31/8/23 Microsoft Power BI Certification : DA-100.

The main objectives: Prepare the data, Model the data, Visualize the data, Analyse the data & Deploy the data.

Power BI Service & Power BI Desktop: Requires work account to sign up.
Runs on browser. Windows Application.

Power BI Desktop: → Import data from various data sources.
→ Navigator pops up: Load / Transform data.

Views: Data view, Report view, Model view. } For Navigation.
- Data view: used to see our actual data
- Report view: create reports of our data
- Model view: creates relationships.

→ Types of Visualizations: Table, Matrix, Graphical.
↳ we've to pick the correct visualization type for our data set.
↳ Column & Bar graphs.
↳ Stacked graphs: Important for Data Analysis.
↳ 100% Stacked Column graphs: shows the result in percentage.

→ Data Analytical eXpression (DAX): Formula.
(eg) Cost = [unitPrice] * [orderQty]. // New column.
↳ DAX Functions are generally used in formulas for more powerful aspects in power BI. (eg) YEAR, MONTH, DAY, WEEKDAY.

(eg) Weekday Name = FORMAT ([SalesDate], "DDDD")
- FORMAT: Function
- [SalesDate]: Column name
- "DDDD": returns the names of the weekdays.

(eg) Month Name = FORMAT ([SalesDate], "mmmm")
- "mmmm": returns the names of the months.

↳ DAX Measures: Sum, Average, min, max, DistinctCount, Divide, Ratio calculations. → Functions.
→ These measures does not create any new columns.
→ we can't change the type of aggregation.
(eg) can't change from sum to average.
→ These are simply just calculations.
→ Can be linked to any table.
→ click on "New Measure" from Measure tools.

→ Workspaces: 2 types of workspaces in Power BI Service:

My Workspace	Workspaces.
↳ Personal, free or paid, only you can access.	↳ shared, multiple users, content can be created & shared across our organisation. ↳ Contains Dataflows.

Exam topics:

→ Connect to Data: Power BI can connect to virtually any type of data source including:
- Flat files & folders
- Databases
- Azure
- Power Platform
- online services.

"Get Data" button is used.

DATA PREP

→ The Query Editor: Lets us shape & transform the data to meet our needs, then load that model into Power BI Desktop.

→ **Data Source:** To connect to a JSON file or transform a field within a flat file that contains JSON:
  ↳ Convert the JSON list to a table.
  ↳ Expand the attributes in the list to columns.
  ↳ Change the data type for each column.
  } ∴ Transformation Steps.

→ **Data Source: Power BI Data Sources Files:** PBIDS: contains a single set of pre-wired data source connection settings (no data)
  ↳ To create PBIDS file: options → Data source settings → select data source → Export PBIDS → select location → close.
  ↳ Great way to speed up the connecting process.

→ **Data Source: Microsoft Dataverse:** It is a cloud-based storage option for our organization's data that we can connect to business applications.
  ↳ For connection we need: To know the server address & need to have permission to access data, read & write on it.

→ **Data Source: SQL Server Analysis Services Tabular SSAS:** There are 2 methods to connect SSAS tabular models: Import & live connection.
  ↳ Live connection: To establish a connection b/w Power BI tenant and server side SSAS model, we need to have a data gateway installed.

→ **Data Source: Sharepoint online:** Lets companies organize, share and access information via sites. Sites contain document libraries, a special type of folder, that stores files & folders.
  ↳ Connection: We can get data from a sharepoint online site by connecting to a "SharePoint Folder" → Document library.
  Steps: ① Enter the site URL.
  ② Combine & Transform data.
  ③ Filter folder path to correct document library.

→ **Storage Modes:** For data sources:
  ① Import: In-memory within Power BI.
  ② DirectQuery: Tables connected directly to the source & queries executed on-demand at the source.
  ③ Dual: Import + DirectQuery.
  ↳ use DirectQuery when:
  → Dataset is too large (>1GB).
  → Source data changes frequently.
  → Reports must show the recent data.
  → Company policy states data can only be accessed from the original source.

→ Imported Data Vs DirectQuery: Refer phone for photo.

→ **Parameters:** The Data Source Settings allows us to manage data connections & permissions. Parameters are a useful way to change data source values dynamically in Power Query.
Home Tab → Manage Parameters → New Parameter

Quiz-1 :

① Lets us connect our data to other business applications : Microsoft Dataverse.

② Source that lets users connect to a set of pre-wired connections :
PBIDS Files - Speeds up the "Get Data" process.

