Data Warehousing and Data Mining in HCIS
Comp 309
week 5
(Reporting)

Fall : 2019

Professor : Mayy Habayeb

Email :mhabayeb@centennialcollege.ca

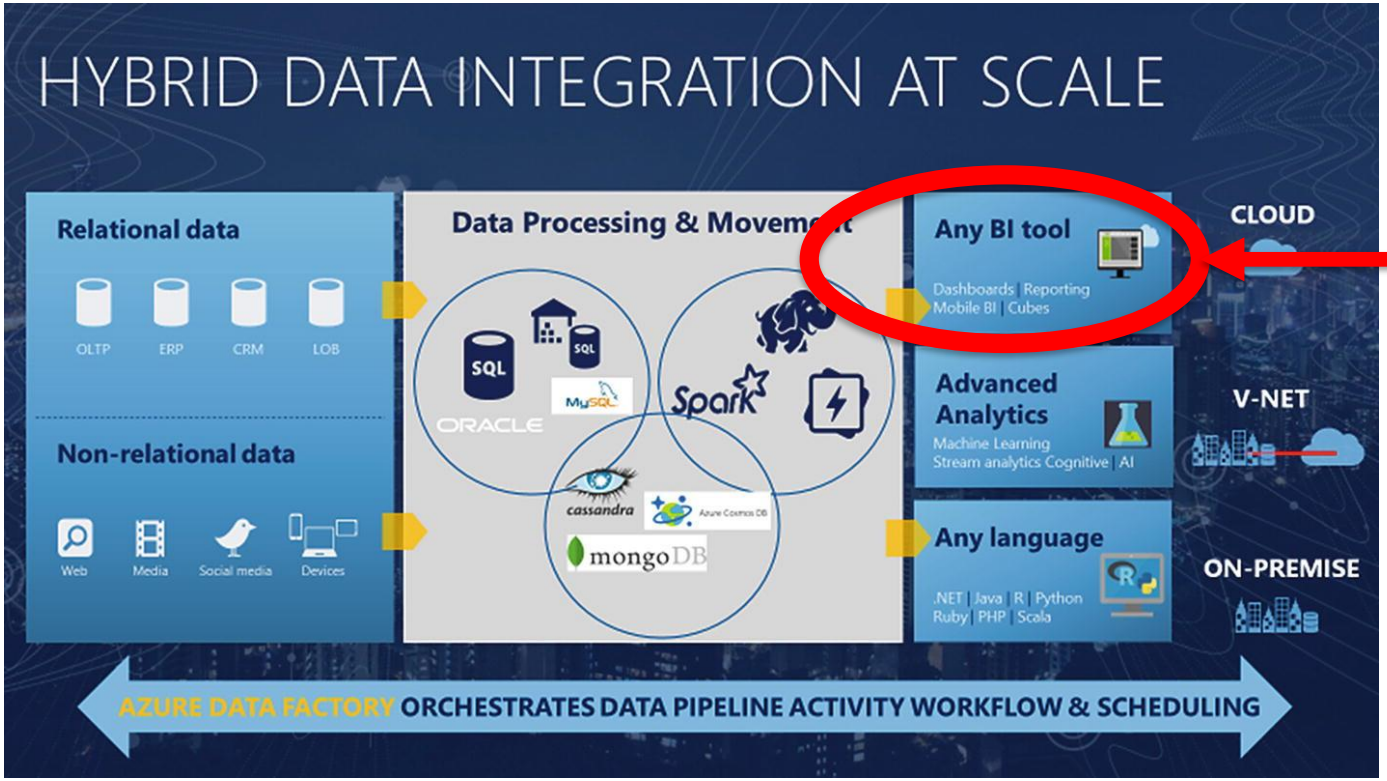
Agenda

- Reporting on the modern data warehouse
- What is Power BI?
- Types of reports

- Comparing Power BI Desktop and the Power BI service.
- Modelling view of Power BI
- Storage modes
- Aggregation
- Reports
- Power BI for developers

Reporting on the modern data Warehouse (BI tools)

Where does visualization and reporting fit ?

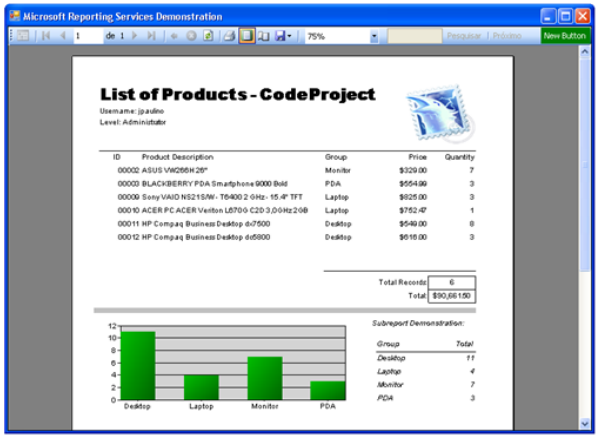


At the end of day you need to present your results

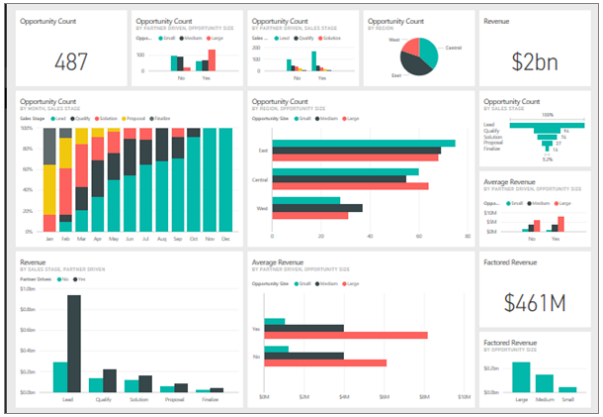
Results can presented as reports, dashboards or visualizations

Reporting on the modern data Warehouse (Information delivery)

Reporting : Reporting provides the ability to create formatted and interactive reports, with highly scalable distribution and scheduling capabilities. In addition, BI platform vendors should handle a wide array of reporting styles (for example, financial, operational and performance dashboards).



Dashboards: This subset of reporting includes the ability to publish formal, Web-based reports, with intuitive displays of information, including dials, gauges and traffic lights. These displays indicate the state of the performance metric compared with a goal or target value. Increasingly, dashboards are used to disseminate real-time data from operational applications.



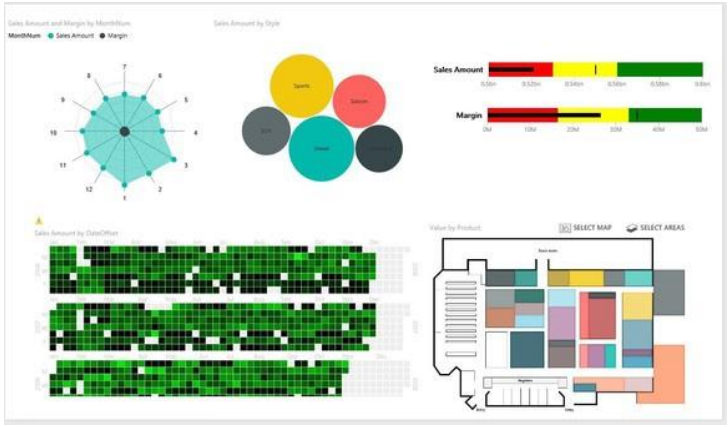
Ref: https://www.gartner.com/it-glossary/information-delivery

Reporting on the modern data Warehouse (Information delivery)

Ad hoc query: This capability, also known as self-service reporting, enables users to ask their own questions of the data, without relying on IT to create a report. In particular, the tools must have a robust semantic layer to allow users to navigate available data sources. In addition, these tools should offer query governance and auditing capabilities to ensure that queries perform well.

Visualization is the illustration of information objects and their relationships on a display. Strategic visualization graphically illustrates the strength of relationships by the proximity of objects on the display.

Advanced technology can make a significant difference in users' ability to interface to large knowledge repositories. These advances use the distance between objects on the display to reflect the similarity of meaning, similarity of content or other relationships (e.g., association with a group).



Ref: https://www.gartner.com/it-glossary/visualization
Ref: https://www.gartner.com/it-glossary/information-delivery

Business intelligence & visualization providers

Many providers

- QlikView
- Tableau
- Microsoft Power **BI**
- IBM Cognos **Analytics**
- Oracle **Analytics** Server
- Sisense
- Dundas **BI**
- Domo
- Zoho
- Google Analytics
- ……

What is Power BI?

Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights.

