Aspecto destacado Nota

Module 1: Introduction to Self-Service BI Solutions

Contents:

Module overview

Lesson 1: Introduction to business intelligence

Lesson 2: Introduction to data analysis

Lesson 3: Introduction to data visualization

Lesson 4: Overview of self-service BI

Lesson 5: Considerations for self-service BI

Lesson 6: Microsoft tools for self-service BI

Lab: Exploring an Enterprise BI solution

Module review and takeaways

Module overview

Business intelligence (BI) is a term that has become increasingly common in recent years. Along with big data, data mining, predictive analytics, data science, and data stewards, BI is now very much part of business vocabulary. Much of the impetus behind this is the need for organizations to cope with ever-increasing datasets. It's normal to have databases that contain millions of rows, requiring gigabytes, terabytes, or even petabytes, of storage space. Data is no longer confined to an on-premises server room—it's hosted in the cloud, feeds are taken from third-party providers, public datasets are freely available, and social media interactions generate ever-expanding datasets.

Reporting and analysis is certainly not a new concept to business, but the difference between how data analysis is done today, compared with five or 10 years ago, is immense. Nowadays, organizations need BI to see not only what was done in the past, but also more of what's to come. There's an overwhelming amount of data to gather and compose into reports. There's also an increasing need for data to offer up-to-the-minute numbers, so business can react faster to changing trends in markets and industries. Businesses that react fast and predict near-term trends to provide products and services where there is consumer demand have the best chance of survival in a modern and highly competitive world. With the rise of big data, there's an increasing need for data analysts who can take this data, and find the critical points within a plethora of information.

Objectives

After completing this module, you will be able to:

- Describe the trends in BI.

- Describe the process of data analysis in Power BI.
- Use the key visualizations in Power BI.
- Describe the rationale for self-service BI.
- Describe considerations for self-service BI.
- Understand how you can use Microsoft® products to implement a BI solution.

Lesson 1: Introduction to business intelligence

This lesson introduces you to the concepts that comprise BI. You will explore scenarios for using BI and how current trends affect the use of BI, project roles and data models.

Lesson objectives

After completing this lesson, you will be able to:

- Understand BI scenarios.
- See how trends in data and reporting solutions have affected BI.
- Describe the project roles within BI.
- Explain how enterprise BI data models work.

Business intelligence scenarios

- Big data is the result of data generated by the internet, social media, and e-commerce:
  - Data is constantly being gathered for commercial use
  - Data is constantly growing in size
- Reporting:
  - Extracting data and presenting it to enable decision-making
  - Show metrics for organizational performance
- Analysis:
  - Evaluating data to discover insights
  - Data should answer questions, but quickly becomes outdated
- Collaboration:
  - Business analysts need to share data for decision-making

Big data is big news—since the rise of the internet, social media, and the rapid growth of e-commerce, more and more data is being generated, gathered, and analyzed. Supermarkets and retail outlets offer store cards, loyalty cards, or reward cards—depending on how they want to label it—because they need to track spending habits and use this data to sell you more. They gather data, analyze what you like to buy, and then offer incentives to entice you to buy more of the same, or similar. Meanwhile, your online habits are monitored by cookies and advertisements show up on websites, tempting you to buy something you might have searched for earlier.

Reporting

Extracting data from your company's database and presenting it in reports is certainly not a new phenomenon. Most organizations, whatever their size, use some form of reporting, as a reflection of performance within their sector. Until recently, most organizations were happy with end-of-month and annual reports, as a backward reflection of their performance. Modern reporting still needs this, but it should also look to the future to predict where and how to sell more, thereby increasing turnover and reducing the bottom line.

Traditionally, reports have been compiled by department heads, and then given to directors to guide their decision-making. Organizational data, or business intelligence, was the privilege of a few. For example, reports show metrics—how much did we sell last month? How many new customers have we acquired this year? How many mentions did our latest promotion receive on social media? A report provides the answers to questions that the organization needs to make decisions. Reports can be contained in spreadsheets, or created using a visual tool, and distributed on a daily, weekly, or other regular schedule. Reflecting on past performance is a worthwhile task, but modern reports must also be forward looking.

Analysis

Analysis is the process of evaluating data to find insights. Data analysis should answer questions, and offer guidance in decision-making. Data is extracted from source, then cleaned, modeled, and transformed until it's presented appropriately in a report. The report can be a simple table in a spreadsheet, or a visual and dynamic,

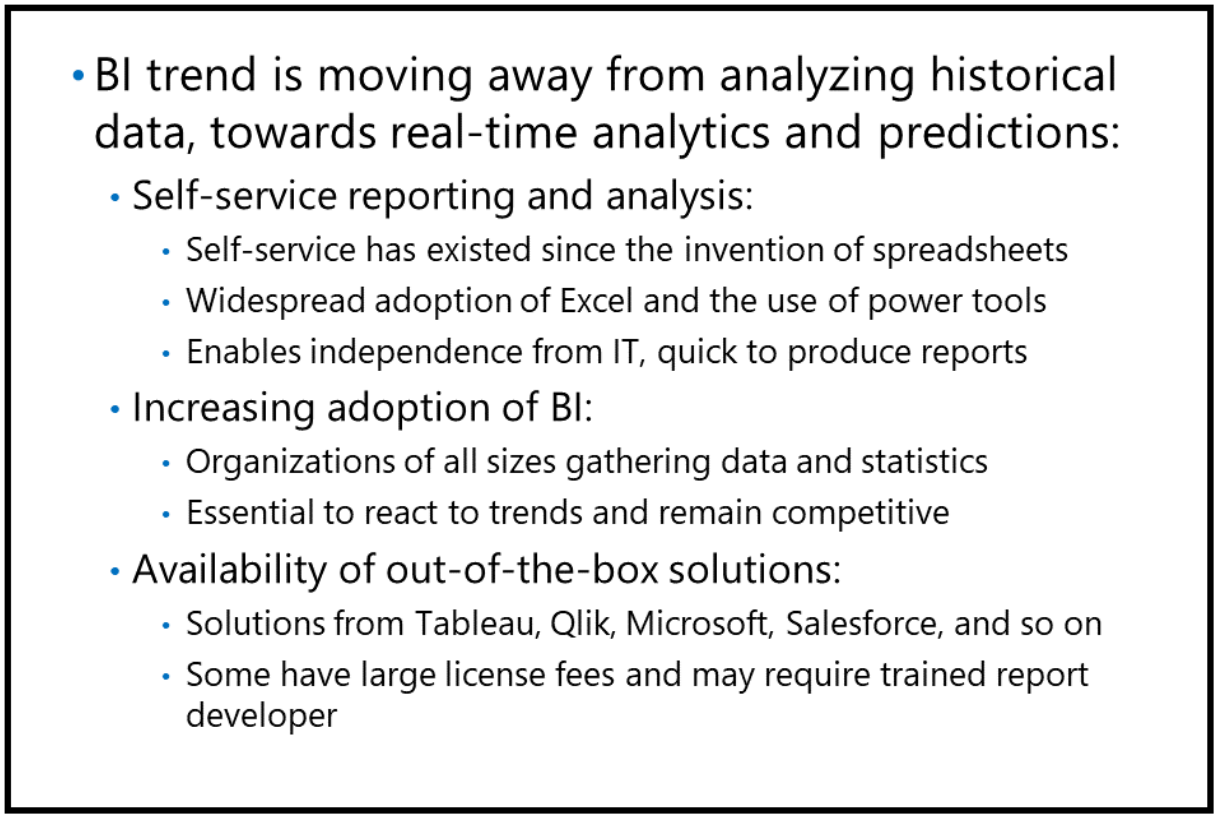
colorful solution. How the data is presented affects the analysis and the conclusions drawn. For example, you can present data in a column chart, but not notice patterns in that data until you use a different type of chart—such as a map or scatter chart—and discover clusters of behavior as a result of geographic location, or outliers that are skewing results.

With so much data to analyze and constant changes in consumer and market trends, modern data has a limited lifespan. Data quickly becomes outdated, so the process of analysis is ongoing. However, with bigger data to analyze, more questions are asked. With an increase in publicly available datasets, including population changes, socioeconomic data, weather patterns, and climate change, you analyze corporate data against a backdrop of relevant statistics.

Collaboration

Data is generated and consumed ubiquitously—it's no longer retained and controlled by a handful of decision makers in an organization. Instead, data is used at all levels, meaning colleagues can react to it, and change the course of their work. Information is critical to companies of all sizes and across industries, with information workers needing to collaborate and share data and results. Microsoft Excel® has long been the dominant tool of the business user—spreadsheets are created, shared, published, altered, emailed, printed, saved, and distributed without version control or adherence to security policies. As spreadsheets are shared and changed, and shared again, analysts work from different datasets, see different results, and reach different conclusions. To collaborate and work cohesively, analysts must be able to synchronize their teamwork.

Trends in business intelligence



The possibilities for analysis grow in line with the increasing number of data sources, and expanding volumes of data. With Microsoft SQL Server® offering in-memory analytics, data doesn't have to be moved outside of the database, and organizations can perform real-time operational analytics. The BI trend is moving away from only analyzing past data, to analyzing real-time data, and using historical data to predict the future.

Self-service reporting and analysis

It could be argued that self-service BI has been around since spreadsheets first entered the software market, enabling users to crunch numbers at their desks. The almost universal adoption of Excel has enabled this trend to continue. With the integration of the four power tools—Power Pivot, Power Query, Power View, and Power Map—Excel users acquire data from a myriad of sources then model, transform, and present that data in sophisticated visualizations. The attraction of Excel and its power tools is the independence it offers to business users. When users access the data they need, they immediately begin shaping and formatting that data, and designing reports to their own specification.

Using a more sophisticated reporting solution generally requires a dedicated report developer, and a lengthy process to submit a feature requirement to IT, then wait for the report to be developed and published—only to find it doesn't deliver the correct data. So begins another lengthy process of submitting a change request, and waiting for the report developer to make the changes. Giving users access to the data means they see what is available for analysis and decide what is useful. The delay in waiting for a report not only frustrates users and holds back their work, but also delays decision-making and the ability for organizations to react to changing circumstances.

Increasing adoption of BI by a wider range of organizations

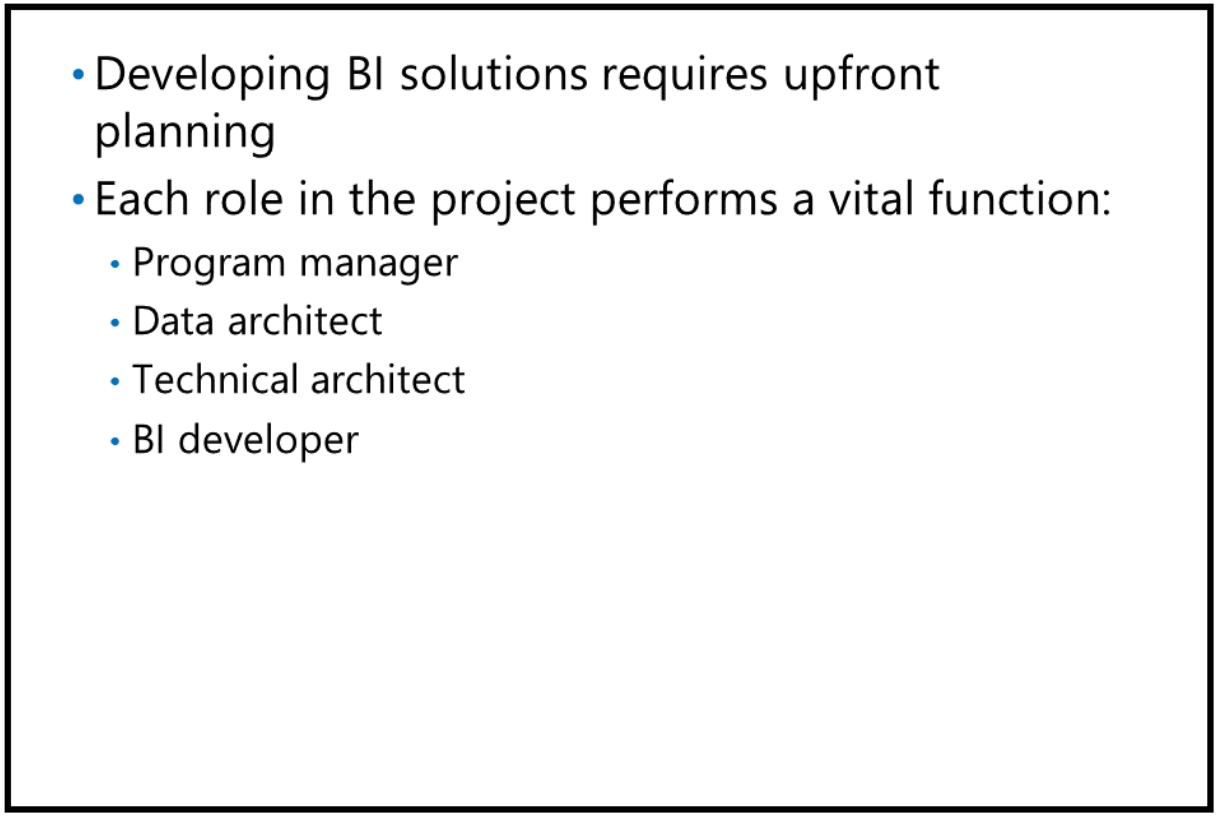
BI is no longer the reserve of big organizations with large budgets to throw at data warehousing projects. Any business operating on the web gathers information about their customers' spending habits, the products they viewed, and their buying decisions. It seems that our online presence, enhanced through the proliferation of mobile devices, is continuously monitored, with all our moves and preferences stored for analysis. To be more efficient, and therefore more competitive, organizations of all

sizes must gather data. However, gathering this data is no use unless it's converted to actionable information. Along with increasing volumes of data, the availability of cheaper, easier to use solutions has helped drive the market, meaning organizations with even the smallest of budgets devote some level of resource to BI.

Availability of out-of-the-box solutions

Organizations can license sophisticated BI solutions from the major vendors in the market, including Tableau, Qlik, Salesforce, Microsoft, Oracle, IBM, SAP, SAS, and more. You can use these solutions to create highly visual reports. With the ability to connect to a variety of data sources, you can create reports and dashboards. However, depending on the vendor, some of the major solutions require expensive server and client licenses, in addition to trained users to create the reports.

Business intelligence project roles



Developing a BI solution requires much upfront planning and designing to ensure the project stays on target and comes to fruition without major issues. The BI project team comprises a number of roles. If it's a new project, the program manager might hire and instruct a data architect and a technical architect—after much of their planning is complete, BI developers will be hired. This depends on the organization, how many projects are in the pipeline, and if contract staff are to provide extra resource.

Program manager

The program manager is responsible for the organizational BI strategy and delivery, often coordinating multiple projects at a time. The program manager is the overall leader of the BI department and, while the role is nontechnical, it does require an understanding of the subject matter, the business requirements, and a comprehension of technical terminology. The main role of a program manager includes:

- Acquiring funding for projects.
- Creating budgets.
- Engaging with stakeholders to determine requirements.
- Analyzing the impact of the project going into production.
- Communicating vision to end users and stakeholders.
- Being responsible for building teams and hiring new employees.
- Undertaking risk assessment.
- Setting standards and ensuring these are met.
- Establishing project priorities, and creating deadlines.
- Managing the expectations of both users and stakeholders.
- Providing status updates.
- Measuring performance.

Data architect

Like the program manager, the data architect is responsible for multiple projects, combining business and technical knowledge to shape the BI solutions. The data must be architected and presented in a design that the organization understands. The main role of the data architect includes:

- Developing the data architecture of the organization.
- Analyzing data requirements and planning for future change requirements.
- Performing logical data modeling.
- Implementing databases.
- Resolving issues between different systems and different data sources.
- Managing master data and liaising with the data steward.

Technical architect

The technical architect must communicate with the BI developers and the operations team to ensure the BI environment is configured correctly. This role is less hands-on than the BI developer, but requires deep technological understanding. The main role of the technical architect includes:

- Assessing the existing BI environment.

- Evaluating development technologies.
- Deciding on appropriate development technologies and justifying the decisions to the program manager.
- Designing the architecture of the extract, transform, and load (ETL) processes.
- Developing the disaster recovery (DR) plan.
- Interfacing with operations and DBA teams.