③ SharePoint Online contains sites, document libraries, & folders.

④ Data Gateways are required for on-premises data sources like SSAS Tabular live connection.

⑤ DirectQuery tables are connected directly to the source.

⑥ Connections to local files reference the exact path, so if the file name or location changes, we will need to change the source via Data source settings.

⑦ Parameters : useful way to change data source dynamically.

→ Data prep : Profiling :

↳ View menu : This tab includes data preview & other advanced editing tools.
→ M code : The language used by Query editors in Power BI.

↳ Column quality : Data profiling tools like column quality, column distribution, column profile provide a visual way to explore data and get a sense of our dataset composition.
(eg) Column quality : % of values within a column that are valid, have errors or are empty.

↳ Column distribution : Provides a sample distribution of the data within the column in the form of graphical representation. (Bar charts).
→ We can use column distribution tool to identify the "Primary key".

↳ Column Profile : Provides a more holistic views of data within a column providing sample distribution of the data & column statistics.
Column profile = Quality + Distribution.

→ ⁂ Note : All of these column profiling tools are based on the first 1000 rows of the data set. To make sure that this distribution represents the entire data set, we need to enable profile based on the entire data set.

Quiz-2 :

① Data profiling tools to check the no. of errors in a column :
Column Quality & Column profile.

② Column Distribution : To identify primary keys.

→ Data prep : cleaning, Transforming & Loading :

↳ Index column : These columns are used when we want to create a set of unique IDs for each row of the table.
Purpose : To create relationship with other tables.

↳ Conditional columns : Allows us to define new fields based on logical rules. (i.e.) IF, else statements.
Add column → Conditional column.

↳ Grouping & Aggregating data : Group by allows us to aggregate our data at a different level.

↳ Pivoting: Turning distinct rows into columns. Unpivoting: Turning columns into rows.

(eg)

2015	A	Pivot →	2015	2016	2017
2016	B		A	B	C
2017	C	← unpivot			

↳ Merging queries: Allows us to join tables based on a common column. Merging adds columns to an existing table.

↳ Appending queries: Allows us to combine tables that share the exact same column structure & data types. Appending adds rows to an existing table.

↳ Modifying queries: Home tab → Transform data.

↳ Data Mashup or M code: The formula language that drives Power Query. Writing or editing M code in the formula bar provides more options & functionalities that using the Power Query UI tools.

Applied Steps → Formula bar → M Code.

→ Advanced Editor: Allows us to see the M code in detail. Which consists of two blocks: ① let: the definition of all variables.
② in: the o/p of our query.

(eg)
```
let // Defines all variables.
    Source = CSV..... → Functions (Actions)
    #"Changed Type" = Table.TransformColumnTypes
    #                  ↘ Variable names (steps)
    #
in // o/p.
```
Home → Advanced Editor.

→ M-function categories: Table, List, Text, Date functions.

Refer phone for photo.

(eg) `Table.SelectRows(#"Reordered Columns", each([Quantity_Sold] = 2))`

Function Name — Previous Step — Function Arguments.

→ "M code is case sensitive"!

Quiz-8:

① The Transform tab includes tools to modify existing columns by overwriting its values.

② What can be useful to create unique IDs & form relationships b/w tables?
Adding an index column.

③ Columns from examples allows us to add new columns by providing an example value.

④ Group By allows us to aggregate our data at a different level.

⑤ Pivot - Row values into column values unpivot - Columns to rows.

⑥ Appending Queries: Adds rows to an existing table.

⑦ Two blocks that make up the M code: let & in.

→ Data Modeling 101:

↳ Data: (or) Facts table contain measurable metrics about the business (quantity, revenue, views, etc). Foreign Keys

↳ Lookup tables contains descriptive attributes about each dimension in our model (customers, products, etc). Primary Keys.

↳ Creating table relationships:

By: option-1: click & drag to connect primary & foreign keys within the Relationships pane.

option-2: Add or detect relationships using Manage Relationships option

↳ Relationship Cardinality: Refers to the uniqueness of values in a column.

↳ Filter Flow: Only downstream tables

→ we should always arrange our lookup tables above our data tables in our model as a visual reminder that filters flow: "downstream".

↳ Automatic Date Tables:

Power BI automatically creates a 'hidden' date table for any table that contains a Date or DateTime column.

↳ Requirements for creating / importing date table:

Ref. phone for photo.

↳ Normalization: Process of organizing the tables & columns in a relational DB to reduce redundancy & preserve data integrity.