The data may be:

- An Excel spreadsheet
- A collection of cloud-based
- On-premises hybrid data warehouses.

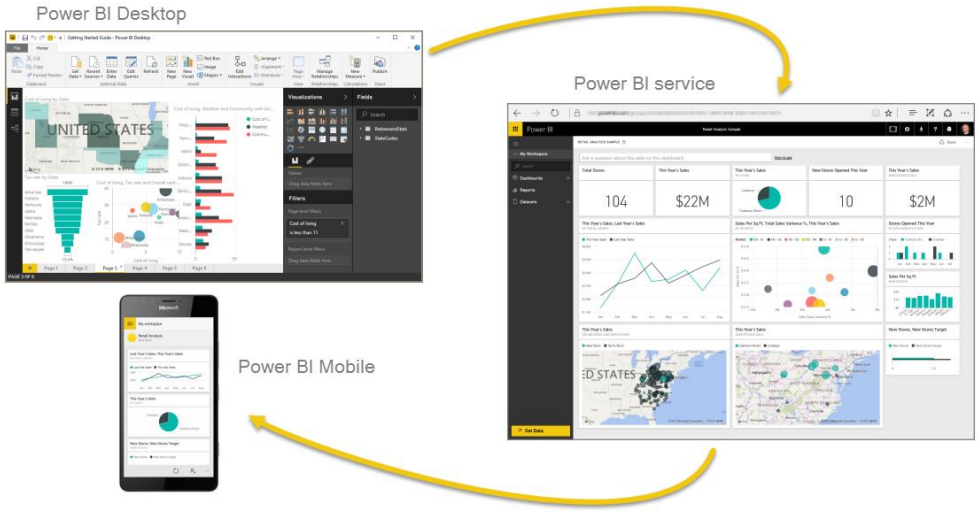
Power BI can be setup to connect to the various data sources, which allows for visualization and discovering what's important. Finally sharing the findings.

Ref:https://docs.microsoft.com/en-us/power-bi/service-service-vs-desktop

The parts of Power BI?

Power BI consists of:

1. A Windows desktop application called **Power BI Desktop**
2. An online SaaS (*Software as a Service*) service called the **Power BI service**
3. Power BI **mobile apps** for Windows, iOS, and Android devices
4. **Power BI Report Server**

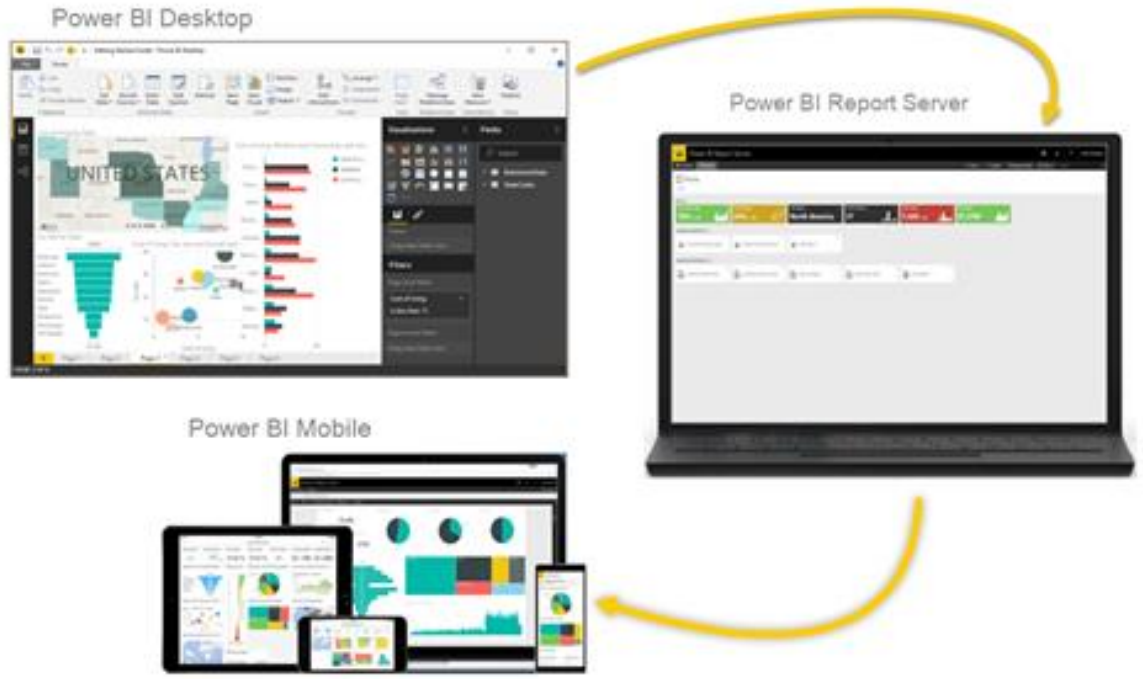


The first three parts are designed to let people create, share, and consume business insights in the way that serves them, or their role, most effectively.

The forth part is designed to allow allows people to publish Power BI reports to an on-premises report server, after creating them in Power BI Desktop.

Ref:https://docs.microsoft.com/en-us/power-bi/service-service-vs-desktop

Power BI report server



Power BI Report Server is a solution that is deployed behind the companies firewall and then delivers the various reports/ dashboards to the right users in different ways, whether that's viewing them in a web browser, on a mobile device, or as an email. And because Power BI Report Server is compatible with Power BI in the cloud, the reports can be moved to the cloud easily.

Ref:https://docs.microsoft.com/en-us/power-bi/service-service-vs-desktop

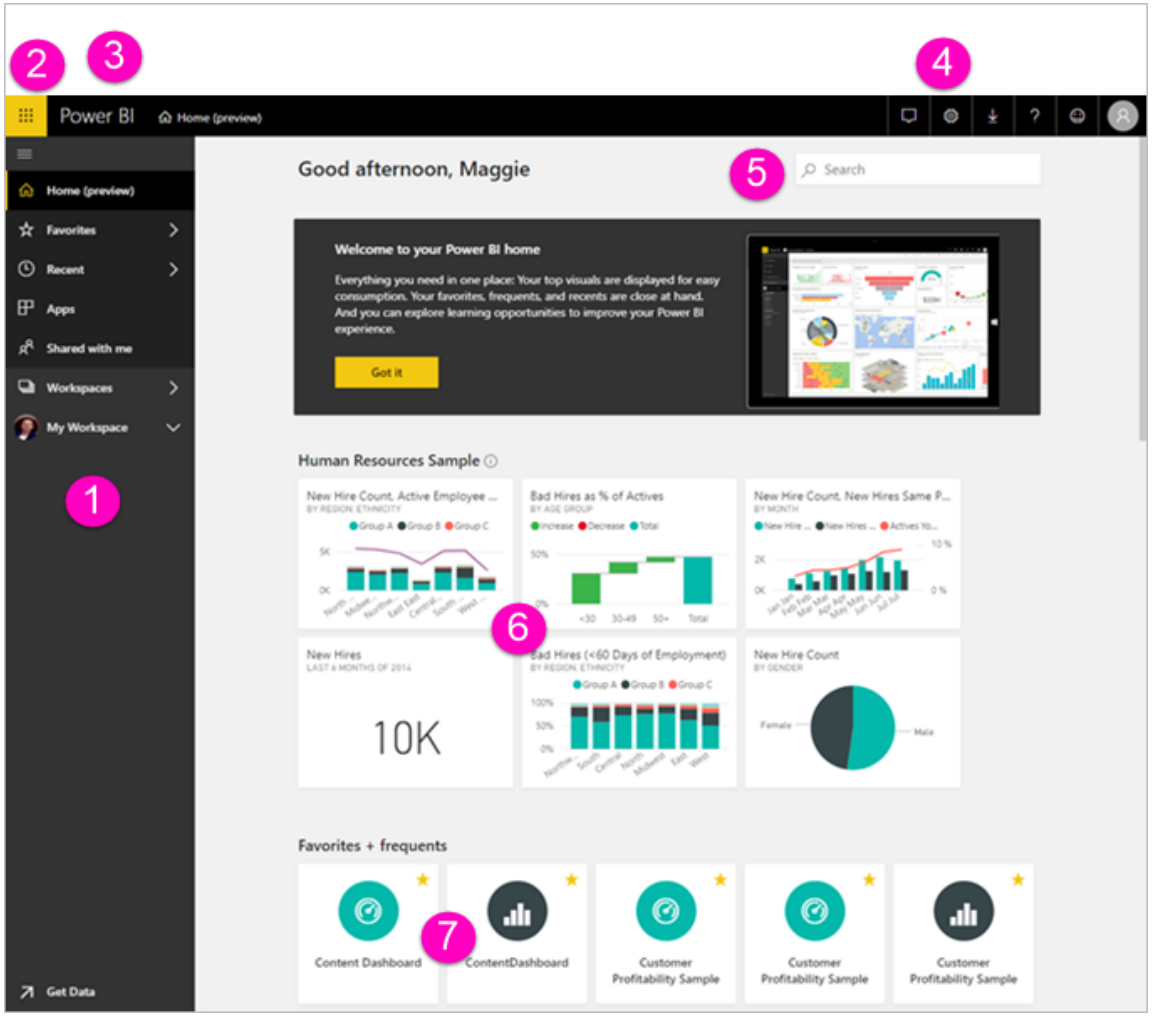
Power BI service

Power BI service is an online SaaS (*Software as a Service*)

https://app.powerbi.com/

When you open the Power BI service in a browser, you start at your Home screen. Here are the elements you may see:

1. Navigation pane (left nav)
2. Office 365 app launcher
3. Power BI home button
4. Icon buttons, including settings, help, and feedback
5. Search box
6. Tiles from a favorite dashboard
7. Favorite and frequent dashboards and reports



Ref :https://docs.microsoft.com/en-us/power-bi/service-basic-concepts

Power BI service concepts

The 4 major building blocks of Power BI are: ***dashboards***, ***reports***, ***workbooks***, and ***datasets***. And they're all organized into ***workspaces***.

Workspaces are containers for dashboards, reports, workbooks, and datasets in Power BI. There are two types of workspaces: *My workspace* and *app workspaces*.

My workspace	*App workspaces*
• Is the personal workspace for any Power BI customer to work with your own content. Only you have access to your My workspace. You can share dashboards and reports from your My Workspace. If you want to collaborate on dashboards and reports, or create an app, then you want to work in an app workspace.	• Are used to collaborate and share content with colleagues. They are also the places where you create, publish, and manage apps for your organization. Think of them as staging areas and containers for the content that will make up a Power BI app. You can add colleagues to your app workspaces and collaborate on dashboards, reports, workbooks, and datasets. • Apps are interactive but end users can't edit them.

Ref :https://docs.microsoft.com/en-us/power-bi/service-basic-concepts

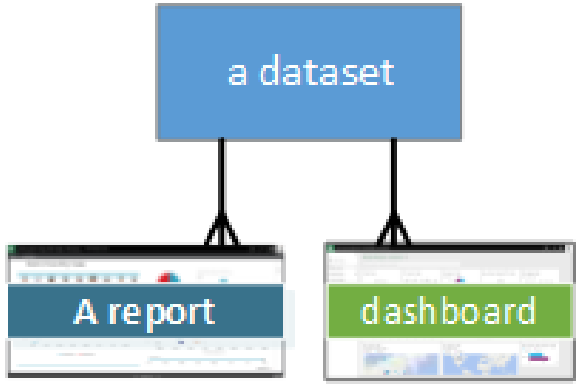
Power BI service concepts

A *dataset* is a collection of data that you *import* or *connect* to. Power BI lets you connect to and import all sorts of datasets and bring all of it together in one place.

Datasets are associated with *workspaces* and a single dataset can be part of many workspaces.

Each listed dataset represents a single source of data, for example, an Excel workbook on OneDrive, or an on-premises SSAS tabular dataset, or a Salesforce dataset.

ONE dataset...
- can be used over and over in one or in many workspaces.
- can be used in many different reports.
- Visualizations from that one dataset can display on many different dashboards.



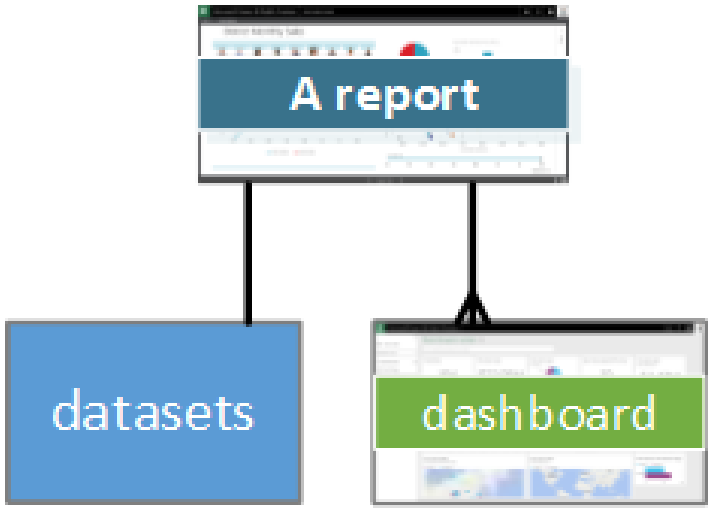
Ref :https://docs.microsoft.com/en-us/power-bi/service-basic-concepts

Power BI service concepts

A Power BI report is one or more pages of visualizations such as line charts, maps, and treemaps. Visualizations are also called ***visuals***. All of the visualizations in a report come from a single dataset.

ONE report...

- is contained in a single workspace.
- can be associated with multiple dashboards within that workspace. Tiles pinned from that one report can appear on multiple dashboards.
- can be created using data from one dataset. Power BI Desktop can combine more than one data source into a single dataset in a report, and that report can be imported into Power BI.



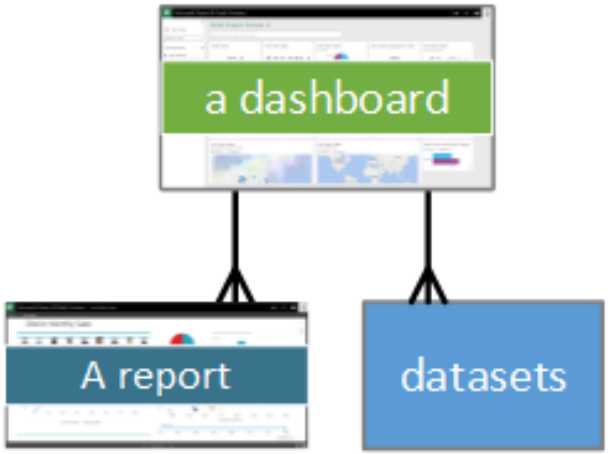
Ref :https://docs.microsoft.com/en-us/power-bi/service-basic-concepts

Power BI service concepts

A *dashboard* is something you create **in the Power BI service** or something a colleague creates **in the Power BI service** and shares with you. It is a single canvas that contains zero or more tiles and widgets. Each tile pinned from a report or from Q&A displays a single visualization that was created from a dataset and pinned to the dashboard. Entire report pages can also be pinned to a dashboard as a single tile.

ONE dashboard...

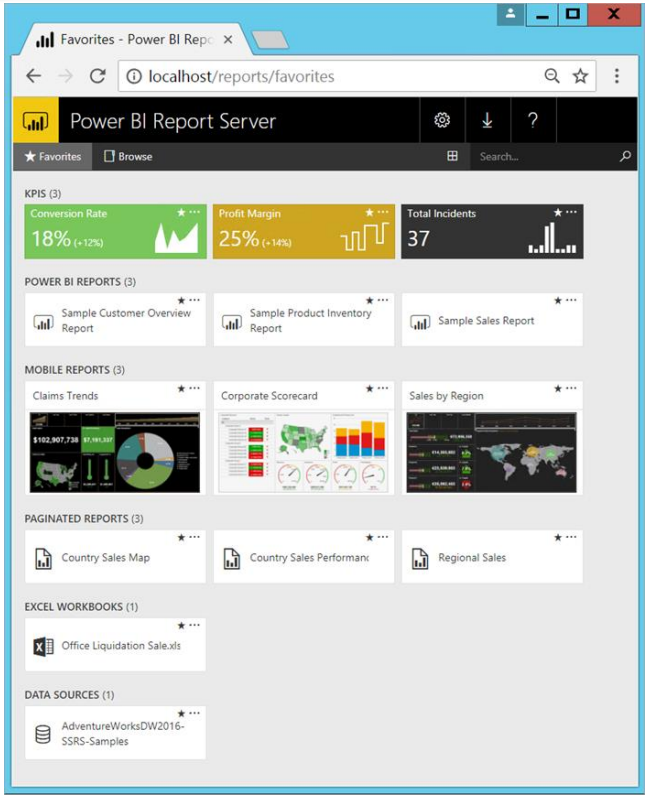
- is associated with a single workspace
- can display visualizations from many different datasets
- can display visualizations from many different reports
- can display visualizations pinned from other tools (for example, Excel)
-



Ref :https://docs.microsoft.com/en-us/power-bi/service-basic-concepts

Power BI report server

Power BI Report Server is an on-premises report server with a web portal in which reports and Key performance indicators KPIs can be display and managed.