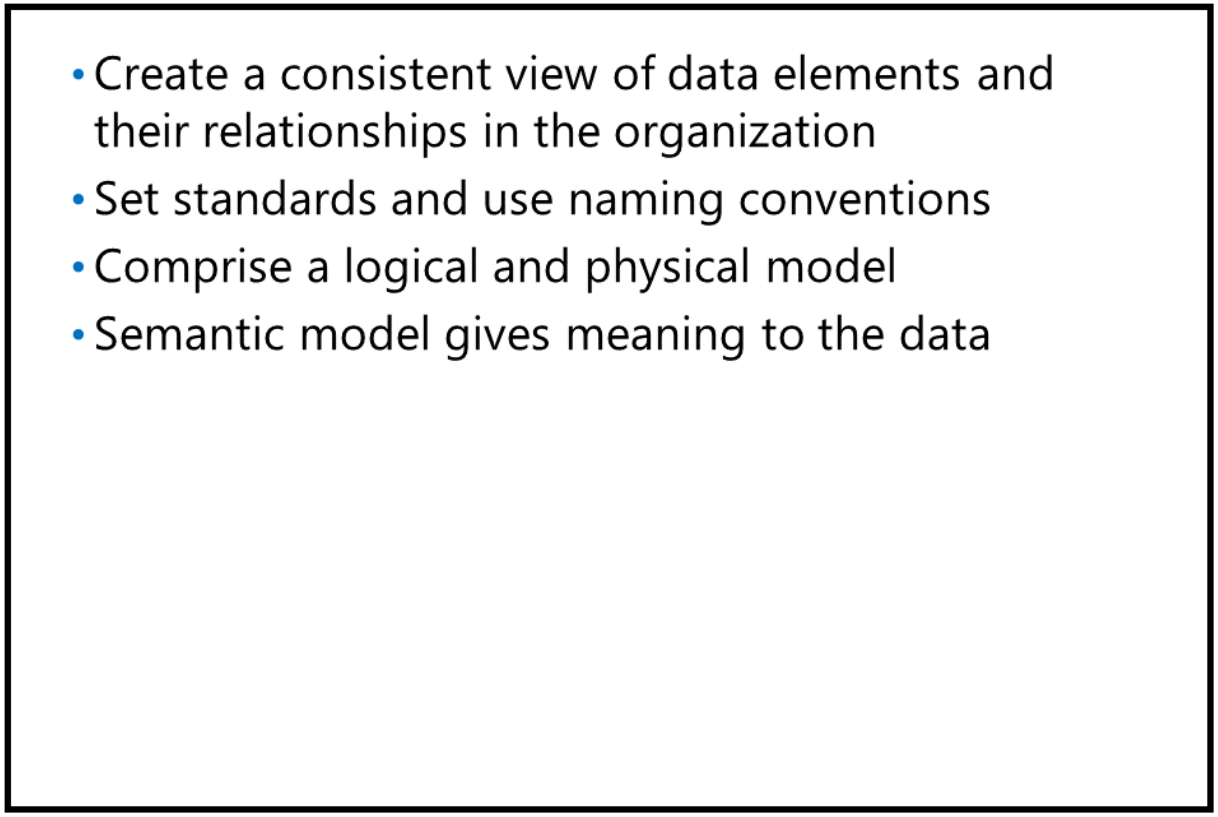
BI developer

The BI developer role can comprise ETL, data warehouse (DW), and report development. Depending on the size of the organization and the structure of the team, a developer might specialize in one aspect, or may perform one or more roles, but there is likely to be an overlap between at least two. The main role of a BI developer includes:

- Designing ETL packages to load data into a staging area.
- Building ETL packages that perform data transformations in the staging area.
- Writing ETL packages to load the transformed data into the data warehouse.
- Creating and managing ETL job schedules.
- Monitoring the ETL process for performance issues and failures.

- Debugging issues in the ETL process.
- Developing the data warehouse database.
- Resolving data issues.
- Building cubes.
- Designing and developing reports.
- Writing code to extract data from the data warehouse.
- Creating a schedule for publishing and distributing reports.

Enterprise BI data models



Enterprise data modeling is the creation of a consistent view of data elements and their relationships in the organization. When more than one data modeler is working on the model, it's important that standards and naming conventions are created and adhered to. Data might be imported from different systems, so naming conventions are likely to vary across sources. This inconsistency should be addressed during the modeling process. If the model comprises a data warehouse, naming conventions should be used for fact, lookup, and history tables. Conventions can also be applied to columns to denote keys, codes, and identifiers. The model might consist of a number of subject areas, reflecting different departments in the organization.

Data modeling

A data model is a visual representation of how the data will be structured in a database. In an OLTP database, the data will be normalized to reduce repeating values and ensure an entity only has the attributes that belong to it. This leads to the best performance for random, small, and isolated transactions. A data warehouse denormalizes the data, so the database performs optimally for reporting.

A data model comprises a logical design and a physical design. There are two approaches to data modeling: a top-down approach, or a bottom-up approach. In a top-down approach, the model is created by gaining an understanding of the business requirements. The bottom-up approach creates a model from existing databases. A model is only a representation of the database, so it will contain objects such as tables, columns, and relationships that can be visualized. A database developer uses the model to develop the physical database.

Semantic models

A semantic model is a data model that includes information to give meaning to the data. The semantic information should enable the model to describe itself. Semantic models help to create consistency. The dataset of a semantic model uses inherent structures; in a database, the context of data is defined through its relationships with other data. Semantic data models give representation to real-world entities such as a Customer, Store, or Employee. A relational model breaks entities into parts, whereas the semantic model uses the entity to represent itself fully.

Check your knowledge

Discovery

How does your organization approach BI? Is this a major part of the corporate strategy?

What BI solutions does your organization use? Is Excel used as a self-service tool?

What do you think are the major issues with your organization's approach to BI?

Show solution Reset

Lesson 2: Introduction to data analysis

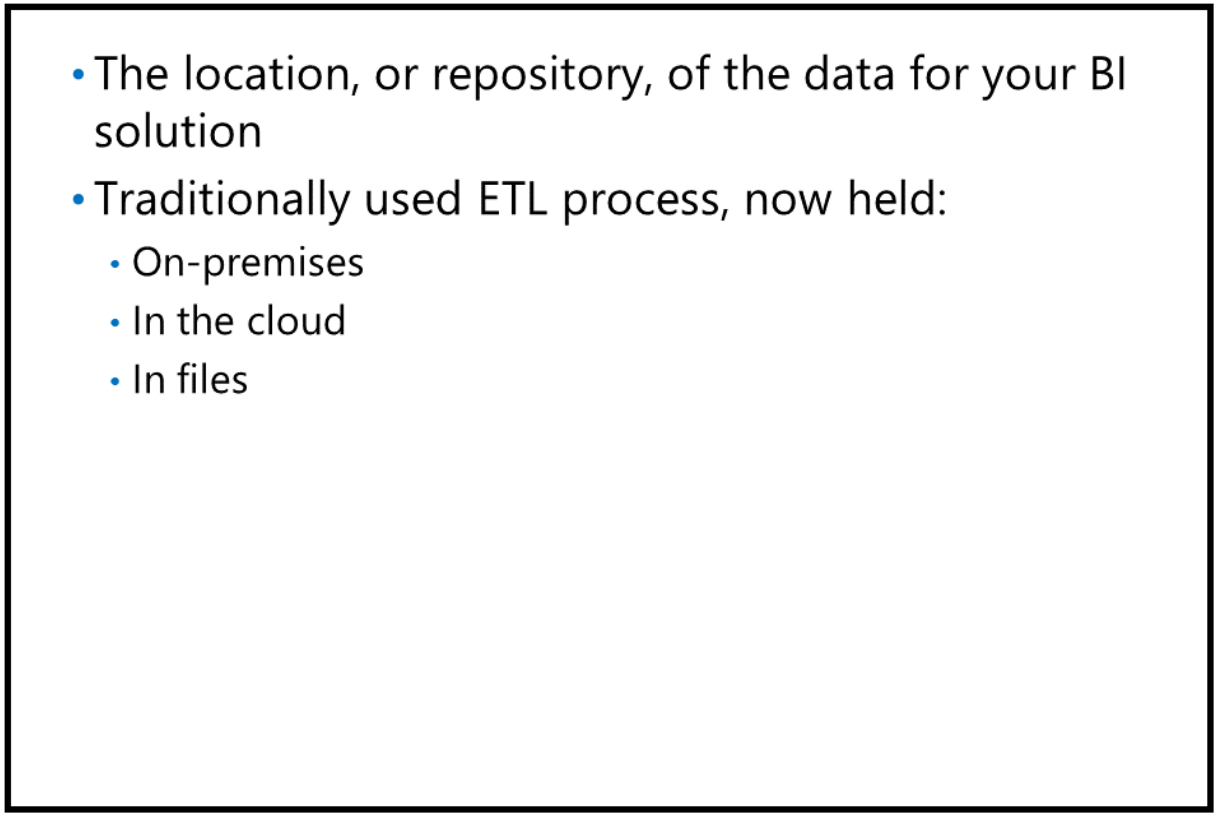
This lesson breaks down the components of data analysis. It looks at using queries to extract data from a variety of data sources, using transformations to make imported data easier to work with, and using visualizations to present data.

Lesson objectives

After completing this lesson, you will be able to:

- Describe how to use data sources in BI.
- Understand how to use queries for extracting data from data sources.
- Explain why transformations are needed.
- Use visualizations to present data.

Data sources



A data source is the location or repository for the data you import into your data warehouse or reporting tools. In a traditional data warehousing scenario, an ETL package extracts the changed data from the operational database and loads it into a staging area, before applying transformations to prepare the data for loading into the data warehouse. Online transactional processing (OLTP) databases are designed for random access and are extremely fast for small transactions. They perform much less well at aggregations, whereas a data warehouse is designed to make this a faster process. Extracting data from operational systems, remodeling, transforming, and applying aggregations in the data warehouse is a lengthy process that requires considerable funding and resources in an organization of any size. In-memory data and real-time operational analytics have the advantage that the data does not need

to be extracted to a secondary location—in-memory processing is designed for optimal performance and can better handle aggregations.

However, the data an organization wants to analyze is typically not confined to an on-premises database server. The online world in which third-party services and publicly available datasets interact with business operations is now very much part of the regular data landscape. The boundaries of data have expanded to disparate locations in the cloud. Data sources you are likely to add to your reports, include:

- **On-premises databases**

  Despite the current trend of moving databases to the cloud, most organizations hold some data on-premises. These might include your Microsoft SQL Servers, such as SQL Server Analysis Services (SSAS), Active Directory® (AD), Exchange, and Access® databases. Your organization might also use other main industry databases, including Teradata, Oracle, MySQL, Sybase, IBM DB2, SAP HANA, and PostgreSQL.

- **Cloud databases**

  Cloud is an increasingly popular choice, with Microsoft offering a wide range of Azure cloud services. These include Azure® SQL Database, Azure SQL Data Warehouse, Azure Marketplace, Azure HDInsight®, Azure Blob Storage, Azure Table Storage, Azure DocumentDB, and Azure Data Lake Store.

- **Software as a service providers**

Organizations are increasingly turning to software as a service (SaaS) providers as a more cost effective option than the development of in-house solutions. Your organization might use third-party solutions such as Facebook, Marketo, and MailChimp, alongside Bing®, Google Analytics, GitHub, and Zendesk. Having the ability to use the data generated from these services is important for gathering a complete picture of activity in your data.

- **Files**

  Most organizations hold data in spreadsheets and are likely to have data stored in Excel or CSV format. JSON and XML are popular languages for exchanging data between systems and should be supported by your BI solution as a data source. In addition, business users might have data stored in text format, which requires importing into the BI solution.

Queries

- Commands you run against the data source to specify the data to extract:
  - Return entire tables or run a query against the source
  - Use stored procedures against SQL Server databases
  - Only return the data that you need
- Expressions used to transform data:
  - M Query Language:
    - Use in Power Query Editor
    - Generate using menu options or edit query directly
  - DAX:
    - Use in Power BI Desktop
    - Derived from MDX and Excel formulas
    - Straightforward to use but very powerful

You use queries to extract data from your data sources. If you have connected to a database, queries specify the tables and columns that you want to export into your BI solution. Your BI solution might offer the choice of importing entire tables, or writing a query to specify the columns you want. If you are connecting to a database such as SQL Server, then using stored procedures to query the data is preferable. A stored procedure is a query that's stored on the server. Stored procedures are more efficient than specific, one-off queries, because SQL Server creates an execution plan, which it reuses each time the procedure is called. This plan works out the optimal way to retrieve the data, resulting in the fastest possible return of results. Stored procedures can also be used by other colleagues—sharing code prevents duplication of effort.

Depending on your role within the organization, you might be dependent on a database developer to write the queries or stored procedures for you and, for security reasons, you might not have access to all objects in the database. It's important that you only return rows and columns from the database that you intend to use in your reports. Not only does importing unnecessary data create additional network traffic, it also makes larger datasets more cumbersome to work with.

After you've imported your data into Power BI Desktop, you might still need to perform further customizations on the dataset.

- Using M in Power Query Editor

  In Power Query Editor, you can use menu options or the M query language to transform your data. Transformations that you make using the menu options are written into your dataset query in M. This is particularly useful when you are unable to use queries to determine the data to import into your original dataset. For example, when you have limited access to the data source or are importing data from flat files.

- Using Data Analysis Expressions (DAX) in Power BI Desktop

  You also use DAX in Power BI Desktop to transform the data in your reports. For example, your dataset might contain first name and last name fields—you can use DAX to concatenate them into a name field.

  If you're an advanced Excel user who is familiar with Excel formulas, you will find Data Analysis DAX to be very much the same. Whereas Excel formulas operate at a row level, DAX is used with relational datasets. You might have already used

DAX in Power Pivot or SQL Server Analysis Services tabular models. This powerful formula language has evolved from the Multidimensional Expression (MDX) language used for querying cubes, and has been merged with Excel functions. DAX offers a library of more than 200 functions, operators, and constraints that mean you can perform sophisticated transformations on your datasets.

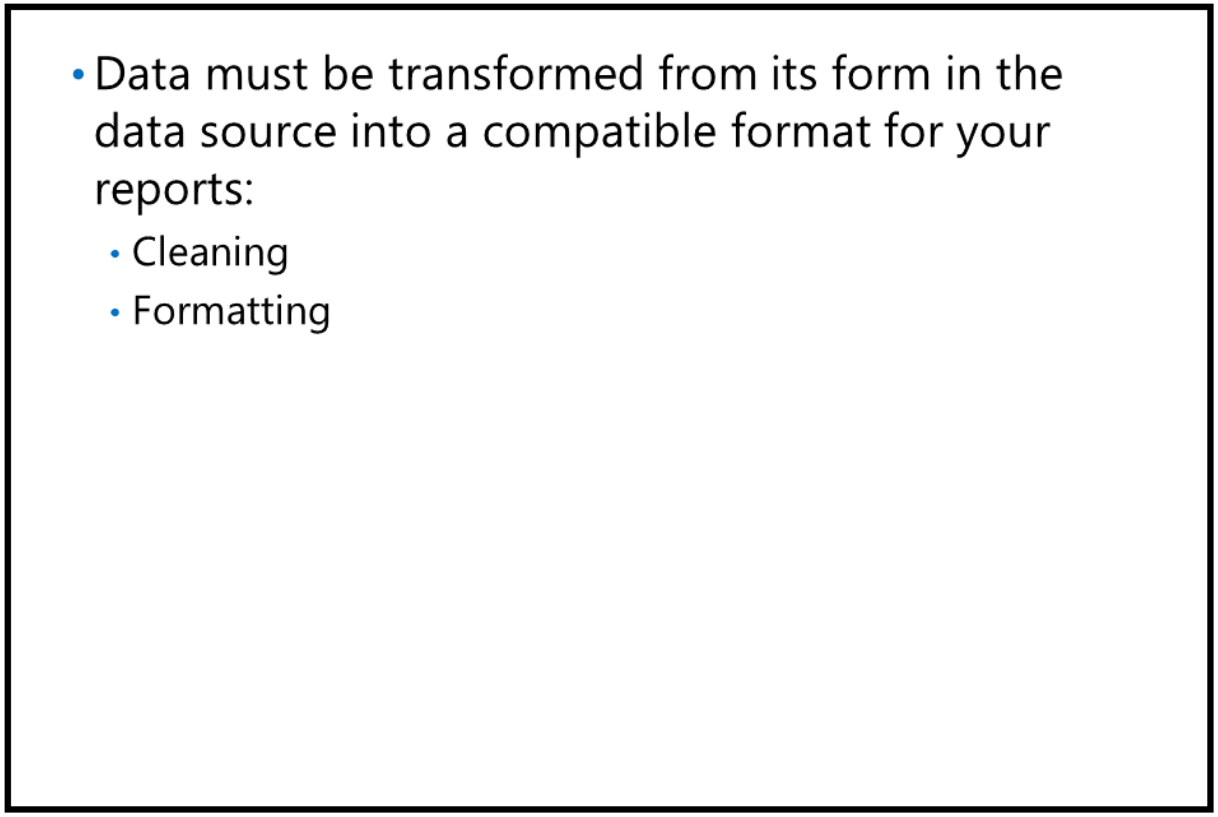
If you're using Power BI Desktop for self-service BI solutions, you can use DAX to enhance the data you import, without having to depend on developers to do this. If you're importing data that cannot be altered until after it has been imported, DAX again comes in useful.

Concatenate the FirstName and LastName Fields Using DAX

The following code uses DAX to concatenate the FirstName and LastName fields to create a new column called FullName:

```
FullName = [FirstName] & " " & [LastName]
```

Data transformations



After you've loaded your data into Power BI, you might need to transform it into a different format for your BI reports. This is often the case when using multiple data sources but, in even the most straightforward systems, it's likely that some transformations are required. To accurately report on the data, you must ensure values are consistent if you intend to use them for filtering.

The following transformations are typically applied to data:

Cleaning

Before applying any transformations to your data, it's a good idea to clean, or cleanse, the data first. This process corrects dirty data or removes it to another area for investigation. You want the quality of your data to be as high as possible. Typical cleansing operations might include:

- Detecting dirty data as it is loaded and either applying a transformation to data that can be cleaned or filtering the dirty data into a separate table for further investigation into why it's incorrect in the source system.
- Removing duplicate rows.
- Eliminating incomplete rows.
- Performing logic tests to check date fields, such as seeing if a date is earlier or later than should be possible—for example, if the **Ship Date** is before the **Order Date**, then this data is dirty.
- Checking address and postal code fields are correct.
- Performing character pattern testing to ensure phone numbers and email addresses are in the correct format.
- Logging missing values.
- Checking that data matches the business rules. For example, only one Sales Person manages a single customer.

