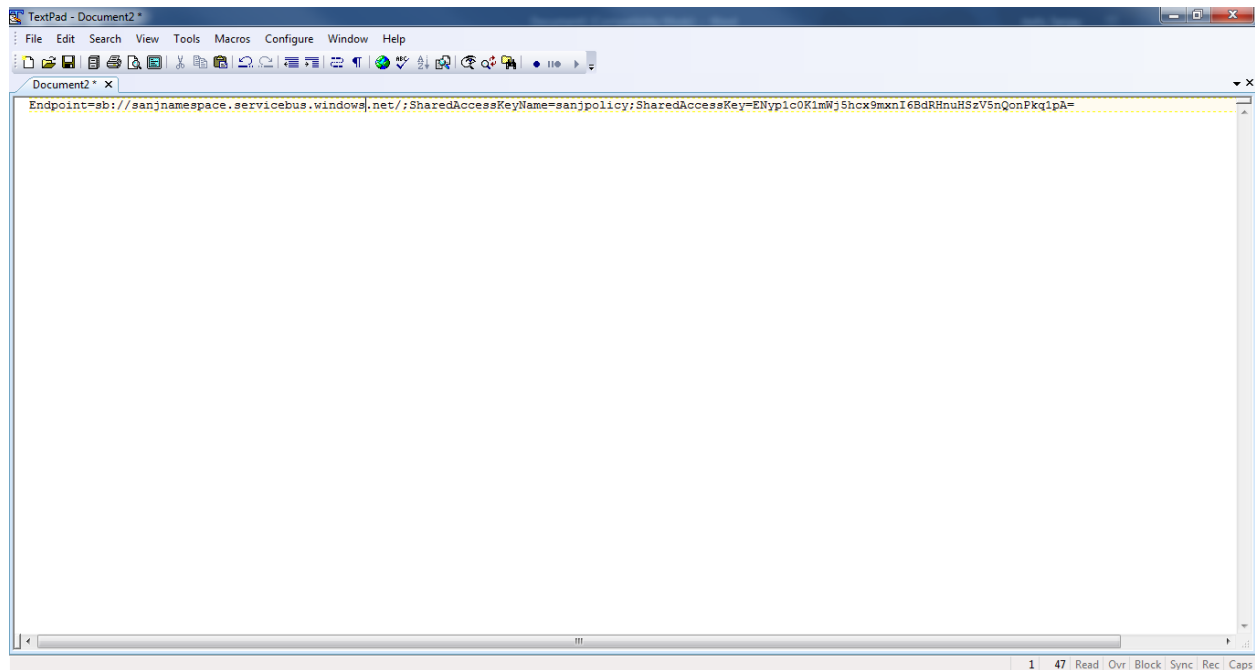
Provide Name and click on 'Mange' checkbox to get all (send and listen access) and hit the Create button



It will create new policy. Click on the policy you created to capture the Primary connection string value and save it to notepad like this



String is stored in notepad, we need this value later to configure the client App



Now setup the job that will generate the fraud telecom data . Download the TelecoGenerator.zip from flowing link.

<http://download.microsoft.com/download/8/B/D/8BD50991-8D54-4F59-AB83-3354B69C8A7E/TelcoGenerator.zip>

extract the file to a folder and open the '**telcodatagen.exe.config**' file in notepad or textpad and replace the value for **eventhubname** and **connection** string , you can get these value we saved in notepad

```
<?xml version="1.0" encoding="utf-8"?>
```

```
<configuration>
```

```
<appSettings>
```

```
<!-- Service Bus specific app setings for messaging connections -->
```

```
<add key="EventHubName" value="sanjeventhub"/>
```



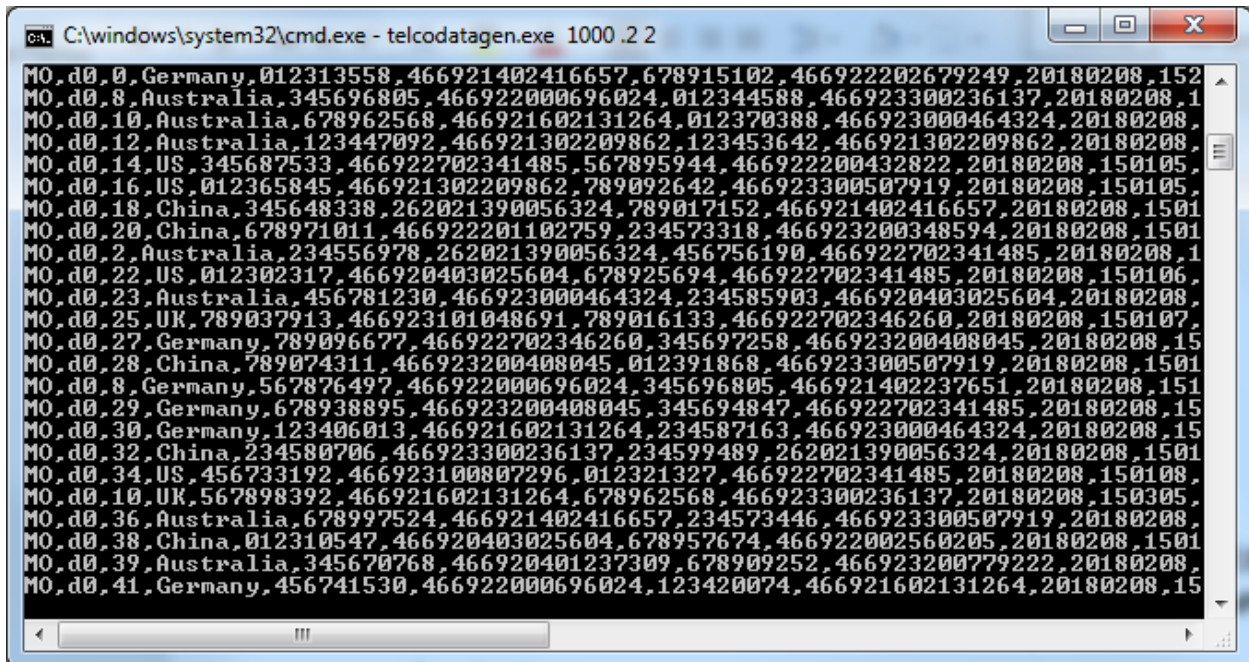
```
<add key="Microsoft.ServiceBus.ConnectionString"
value="Endpoint=sb://sanjamespace.servicebus.windows.net/;SharedAccessKeyName=sanjpolicy;SharedAccessKey=ENyp1c0K1mWj5hcx9mxnI6BdRHnuHSzV5nQonPkq1pA="/>

</appSettings>
```

Save the config file and run command to generate the data for 2 hours with .2 seconds interval

telcodatagen.exe 1000 .2 2

It will generate the sample data like this and this data is being ingested by evenhub we created

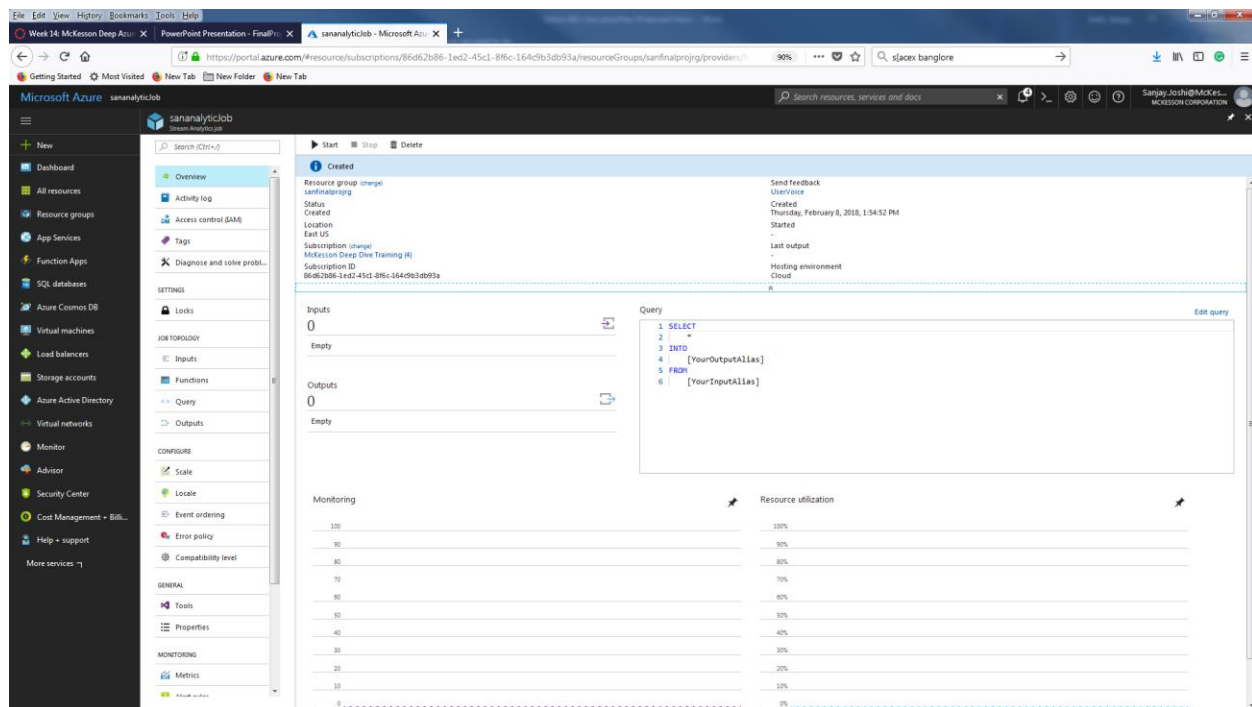


```
C:\windows\system32\cmd.exe - telcodatagen.exe 1000 .2 2
MO,d0,0,Germany,012313558,466921402416657,678915102,466922202679249,20180208,152
MO,d0,8,Australia,345696805,466922000696024,012344588,466923300236137,20180208,1
MO,d0,10,Australia,678962568,466921602131264,012370388,466923000464324,20180208,1
MO,d0,12,Australia,123447092,466921302209862,123453642,466921302209862,20180208,1
MO,d0,14,US,345687533,466922702341485,567895944,466922200432822,20180208,150105,
MO,d0,16,US,012365845,466921302209862,789092642,466923300507919,20180208,150105,
MO,d0,18,China,345648338,262021390056324,789017152,466921402416657,20180208,1501
MO,d0,20,China,678971011,466922201102759,234573318,466923200348594,20180208,1501
MO,d0,2,Australia,234556978,262021390056324,456756190,466922702341485,20180208,1
MO,d0,22,US,012302317,466920403025604,678925694,466922702341485,20180208,150106,
MO,d0,23,Australia,456781230,466923000464324,234585903,466920403025604,20180208,1
MO,d0,25,UK,789037913,466923101048691,789016133,466922702346260,20180208,150107,
MO,d0,27,Germany,789096677,466922702346260,345697258,466923200408045,20180208,15
MO,d0,28,China,789074311,466923200408045,012391868,466923300507919,20180208,1501
MO,d0,8,Germany,567876497,466922000696024,345696805,466921402237651,20180208,151
MO,d0,29,Germany,678938895,466923200408045,345694847,466922702341485,20180208,15
MO,d0,30,Germany,123406013,466921602131264,234587163,466923000464324,20180208,15
MO,d0,32,China,234580706,466923300236137,234599489,262021390056324,20180208,1501
MO,d0,34,US,456733192,466923100807296,012321327,466922702341485,20180208,150108,
MO,d0,10,UK,567898392,466921602131264,678962568,466923300236137,20180208,150305,
MO,d0,36,Australia,678997524,466921402416657,234573446,466923300507919,20180208,1
MO,d0,38,China,012310547,466920403025604,678957674,466922002560205,20180208,1501
MO,d0,39,Australia,345670768,466920401237309,678909252,466923200779222,20180208,1
MO,d0,41,Germany,456741530,466922000696024,123420074,466921602131264,20180208,15
```