Along with it come the tools to create Power BI reports, paginated reports, mobile reports, and KPIs.

Users can access those reports in different ways:

1. Viewing them in a web browser
2. Viewing them in a mobile device
3. Viewing them as an attachment to an e-mail

Power BI Report Server is available through two different licenses:

1. Power BI Premium
2. SQL Server Enterprise Edition with Software Assurance.

With a Power BI Premium license, a hybrid deployment is possible mixing cloud and on-premises.

Ref:https://docs.microsoft.com/en-us/power-bi/service-service-vs-desktop

Comparing Power BI Desktop and the Power BI service

Power BI Desktop is a complete data analysis and report creation tool. It can be installed free of charge on local computers. It can be setup to connect to many different sources of data, and combine them (often called modeling) into a data model.

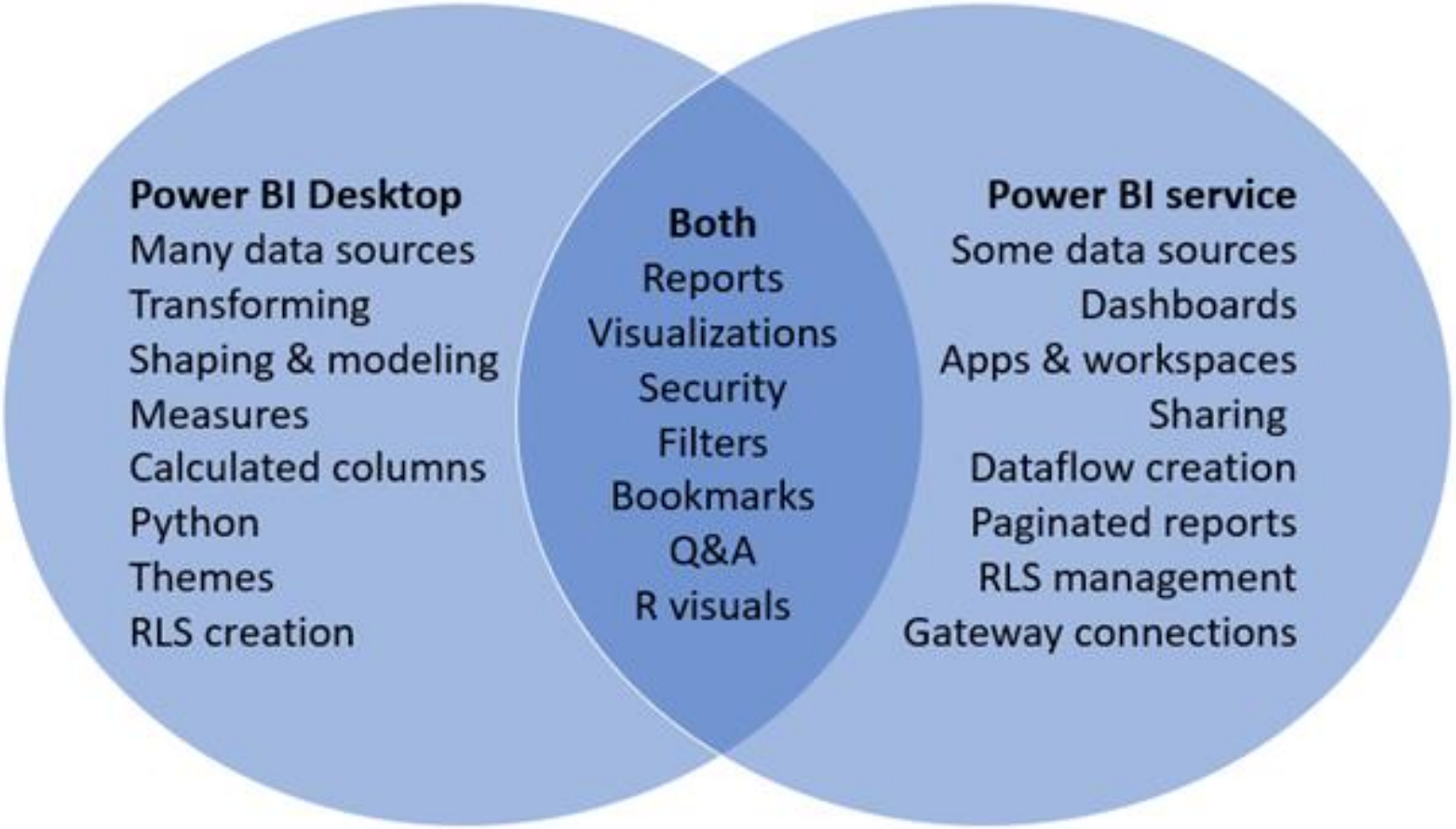
The **Power BI service** is a cloud-based, online service for light report editing and collaboration for teams and corporations. Through the service connections to several data sources is possible. but modeling is limited.

Most report designers who work on business intelligence projects use **Power BI Desktop** to create reports, and then use the **Power BI service** to share their reports with others.

Ref:https://docs.microsoft.com/en-us/power-bi/service-service-vs-desktop

Comparing Power BI Desktop and the Power BI service

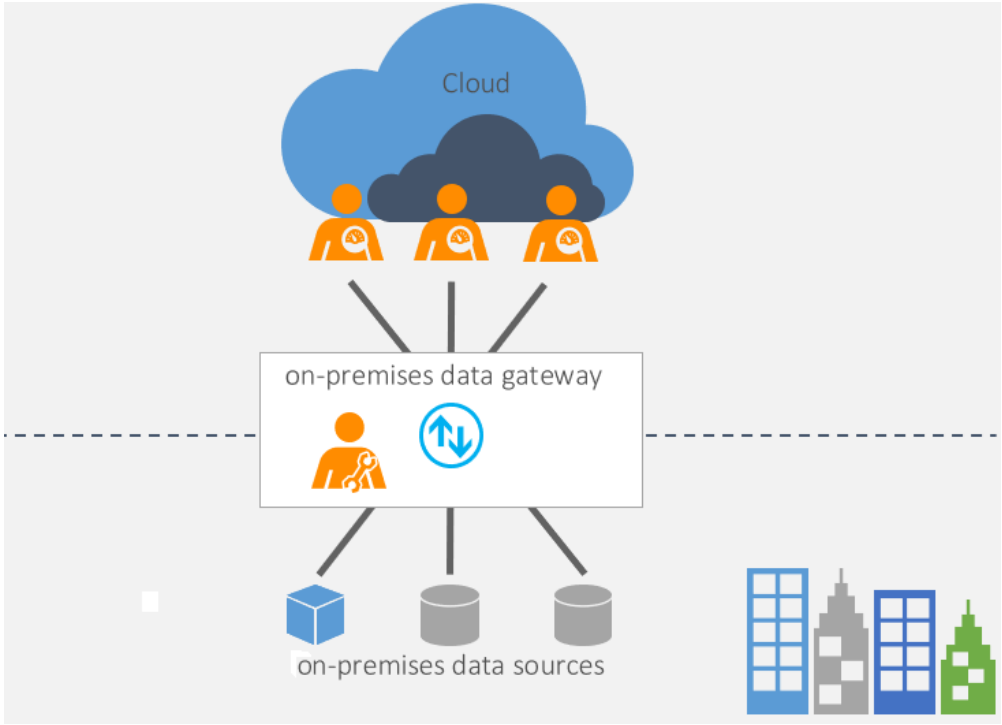
In a Venn diagram comparing Power BI Desktop and the Power BI service, the area in the middle would show how the two overlap..



Ref:https://docs.microsoft.com/en-us/power-bi/service-service-vs-desktop

Gateways

The on-premises data gateway acts as a bridge, providing quick and secure data transfer between on-premises data (data that is not in the cloud) and several Microsoft cloud services, including Power BI, PowerApps, Microsoft Flow, Azure Analysis Services, and Logic Apps. By using a gateway, organizations can keep databases and other data sources on their on-premises networks, yet securely use that on-premises data in cloud services.