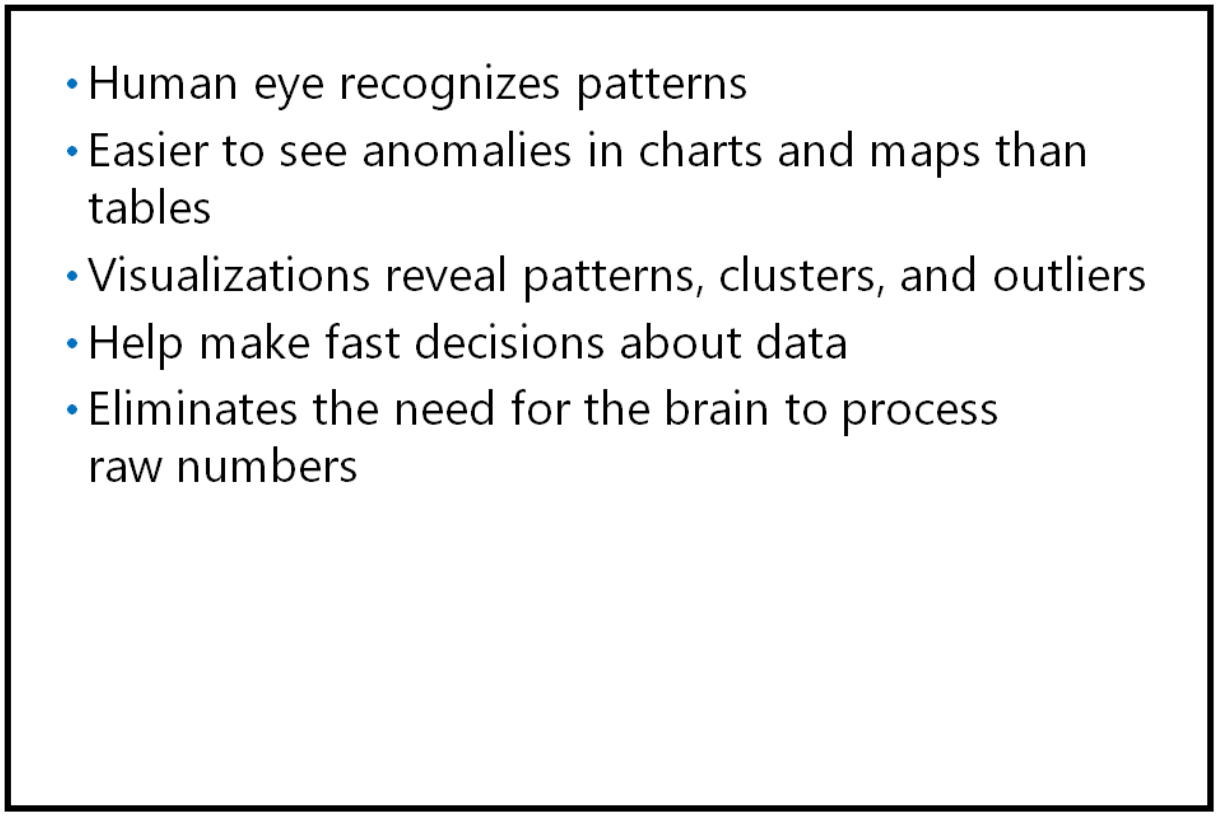
Formatting

When the data is clean, you apply formatting to ensure the source data is in a useful format for your end users. Typical formatting operations include:

- Concatenating columns. For example, combining **First Name** and **Last Name** into a **Full Name** column, or concatenating **Address1**, **Address2**, **City**, **Country**, and **Postal Code** into a **Full Address** column.
- Replacing shorthand values with full words to enable better filtering. For example, you could change **M**, **F**, and **U** values to **Male**, **Female**, and **Unknown**, or **S**, **M**, **D**, **W**, to **Single**, **Married**, **Divorced**, and **Widowed**. **True** and **False** values are frequently stored as **1** and **0** values in the source database, and should be converted.
- Changing the casing on text values. You might want to ensure country or state codes are all uppercase, and names and address all have title case, with the first letter of each word in uppercase, the rest in lowercase.
- Dates might need to be formatted to full date time values to enable filtering at a low level of granularity. The format of dates generally varies quite widely across systems, with no consistency, so you need to be aware of formats and ensure that datetime values are converted to the same format and locale.
- Currency and number fields should be formatted and handled carefully. Ensure decimal columns that undergo any rounding up or down do not skew figures and produce unexpected results. If accuracy is critical, you must ensure that values are entered correctly into the destination database. If decision makers are not

concerned about precision and are happy with an approximate figure in aggregations, you will have more freedom to apply some formatting.

Visualization



Evolution has given humans the ability to recognize patterns—this means we instantly read and deal with dangerous situations, helping us to survive. We very quickly identify irregularities, which means we recognize when a situation is no longer regular—something has changed and could be life threatening. Although we are no longer presented with the same dangers that early humankind endured, we have

retained the ability to visually assess and make judgements within incredibly small timeframes. In the modern world of information, you might apply this innate ability to different scenarios, primarily including the reading of data.

The way in which data is presented affects how quickly and efficiently you can process and understand it. If you're presented with a table of numbers in a spreadsheet, it's likely that you would need to reorder the data and take time to work out the highest and lowest values; you might not notice clustering, outliers, or other patterns within the data. If you present the data on a map, or in a column or scatter chart, you might instantly see the high and low values, such as customers who spend most on products within a particular category live by the coast—or that males over 45 are the most popular return customers. The context within which you place the data affects its interpretation.

The power tools within Excel have no doubt increased its popularity as a data analysis tool. This is because users quickly take data that's in a table format and difficult to comprehend, and convert it into colorful charts and maps, which become instantly readable. Tables of data, even when ordered so values run from high to low or vice versa, still require us to read the numbers and compare rows of values. For example, when we view a colored pie chart, we can instantly see how the values are distributed by the size of the portions. Initially, we don't need to know the values behind the portions; we can make an instant assessment, and then start drilling down to obtain further detail. Visualizations are vital for helping us make fast decisions about business data. They effectively eliminate the need for the human brain to process raw numbers, search for patterns, or dig for outliers by manipulating the data.

Demonstration: Importing data with Power BI Desktop

In this demonstration, you will see how to:

- Import data warehouse data into Power BI Desktop.
- Remove columns.
- Format a column.
- Create a new column using a DAX expression.

Check your knowledge

Discovery

How much data does your organization gather? Have you noticed an increase in the volume of data that you have to work with? Do you have a mix of data sources, such as on-premises databases, cloud services, and SaaS providers?

Show solution Reset

Lesson 3: Introduction to data visualization

Data visualizations bring data to life, using colors and shapes to present data that would otherwise remain as text and numbers. This lesson explores how visualizations help you discover insights into your data that you would not otherwise

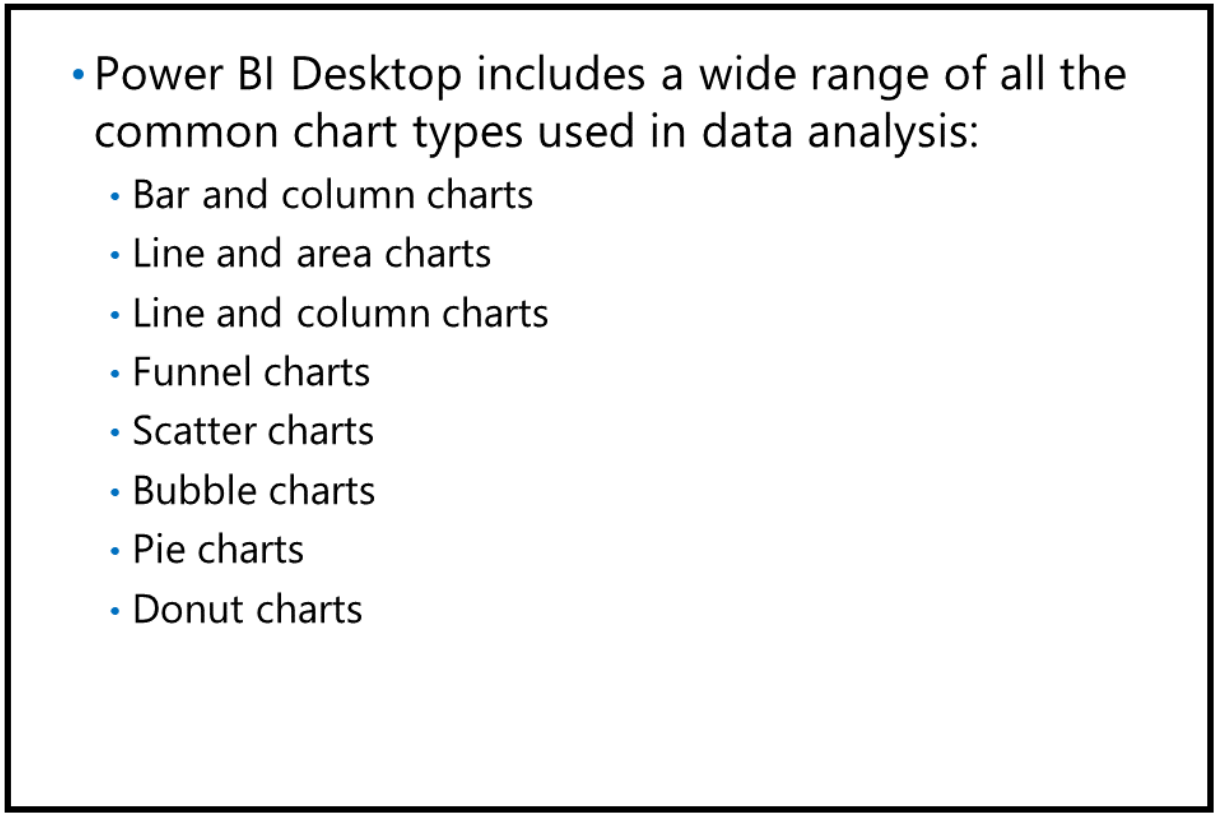
find. The chart types in this lesson focus on the charts available in Power BI Desktop; however, the principles of charting components are generally standard across BI solutions and vendors.

Lesson objectives

After completing this lesson, you will be able to:

- Describe the different types of chart available for presenting data.
- Use cards to display data.
- Use maps to show the spread of data in a geographic area.
- Use tables to organize data.
- Explain how the tree map works.
- Format charts.

Charts



You use the chart visuals in Power BI Desktop, to quickly create visually stunning and interactive reports and dashboards. You can select a chart from the **Visualizations** pane to add to the report canvas, or drag a data field onto the report to automatically create a table visual—that can then be converted to another chart type. For example, you could drag the **Categories** field onto the report, which automatically creates a table. You could then drag **Total Sales** onto the table, to add another column. Then you could click one of the chart icons in the **Visualizations** pane, and quickly switch between a bar or pie chart.

Bar and column charts

Stacked bar and column charts are identical, except that the bars on a stacked bar chart span horizontally, rather than vertically, as in a column chart. Each chart accepts an axis field, such as Sales Person, and a Value; for example, Sales YTD.

Clustered bar and column charts are similar to stacked charts, but they include two data fields for the Value, which results in two bars or columns for each axis.

Again, 100 percent stacked bar and column charts are similar to stacked and clustered charts, except the bars and columns stretch the width or length of the chart area, and display the progress of each axis against a value. You add two data fields to the Value, such as Sales YTD and Sales Quota. If you need to display progress in attempting to meet a target figure, 100 percent stacked charts are useful.

Line and area charts

The line and area charts are fundamentally the same. However, the area chart is filled in, so the area below the line values appears as a solid block. Line and area charts are useful for displaying data over a period of time—such as financial data.

Line and column charts

The line and stacked column chart combines columns and lines. The columns and lines share the same data field for the axis—for example, Year. The column value could be Gross Sales, with a line value for Share Price. You might include multiple lines on a line and stacked column chart. You use the line and clustered column chart to include multiple columns for each shared axis.

> **Note:** If your data creates a large number of data points—for example, hundreds of bars on a bar chart—the scrollbar will adjust so that it does not become too small. Instead, as you scroll to the end, more data is loaded, but the scrollbar remains a viewable size.

Other chart types

Power BI Desktop provides other types of graphical chart, including:

- Funnel charts
- Scatter charts
- Bubble charts
- Pie charts
- Donut charts

Cards

- Present most important data first:
  - If users normally read left to right and top to bottom, show most important data in top left
  - Use card, multirow card, and KPI visuals to present important figures clearly and efficiently
- Card chart:
  - Displays a single numeric value, such as Total Sales
  - Optionally displays data label and title
- Multirow card chart:
  - Shows multiple numeric values, useful for small datasets, such as Main Category and Total Sales
  - Optionally include the data labels and a chart title
- KPI
  - Visualize a business objective and show progress towards the goal

When you present data in a report or dashboard, you should take care to ensure the most important information is easy to find. If your audience normally reads from left to right, top to bottom, then displaying the most critical data in the top left, flowing through to less important content at the bottom right, is helpful. If you have important figures that need to be presented clearly, so that they are easily read, then the card and multirow card charts suit this purpose.

Card chart

The card chart displays a single value and a description. The numeric column values are aggregated to show the total value, such as Total Sales; the data label is the

name of the field. Before using the card chart, ensure that the field to be aggregated is formatted correctly, especially if this represents financial data. If the Value column is not specified as a currency data type, then it shows only a number without the currency symbol. This should be included to make clear that it is a monetary figure. The data label can be turned off, but unless it's entirely clear what the figure refers to, this is best included. You rename the field by right-clicking it in the Fields pane, then selecting Rename. Again, be as clear as possible as to what this refers to. If you cannot change the name of the field, you can hide the data label and add a title instead. You format the card to change the background color and transparency, format the card border, and change the font properties of both the data value, and the label and title.

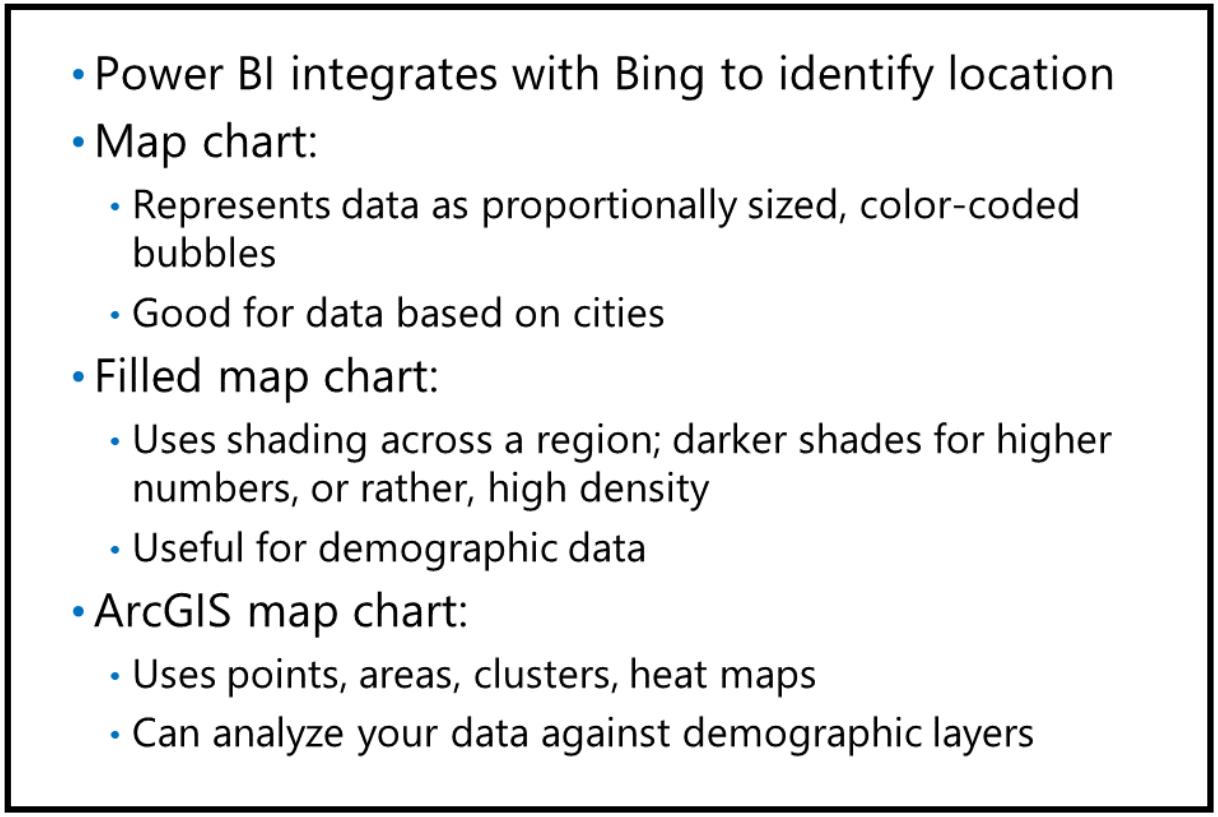
Multirow card chart

The multirow card chart is a useful way to clearly present numbers, without using the format of a table or matrix chart—which are difficult to digest. Like tables, the multirow card chart works best for smaller data sets; otherwise, there is too much data and text to read. For example, a multirow card chart is useful for displaying main categories, and sales. You can also add a title to the multirow card chart and turn off the category label. Use the Format options to customize all aspects of the card, including adding a border, changing the background color, modifying font properties, and adding a back color to each data value.

KPIs

Key Performance Indicators enable companies to measure their progress towards a business objective or goal. You create KPIs at a high level to measure the overall performance of the company; you can also set KPIs at lower levels, such as by departments—for example, sales, call center, or warehouse. You add a KPI visual to your report in Power BI to track progress towards a target. Similar to the card visual, the KPI displays a single value such as TotalSales for the current year—this is the Indicator. The Target value is the goal, such as TargetSales. Add a data value such as Year to the Trend axis to display how well the target is being met. This is represented as a filled line chart and Power BI automatically colors the filled area using green, yellow, or red to show if progress is good, neutral, or bad. You use the Format options to change these colors.