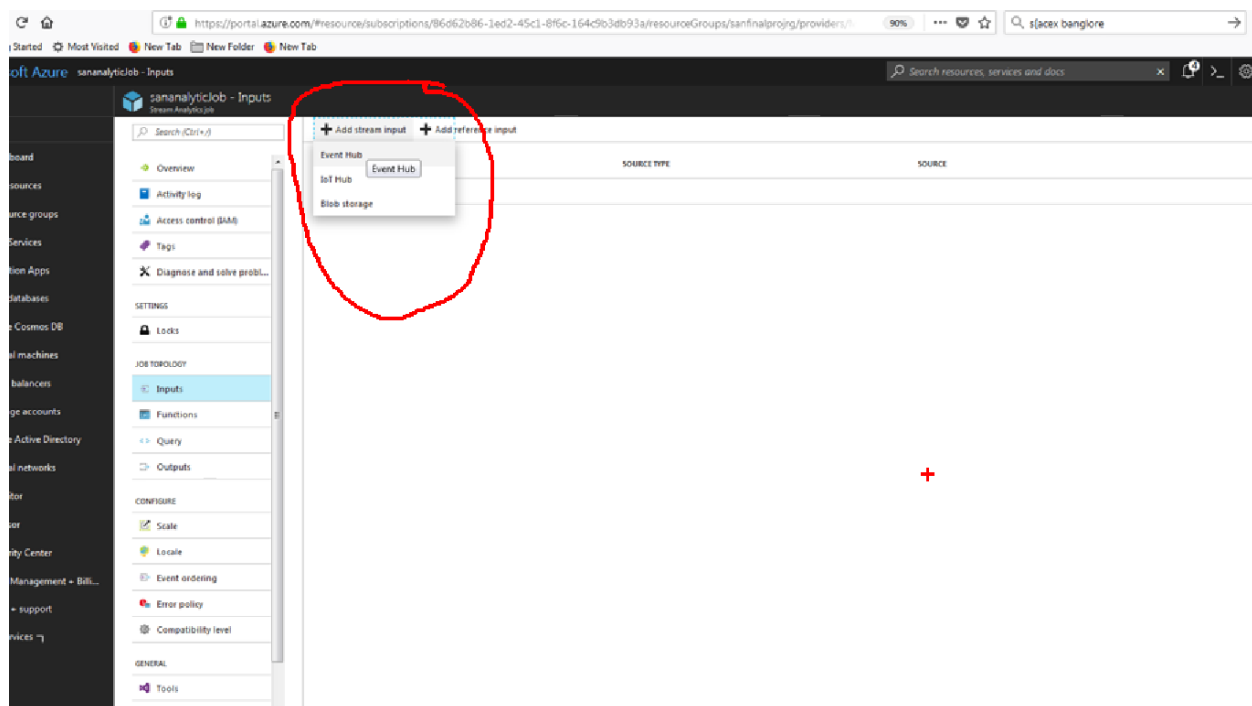
Next step is to create Stream Analytic job

In the Azure portal, click **New > Internet of Things > Stream Analytics** job. Provide job name, resource group and hit create button .

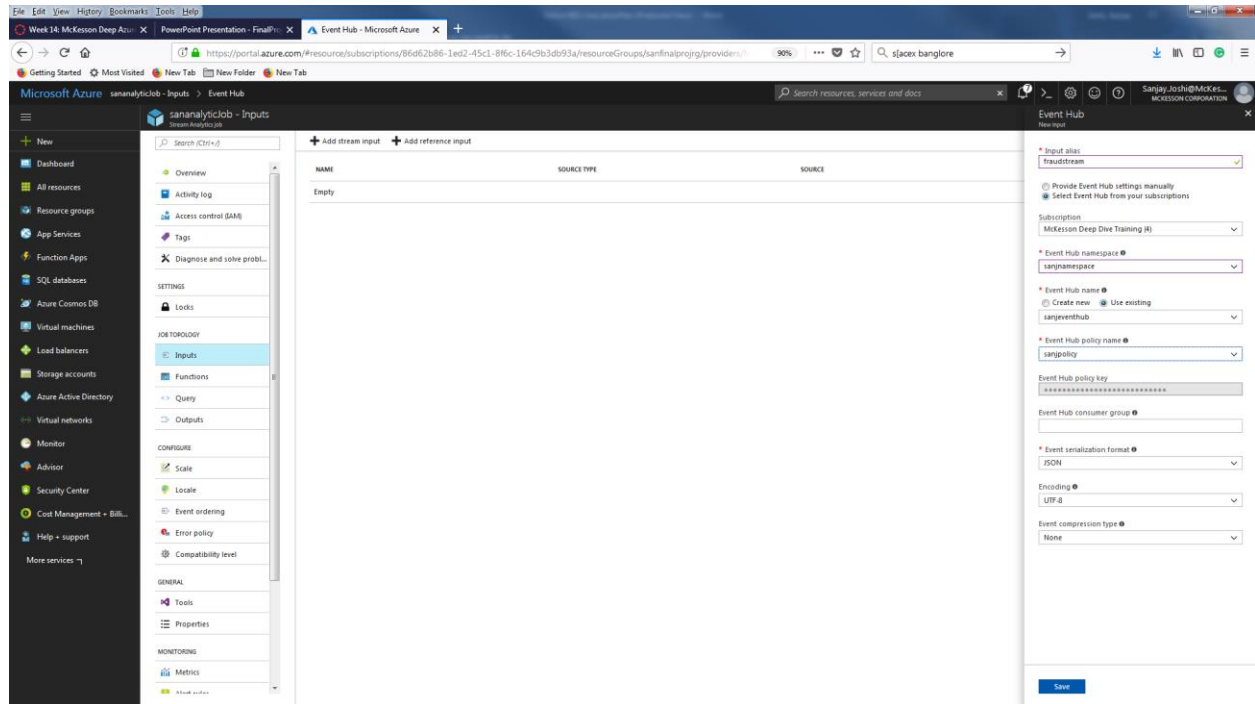
The following screen will appear



Under Job category click on input and click on +Add stream input and select eventhub



Provide input alias In this example 'fraudstream', select your event hub namespace you created before ('sanjamespace') choose your resource group and also select the policy you created when creating the event hub and hit save button . The screen will look like this