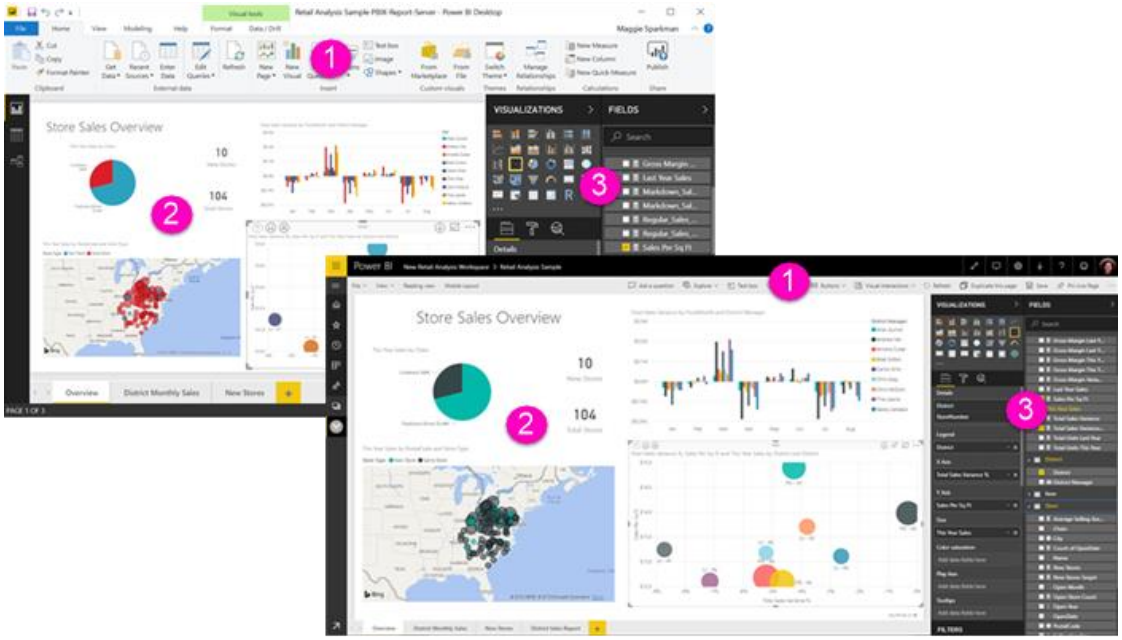
Ref:https://docs.microsoft.com/en-us/power-bi/service-gateway-onprem

Report editing

Both the application and the service, allow for building and editing *reports*.
A report can have one or many pages, with visuals and collections of visuals.

The report editors in Power BI Desktop and in the service are similar. They're made up of three sections:

1. The top navigation bars, different in Power BI Desktop and the service
2. The report canvas
3. The **Fields**, **Visualizations**, and **Filters** panes



Ref:https://docs.microsoft.com/en-us/power-bi/service-service-vs-desktop

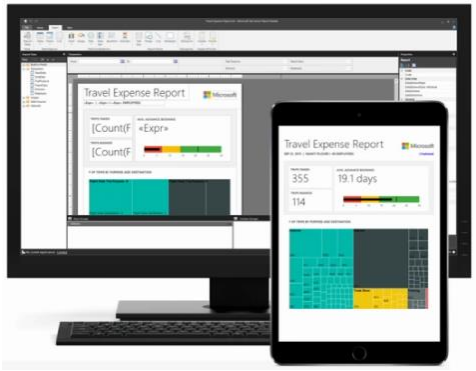
Types of reports

The following types of reports can be generated through power BI:

1. **Power BI reports**
2. **Paginated reports**
3. **Mobile reports**

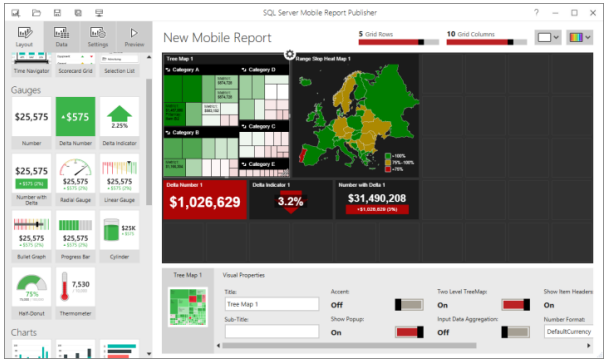
Power BI reports

. PBIX



Paginated reports

.RDL



Mobile reports

Ref:https://docs.microsoft.com/en-us/power-bi/service-service-vs-desktop

Types of reports – Power BI reports

1. Power BI reports

A Power BI report is a multi-perspective view into a data model, with visualizations that represent different findings and insights from that data model. A report can have a single visualization or pages full of visualizations. Depending on your role, you may read and explore reports, or you may create them for others.

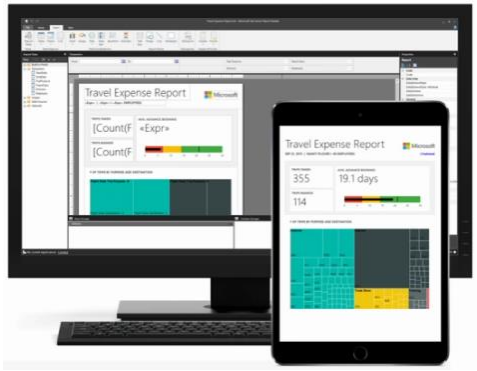
Power BI reports (.PBIX) are created with the version of Power BI Desktop optimized for the report server. Then published to the web portal in an environment.

Ref:https://docs.microsoft.com/en-us/power-bi/service-service-vs-desktop

Types of reports – Paginated reports

Paginated reports

Paginated reports (.RDL) are document-style reports with visualizations, in which tables expand horizontally and vertically to display all their data, continuing from page to page as needed. They're great for generating fixed-layout, pixel-perfect documents optimized for printing, such as PDF and Word files.



Paginated reports
.RDL

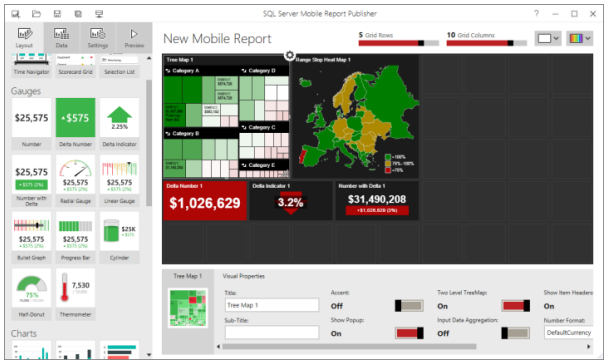
These are created using the Report Builder or the report designer in the SQL server tools SSDT.

Ref:https://docs.microsoft.com/en-us/power-bi/service-service-vs-desktop

Types of reports

Mobile reports

Mobile reports connect to on-premises data and have a responsive layout that adapts to different devices and the different ways you hold them. You create them with SQL Server Mobile Report Publisher.



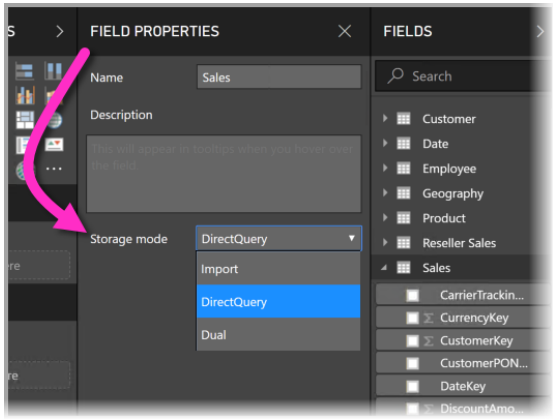
Mobile reports

Mobile reports feature an assortment of visualizations, from time, category, and comparison charts, to treemaps and custom maps.

Ref:https://docs.microsoft.com/en-us/power-bi/service-service-vs-desktop

Storage mode in Power BI Desktop

In Microsoft Power BI Desktop, the *storage mode* of tables can be specified. *Storage mode* allows the control of whether Power BI Desktop caches table data in-memory for reports.



There are three values for storage mode:

Import: When the value is set to **Import**, imported tables are cached. Queries submitted to the Power BI dataset that return data from Import tables can be fulfilled only from cached data.

DirectQuery: With this setting, DirectQuery tables aren't cached. Queries that are submited to the Power BI dataset and that return data from DirectQuery tables can be fulfilled only by executing on-demand queries to the data source. Queries that you submit to the data source use the query language for that data source - for example, SQL.

Dual: Dual tables can act as either cached or not cached, depending on the context of the query that's submitted to the Power BI dataset. In some cases, you fulfill queries from cached data. In other cases, you fulfill queries by executing an on-demand query to the data source.