Maps



Power BI Desktop includes a map chart and a filled map chart. You use these charts to map your data visually, both regionally and globally. Power BI integrates with Bing maps to find default coordinates for locations, based on a string value, in a process known as geo-coding. This integration means you do not need to provide longitude and latitude coordinates in your data—this is optional, because Bing makes a best guess at the location.

Map chart

The map chart accepts data for the Legend, Longitude, Latitude, Values, and Color saturation. The Legend property accepts fields such as City, County, and Province,

and the Values property accepts numeric values such as Total Sales, or Number of Customers. The numeric values are presented as colored bubbles on the applicable location specified in the Legend property. The bubbles are sized proportionally to the data they represent within the field in the dataset; that is, the bigger the value, the bigger the bubble. The map chart is useful for presenting data based on cities, rather than wide areas.

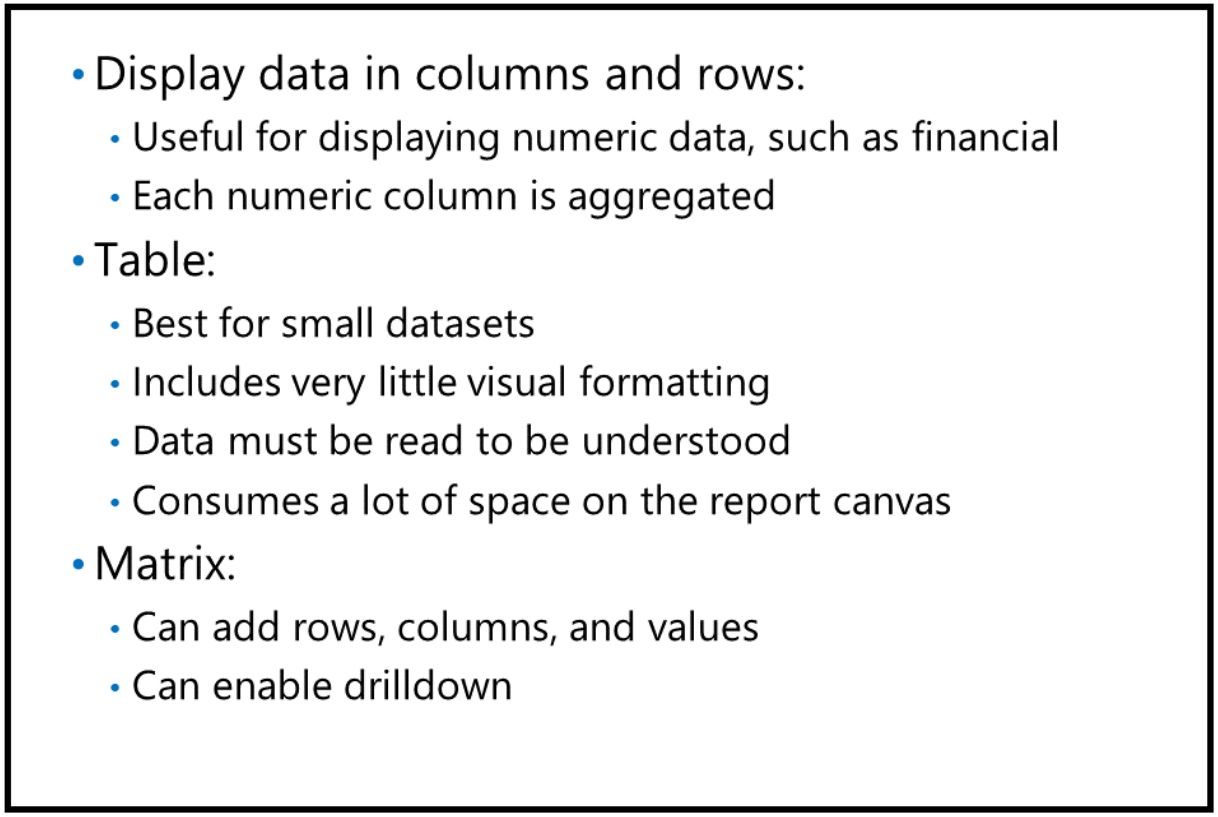
Filled map chart

The filled map chart (also known as a choropleth) uses a slightly different visualization to represent the data. This chart uses shading, tinting, or patterns to represent the data value across a geographic area. The darker the color, the higher the value; the lighter the color, the smaller the value. This is particularly useful for presenting socioeconomic or demographic data, because it provides a visual overview of data across a wide area, such as all the states in the United States.

ArcGIS map

The ArcGIS map chart uses the ArcGIS technology provided by Esri. Before creating or viewing an ArcGIS map, the report builder or user must accept the Esri terms and privacy policy. You use ArcGIS maps to create informative map visualizations by using points, areas, clusters, or heat maps alongside demographic and reference layers.

Tables



You use table and matrix charts to add data fields to create columns and build up a table. Each numeric column is automatically aggregated, with a total at the bottom of the column. Using a table or matrix is useful when you want to display the actual numbers, such as for financial data.

Tables are best used for smaller sets of data because, as the number of rows or columns increases, the information becomes harder to assimilate. To help alleviate this, the table chart includes the option to apply predefined styles, which makes the data easier to read. You set styling, such as alternate row highlighting, and use the predefined styles or select custom colors for the alternate rows, to format the table to your exact requirements.

Consider the following table, which would appear much the same in a Power BI Desktop report. The chart displays the total sales by category and country. It is consuming a lot of space, and requires you to read each of the values in the **Sales Territory Country** column, and then the figures in the **Total Sales** column. Furthermore, the values in the **Sales Territory Country** column are ordered alphabetically, which determines the order of the **Total Sales** column, making it difficult to compare the sales figures. You might be able to order by each column, but not by **Total Sales** within the **Accessories** category only.

Category Name	Sales Territory Country	Total Sales
Accessories	Australia	$81,309.16
Accessories	Canada	$59,758.93
Accessories	France	$37,421.30
Accessories	Germany	$36,908.60
Accessories	United Kingdom	$43,481.35
Accessories	United States	$148,170.91
Bikes	Australia	$2,440,928.44
Bikes	Canada	$581,424.73
Bikes	France	$870,221.82
Bikes	Germany	$1,025,888.91
Bikes	United Kingdom	$1,148,585.76
Bikes	United States	$3,095,275.19
Clothing	Australia	$41,646.69
Clothing	Canada	$32,444.55

Category Name	Sales Territory Country	Total Sales
Clothing	France	$14,535.92
Clothing	Germany	$14,093.26
Clothing	United Kingdom	$18,219.16

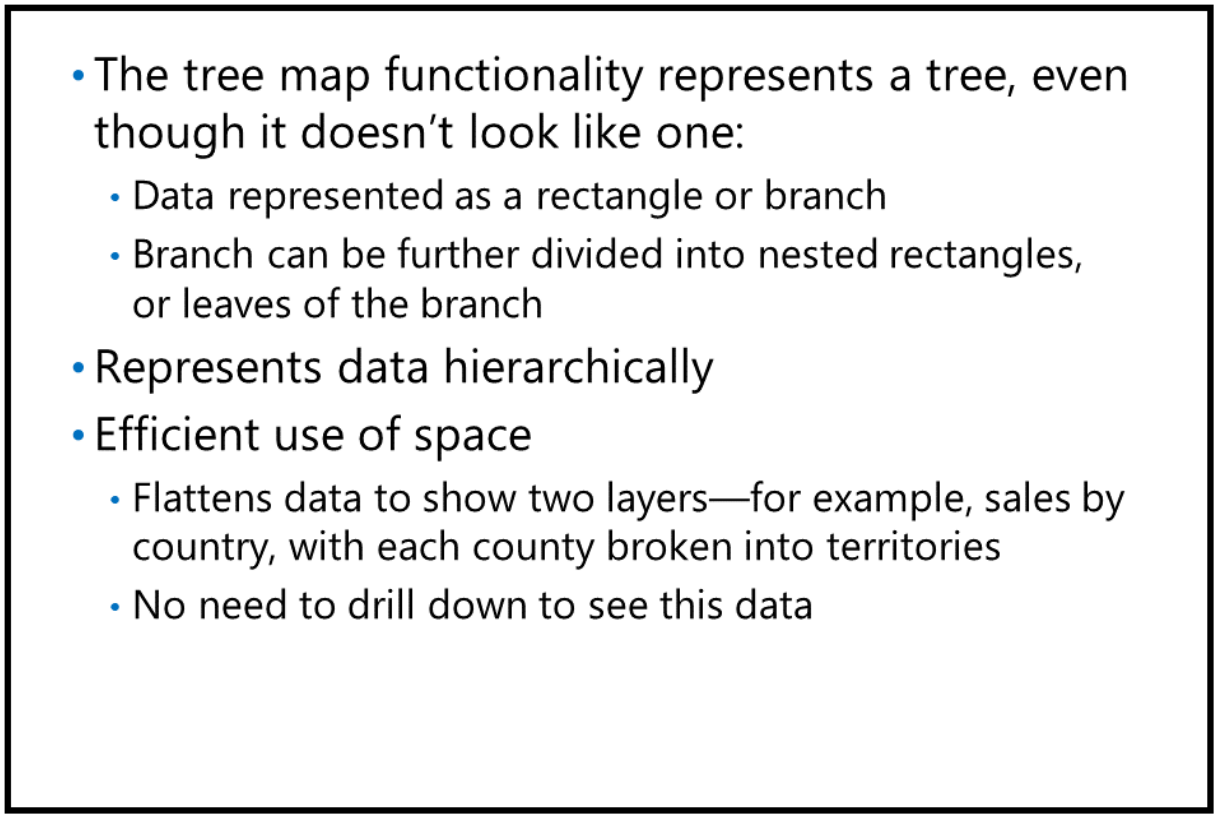
There is little difference in displaying data in a table in Power BI Desktop compared to Excel, or even a SQL Server Reporting Services report. You'll see that the table consumes space and takes time to read.

Visually, the table and matrix charts look quite similar; however, the matrix chart provides more functionality than a table. When using the matrix chart, you add rows, columns, and values to your data, in addition to implementing drilldown capabilities. Displaying the example table as a matrix would enable users to group the information by category or country, simplify the layout and help end users to better understand the data to make more informed business decisions based upon that data.

Conditional formatting

You customize the background color of a cell depending on its value, including the ability to use gradient colors. After you create a table in Power BI, in the **VISUALIZATIONS** pane, in the **Fields** bucket, right-click the field that you want to colorize. From the menu, select **Conditional formatting** and then select whether you want to color the text or the background, or use data bars. You then select the minimum and maximum colors and set the values to be that of the lowest and highest values in the data—or manually set the values. You can optionally add a center, or middle, value and color, by clicking the **Diverging** box.

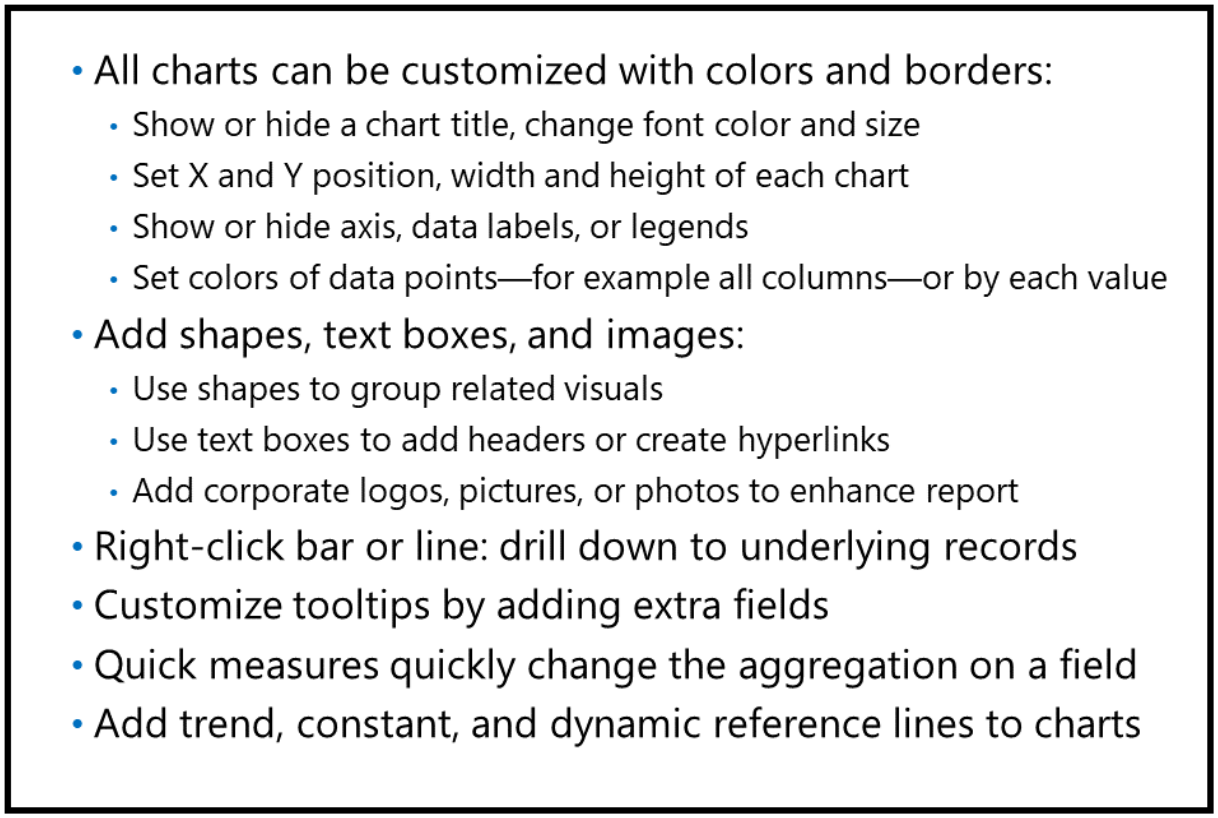
Tree maps



The tree map might not physically represent a tree; however, the principle behind its function is representative of a tree with larger data scaling through to smaller data, as if the data were branches scaling down to twigs. For example, in Power BI Desktop, add the **Country** data field to **Group**, add **Territory** to **Details**, and **Total Sales** to **Values**. Each country is represented by a rectangle that is proportionate to the number of sales, so the countries with the most sales have the largest rectangles. Each country rectangle is subdivided into territory rectangles, with their size again being proportional to the value of sales in that territory. This style of representing data is classed as hierarchical.

Unlike a table or matrix chart, the tree map is more efficient in how it uses the space it consumes in a report. By showing both Country and Territory in the tree map, it has effectively flattened the data—you don't need to drill down to see categories for each territory.

Formatting charts



The visuals in Power BI include extensive options for customizing how your data is displayed. Some of the options available will depend on the type of chart.

Settings

Each visual can be customized with colors and other settings using **Format**, so you can easily use corporate colors to ensure your Power BI reports match the look and feel of business-specific colors. This is particularly useful if you use the embedding tools to include visuals within your own custom applications or websites.

The title of each visual can be customized. Included by default, you turn off the title to hide it completely, or change the text, font color, text size, and background color, and set the alignment of the text to left, right, or center. You can also choose to lock the aspect of the visual. Under the **General** settings, you can configure the X Position and Y Position of each visual, and specify Width and Height, ensuring your visuals are of a consistent size in your reports. All visuals enable you to add a border, which is not included by default, and you can change the color of the border to suit your design requirements. Each group of settings includes a **Revert to default** button to reset the visual and remove any formatting you have applied.

Other settings include the ability to show or hide axis, data labels, or legends, and set the colors of data points. With a column chart, you change the color of all columns, or set them individually, based on each data value. This is helpful if your report shows consistent data, such as sales by department or category, where the department or category is represented by color. For a supermarket, fresh fruit and veg might be represented with green, frozen food with blue, pet food with brown, and so on. For each visual, click **Format** to see the available options.

Shapes, text boxes, and images

In addition to data-bound visuals such as column charts and maps, you can also add static features to further format and customize your reports.

You use shapes to highlight or group items in a report. From the **Insert** group on the **Home** tab, click **Shapes**, and choose from Rectangle, Oval, Line, Triangle, or Arrow. For example, you could use the rectangle shape to group one set of visuals that contain data pertaining to sales, and another to group visuals referring to product returns. You might also use the line shape to divide the report into sections using horizontal and vertical lines. The arrow shape can be used to point to a spot on the report to which you want to draw your colleague's attention. Each shape can be customized, and you can change the border and background colors, or add a title.

Including a text box in your report is a useful feature for adding titles or extra headings to visuals. For example, you could add a main heading to the report, and then a subheader to a group of visuals. To add a text box, from the **Insert** group on the **Home** tab, click **Text box**. Type the text into the main box then you can format the face and size of the font, set bold, underline, italic formatting, and alignment—or add a background color. You can also create a hyperlink using one or more words in a text box.

You add images from the **Insert** group on the **Home** tab, by selecting **Image**. Browse to the image you want to add and click **Open**. The image appears on the report canvas and you can add a title, turn on the background color and set the transparency, and add a border. This is useful for adding logos to your reports so they adhere to corporate design. You can also add a photo to a report.

Drill through

Power BI visuals automatically include the ability to click a data point, such as a bar, line, or portion of a donut chart, and it will display the underlying records. For example, right-click a bar in a bar chart and click **See Records** to show a list of the underlying data, or click **See Data** to display both the visual and the aggregations for each bar. This is available in both Power BI Desktop and the Power BI service.

Customizable tooltips

By default, visuals will display a tooltip that includes the data point's value and category. You add other fields to the tooltip by dragging a field from the **FIELDS** pane, to the **Tooltip** bucket on the **VISUALIZATIONS** pane. Right-click the field in the bucket list to choose from additional aggregations that you apply to the field.