SOURCE

* Input alias
fraudstream ✓

☐ Provide Event Hub settings manually
☒ Select Event Hub from your subscriptions

Subscription
McKesson Deep Dive Training (4) ▼

* Event Hub namespace
sanjnamespace ▼

* Event Hub name ⓘ
☒ Create new ☐ Use existing
sanjevethub ▼

* Event Hub policy name ⓘ
sanjpolicy ▼

Event Hub policy key

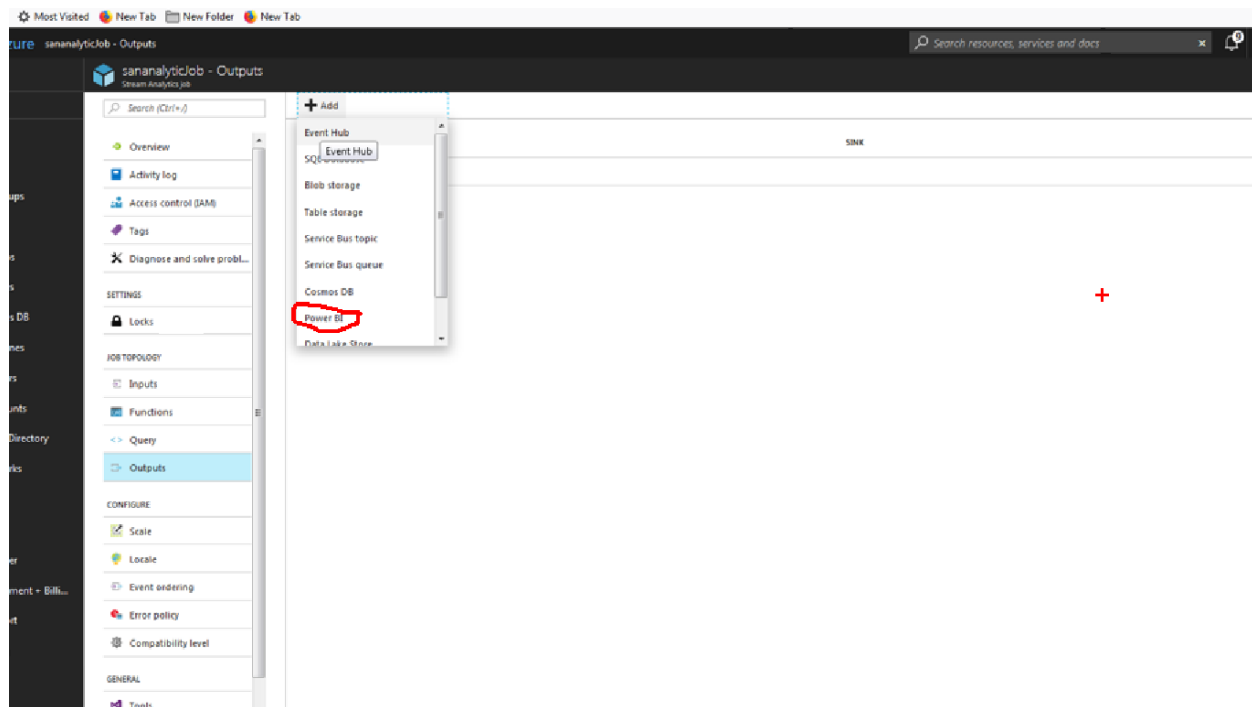
Event Hub consumer group ⓘ

* Event serialization format ⓘ
JSON ▼

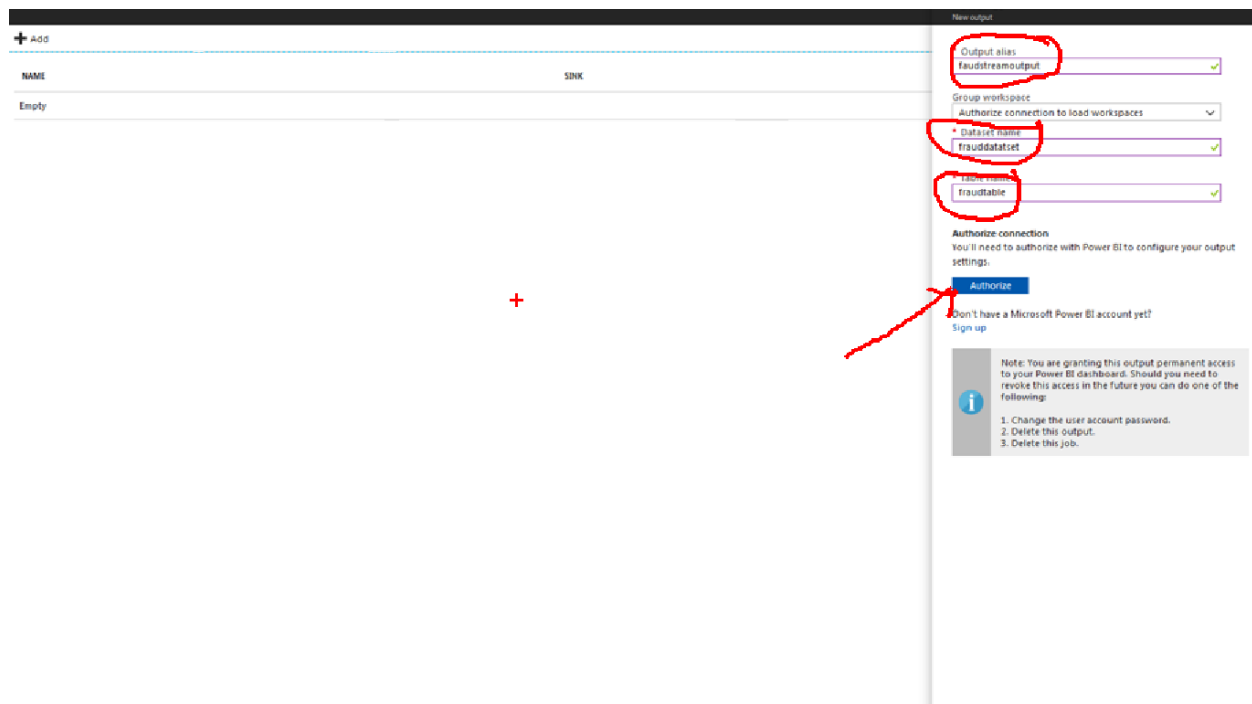
Encoding ⓘ
UTF-8 ▼

Event compression type ⓘ
None ▼

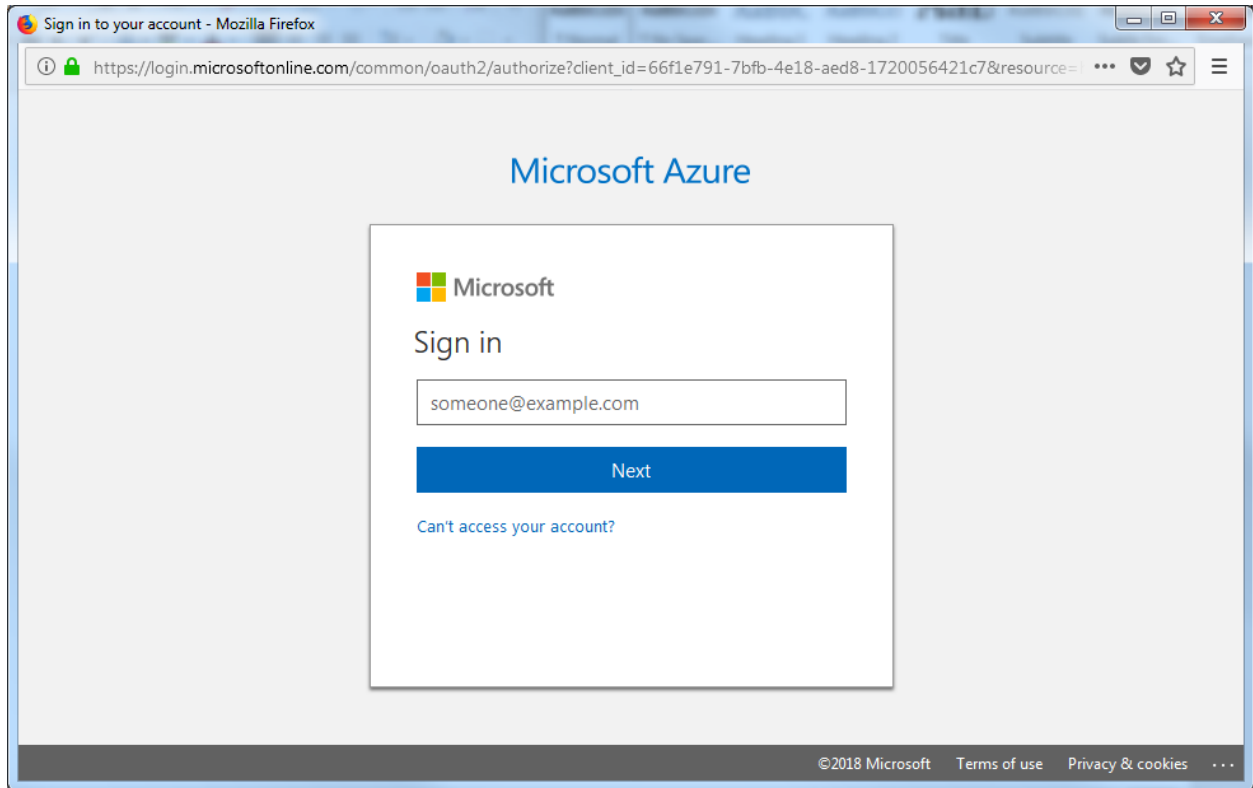
Now click on the output blade under job topology and hit the +add button and select the power BI



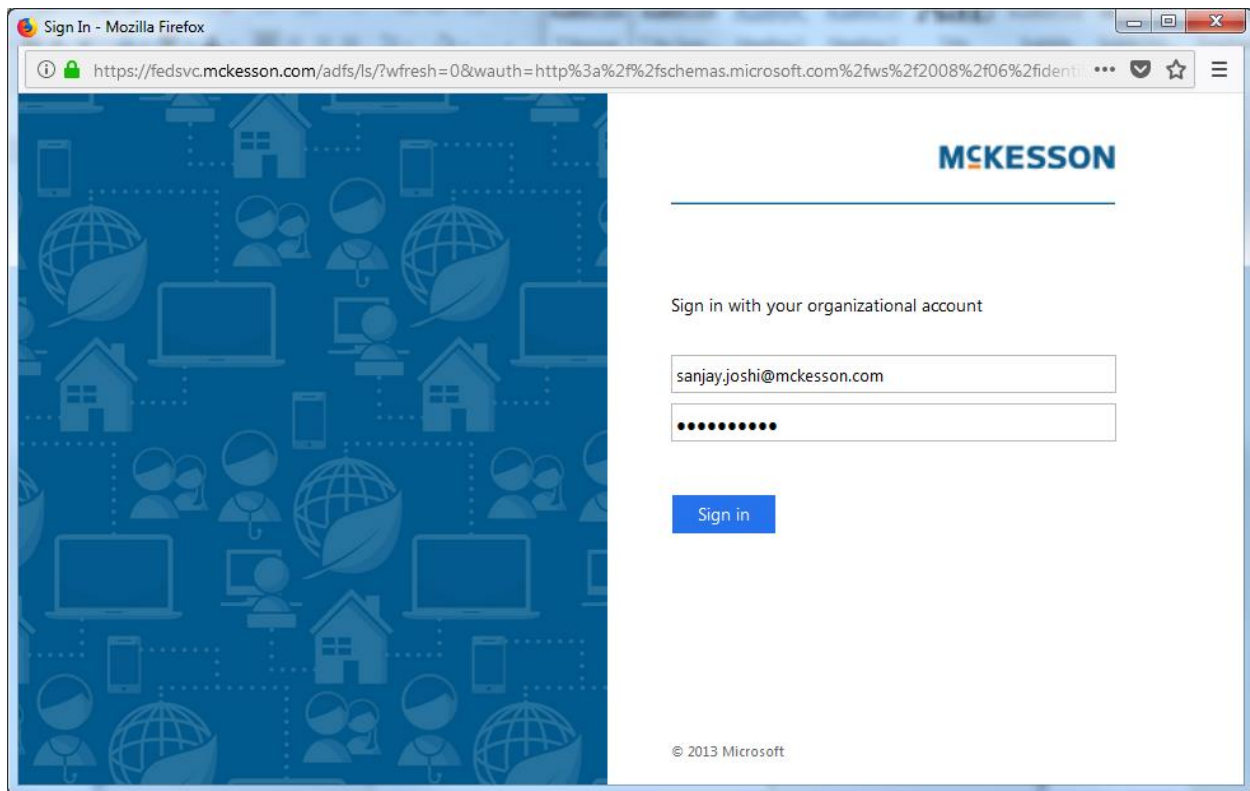
Provide output alias dataset name and table_name and hit the Authorize button to Authorize your Power bi portal credential .