Ref: https://docs.microsoft.com/en-us/power-bi/desktop-storage-mode

Storage mode –Direct query

The best practice for using **import** versus **DirectQuery** is the following

You should **import** data to Power BI wherever possible. Import takes advantage of the high-performance query engine of Power BI, and provides a highly interactive and fully featured experience over your data.
If your goals can't be met by importing data, then consider using **DirectQuery**. For example, if the data is changing frequently and reports must reflect the latest data, DirectQuery may be best. However, using DirectQuery is only feasible when the underlying data source can provide interactive queries (less than 5 seconds) for the typical aggregate query, and is able to handle the query load that will be generated. Additionally, the list of limitations that accompany use of DirectQuery should be considered carefully.

Ref: https://docs.microsoft.com/en-us/power-bi/desktop-directquery-aboute

Storage mode –Direct query

When is DirectQuery useful?

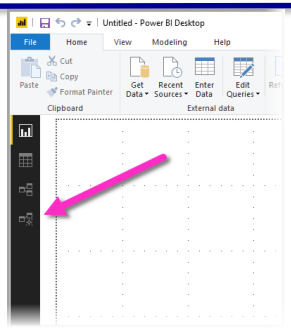
1. Data is changing frequently, and near ‘real-time’ reporting is needed.
2. Data is very large.
3. Security rules are defined in the underlying source.
4. Data sovereignty restrictions apply.
5. Underlying data source is an OLAP source, containing measures.

Implications of using DirectQuery

1. Performance and load on the underlying source
2. Security implications when combining data sources
3. Limited data transformations
4. Modeling limitations
5. Reporting limitations

Ref: https://docs.microsoft.com/en-us/power-bi/desktop-directquery-aboute

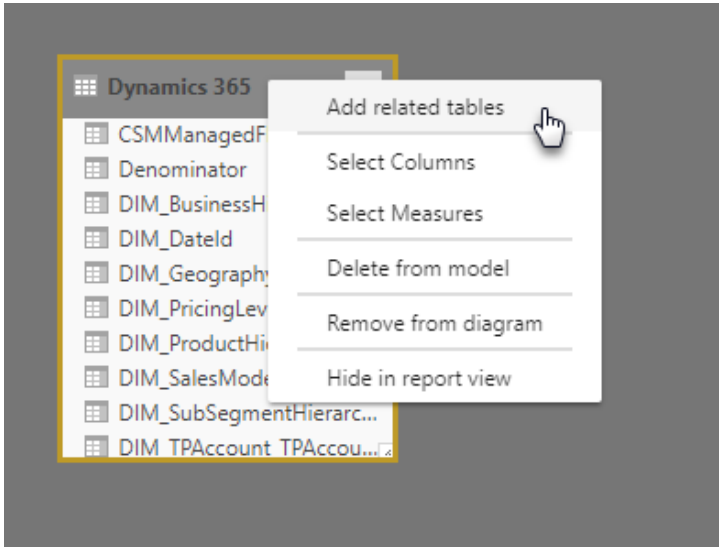
Modelling view of power BI



With the **Modeling view** in **Power BI Desktop** complex datasets that contain many tables can be viewed, relationships can be updated and worked with.

With Modeling view, you can create diagrams of your model that contain only a subset of the tables in your model. This can help provide a clearer view into the tables you want to work with, and make working with complex datasets easier. To create a new diagram with only a subset of the tables, click the **+** sign next to the **All tables** tab along the bottom of the Power BI Desktop window.

You can then drag a table from the **Fields** list onto the diagram surface. Right click the table, and then select **Add related tables** from the menu that appears.

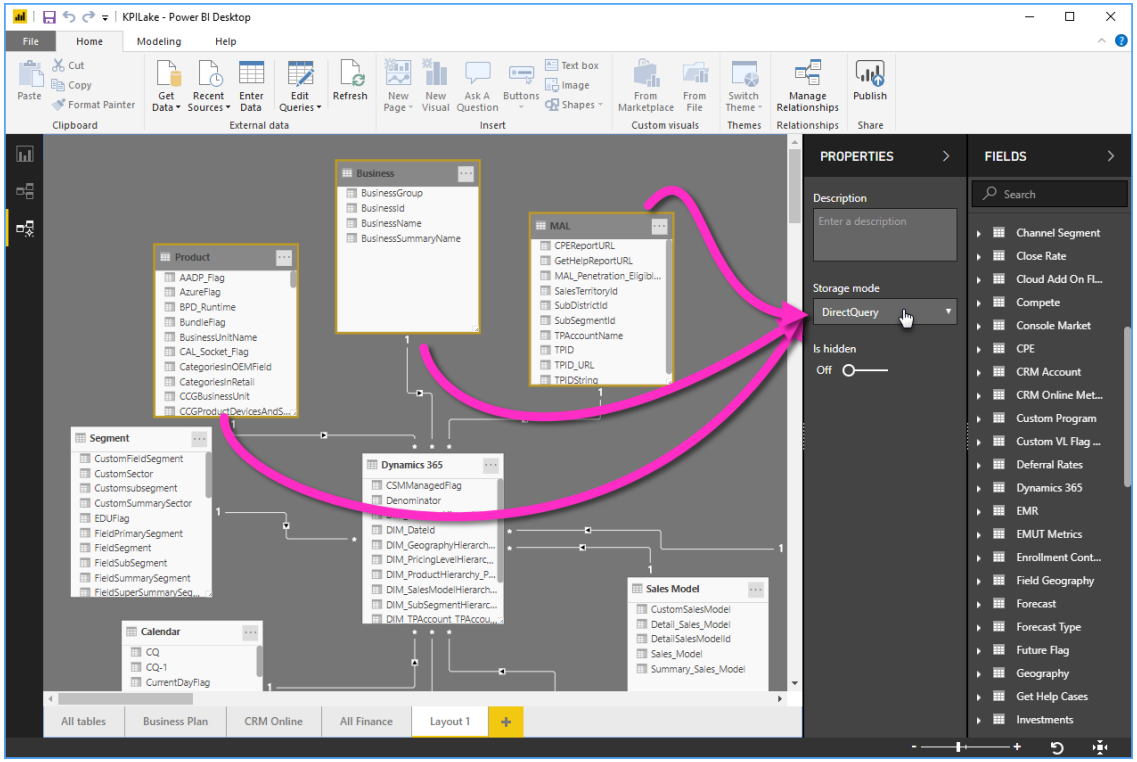


Ref: https://docs.microsoft.com/en-us/power-bi/desktop-modeling-view

Modelling view of power BI

You can select multiple objects at once in Modeling view by holding down the **CTRL** key and clicking multiple tables. When you select multiple tables they become highlighted in Modeling view. When multiple tables are highlighted, changes applied in the **Properties** pane apply to all selected tables.

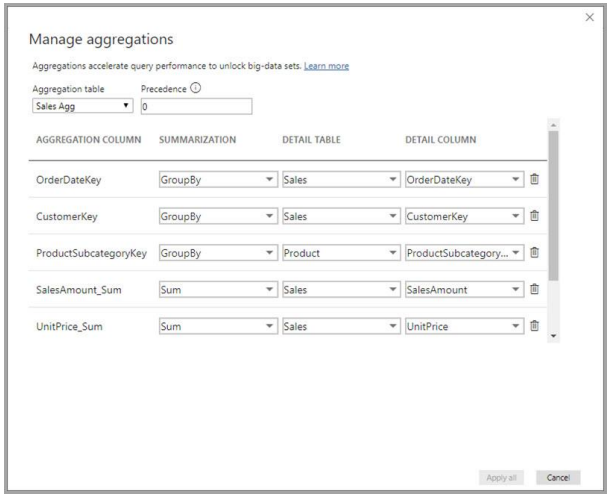
For example, you could change the [storage mode] for multiple tables in your diagram view by holding down the **CTRL** key, selecting tables, then changing the storage mode setting in the **Properties** pane.



Ref: https://docs.microsoft.com/en-us/power-bi/desktop-modeling-view

Aggregations in Power BI Desktop

Using **aggregations** in Power BI enables interactive analysis over big data in ways that previously weren't possible. **Aggregations** can dramatically reduce the cost of unlocking large datasets for decision making



Advantages of using aggregations:

Query performance over big data - as users interact with visuals on Power BI reports, DAX queries are submitted to the dataset. Boost query speeds by caching data at the aggregated level, using a fraction of the resources required at the detail level. Unlock big data in a way that would otherwise be impossible.

Data refresh optimization - reduce cache sizes and refresh times by caching data at the aggregated level. Speed up the time to make data available for users.