Quick measures

Use the quick measures feature to quickly change the aggregation that is applied to the data in a visual. The default aggregation function is Sum, but you change this by right-clicking the **Values** field in the VISUALIZATIONS pane, and choosing a different function, such as Average, Minimum, Maximum, or Count. You can also show the values as percentages of subtotals or grand totals already included in the dataset.

Reference lines

Use the **Analytics** pane to create trend, constant, and dynamic reference lines on selected visuals. A constant reference line is located at the value you specify—for example, 10 million on a sales bar chart—regardless of the underlying data. You use dynamic reference lines to add lines based on minimum, maximum, or average, which change dynamically depending on the underlying data. You can also have multiple lines on one chart, including more than one constant line. You customize each line by changing the color, transparency, and dash type—and whether the line sits in front or behind the data points. The lines that you can add depend upon the visual that you use.

The following visuals can include all lines:

- Area chart
- Line chart
- Scatter chart
- Clustered column chart
- Clustered bar chart

The following visuals can only include a constant line:

- Stacked area
- Stacked bar

- Stacked column
- 100 percent stacked bar
- 100 percent stacked column

The following visuals can only include a trend line:

- Nonstacked line
- Clustered column chart

Demonstration: Visualizing data with Power BI Desktop

In this demonstration, you will see how to:

- Add visualizations to a Power BI report.
- Apply basic formatting to the visualizations.

Check your knowledge

Select the best answer

Which of the following is not a real chart type?

100 percent stacked bar chart

Line and column chart

Multirow card chart

Donut chart

Pie and line chart

Check answer Show solution Reset

Lesson 4: Overview of self-service BI

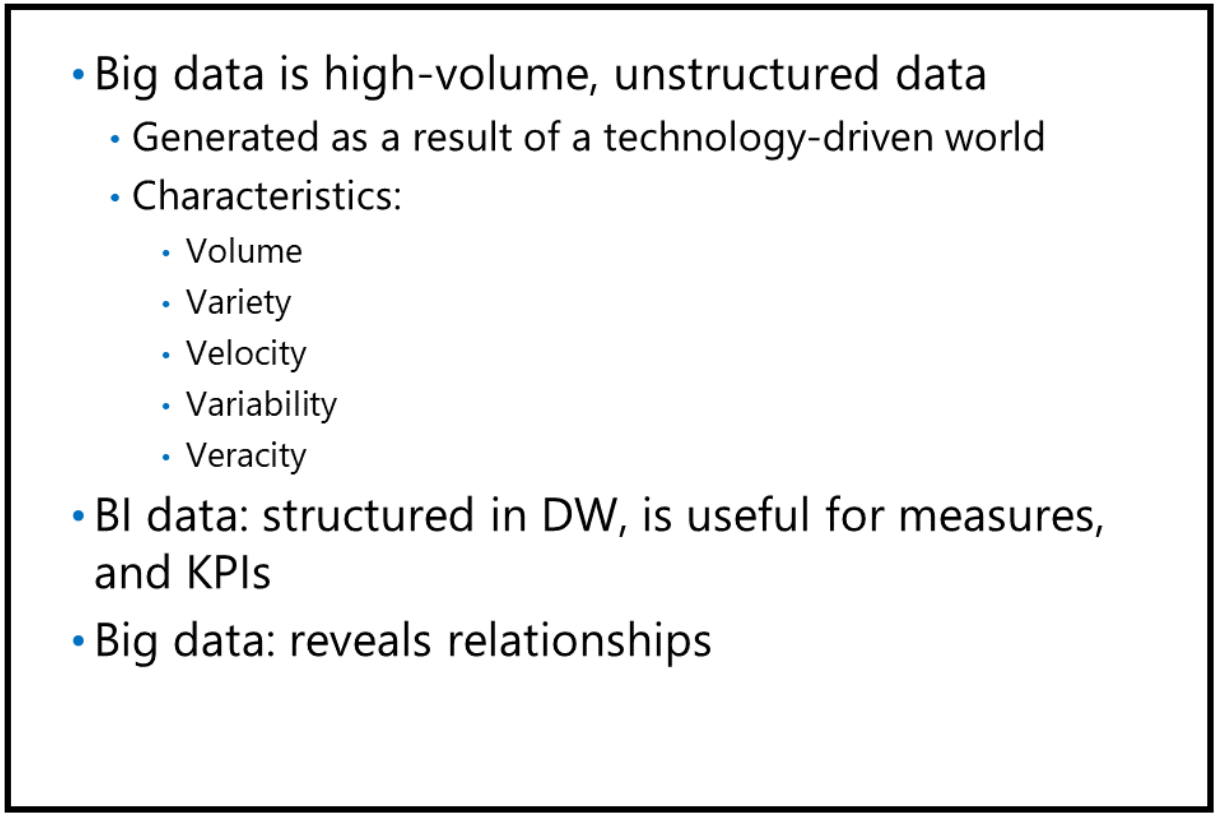
This lesson describes how the recent growth in data has driven the need for self-service BI solutions, and compares managed enterprise BI to self-service BI solutions.

Lesson objectives

After completing this lesson, you will be able to:

- Understand how prolific data growth has affected and driven the BI market.
- See how managed enterprise BI solutions limit users.
- Explain why self-service BI has become such a popular choice.

Data explosion



The term “big data” was recently plunged into the limelight to describe the vast quantities of unstructured data being generated in our technology-driven world. It is now a common term used not only by the CTOs and by CIOs in the boardrooms of major global organizations—such as Microsoft, Amazon, and Facebook—but also by organizations in all sectors and of all sizes. It seems that big data is unavoidable. Big data is too large for traditional software programs to capture, store, and manage, and presents a challenge to businesses wanting to analyze this data.

Big data is described using the following characteristics:

- **Volume**: this is the quantity of data generated and stored. The data must be large

enough to be considered big data, and the size is also a determining factor of the value—and whether insights can be gained from it.

- **Variety**: this refers to the type of data. For example, data taken from a Facebook feed would gather text, photos and images, and video.
- **Velocity**: this is the speed at which data is generated and processed. Big data can be available in real time, using in-stream technology to view it as it's in motion.
- **Variability**: this refers to the consistency of the data; that is, how much does it vary? Inconsistency causes issues with data processing and management.
- **Veracity**: this is the quality of the captured data. The higher the quality, the better the results.

Organizations already have a lot of data, and the volume is constantly growing, with big data expanding from terabytes to petabytes. It is not easy for business to cope, especially if an organization considers all data to be valuable—and does not know how to separate any data that is not useful. However, big data does have a shelf life, and before too long, becomes worthless. In addition, there is a cost associated with storing and managing the data.

Difference between big data and BI data

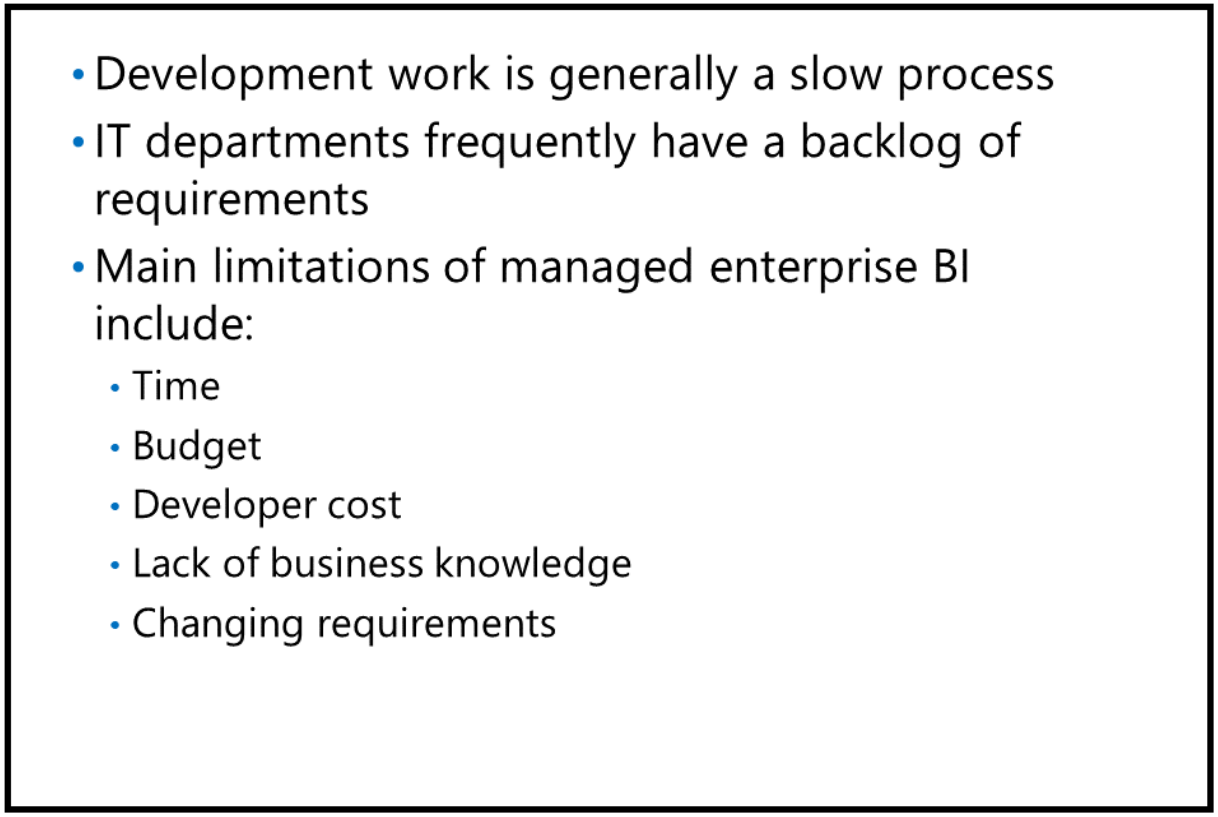
BI data is extracted from operational systems and processed using ETL. The staging area enables the data to be highly structured, consistent, and organized, ready for loading into the data warehouse. The data is highly dense, trends can be highlighted,

and data can be measured. Because of its size and unstructured format, big data requires a new approach when it comes to processing and analyzing. The data is not dense, but is a patchwork of clustered information. Rather than using measures and KPIs, the nonlinear format of big data reveals relationships and dependencies, and predicts behaviors.

Cause of big data

The Internet of Things (IoT), and social media—with their usage facilitated by mobile devices—are major contributors to the generation of big data that is unstructured and difficult to process. The IoT is a network of objects embedded with software, electronics, and sensors. Built-in network connectivity enables devices, buildings, and vehicles to communicate and exchange data. Increasingly, IoT technology is entering our homes, built in areas such as fridges, thermostats, fitness wristbands, and AV equipment. Not only are these devices gathering data, we can also often control them remotely. Social media websites such as Facebook, Yammer, Twitter, and LinkedIn, all operate on the connection of interpersonal relationships, generating data containing a variety of text, images, photos, hyperlinks, and video.

Limitations of managed enterprise BI



The nature of software development—for example, web applications, database development, or report creation—means a project can take a long time to come to fruition. IT departments are frequently overloaded with user requests for new features, or changes that need to be made to existing systems. This can be obstructive to users wanting to do their work, because they are waiting on a developer as an available resource to complete the task. IT departments, especially development teams, often have a backlog of work.

The main limitations of managed enterprise BI include:

- **Time**: one of the biggest factors in managed enterprise BI is the time taken to

develop the ETL system, build the data warehouse and cubes, write code to query the data, and design, develop, and publish reports. Even in a small organization, this is not a quick process—it requires planning, and a team of skilled developers. Much of the work is often centered on transforming the data in the staging database after extracting it from the source systems. This is ongoing work, because anomalies that arise from the source systems must be continuously monitored and fixed. Furthermore, the design and development of reports can be a slow process, especially if there is a lot of detailed information over several pages.

- **Budget**: the budget is linked to the time it takes to build the BI infrastructure and associated code base. The amount of work required up front before anything tangible is delivered is often a concern for stakeholders. Developers might be working hard creating the ETL and data warehouse, but until reports are delivered, stakeholders and users do not see that anything is actually being done. This can be difficult for nontechnical users to understand—why must they wait so long for what they consider a straightforward report? The cost of hiring BI developers is also expensive, especially if contract staff are required solely for the length of time it takes to deliver the project.

- **Developer cost versus business user cost**: in many instances, the cost of employing a report developer is costlier than that of business users. It makes financial sense, therefore, to empower the business user to create their own reports.

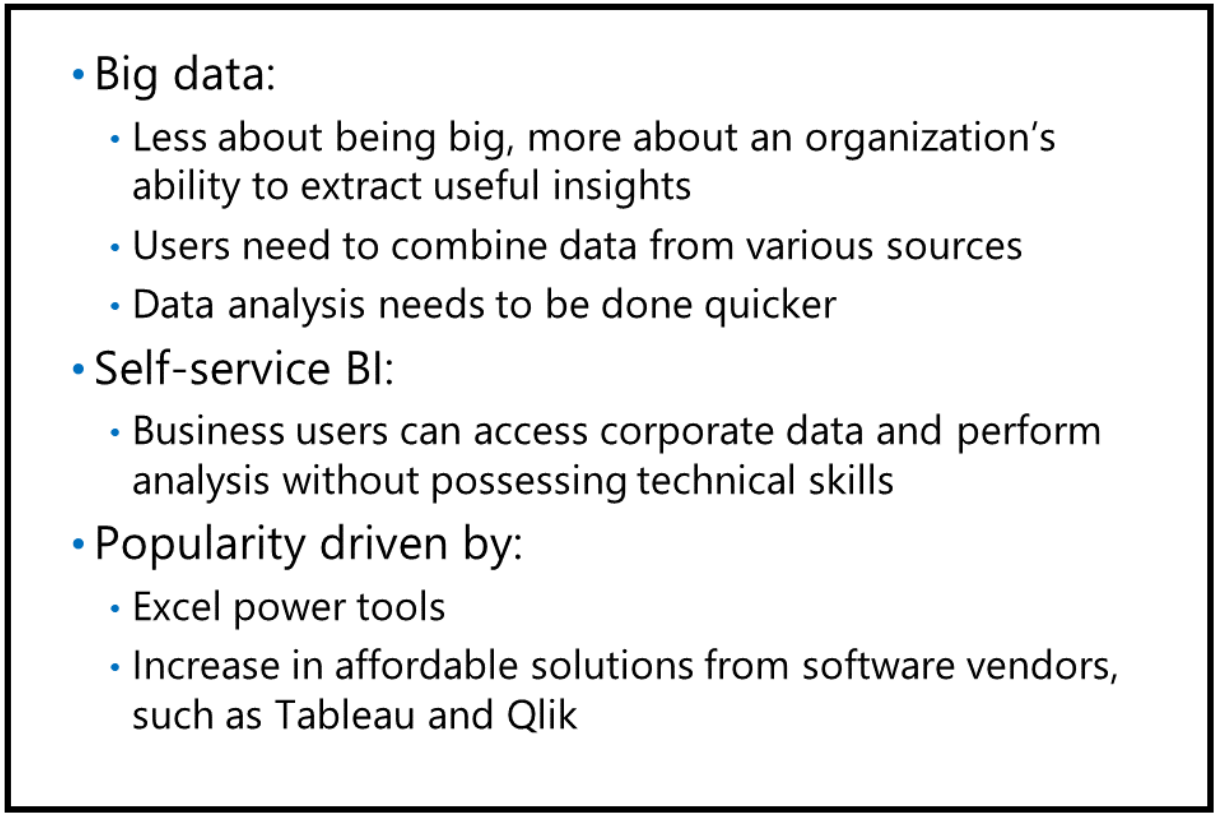
- **Lack of developer knowledge**: while a report developer might be highly technically skilled, they do not necessarily possess an understanding of the

business, or the data. If this is the case, the developer is unlikely to produce a report that details exactly what the user needs. This can be frustrating when a user has been waiting for a developer to be available to create the report, only to find it is not what they need. A request for change is then submitted, and the user must wait for this work to be done. However, a further request does not guarantee that the developer will produce what they need.

- **Changing requirements**: in addition to user requests to change reports that do not actually give the user the data they need, developers must cope with new requirements, and increasing volumes of data. For example, with sales, marketing, finance, and support departments all using SaaS data sources—requiring publicly available datasets to be included in their analysis, and statistics from customer data and internet usage—the developer must continuously integrate new data.

However, even if an organization handed over all report development to the business users, there would still be a requirement to build the ETL and data warehouse, provide access to the databases, ensure security is properly implemented, and assist users with complex queries.

Self-service BI trend



Nowadays, big data is less about it being big, and more about an organization's ability to extract useful insights from it, to improve company performance. Many SaaS providers, such as MailChimp and Google Analytics, already offer some level of data analysis to their customers. However, this usually involves the customer logging into the SaaS portal to view the data. Having the ability to download data from MailChimp, Twitter, and Facebook—and combine this with a marketing campaign created in Marketo—offers more cohesive insights. Being able to analyze data, and react to it quickly, requires a quick turnaround time for processing data. Dependency on an IT department delays this considerably, so business analysts utilizing a self-service BI approach have greater gains from their data.

What is self-service BI?

Self-service BI is an environment in which business users access corporate data to produce their own reports, without dependency on IT. Until quite recently, BI was held tightly in the realm of specialists, who were highly skilled in the use of the tools on offer. Now, with modern self-service BI tools, users do not need to have IT skills in writing complex database query code, developing data warehouses, reports, or data mining. Self-service BI tools do most of the hard work, enabling the user to quickly produce data that's suitable for analysis and can be shared with colleagues.

Why is self-service BI so popular?