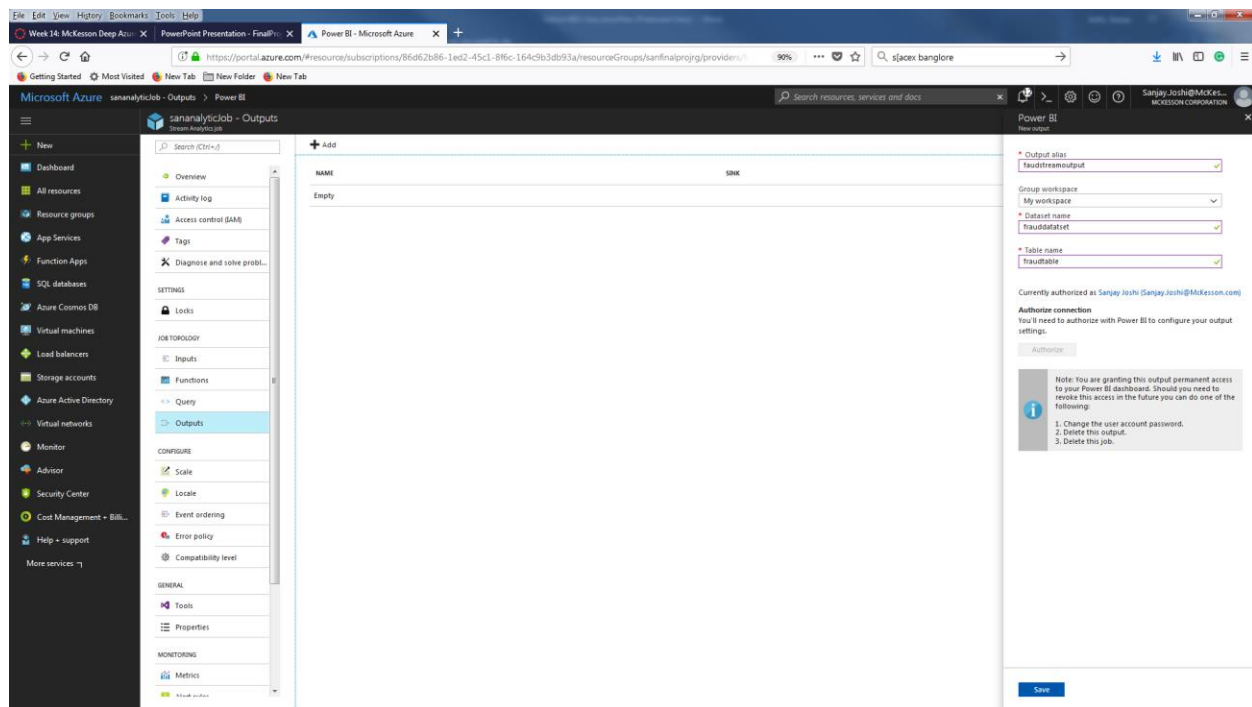
It will open a login window for power bi portal



Provide login as password and login to portal



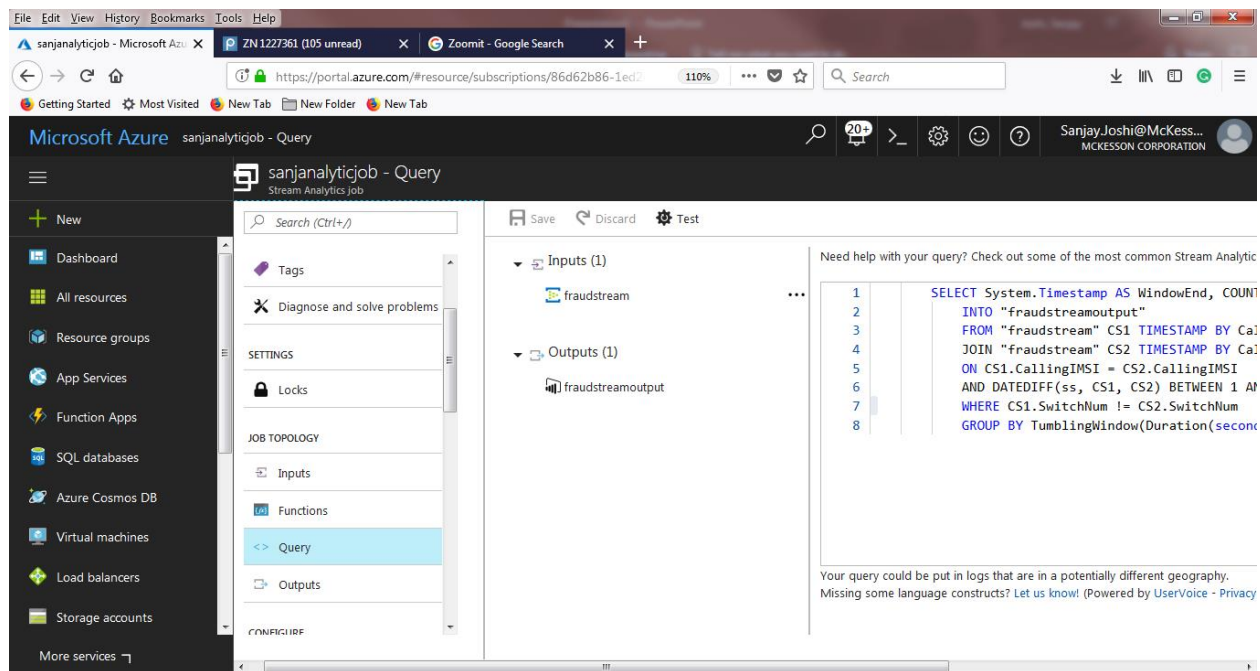
And following screen will appear . Hit the save button after it authorized for powerBI account.



Now click on the query page and paste the following query

```
SELECT System.Timestamp AS WindowEnd, COUNT(*) AS FraudulentCalls
INTO "fraudstreamoutput"
FROM "fraudstream" CS1 TIMESTAMP BY CallRecTime
JOIN "fraudstream" CS2 TIMESTAMP BY CallRecTime
ON CS1.CallingIMSI = CS2.CallingIMSI
AND DATEDIFF(ss, CS1, CS2) BETWEEN 1 AND 5
WHERE CS1.SwitchNum != CS2.SwitchNum
GROUP BY TumblingWindow(Duration(second, 1))
```

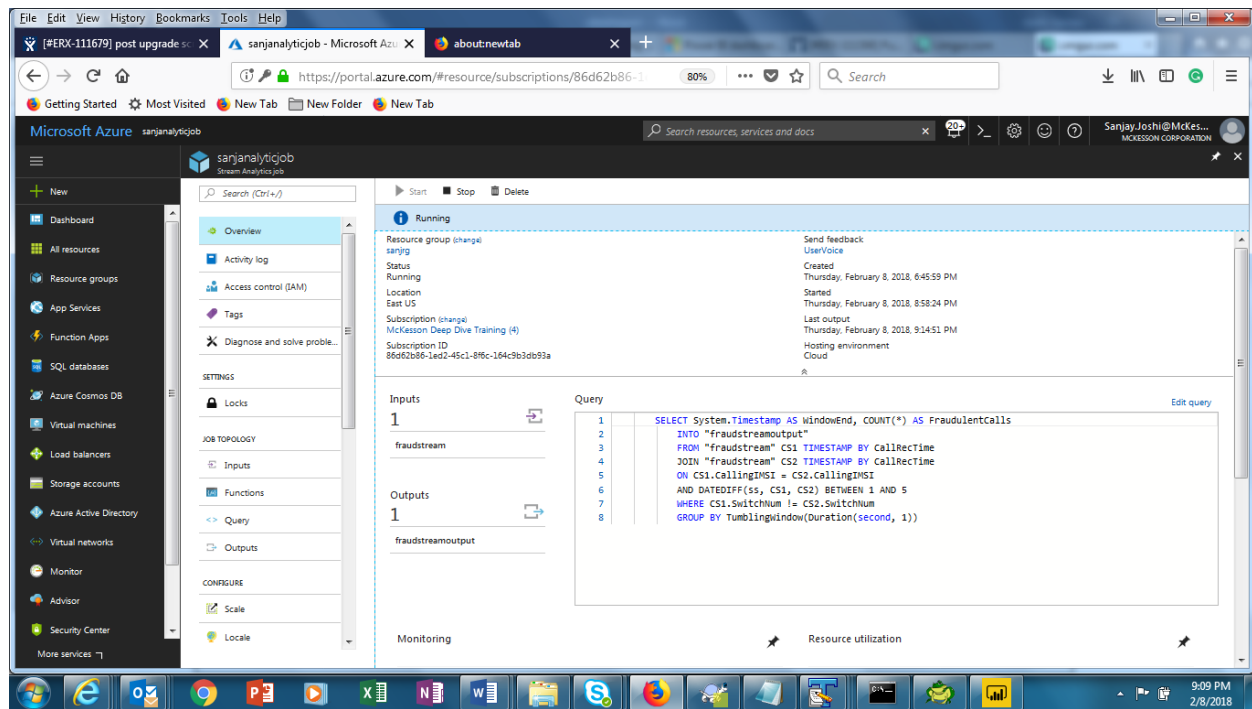
And it will look like this



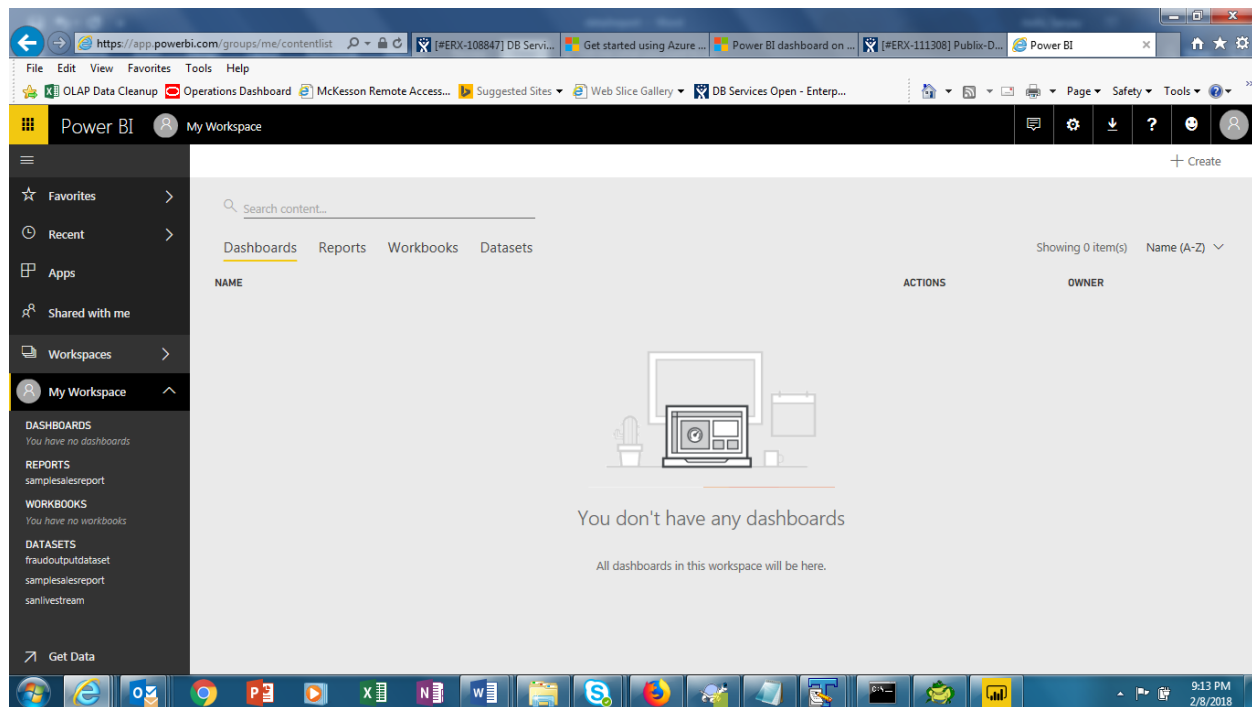
Hit the save button

Now go back to portal and select your Analytic stream job and start the analytic job by clicking start button