Achieve balanced architectures - allow the Power BI in-memory cache to handle aggregated queries, which it does effectively. Limit queries sent to the data source in DirectQuery mode, helping stay within concurrency limits. Queries that do get through tend to be filtered, transactional-level queries, which data warehouses and big-data systems normally handle well.

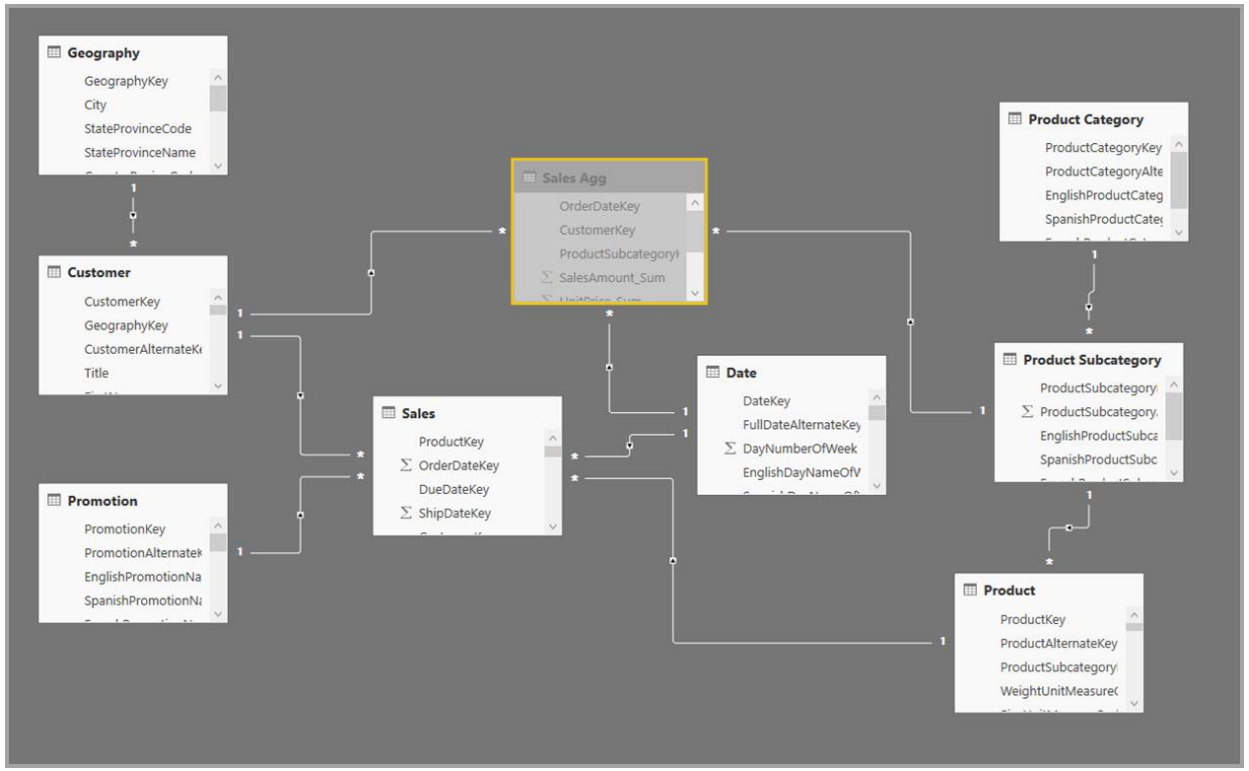
Ref: https://docs.microsoft.com/en-us/power-bi/desktop-aggregations

Aggregations based on relationships

Example… continued

Instead, we create the **Sales Agg** table as an aggregation table. It's at a higher granularity than **Sales**, so it'll contain far fewer rows. The number of rows should equal the sum of **SalesAmount** grouped by **CustomerKey**, **DateKey**, and **ProductSubcategoryKey**. Instead of billions, it might be millions of rows, which are much easier to manage.

Assuming that the following dimension tables: Geography, Customer, Date, Product Subcategory, Product Category are the most commonly used for the queries with high business value. They're the tables that can filter **Sales Agg** using *one-to-many* relationships.



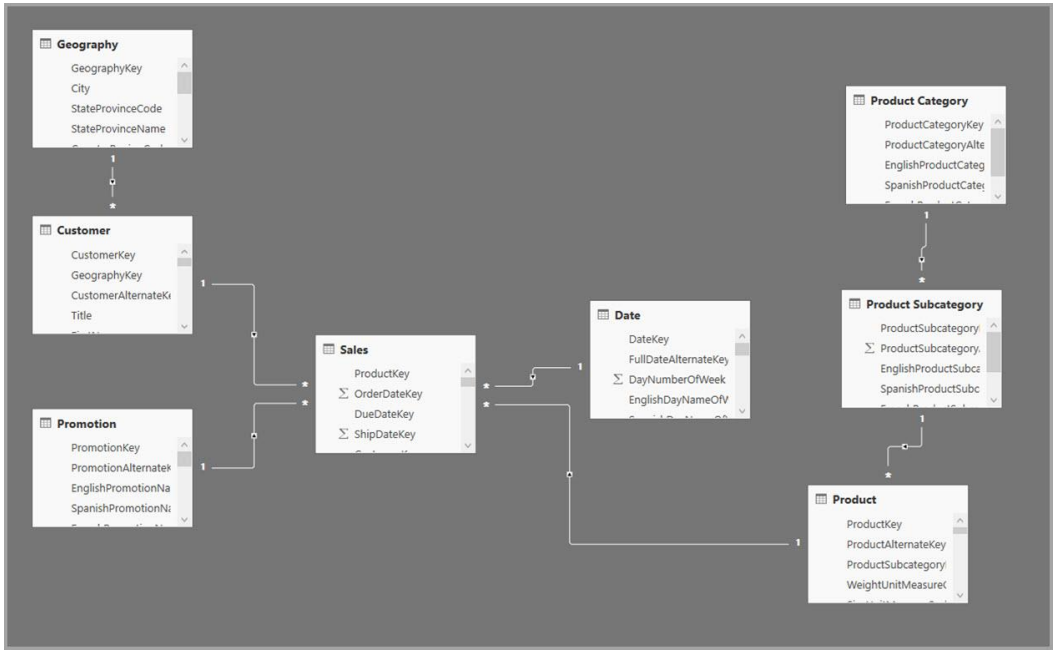
Ref: https://docs.microsoft.com/en-us/power-bi/desktop-aggregations

Aggregations based on relationships

Aggregations based on relationships are typically used with dimensional models. Power BI datasets that source from data warehouses and data marts resemble star/snowflake schemas with relationships between dimension tables and fact tables.

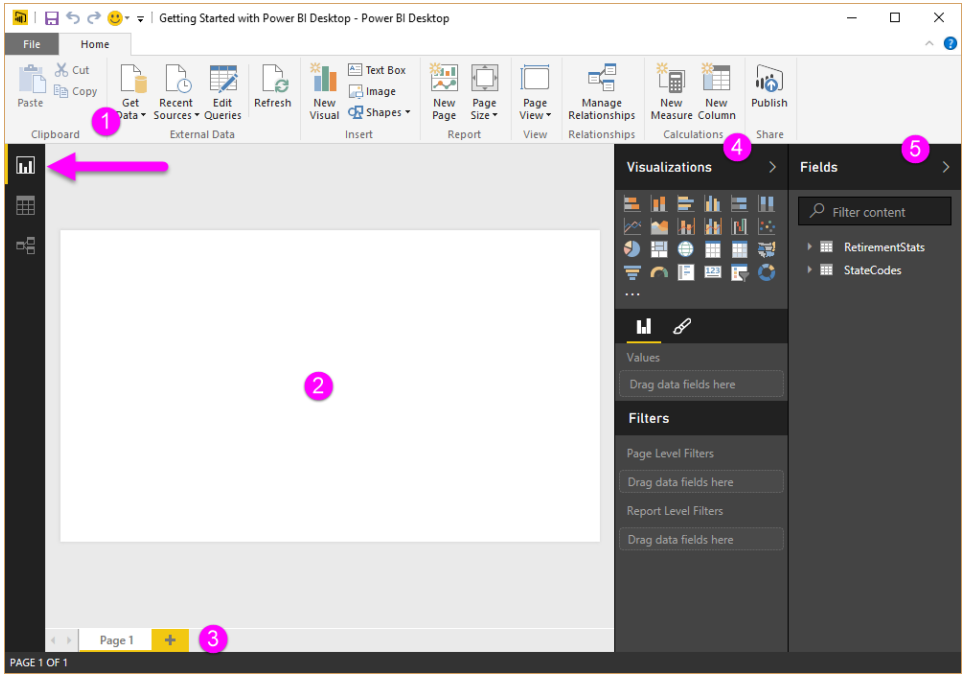
Example

Consider the following model, which is from a single data source. Let's say all the tables are using DirectQuery to start with. The **Sales** fact table contains billions of rows. Setting the storage mode of **Sales** to **Import** for caching would consume considerable memory and management overhead.