Quiz-4:

① In a data model, tables are connected via relationships, based on their common fields.

② A well designed model is critical & ideally should use a star schema with one-to-many (*) relationships, NOT many-to-many.

③ Data tables - measurable metrics Lookup tables - descriptive attributes.

④ Cardinality: Refers to the uniqueness of values in a column.

↳ One instance of unique value: PRIMARY KEY.

↳ multiple instances: FOREIGN KEY.

⑤ USERELATIONSHIP: A DAX function that is used to activate inactive relationships.

→ Data Modeling: DAX CALCULATIONS:

↳ DAX is the formula language that drives front-end calculated tables, columns & measures in Power BI. Data Analysis Expressions.

↳ Two ways to use DAX:

* Calculated columns * Measures.

↳ Calculated columns: Allows us to add new, formula-based columns to tables. This is not useful for aggregate functions.

DAX SYNTAX = Month Name =

Column Name ↙ FORMAT (↗ Function

Function ↙ DATE (1, 'Maven Cycles Calendar' [Month Number], 1),

"mmm" ↘ Arguments.

↳ Measures: DAX formulas used to generate new calculated values. This is used for data aggregations.

↳ Quick Measures: Pre-built formula templates that allow us to drag & drop fields rather than writing DAX from scratch.

Calculation → Quick Measure.

↳ Common DAX Function Categories:

* Math & stats * Logical functions * Text * Filter * Date & Time

→ Refer phone for photo.

← Parameters →

(eg) Total Quantity: = SUM (Transactions [quantity])

Measure Name | Function Name | Referenced Table name | Referenced Column Name.

→ Iterator functions:

Also called X functions: Allows us to loop through the same calculation on each row of a table, & then apply some sort of aggregation to the results (SUM, MAX, etc).

Syntax: =SUMX (Table, Expression).

Aggregation to apply | Table Name | To be evaluated for each row.

→ Divide function:

= DIVIDE (Numerator, Denominator, [AlternateResult]).

- X. Function

→ Calculate function:

= CALCULATE (EXPRESSION, [Filter 1], [Filter 2], ...)

(eg) Australian Orders = CALCULATE ([Total Orders], 'Maven Regions' [Country] = "Australia")

Measure | Function | Calculated measure Expression. | Filter.

↳ Calculate function always overrides any filter values present.

↳ Calculate modifiers: Added as filter arguments within a CALCULATE.

↳ USERELATIONSHIP: one of the calculate modifiers.

= USERELATIONSHIP (ColumnName1, ColumnName2).

Foreign key | Primary key of the relationship.

→ used to create measures based on INACTIVE relationships.

↳ ALL(): Returns all rows in a table, or all values in a common, column ignoring any / all filters applied.

= ALL (Table or ColumnName, [ColumnName], ...).

↳ FILTER():

= FILTER (Table, FilterExpression).

Table to be filtered | A Boolean filter expression.

↳ TOPN():

= TOPN (N_value, TableName, [OrderBy Expression], [Order]).

Function | The no. of rows to be returned | Name of the table.

→ Time Intelligence functions: Allows us to easily calculate common time comparisons.

* DATESYTD: Dates Year to Date.
* DATESQTD: Dates Quarter to Date.

Quiz-5:

① DAX - Data Analysis Expression

② Creating calculated columns as close to the source as possible helps reduce data model size & improve performance.

③ Measures are not evaluated based on row content.

④ Measures are numerical, calculated values that can be analyzed in the value fields of a report visual.

⑤ CALCULATE modifies & overrules any competing filter context.

⑥ ALL function removes filter context.

⑦ QUICK MEASURES allows us to drag & drop fields rather than write DAX from scratch.

⑧ DATESINPERIOD: Allows us to calculate running totals.
DATESYTD: Allows us to calculate performance to date.
DATEADD: Allows us to calculate for a previous period.

→ Data Visualization: Creating reports & Dashboards:

↳ Small multiples: (OR) Trellising, splits a visual into multiple versions of itself that are presented side by side, with the data divided across a dimension.
→ It can only be created on bar, column, line & Area charts.

↳ Tooltips: Way to add additional data when you hover over a visual.

↳ Importing custom visuals: From files, organization or the AppSource marketplace directly into Power BI.

↳ We can create visuals from queries & datasets generated in R or Python.
↳ Install Python on local machine.
↳ Install necessary packages / libraries (Matplotlib, NumbPy).

Quiz-6:

① Report interactions allows us to define how filters applied to one visual impact the others. The options are: Filter, Highlight, None.