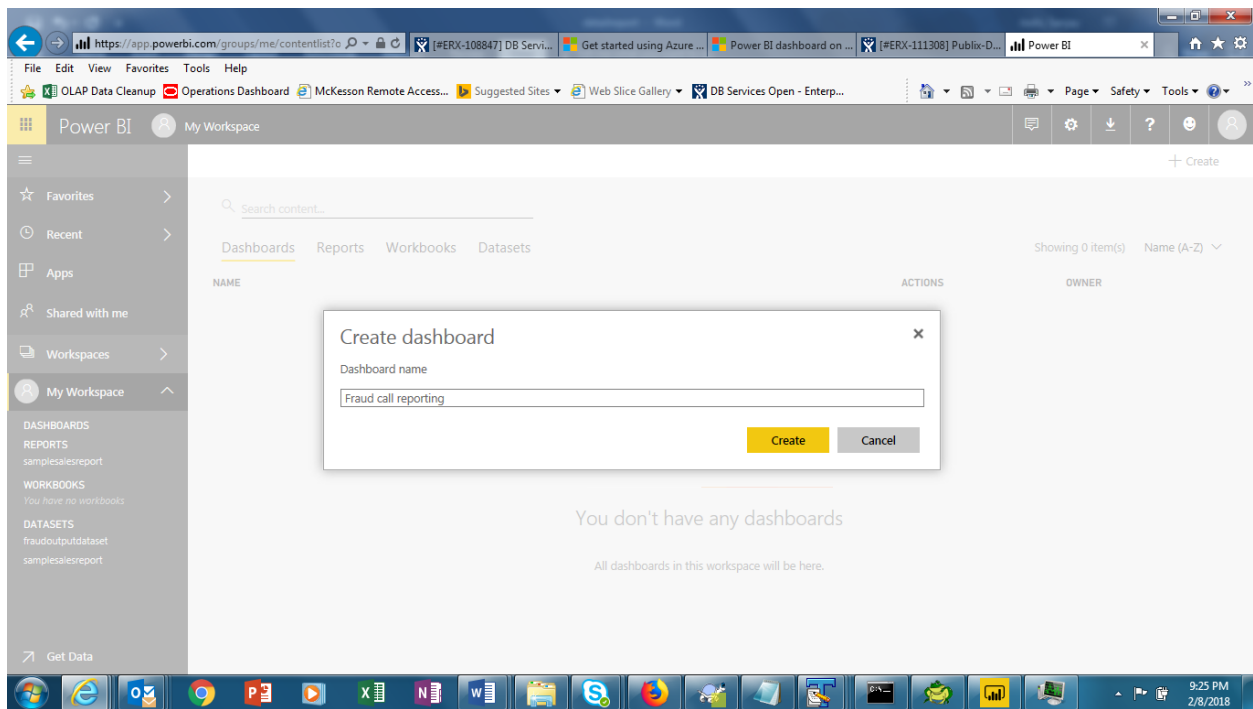
It will say starting and finally give you 'running' message like this



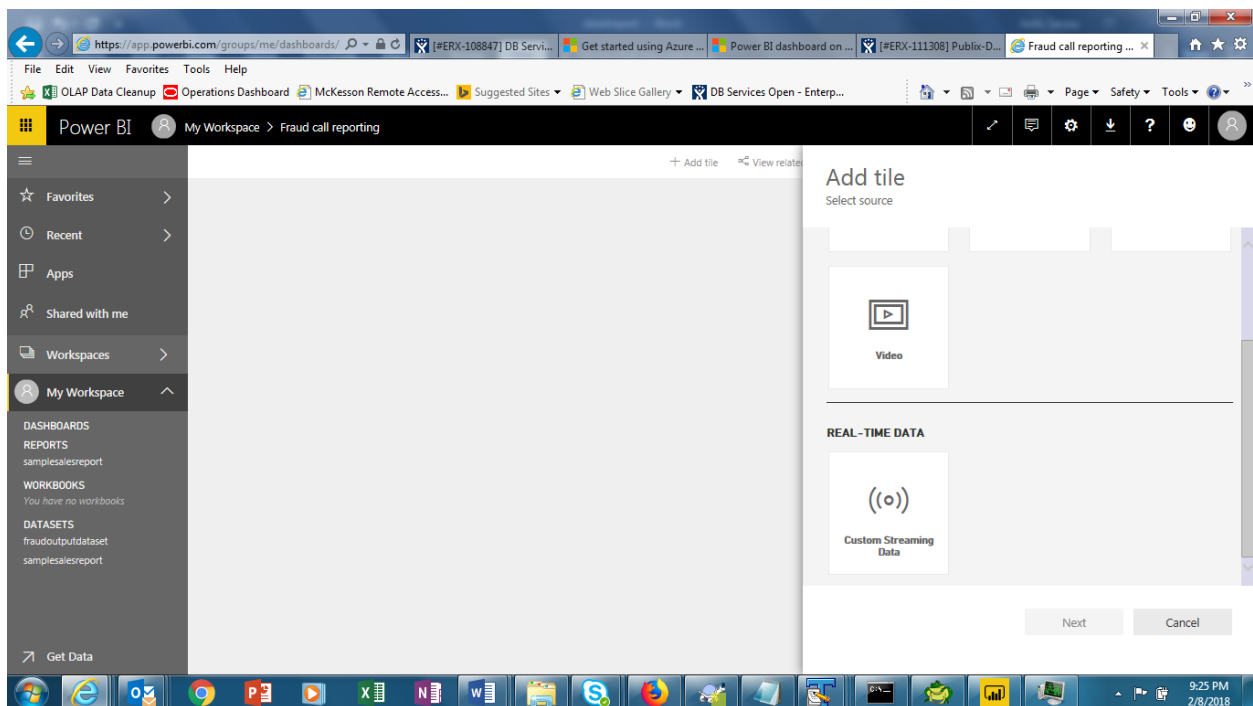
Login to powerbi portal



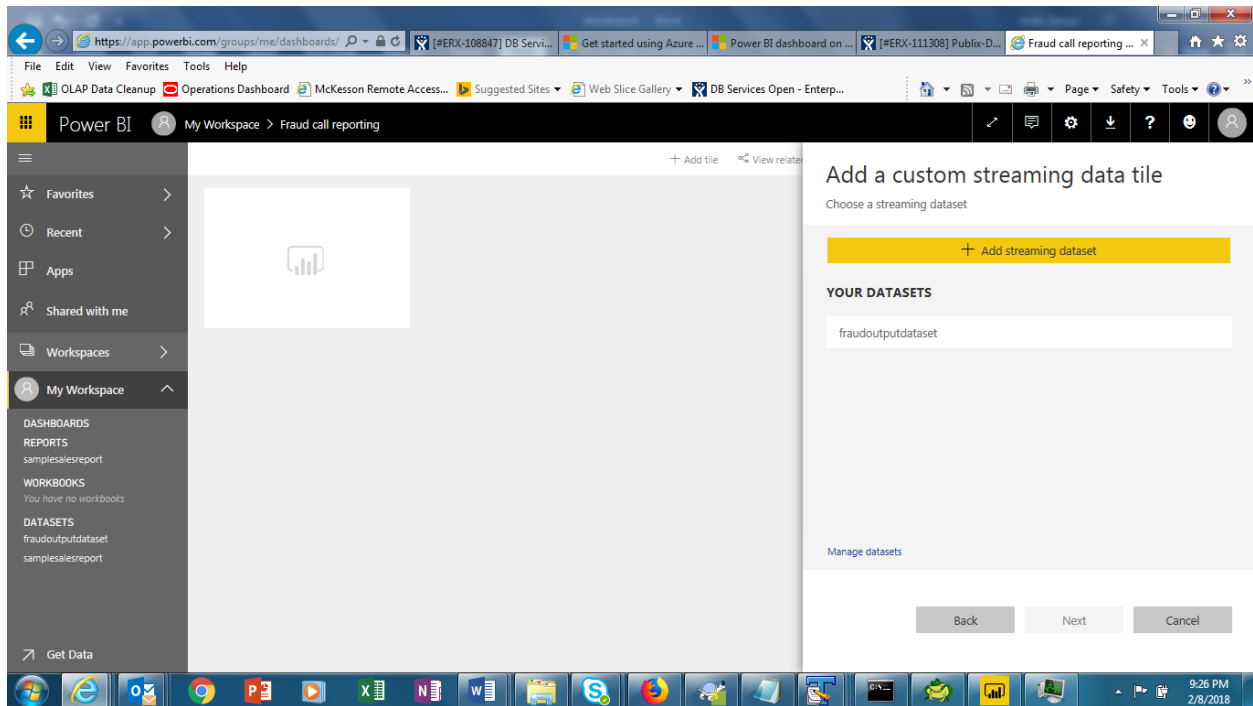
In Dashboard click + create and choose dashboard



And click on add tile on next page and choose real time data (custom streaming data)