Ref: https://docs.microsoft.com/en-us/power-bi/desktop-aggregations

Building reports



The **Report** view has five main areas:

1. The ribbon, which displays common tasks associated with reports and visualizations
2. The **Report** view, or canvas, where visualizations are created and arranged
3. The **Pages** tab area along the bottom, which lets you select or add a report page
4. The **Visualizations** pane, where you can change visualizations, customize colors or axes, apply filters, drag fields, and more
5. The **Fields** pane, where query elements and filters can be dragged onto the **Report**view, or dragged to the **Filters** area of the **Visualizations** pane

Power BI consumption

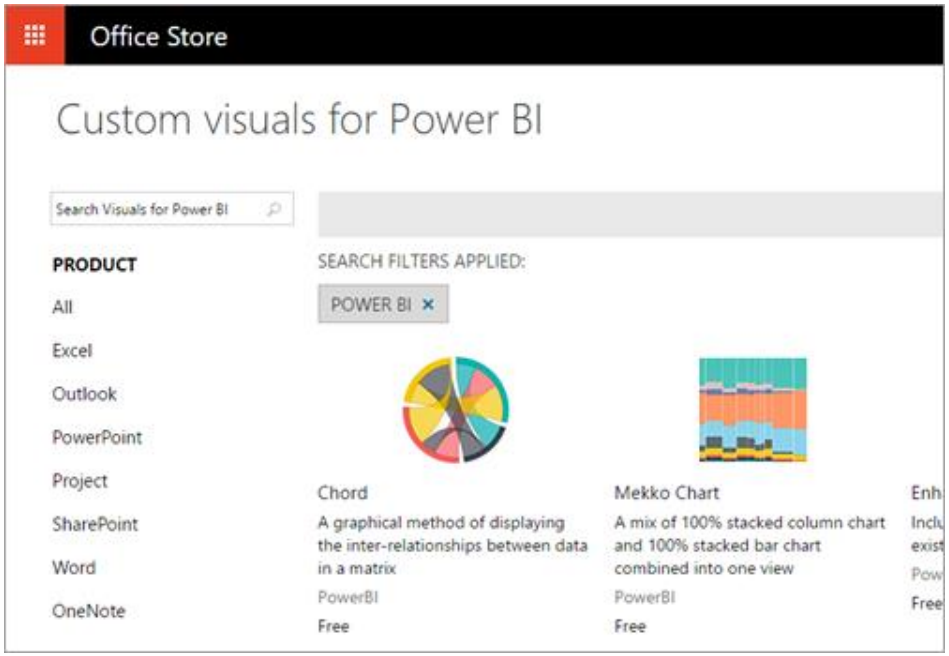
There are several ways to consume Power BI dashboards and reports:

Web browser: This the most common way to consume the data when we are provided a valid URL for the reports.
Power BI mobile: This is a client application (app) that is available for Windows 10, iOS, and Android. It has the capability to connect to Power BI on the cloud or on premises to a Power BI report Server. It is optimized for tablet and touchscreen usage.
Power BI embedded: This allow us to integrate Power BI dashboards and reports in custom applications. They are mainly used by ISVs and developers to distribute their reports from inside their applications.

Power BI for developers

The Power BI service (SaaS) and the Power BI Embedded service in Azure (PaaS) have **APIs** for embedding dashboards and reports.
This feature means allows access the latest Power BI features – such as dashboards, gateways, and app workspaces – when embedding content.

Developing custom visuals
Custom visuals are written in TypeScript, which is a superset of JavaScript. TypeScript supports some advanced features and early access to ES6/ES7 functionality. Visual styling is handled using cascading styles sheets (CSS).



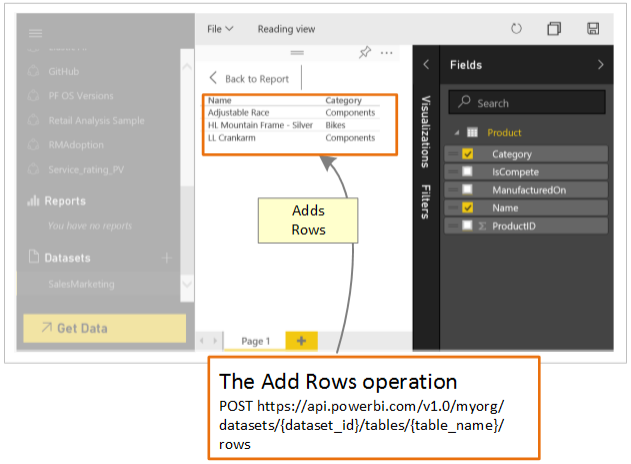
Ref: https://docs.microsoft.com/en-us/power-bi/developer/what-can-you-do

Power BI for developers

The Power BI service (SaaS) and the Power BI Embedded service in Azure (PaaS) have **APIs** for embedding dashboards and reports.
This feature means allows access the latest Power BI features – such as dashboards, gateways, and app workspaces – when embedding content.

Using API automation

Using any programming language that supports REST calls, you can create apps that integrate with a Power BI dashboard in real time. Developers can also integrate Power BI tiles and reports into apps. Developers can also build their own data visualizations that can be used in interactive reports and dashboards.



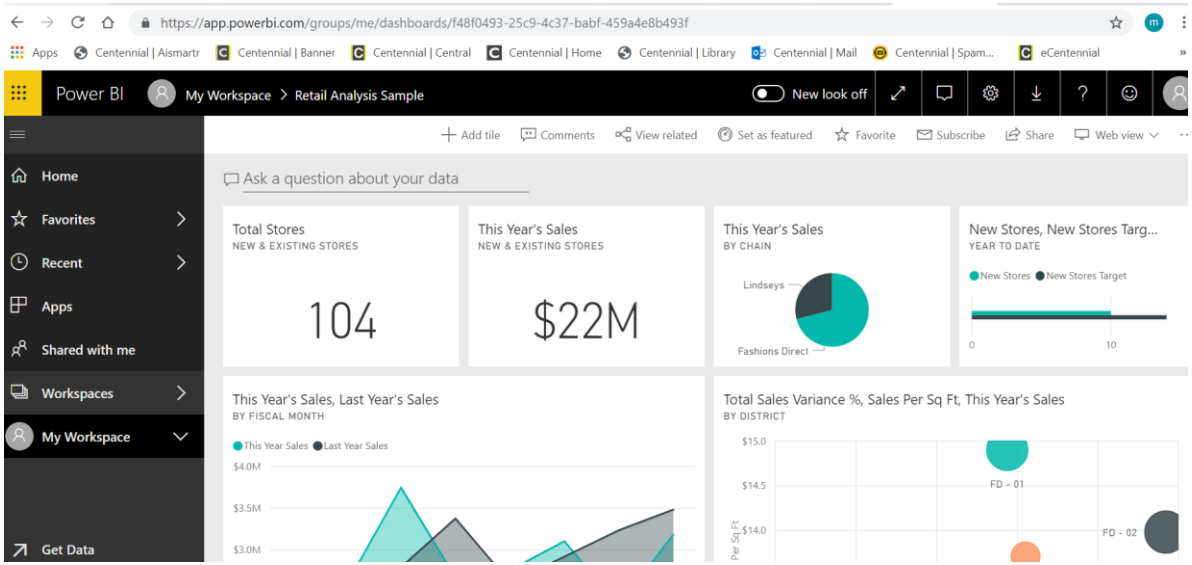
With Power BI REST API, it is possible to perform management tasks on Power BI objects like reports, datasets, and workspaces.

Ref: https://docs.microsoft.com/en-us/power-bi/developer/what-can-you-do

Demo

Demo

https://docs.microsoft.com/en-us/power-bi/sample-retail-analysis#get-the-excel-workbook-for-this-sample



References

Hands-On Data Warehousing with Azure Data Factory, by Cote, Christian; Gutzait, Michelle; Ciaburro, Giuseppe. Published by Packt Publishing Ltd. ISBN 978-1-78913-762-0

Microsoft Tutorials

Gartner IT glossary https://www.gartner.com/it-glossary/