Using a self-service BI tool frees up IT departments, and means business users can generate reports exactly how they want them. When thinking of self-service BI, Microsoft Excel initially comes to mind. Its popularity as a spreadsheet program, ideal for day-to-day number crunching, was boosted by the inclusion of the four additional power tools—Power Query, Power Pivot, Power View, and Power Map. These tools take data from a tabular format that is difficult to read, enabling external data connections, data formatting and manipulation, and a whole host of charts and maps to present the data, and perform deeper analysis. Adding these tools into a program with which millions of users were already familiar, takes BI from the boardroom, and gives the power of analysis to the business user.

Furthermore, a wide range of tools are on offer in the self-service BI solutions marketplace, ranging from Microsoft's Power BI suite of tools—which is available on a free license—to solutions from popular vendors such as Tableau, and Qlik, that are

priced considerably higher. Yet, while the license fees may initially appear steep, return on investment (ROI) of this initial financial cost is recouped when compared to the time cost of employing a report developer to manually create equivalent reports, and manage them. These tools can also deal with unstructured data better than spreadsheets, which need data in tabular format before any visualizations are applied. With the ubiquity of big data in business, it is fast becoming a requirement that a BI tool should cope with the challenge.

Check your knowledge

Discovery

Given what you have learned so far in this module, regarding the limitations of managed BI and the uptake of self-service BI with all its advantages, do you think there is a future for managed BI?

Show solution Reset

Lesson 5: Considerations for self-service BI

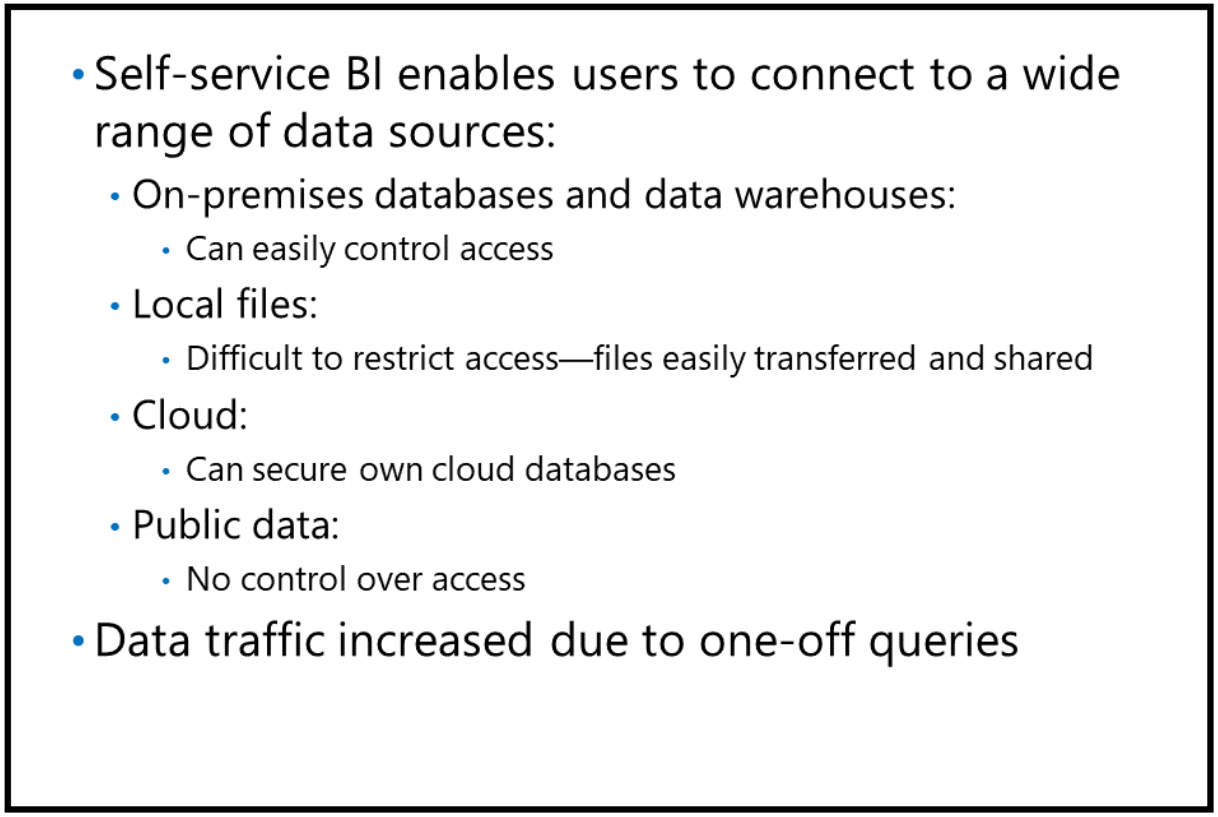
This lesson looks at some of the important aspects to consider when planning a self-service BI solution. This includes issues users might have when accessing data, the importance of data source reliability, how users require analysis skills, and how a data steward can help.

Lesson objectives

After completing this lesson, you will be able to:

- Explain issues that arise when accessing data in a managed and a self-service BI solution.
- Understand why the reliability of data sources is important.
- Describe how users need some expertise in data analysis.
- Explain the role of the data steward.

Data access



By using self-service BI, users connect to a wide variety of data sources, including on-premises databases and data warehouses, local files, cloud services, SaaS hosted solutions, and public datasets. While managed BI solutions tend to be more highly controlled by policies maintained by IT, self-service opens up the possibilities for importing data from anywhere, outside the control of IT.

On-premises data

Self-service access to on-premises data is generally controlled by IT. Data can be controlled in how it is shared with users, through database security rules to restrict access to sensitive data. For example, users of SQL Server databases are given

access to data views, which provide selective fields, without giving full access to other sensitive data. It is imperative that data is protected and controlled, and that business users have access to the data they need to do their job.

Data from files such as Excel, CSV, text, and XML, are emailed, shared, and imported into a self-service BI solution. It is harder for IT to control and secure access to this data, because it is easily transferable, both within the organization and externally.

Cloud and public data

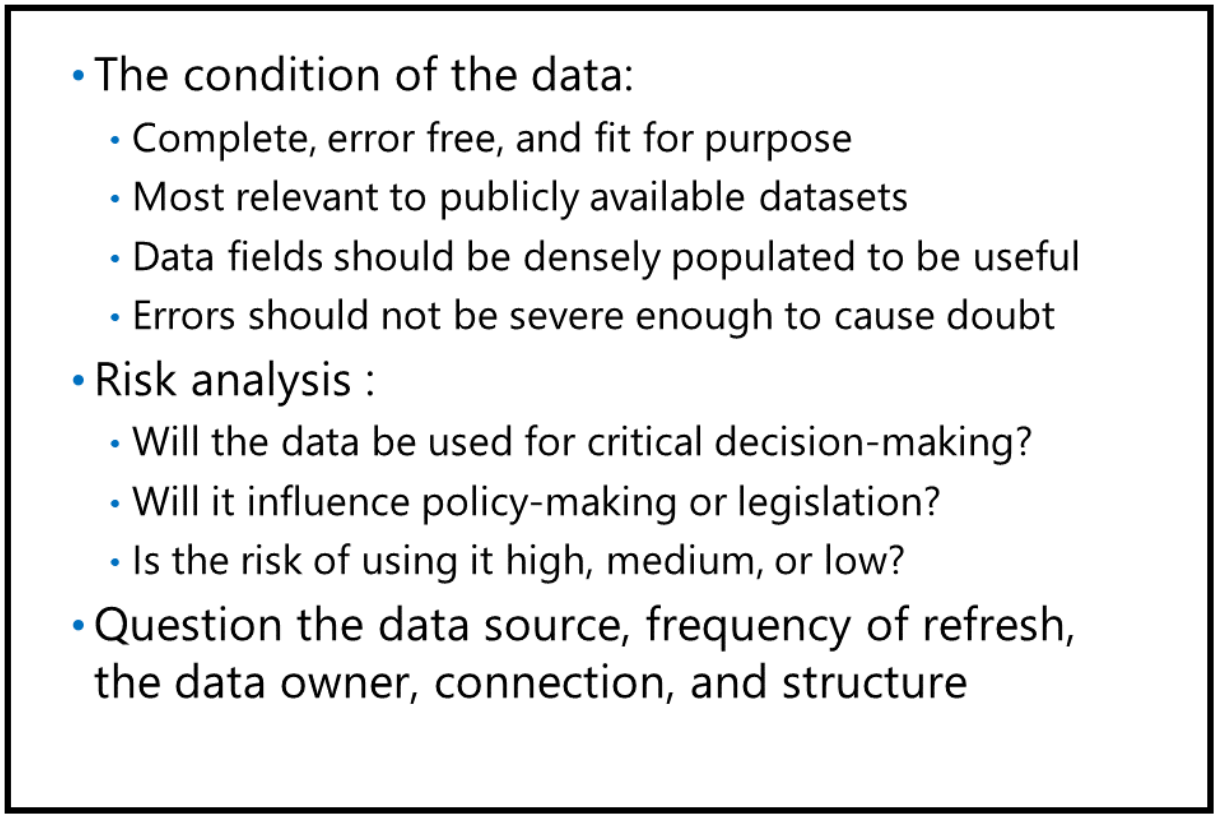
Self-service BI enables business users to take advantage of publicly available data. Data repositories, such as Microsoft Azure Marketplace, Amazon Web Services, and Wikipedia, all provide datasets, some of which are free. These can easily be incorporated into a self-service BI solution, by downloading the data, or by connecting directly to the source using a URL from within the self-service BI solution. This provides quick and easy access to very useful data that enhances the analysis of existing corporate data.

Databases stored in the cloud, including Microsoft Azure SQL Database and Microsoft Azure SQL Data Warehouse solutions, can be managed by IT with the same security principles applied. Users connecting to cloud-based data stored by SaaS providers require a username and password.

Data traffic

Enabling users to access large datasets and transfer data by sharing reports can cause issues with the volume of data moving around the network. With many users accessing data in a specific, one-off fashion, the load on the network increases. IT needs to monitor the performance of servers and networks to prevent bottlenecks. For managed BI solutions, this is less of an issue because precompiled queries executed against the database provide better performance, and data is cached.

Data reliability



Data reliability refers to the condition of data, and whether it is complete and sufficiently free of errors, so that it’s fit for purpose. This is particularly relevant to data

imported from public sources. To be complete, the data fields must be sufficiently populated. A dataset with a sparse population of data across many fields and rows cannot provide suitable results. The data need not be entirely free of errors, but the errors that do exist must not be severe enough to make the user doubtful of the results and question their validity. The data within each field should accurately represent the field, be correct, and of the correct data type. This ensures that the data is analyzed with confidence.

Risk analysis

Risk analysis is a useful and often essential exercise to perform on data that is imported from sources external to the organization. If you need to make serious decisions after analyzing the data, then consideration must be given to the reliability of the data. In such circumstances, the following questions should be considered:

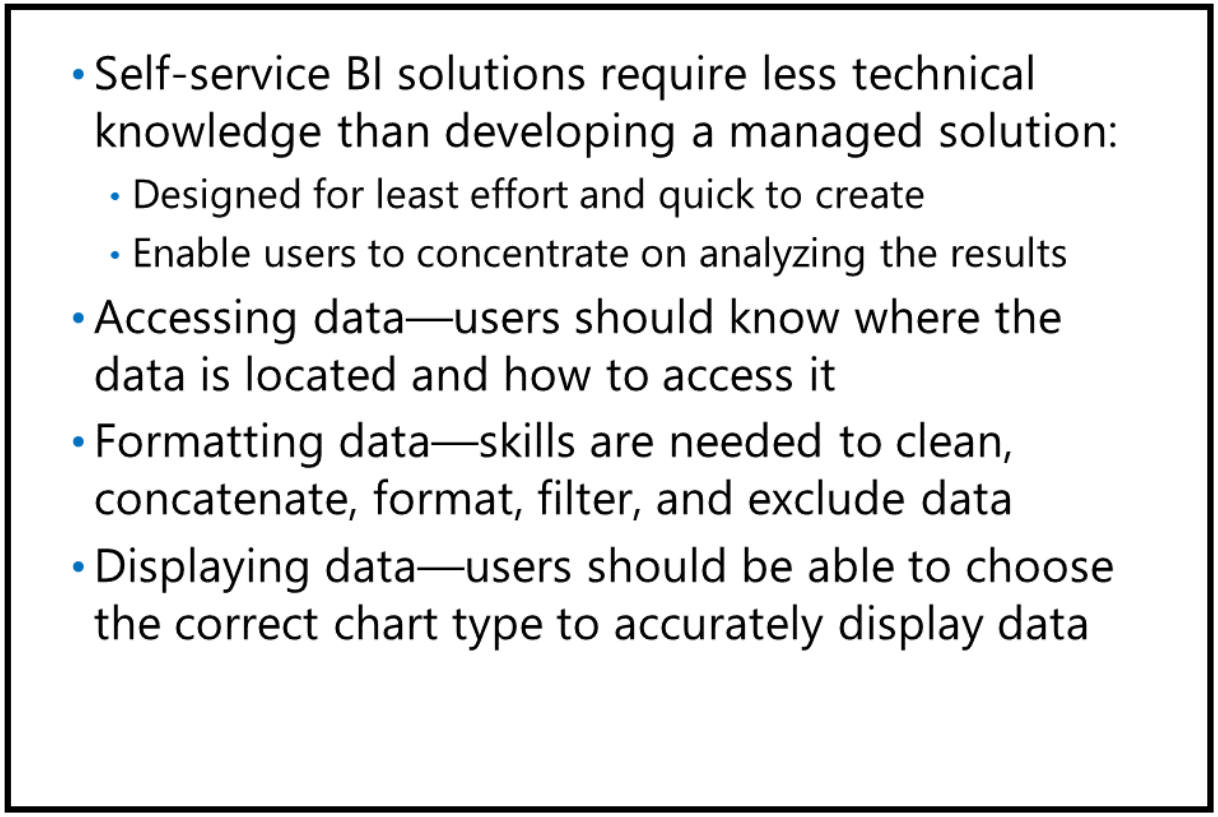
- Is the data to be used for critical decision-making by an organization or individual?
- Will the figures be used to influence policy-making or legislation?
- Is the risk of using the data high, medium, or low?
- Is the data of a sensitive nature?
- Will the results of the data be made available publically?

When performing risk analysis to determine the reliability of the data, the following questions should be answered as part of the assessment:

- **Data source**: where has the data come from? Is the data provided by a reputable organization?
- **Data refresh**: how often is the data refreshed? Does the analysis that uses the data require it to be kept up to date, for the reported results to be useful and accurate?
- **Data owner**: who owns the data? Does the organization require any permission to use the data? Is it permissible to publish reports that include the data?
- **Connection**: are there likely to be any issues connected with the data? What is the up time of servers on which the data is stored? Will the data always be available, or is there a time limit?
- **Structure**: will the structure of the data change, thereby requiring the dataset to be reimported?

Data from on-premises databases that store corporate information do not need to undergo such extensive risk assessment. Data should already be qualified, especially if it is derived from a data warehouse that has been designed and managed in-house.

User expertise



Self-service BI solutions require less technical knowledge than is needed to produce a managed BI solution. The purpose is for users to create reports as quickly as possible, with the least amount of effort, so that time and energy can be spent on analyzing the results of the reports. However, having knowledge of the business, formatting data, and understanding which visualizations best display the data, are useful for making the most of the BI solution.

Accessing data

Users need to know where data is located, and how to access it. Data stored in on-premises databases or data warehouses are supported by IT, so there is likely to be

scope for a developer to provide queries, or offer advice on exporting data. External guidance might be required for accessing data held by third parties, including SaaS providers, and publicly available datasets.

Formatting data

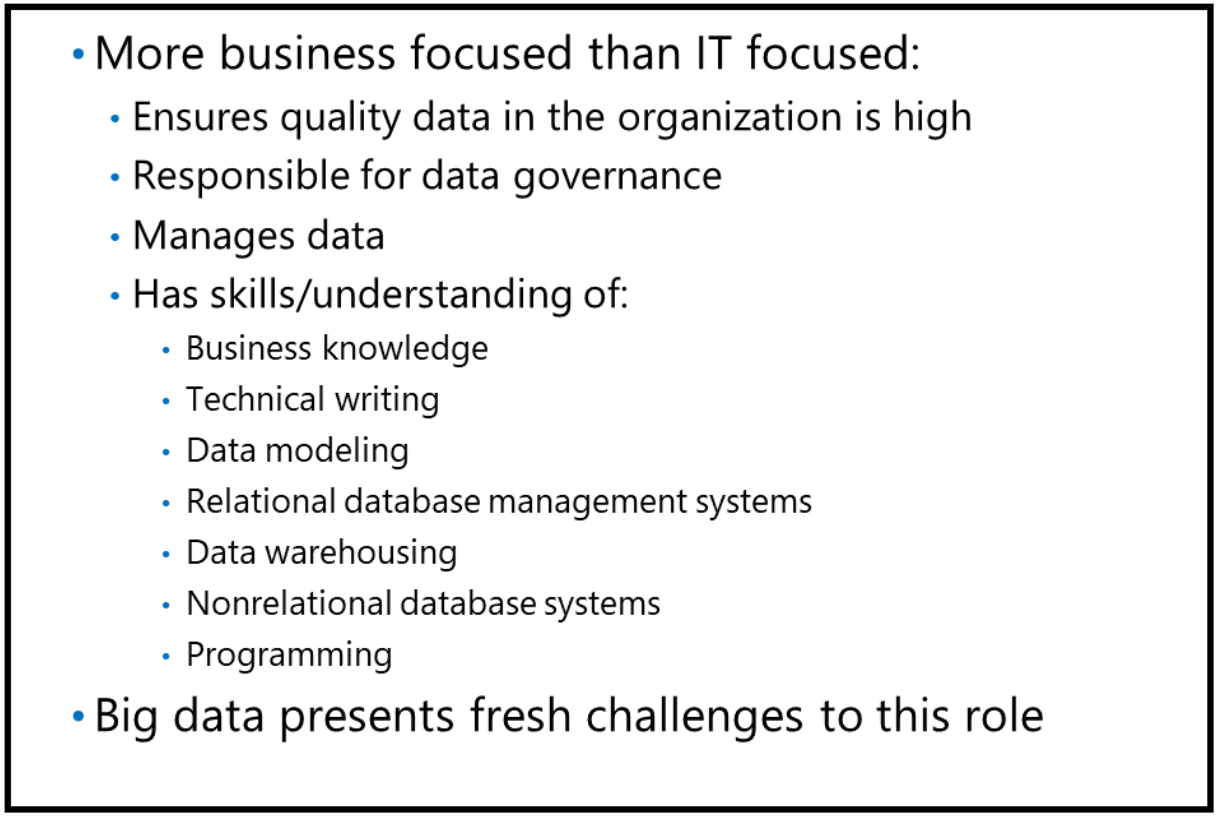
Transforming and formatting data is an important step in the process of building reports. If the data is not right, then the results will not be right. Users must understand the principles and structures of data that is sourced from a relational database, a data warehouse, or an unstructured big data source, such as a social media site. Skills are required to:

- **Perform data cleaning**: remove duplicate rows, handle dirty data, and errors.
- **Concatenate data**: create new columns by combining existing columns.
- **Format data types**: ensure currency, number, and datetime columns have the correct data type.
- **Apply adequate filtering**: ensure data can be filtered to the expected granularity. How do sales need to be measured? Do “days” represent a fine enough granularity or does the report need to show online sales by the hour?
- **Exclude redundant columns**: columns and rows that are not needed in the dataset should be removed, to make the dataset easier to manage and understand.