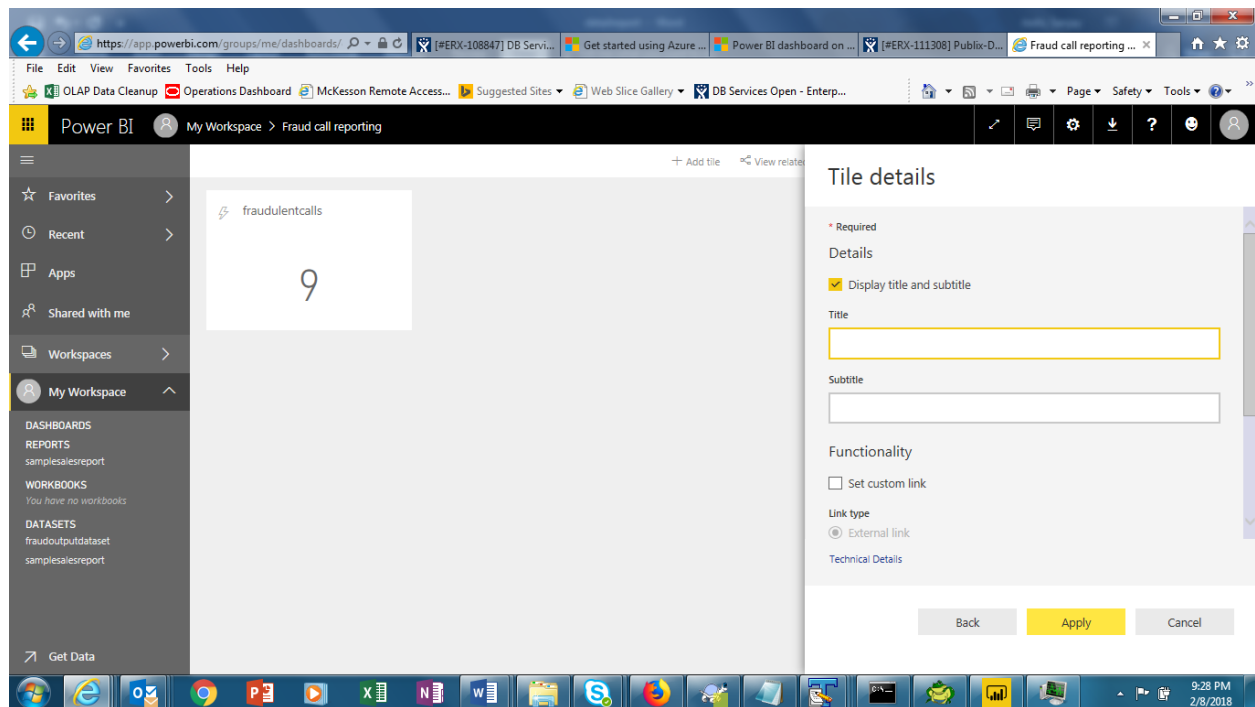


In next screen select your data set



And click on the value field and select the fraudulentcalls

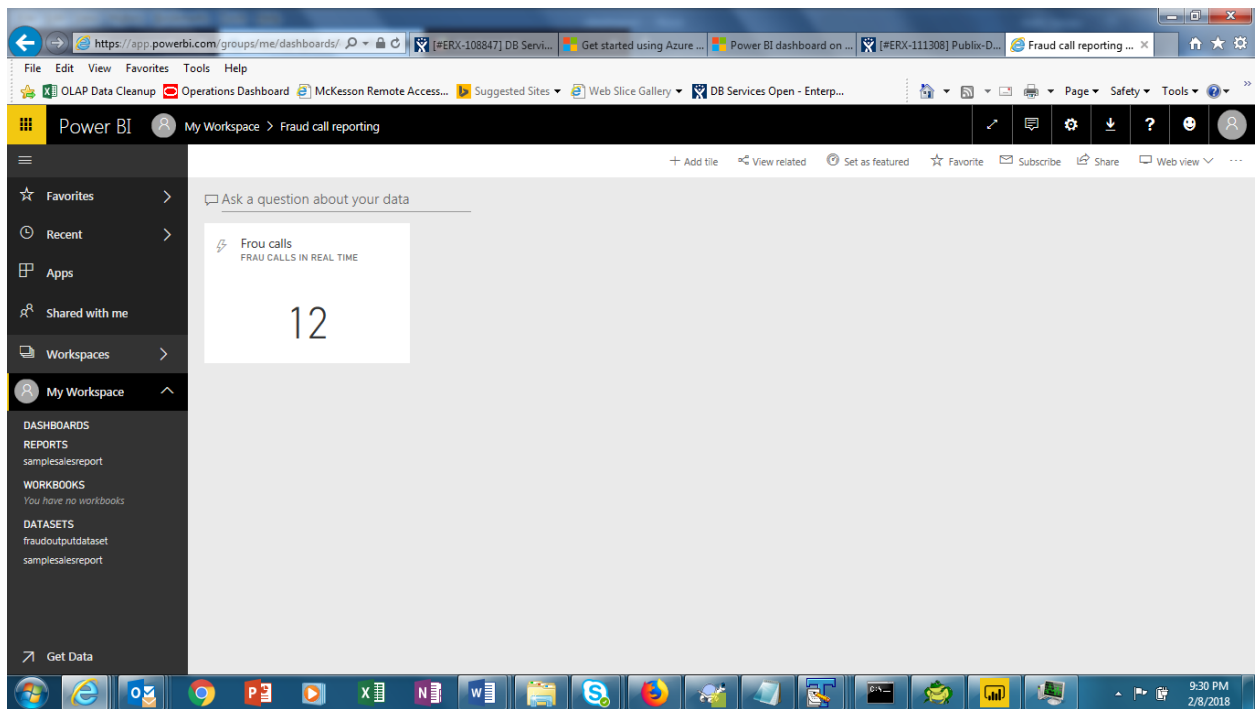
Provide tittles and subtitles in next screen



And click apply

It will show the fraud calls getting generated in real time in dashboard

Another screenshot immediately taken after 2 seconds



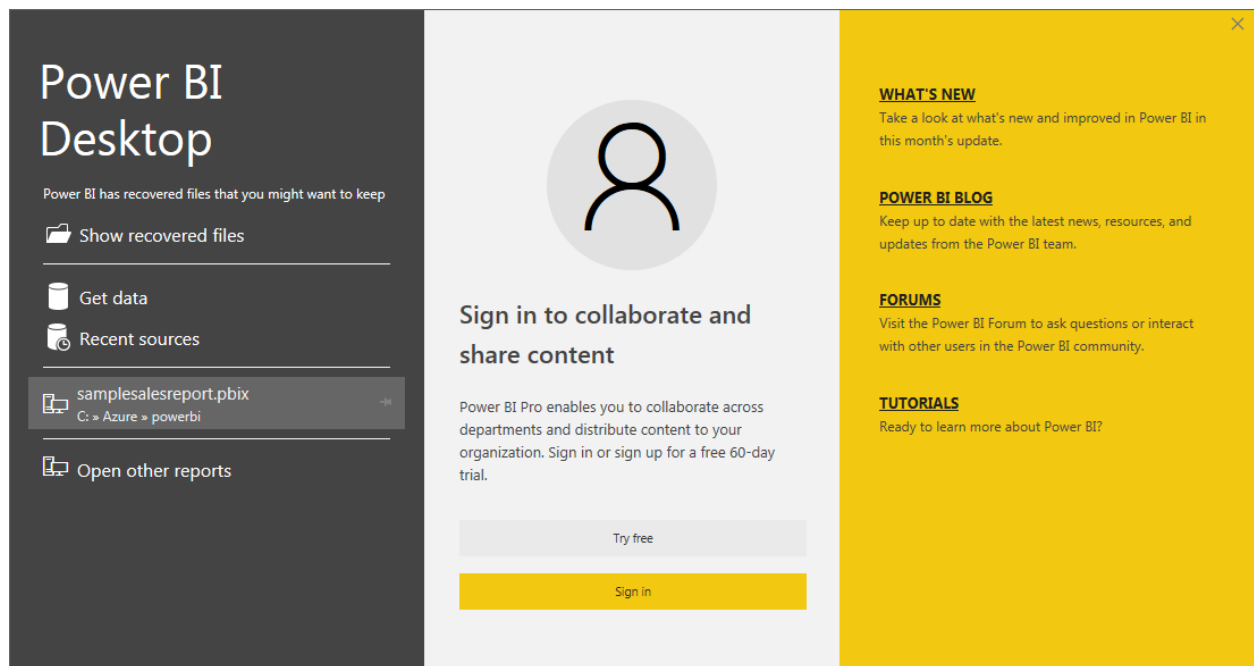
Demo2

(Desktop version with static data)

Another use of powerBI to create beautiful and easy to access dashboard with drill down capability .

Download the powerbi desktop version from Microsoft and install it

Launch the powerbi desktop



Get the sample sales data from <http://go.microsoft.com/fwlink/?LinkID=521962>

Save it to local folder and CSV file.

Then click on the get data on powerBI desktop and select your downloaded csv file

The data will look like this. This is sales data is by product, by county .

Auto recovery contains some recovered files that haven't been opened. View recovered files

Segment	Country	Product	Discount Band	Units Sold	Manufacturing Price	Sale Price	Gross Sales	Discounts	Sales	COGS	Profit	Date
Government	Canada	Carretera	None	1618.5	\$3	\$20	\$32,370		\$32,370	\$16,185	\$16,185	Wednesd
Government	Germany	Carretera	None	1321	\$3	\$20	\$26,420		\$26,420	\$13,210	\$13,210	Wednesd
Midmarket	France	Carretera	None	2178	\$3	\$15	\$32,670		\$32,670	\$21,780	\$10,890	Si
Midmarket	Germany	Carretera	None	888	\$3	\$15	\$13,320		\$13,320	\$8,880	\$4,440	Si
Midmarket	Mexico	Carretera	None	2470	\$3	\$15	\$37,050		\$37,050	\$24,700	\$12,350	Si
Government	Germany	Carretera	None	1513	\$3	\$350	\$529,550		\$529,550	\$393,380	\$136,170	Monday
Midmarket	Germany	Montana	None	921	\$5	\$15	\$13,815		\$13,815	\$9,210	\$4,605	Satur
Channel Partners	Canada	Montana	None	2518	\$5	\$12	\$30,216		\$30,216	\$7,554	\$22,662	Si
Government	France	Montana	None	1899	\$5	\$20	\$37,980		\$37,980	\$18,990	\$18,990	Si
Channel Partners	Germany	Montana	None	1545	\$5	\$12	\$18,540		\$18,540	\$4,635	\$13,905	Si
Midmarket	Mexico	Montana	None	2470	\$5	\$15	\$37,050		\$37,050	\$24,700	\$12,350	Si
Enterprise	Canada	Montana	None	2665.5	\$5	\$125	\$333,187.5		\$333,187.5	\$319,860	\$13,327.5	Ti
Small Business	Mexico	Montana	None	958	\$5	\$300	\$287,400		\$287,400	\$239,500	\$47,900	Fri
Government	Germany	Montana	None	2146	\$5	\$7	\$15,022		\$15,022	\$10,730	\$4,292	Monday
Enterprise	Canada	Montana	None	345	\$5	\$125	\$43,125		\$43,125	\$41,400	\$1,725	Tuesd
Midmarket	United States of America	Montana	None	615	\$5	\$15	\$9,225		\$9,225	\$6,150	\$3,075	Monday
Government	Canada	Paseo	None	292	\$10	\$20	\$5,840		\$5,840	\$2,920	\$2,920	Saturd
Midmarket	Mexico	Paseo	None	974	\$10	\$15	\$14,610		\$14,610	\$9,740	\$4,870	Saturd
Channel Partners	Canada	Paseo	None	2518	\$10	\$12	\$30,216		\$30,216	\$7,554	\$22,662	Si
Government	Germany	Paseo	None	1006	\$10	\$350	\$352,100		\$352,100	\$261,560	\$90,540	Si
Channel Partners	Germany	Paseo	None	367	\$10	\$12	\$4,404		\$4,404	\$1,101	\$3,303	Ti
Government	Mexico	Paseo	None	883	\$10	\$7	\$6,181		\$6,181	\$4,415	\$1,766	Fri

TABLE: financial sample (700 rows)

To allow drill down analysis click on the modelling and select new table to create a new table and paste the following code in expression window

```
Date_dim = ADDCOLUMNS ( CALENDAR (DATE(2000,1,1), DATE(2025,12,31)), "DateAsInteger", FORMAT
( [Date], "YYYYMMDD" ), "Year", YEAR ( [Date] ), "MonthNo", FORMAT ( [Date], "MM" ),
"YearMonthNo", FORMAT ( [Date], "YYYY/MM" ), "YearMonth", FORMAT ( [Date], "YYYY/mmm" ),
"MonthShort", FORMAT ( [Date], "mmm" ), "MonthLong", FORMAT ( [Date], "mmmm" ), "WeekNo",
WEEKDAY ( [Date] ), "WeekDay", FORMAT ( [Date], "dddd" ), "WeekDayShort", FORMAT ( [Date], "dddd"
), "Quarter", "Q" & FORMAT ( [Date], "Q" ), "YearQuarter", FORMAT ( [Date], "YYYY" ) & "/Q" & FORMAT
( [Date], "Q" ))
```


samplesalesreport - Power BI Desktop

Home Modeling Help

Manage Relationships New Measure New Column New Table Sort by Column Sort Data type: Format: Home Table: Data Category: Uncategorized Default Summarization: Don't summarize

Auto recovery contains some recovered files that haven't been opened. View recovered files

Date_dim = ADDCOLUMNS (CALENDAR (DATE(2000,1,1), DATE(2025,12,31)), "DateAsInteger", FORMAT ([Date], "YYYYMMDD"), "Year", YEAR ([Date]), "MonthNo", FORMAT ([Date], "MM"), "YearMonthNo", FORMAT ([Date], "YYYY/MM"), "YearMonth", FORMAT ([Date], "YYYY/mm"), "MonthShort", FORMAT ([Date], "mmm"), "MonthLong", FORMAT ([Date], "MMMM"), "WeekNo", WEEKDAY ([Date]), "WeekDay", FORMAT ([Date], "dddd"), "WeekDayShort", FORMAT ([Date], "ddd"), "Quarter", "Q" & FORMAT ([Date], "Q"), "YearQuarter", FORMAT ([Date], "YYYY") & "/" & "Q" & FORMAT ([Date], "Q"))

Date	DateAsInteger	Year	MonthNo	YearMonthNo	YearMonth	MonthShort	MonthLong	WeekNo	WeekDay	WeekDayShort	Quarter	YearQuarter
7/2/2000 12:00:00 AM	20000702	2000	07	2000/07	2000/Jul	Jul	July	1	Sunday	Sunday	Q3	2000/Q3
7/9/2000 12:00:00 AM	20000709	2000	07	2000/07	2000/Jul	Jul	July	1	Sunday	Sunday	Q3	2000/Q3
7/16/2000 12:00:00 AM	20000716	2000	07	2000/07	2000/Jul	Jul	July	1	Sunday	Sunday	Q3	2000/Q3
7/23/2000 12:00:00 AM	20000723	2000	07	2000/07	2000/Jul	Jul	July	1	Sunday	Sunday	Q3	2000/Q3
7/30/2000 12:00:00 AM	20000730	2000	07	2000/07	2000/Jul	Jul	July	1	Sunday	Sunday	Q3	2000/Q3
7/1/2001 12:00:00 AM	20010701	2001	07	2001/07	2001/Jul	Jul	July	1	Sunday	Sunday	Q3	2001/Q3
7/8/2001 12:00:00 AM	20010708	2001	07	2001/07	2001/Jul	Jul	July	1	Sunday	Sunday	Q3	2001/Q3
7/15/2001 12:00:00 AM	20010715	2001	07	2001/07	2001/Jul	Jul	July	1	Sunday	Sunday	Q3	2001/Q3
7/22/2001 12:00:00 AM	20010722	2001	07	2001/07	2001/Jul	Jul	July	1	Sunday	Sunday	Q3	2001/Q3
7/29/2001 12:00:00 AM	20010729	2001	07	2001/07	2001/Jul	Jul	July	1	Sunday	Sunday	Q3	2001/Q3
7/7/2002 12:00:00 AM	20020707	2002	07	2002/07	2002/Jul	Jul	July	1	Sunday	Sunday	Q3	2002/Q3
7/14/2002 12:00:00 AM	20020714	2002	07	2002/07	2002/Jul	Jul	July	1	Sunday	Sunday	Q3	2002/Q3
7/21/2002 12:00:00 AM	20020721	2002	07	2002/07	2002/Jul	Jul	July	1	Sunday	Sunday	Q3	2002/Q3
7/28/2002 12:00:00 AM	20020728	2002	07	2002/07	2002/Jul	Jul	July	1	Sunday	Sunday	Q3	2002/Q3
7/6/2003 12:00:00 AM	20030706	2003	07	2003/07	2003/Jul	Jul	July	1	Sunday	Sunday	Q3	2003/Q3
7/13/2003 12:00:00 AM	20030713	2003	07	2003/07	2003/Jul	Jul	July	1	Sunday	Sunday	Q3	2003/Q3
7/20/2003 12:00:00 AM	20030720	2003	07	2003/07	2003/Jul	Jul	July	1	Sunday	Sunday	Q3	2003/Q3
7/27/2003 12:00:00 AM	20030727	2003	07	2003/07	2003/Jul	Jul	July	1	Sunday	Sunday	Q3	2003/Q3
7/4/2004 12:00:00 AM	20040704	2004	07	2004/07	2004/Jul	Jul	July	1	Sunday	Sunday	Q3	2004/Q3
7/11/2004 12:00:00 AM	20040711	2004	07	2004/07	2004/Jul	Jul	July	1	Sunday	Sunday	Q3	2004/Q3
7/18/2004 12:00:00 AM	20040718	2004	07	2004/07	2004/Jul	Jul	July	1	Sunday	Sunday	Q3	2004/Q3

TABLE Date_dim (9497 rows)

FIELDS

Search

Date_dim

- Date
- DateAsInteger
- MonthLong
- MonthNo
- MonthShort
- Quarter
- WeekDay
- WeekDayShort
- WeekNo
- Year
- YearMonth
- YearMonthNo
- YearQuarter

financial sample

- Sales
- COGS
- Discount Band
- Discounts
- Gross Sales

It will create a date_dim table

We need to create a relationship between date_dim and sales table .Click on the manage relationship and click new to create relationship and select the date column from sales sample table to date_dim table and choose many to 1 (many records from sales table mapped to one record in date dim)

×

Edit relationship

Select tables and columns that are related.

financial sample

Manufacturing Price	Sale Price	Gross Sales	Discounts	Sales	COGS	Profit	Date
\$3	\$20	\$32,370	null	\$32,370	\$16,185	\$16,185	Wednesday, January 01, 2014
\$3	\$20	\$26,420	null	\$26,420	\$13,210	\$13,210	Wednesday, January 01, 2014
\$3	\$15	\$32,670	null	\$32,670	\$21,780	\$10,890	Sunday, June 01, 2014

Date_dim

Date	DateAsInteger	Year	MonthNo	YearMonthNo	YearMonth	MonthShort	MonthLong
7/2/2000 12:00:00 AM	20000702	2000	07	2000/07	2000/Jul	Jul	July
7/9/2000 12:00:00 AM	20000709	2000	07	2000/07	2000/Jul	Jul	July
7/16/2000 12:00:00 AM	20000716	2000	07	2000/07	2000/Jul	Jul	July

Cardinality

Cross filter direction

Many to one (*:1)

Single

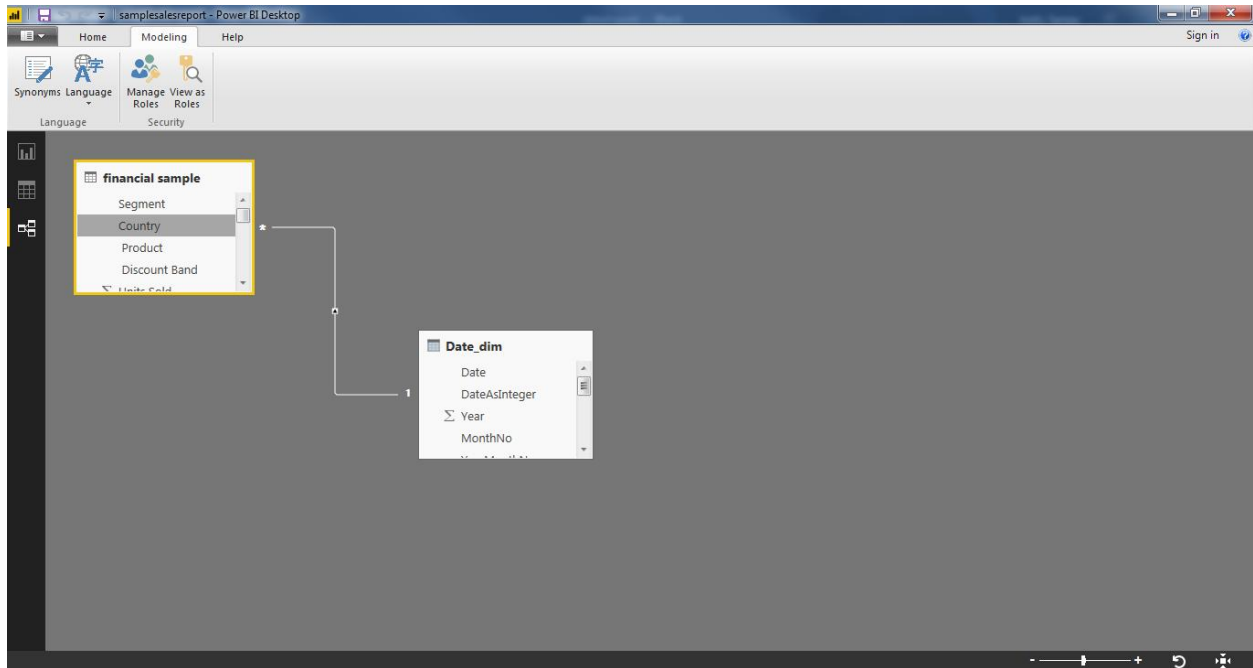
☒ Make this relationship active
 ☐ Apply security filter in both directions