Displaying data

Users should be familiar with all the major chart types and understand how to use them to display data most effectively so that decisions can be made. For example, geographic data is best presented using a map chart; a scatter chart should be used to show overlaps in data, clusters, and outliers. Financial data, such as a share price, is best displayed using a line chart. There are plenty of free internet resources that show examples of all the chart types and how they can be used. This will help self-service BI users quickly become familiar with chart types. Users should also understand how to create and use measures and KPIs.

Data stewards



The data steward role is aligned more with the business than with IT. A data steward ensures that the quality of data in an organization is high and is responsible for data governance. With the proliferation of data in organizations, a data steward is considered less of a luxury, and more of an essential role. A data steward has a varied role in managing data, and is responsible for:

- Master data management.
- Ensuring the consistency of data between systems.
- Mapping data between different systems.

The data steward is responsible for managing data in the following ways:

- Removing duplicate data, particularly lookup data, or data that should be stored once.
- Removing unused, out of date data; for example, a product category that is never used.
- Removing ambiguous data.
- Checking data is fit for purpose.
- Securing data to ensure only authorized users can make amendments.
- Documenting metadata.
- Ensuring the organization adheres to data-related legislation.
- Determining data security requirements.
- Monitoring the quality of data.
- Developing data definitions.
- Establishing naming standards and conventions.
- Documenting business rules.

The data steward should either possess skills in, or a thorough understanding of, the following areas:

- **Business expertise**: the role of the data steward sits more with the business side than with IT. It is crucial that a data steward understands how the business functions and has departmental knowledge of all business areas, such as finance, marketing, sales, enterprise resource planning (ERP), manufacturing, retail, and supply chain.
- **Technical writing**: the data steward is responsible for documenting the metadata and should be able to write clearly, and with accuracy. The documentation spans multiple departments within the organization and must be clear to all who read it.
- **Data modeling**: although data modeling experience is not necessary, the data steward works closely with the technical architect and, at the very least, needs an understanding of terminology.
- **Relational database systems**: an understanding, or preferably first-hand experience, of relational database management systems is vital for the data steward. This role works closely with database developers, so knowledge is crucial.
- **Data warehousing**: understanding data warehouse concepts, including ETL, is also essential for the data steward to communicate effectively with BI developers.
- **Nonrelational database systems**: the emergence and pervasiveness of big data requires an understanding of unstructured, large volume datasets, the issues of managing them, and the technology required to process them.

- **Programming skills**: understanding programming and being able to directly manage data in the database is a useful skill for the data steward.

Data that is managed by a data steward will be of a higher quality than data that is not. This quality will be reflected in the data that is presented to customers, and data used in reporting and analysis. The growth of data provides continuous challenges to the data steward—the rise of big data demands another element of management that is less easy to apply, given the size and lack of structure.

Check your knowledge

Discovery

Discuss the role of the data steward. Does your organization have a data steward? If not, do you think one is necessary? Discuss some of the issues your organization faces, that your existing data steward manages, or that the addition of one could solve.

Show solution Reset

Lesson 6: Microsoft tools for self-service BI

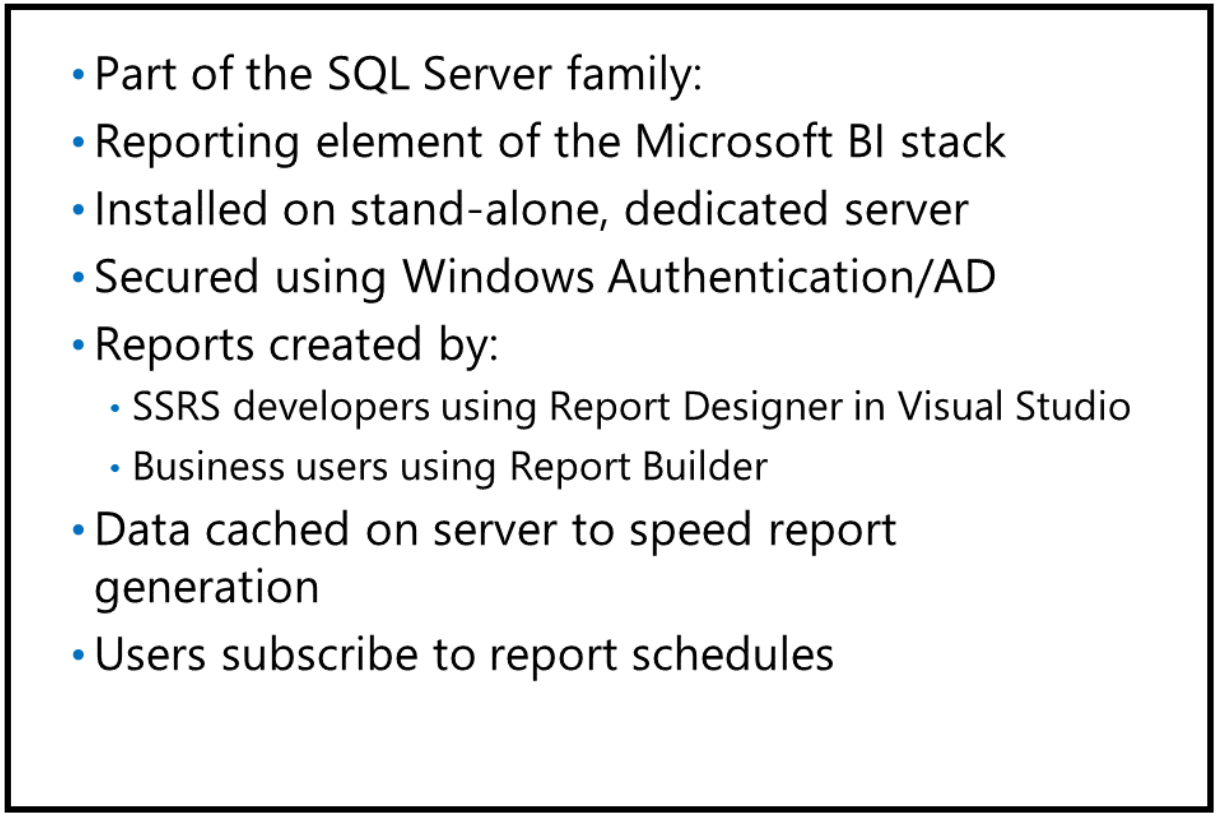
This lesson reviews the different self-service BI solutions currently offered by Microsoft, and looks at the benefits and restrictions of using each solution.

Lesson objectives

After completing this lesson, you will be able to:

- Describe the main features of SQL Server Reporting Services.
- Understand how Excel is currently used as a self-service BI solution.
- Explain how SharePoint® Online is used for sharing and collaborating.
- See the benefits of using Power BI Desktop as your self-service BI solution.
- Understand the capabilities of Power BI Report Server.

SQL Server Reporting Services



SQL Server Reporting Services (SSRS) is part of the SQL Server family, comprising the reporting component of the Microsoft BI stack. SSRS was introduced in 2004 as an add-in to SQL Server 2000. Since then it has grown to be a popular reporting choice for organizations running SQL Server. Reporting Services is generally installed as a stand-alone instance, as report generation requires much hardware resource, and SSRS works best on a dedicated server. Servers exist on-premises and security is managed using Windows authentication and Active Directory (AD).

Developing reports

Organizations using Reporting Services usually have a dedicated report developer to create and update the organizational reports. The developer will have skills to query the relational database (OLTP) system, and the data warehouse if one is used. Report Designer in Visual Studio® is the main development environment for creating reports for SSRS. Usually, the developer accepts user requests to create a report based on a specification.

Reports can span multiple pages, and SSRS reports are particularly adept at managing data tables that expand to fit the size of the data, which might be unknown at design time. SSRS also supports HTML5 rendering and mobile reports, in addition to a wide range of charts, including sunburst and tree maps.

Business users with more advanced skills can create their own reports using Report Builder.

Deploying reports

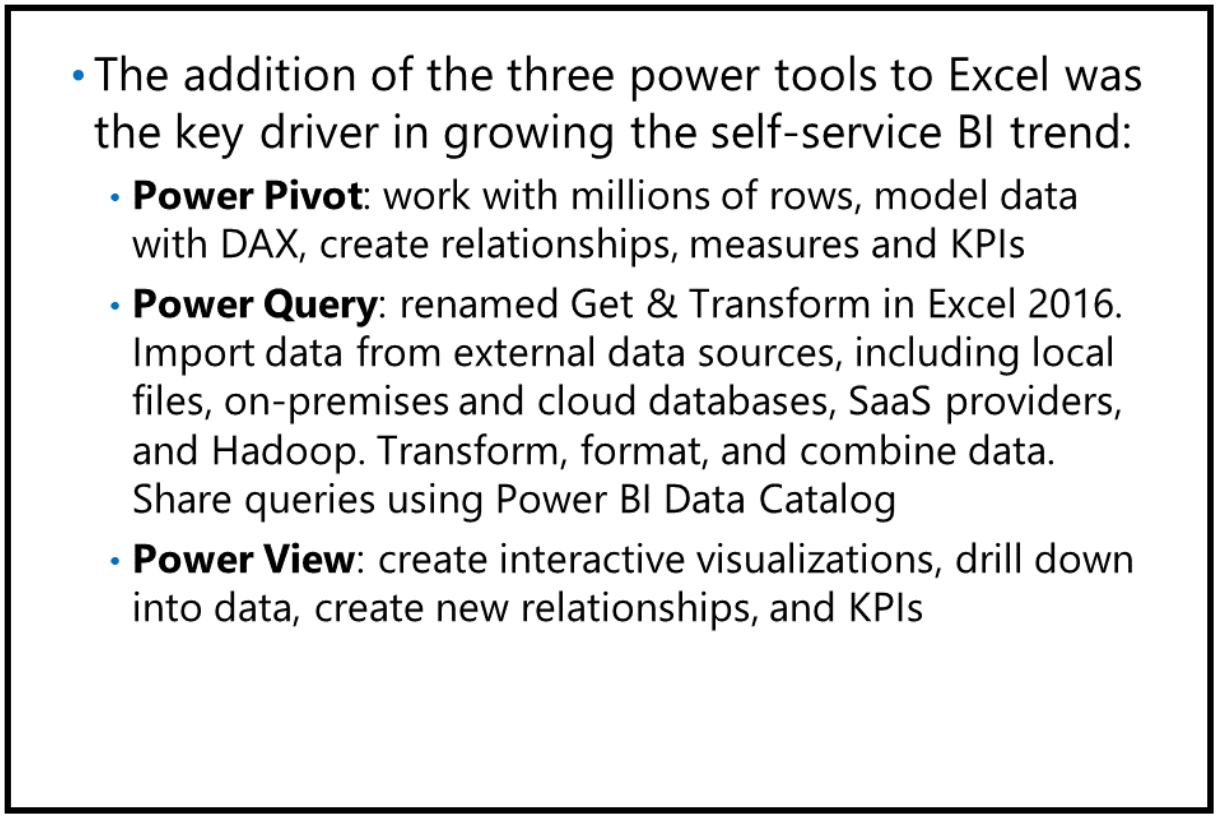
After developing a report, it is deployed to the Report Server. The dataset is deployed alongside the report, and the data can be cached for faster report generation. This is useful when multiple users access the report, but the data is not frequently updated, because it delivers a faster experience.

Report subscriptions

By subscribing to scheduled reports, users receive an email message with a report attached. With the right permissions, users generate reports using the Report

Manager portal, and by subscribing to report subscriptions. Reports are delivered as soon as data is updated, or can be emailed after the data warehouse load has completed overnight, so that the report is available at the start of the business day. Reports can be sent to users outside of the organizational domain.

Excel



Microsoft Excel has a loyal following of business users, and its leadership in the spreadsheet software market has long remained unchallenged. The addition of the power tools—Power Pivot, Power Query, and Power View—moved Excel to new heights, bringing self-service BI to its massive fan base. A key driver in the recent

uptake of self-service BI was the enabling of business users to analyze and report on data without dependency on a managed BI solution. These power tools have liberated business users, and reduced the workload on IT to develop and manage a time-consuming BI solution.

Power Pivot

Power Pivot was launched as an add-in to Excel in 2010, but since Office 2016, this is now included as part of the standard installation. This feature enables advanced data modeling, and data analysis—much of Power Pivot's strength lies in its ability to handle large datasets that have been imported from different data sources. You use Power Pivot to convert raw data into useful, visual charts and maps, helping you discover business insights, and trends. Using Power Pivot, you can:

- Import millions of rows of data from different data sources, including external sources.
- Model data using DAX functions.
- Create relationships between tables of data, including tables from different sources.
- Integrate with the other power tools to create charts, pivot tables, maps, and interactive Power View visualizations.
- Add measures and KPIs to your data model.

> **Note:** To use Power Pivot in Office 2016, open Excel, on the **File** menu, point to **Options**, and then click **Add-ins**. In the **Manage** dialog box, click **COM Add-ins**, and then click **Go**. Select **Microsoft Power Pivot for Excel**, and click **OK**.

Power Query

Since Excel 2016, Power Query is known as **Get & Transform**, and the tools are located on the **Data** tab in Excel. You can use Get & Transform to search for data sources, connect to the data source and import the data, and then shape the data ready for visualizing. With Get & Transform, you can:

- **Connect**: you connect to local files including an Access database, CSV, or Excel file, data stored in the cloud, and data located on the internet, in addition to on-premises SQL Server, and SQL Server Analysis Services databases, Oracle, and MySQL. You can also connect to SaaS providers such as Facebook, and Salesforce, and big data sources including Hadoop.
- **Transform**: you can transform data using the Query Editor. Transformations enable you to shape your data so it is in the structure and format required to fulfill your reporting and analysis objectives. You can create new columns, remove columns and rows, and split columns. You alter data types to ensure number and currency values are aggregated and displayed correctly. Text data can be cleaned and trimmed, and the case changed to upper, lower, or title. You can also write your own transformations using the M Language.
-

Combine: you combine rows from different tables to create a new table, and you can append rows from one table to the end of the rows in another table.

- **Share**: rather than saving your workbooks and distributing them to colleagues using email, you can share the queries in your workbooks to the Power BI Data Catalog. You can also Merge and Append queries.

> **Note:** The data sources you connect to depend on the license you have. Some sources are only available to Professional, and Professional Plus license holders.

Each of the steps you perform as part of Get & Transform records in the Query Editor, enabling you to undo, redo, reorder, and even modify steps using the M Language.