☐ Assume referential integrity

OK

Cancel

And it will look like this



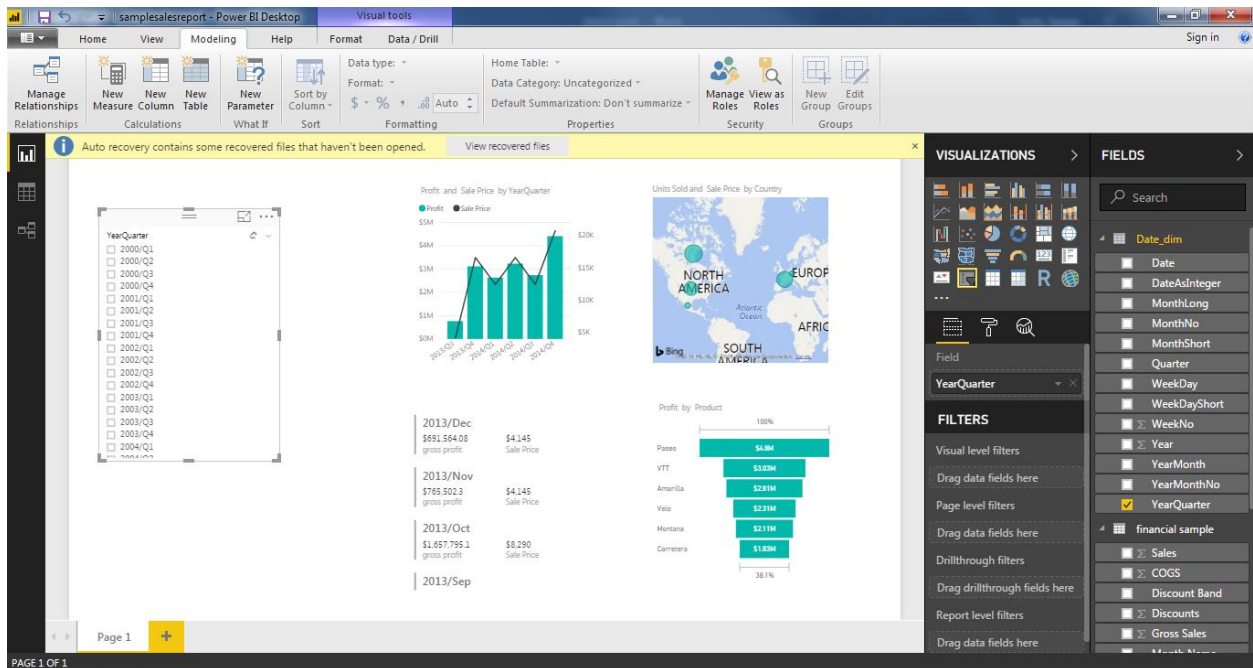
Now click on the graph icon and we have many options to create different graph under visualization , select map graph since we have data available for different country . On location put country and under size put sales.

Then choose funnel graph and under group select product and under value select profit .

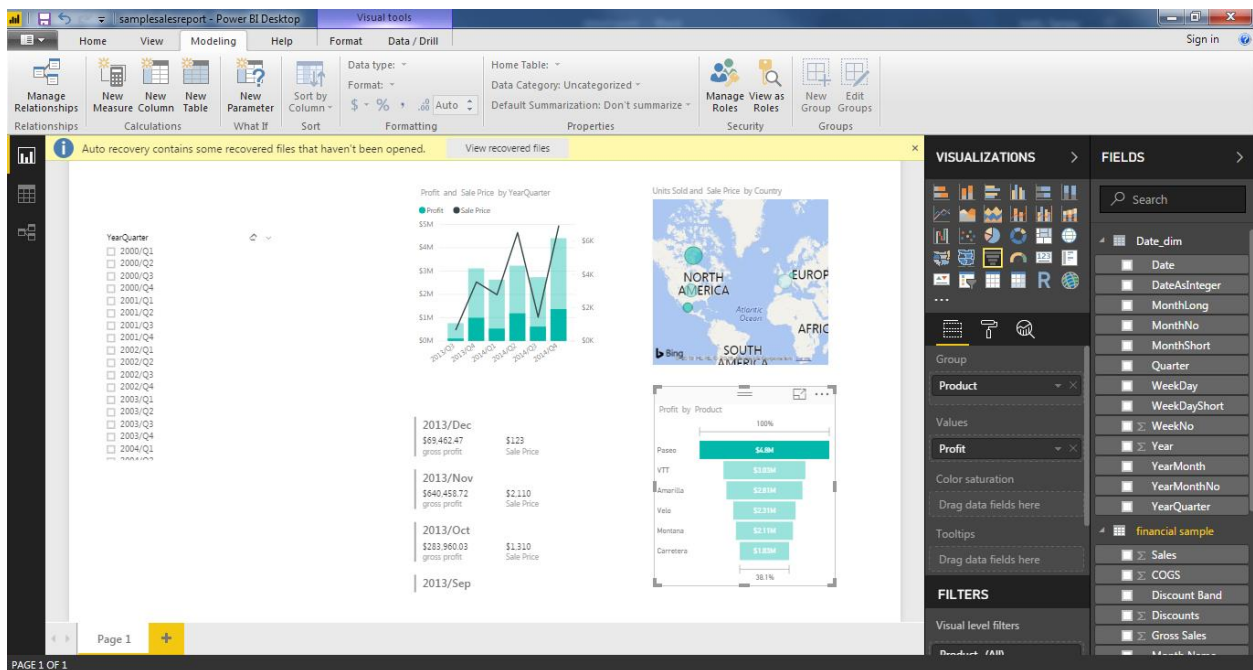
Similarly you can add other graphs in your dashboard by selecting different values.

Final choose the slicer to slice and dice the data and choose yearquarter from date_dim.

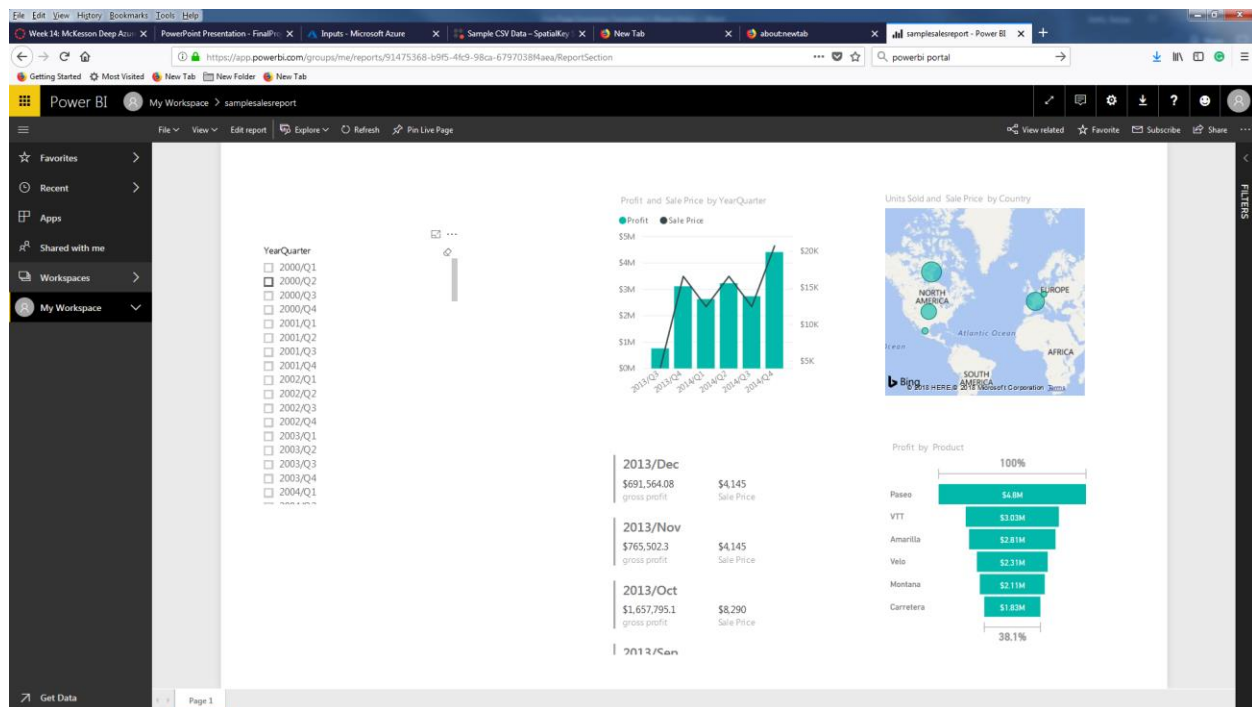
Your graph will finally look like this .



All the graph are interconnected means if you click on one particular value on any graph other graphs will change that value . for example on funnel bar I clicked on topmost product (paeso) and other graph changed



Finally you can click on file and click on publish to PowerBI portal .



YouTube URLs, GitHub URL, Last Page

- Two minute (<https://youtu.be/LD0UTGJpCoQ>):
- 15 minutes (<https://youtu.be/9eayxqI06UA>):
- GitHub Repository with all artifacts: (<https://github.com/sjoshi71/project>)