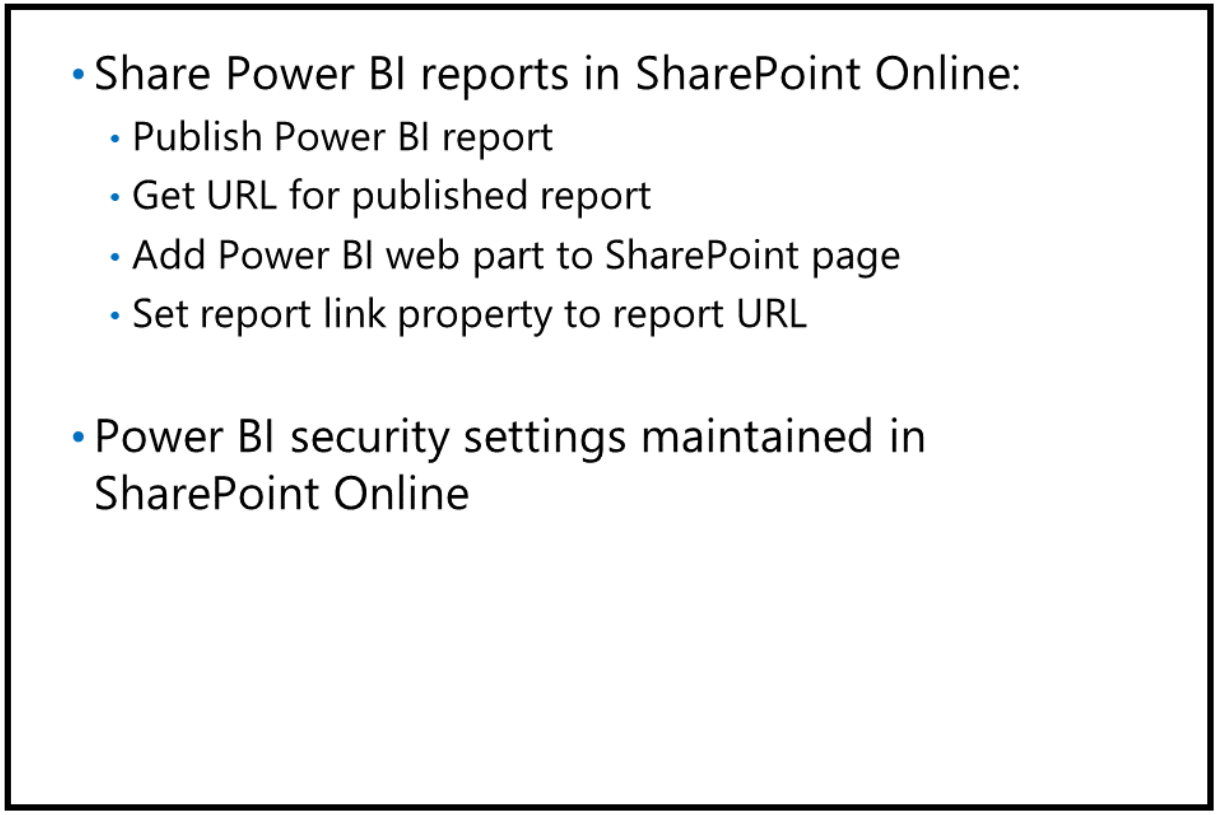
Power View

Power View is an interactive visualization tool that you can use to quickly build a model, using the drag-and-drop interface. You can use advanced pie charts, maps, and data hierarchies that enable drill-down into your data. In addition, you can create new relationships and add KPIs based on these new relationships.

> **Note:** To use Power View in Office 2016, open Excel, and on the **File** menu, point to **Options**, and then click **Add-ins**. In the **Manage** dialog box, click

COM Add-ins, and then click **Go**. Select **Microsoft Power View for Excel**, and click **OK**.

SharePoint Online



In addition to using SharePoint data in your Power BI reports, you can also share Power BI reports in SharePoint Online.

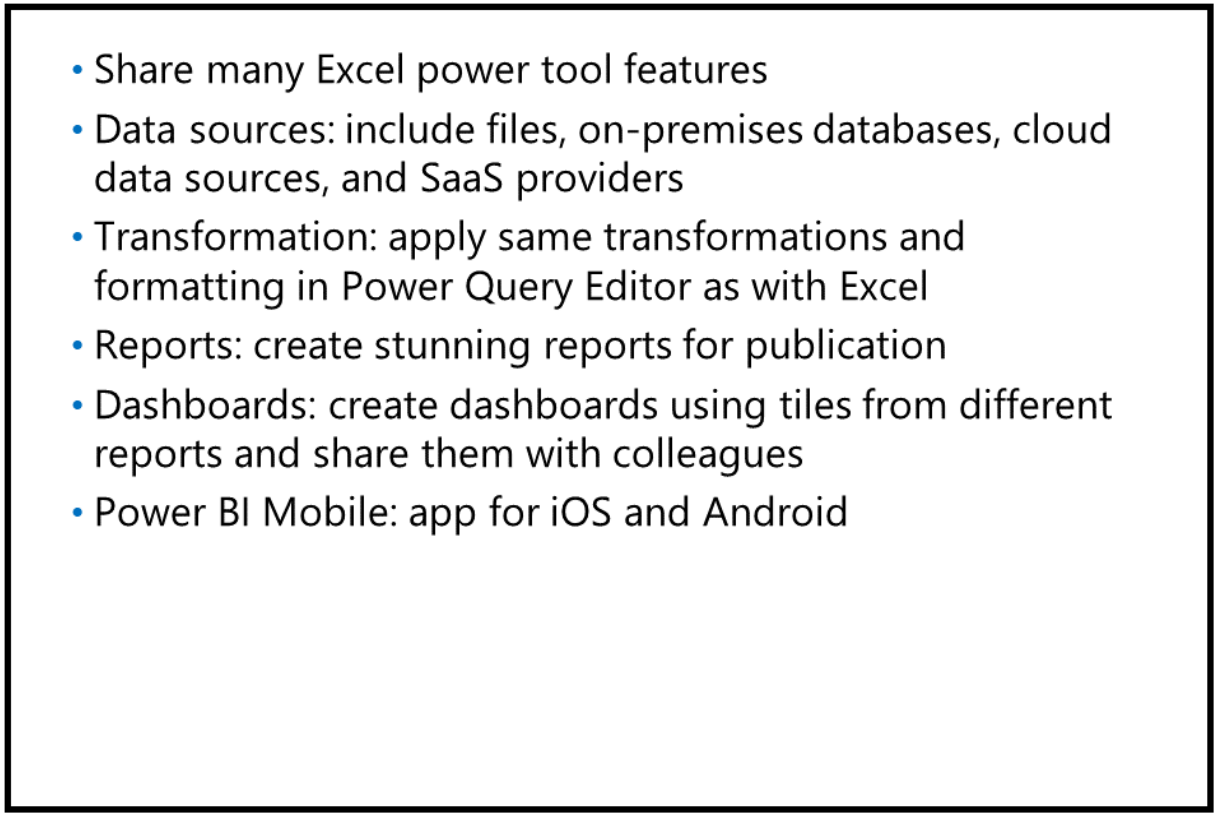
After publishing a report to the Power BI Service, you locate the URL for that report. When you add a Power BI web part on SharePoint Online, you specify the URL of the report, which embeds the report into your SharePoint page. Only users who have

been granted access to the report in the Power BI Service are able to view the report, so security is maintained in Power BI and there's no risk of unintended access being given.

The Power BI web part requires the user to have a Power BI Pro license. If they do not have one, when they access the page containing the web part, they will be directed to the Power BI website to purchase a license or sign up for a trial.

> **Note:** You can only use the Power BI web part in SharePoint Online, not on-premises SharePoint Server.

Power BI Desktop



Power BI Desktop shares many of the features offered by the Excel power tools, so business users will find transitioning between the two tools to be a straightforward process. Power BI Desktop is a stand-alone tool that enables you to import data, model and apply transformations to your data, and then create stunning, interactive reports.

You can download the 32-bit or 64-bit version of Power BI Desktop from **https://www.microsoft.com/en-us/download/details.aspx?id=45331**. It is updated monthly with new features and capabilities. A message in the user interface will notify you when an update is available to install directly from the product.

The original Power BI Desktop application enables you to upload reports to the online Power BI service, where colleagues share reports, and create dashboards. Power BI is available on a Free license, or a Professional license that offers extra features, and supports a higher volume of data for a small monthly fee.

Power BI Desktop, optimized for Power BI Report Server, provides the same report creation functionality, but enables you to upload reports to the on-premises Power BI Report Server. Power BI Report Server is available through either a Power BI Premium or SQL Server Enterprise Edition with Software Assurance purchase. For more information about Power BI Desktop optimized for Power BI Report Server, see *Install Power BI Desktop optimized for Power BI Report Server* in the Power BI documentation.

Install Power BI Desktop optimized for Power BI Report Server

http://aka.ms/AA52x7l

Data sources

From Power BI Desktop, you can connect to a wide range of data sources including:

- **Files**: you can import from Excel, CSV, XML, Text, JSON, a folder, or a SharePoint folder.
- **Databases**: all the main industry databases are supported—SQL Server, Access, SQL Server Analysis Services, Oracle, IBM DB2, MySQL, PostgreSQL, Sybase,

Teradata, SAP HANA, Amazon Redshift, Impala, Google BigQuery and Snowflake.

- **Azure**: Microsoft Azure SQL Database, Microsoft Azure SQL Data Warehouse, Microsoft Azure HDInsight, Microsoft Azure Blob Storage, Azure HDInsight Spark, Microsoft Azure Cosmos DB, Microsoft Azure Data Lake Storage.
- **Online services**: the main SaaS providers are supported, including Dynamics 365, Facebook, Google Analytics, Salesforce, GitHub, MailChimp, Marketo, QuickBooks Online, Webtrends, and Zendesk.
- **Others**: you can also import from a webpage, an OData feed, Hadoop, Active Directory, Microsoft Exchange, ODBC, and R Script.

Power BI Desktop supports DirectQuery, which you can use to query the data source, rather than importing the data. This is helpful when analyzing very large datasets.

Transformations

You can use Power BI Desktop to transform your data, and the Power Query Editor feature includes the same functionality as Get & Transform in Excel. With DAX for Power BI, you can choose from more than 200 functions, constants, and operators, to help shape data exactly how you need it. DAX for Power BI is slightly different to Excel, as it works at the column, rather than the row level. You can also create calculated columns, calculated tables, and measures, in addition to using the measures in your functions.

Reports

After importing and transforming your data, you can drag visuals or fields onto the report designer, to begin building reports. The visuals are customized with colors, titles and text, and other settings applicable to each type of chart. You can also create or download custom visuals for your reports.

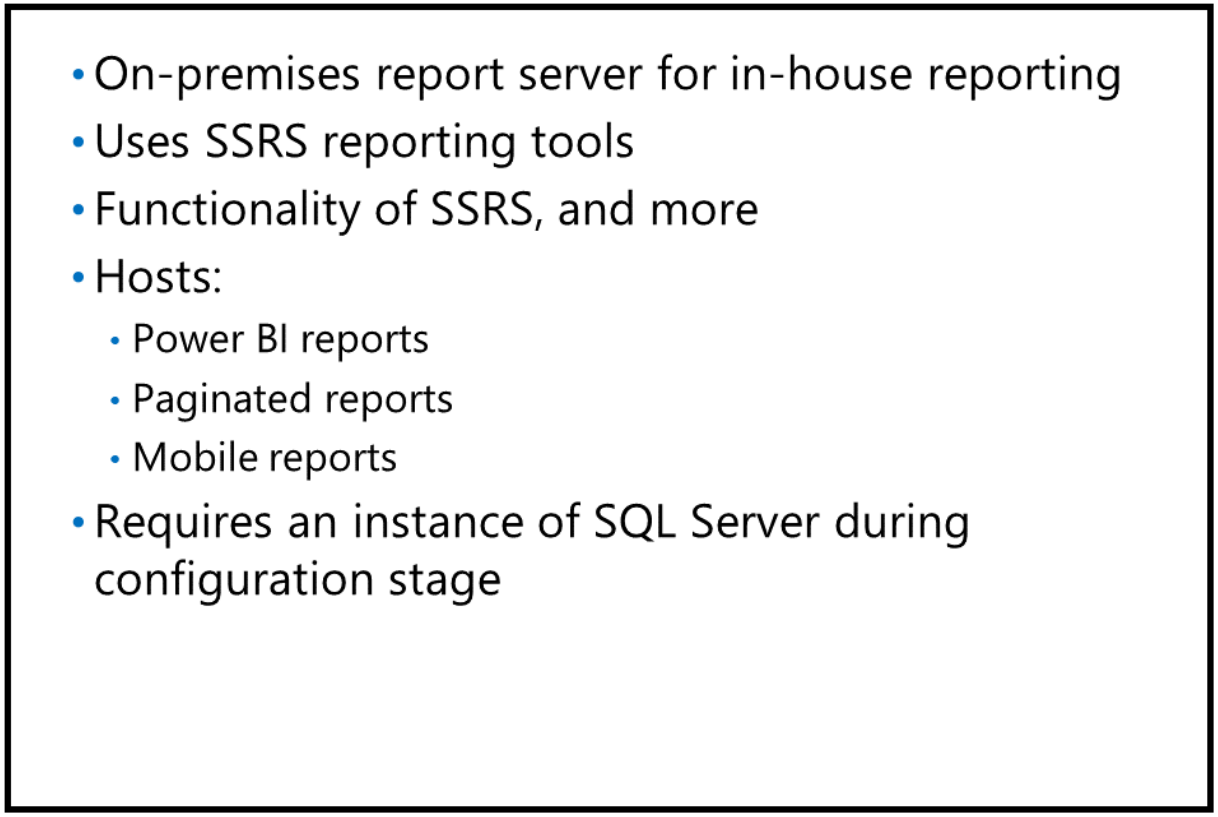
Dashboards

One of the most powerful features of Power BI is the ability to quickly and easily share reports, dashboards, and datasets. After publishing a report, you can use the report items, known as tiles, to create a new dashboard, even combining charts, maps, and KPIs from different reports. With the Power BI Service, Professional license holders can create content packs. A content pack is a bundle of reports, dashboards, and datasets that make it easy for colleagues to share their work. Users on a Free license can download and view content packs. Reports are published to the Power BI service, or Pyramid Analytics.

Power BI Mobile

Power BI offers a mobile app for iOS and Android devices. Reports and dashboards automatically adjust their size to fit the screen of the device, so you need not worry about creating mobile versions of your work. The apps are free to download, and reports and dashboards are fully interactive.

Power BI Report Server



Power BI Report Server is an on-premises report server that integrates with Power BI Desktop to create an easy to use in-house reporting platform. In addition to hosting reports with Power BI Report Server, you can use the reporting tools from SSRS to track reporting data. You use Power BI Report Server to do everything that SSRS does—and more. You can migrate your SSRS instance to Power BI Report Server to take advantage of this.

Power BI Report Server provides a web portal through which users access your Power BI reports (.pbix), paginated reports (.rdl), and mobile reports. Because users often want to view reports on demand, Power BI Report Server enables them to access their reports in a web browser, on a mobile device, or directly from an email.

Installing Power BI Report Server

Power BI Report Server is supported only on x64 processors. It uses an instance of SQL Server to host the report server databases, although no checks are made for this during the installation process.

Configuring Power BI Report Server

The final stage of the installation wizard enables you to start the report server configuration process using SQL Server Reporting Services Configuration Manager. During configuration, you need to specify the instance of SQL Server on which to create the report server database. If this is a remote instance, you will need to perform additional steps to enable remote access to this database.

Licensing Power BI Report Server

The report server is available with SQL Server Enterprise Edition with Software Assurance or as an extension to Power BI Premium. When using the Power BI Premium license, you create deployments spanning on-premises and hosted cloud.

Demonstration: Publishing a report to the Power BI Service

In this demonstration, you will see how to:

- Publish a report to the Power BI Service.
- Create a dashboard.

Check your knowledge

Select the best answer

Which of the following is not an Excel power tool?

- Power Chart
- Get & Transform
- Power Pack
- Power Pivot
- Power View

Check answer Show solution Reset

Lab: Exploring an Enterprise BI solution

Scenario

Adventure Works employees are increasingly frustrated by the time it takes to implement managed BI services. The existing managed BI infrastructure, including a data warehouse, enterprise data models, and reports and dashboards, are valued sources of decision-making information. However, users increasingly want to explore relationships with other, currently unmanaged data—and it takes too long for the IT department to incorporate these requirements into the corporate BI solution.

As a BI professional, you have been asked to explore ways in which Adventure Works can empower business users to augment their managed enterprise BI solution with self-service BI.

Objectives

After completing this lab, you will be able to:

- View reports in SharePoint Server.
- Create a Power BI report.
- Create a Power BI dashboard.

> **Note:** Because of updates to Microsoft Power BI, the lab steps for this course change frequently. Microsoft Learning regularly updates the lab steps, so they are not available in this manual − but you can access them on GitHub.

Lab setup

Estimated time: 60 minutes

Virtual machine: **20778C-MIA-SQL**

User name: **ADVENTUREWORKS\Student**

Password: **Pa55w.rd**

All the lab steps are contained in 20778C_LAB_01.md.

Exercise 1: Viewing reports

Scenario

You have been asked to compare Excel Services in SharePoint with Power BI Desktop and Power BI Service to see which offers the best self-service BI solution. You will share an Excel file on SharePoint to determine how user friendly this experience is.

The main tasks for this exercise are as follows:

1. Prepare the lab environment.
2. View reports in SharePoint Server.

Result: At the end of this exercise, the Adventure Works Sales workbook will be published on SharePoint.

Exercise 2: Creating a Power BI report

Scenario

You have published an Excel workbook to SharePoint, and you next need to see how this compares to Power BI. You will create a report and add data, and then add visualizations to the report.

The main tasks for this exercise are as follows:

1. Import data into Power BI Desktop.
2. Add visualizations to the report.

Result: At the end of this exercise, you will have a new Power BI report.

Exercise 3: Creating a Power BI dashboard

Scenario

Your Power BI report is ready to be published to the Power BI Service. Next, you will publish the report and create a dashboard, and then use the Natural Query Language to ask questions of your data.

The main tasks for this exercise are as follows:

1. Create a Power BI dashboard.

2. Ask questions of your data.

Result: At the end of this exercise, you will have published a report to create a dashboard.

Review question(s)

Check your knowledge

Discovery

Discuss using Power BI Desktop and Power BI Service, compared to using Excel and Excel Services in SharePoint. Which do you think is the best, and why?

Show solution Reset

Check your knowledge

Discovery

Has your organization started using Power BI? If not, how easy do you think it will be to implement, and convert existing business users from Excel, or other BI solutions? If you have already started using it, how do users find the experience compared to the previous solution?

Show solution Reset

Module review and takeaways

In this module, you have learnt about the basics of BI and data analysis. You have considered the emergence of self-service BI and looked at the tools available for creating self-service BI solutions.