② When the interaction mode is set to Highlight, the relevant subsegments in the chart are highlighted because of the selection.

③ Drill through filters allows users to jump to different report pages, while simultaneously filtering based on the specific item selected.

④ Bookmarks allows us to create pre-filtered views of our reports.

⑤ Power BI's Accessibility features:

↳ Paginated Reports: These are called Paginated because they're formatted to fit well on a page & are designed to be printed / shared.
→ Premium Capacity needed to publish & share. ⁂ CONCEPT ⁂.

↳ Paginated Reports when exported as PDF, it provides complete detail of all the data in our report. Whereas Power BI reports only gives us images instead of all the data when exported.

→ Data Analysis: Enhancing Reports:

↳ chart types based on Analysis:

→ Scatter charts: Shows the relationship b/w two numerical values.

↳ uses: * shows patterns in large sets of data.
* Show linear & non-linear trends.
* cluster analysis.
* Outlier identification.

→ Line charts: Used to track changes over periods of time.

↳ uses: * Add multiple lines to compare trends b/w series.

→ clustered column charts: compares values across different categories.

↳ uses: * Show distribution of data points.

↳ Filtering options: There are 4 primary filter options in Power BI:

Visual level: Applies only to the specific visual.

Page level: Applies to all the visuals on the specific page.

Report level: Applies to all visuals across all pages of the report.

Drill through: Applies to specific pages, updates based on the item clicked.

↳ slicers: Provide an interactive way for users to sort & filter a report.

↳ AI VISUALS (Key Influencers):

·⁘· Concept ·⁘· The key influencer visual helps us understand the factors that drive a specific metric. This is an AI generated visual. It gives a detailed information about products by predictions & intelligence.

↳ AI VISUALS (Decomposition tree):

·⁘· Concept ·⁘· The decomposition tree visual allows us to perform exploratory analysis by successively breaking down a measure across multiple dimensions. This is a great choice when we want to perform a root cause analysis or Ad hoc exploration.

↳ Clustering:

Allows us to group observations in a dataset with similar characteristics.

→ It is a form of unsupervised machine learning.

→ Power BI's default clustering algorithm is the Expectation Maximization (EM) method.

use cases:

→ Identify relationships that may not be derived through casual observation.

→ Detection of anomalies in the dataset.

→ Identify performance by segments.

→ clustering dimensions based on factors.

↳ Grouping & Binning:

Grouping is the process of creating logical categories for text data, while binning is the process of creating logical ranges for numerical data.

Quiz-6:

① Only line charts have forecast in the chart analytics options.

② Q+A lets us explore our data "in our own words", using Natural language queries.

③ Best practices when using Q&A visuals:
→ Fix incorrect data types.
→ Add missing relationships b/w tables.
→ Add synonyms to tables & columns.
→ Normalize our model.

④ Report level filters apply to all visuals across all pages of the report.

→ Deploying & Maintaining Deliverables:

↳ Scheduled Refresh: Allows us to keep our Power BI reports up to date by automatically refreshing datasets based on a given frequency & time of day.
→ A data gateway is required to refresh on-premises & online data sources.

↳ Static Row-Level Security:
→ Static roles allow us to define filtered views for specific audiences (managers, leads, execs, etc.) using simple DAX statements. Roles filter data out of our model & limit what audiences can access.
Modeling → Security → Manage Roles.

↳ Dynamic Row-Level Security:
→ Dynamic roles allow us to define filtered views for a specific list of users with DAX functions USERNAME() (or) USERPRINCIPALNAME()

Dynamic role will filter by the username in the table

Dynamic role will filter by the Email Address in the table.

USERNAME: Doesn't have any parameters. Returns the domain & user's username in the format domain-name/user-name.

USERPRINCIPALNAME: No parameters. Combination of 3 items:
→ Person's Username
→ "@" Symbol.
→ Company domain.

↳ Azure Active Directory Security Groups:
→ It allows us to manage an entire group of users instead of a list of individual users.
→ It also creates specific security policies for diff. groups of users.

↳ Subscriptions: To recieve periodic email updates with a report, dashboard or app snapshot. This is an email update.

↳ Sharing options:
Reports → Individuals
Dashboards → Teams
Apps → Organizations.
} Max. no. of recipients is 100 at a time & 500 in total.

→ When a developer shares a report or dashboard or publishes an app, users must have either a pro license (or) access to premium capacity to view the shared content.

→ user permission: Levels of permissions:
↳ Viewer ↳ Contributor ↳ member ↳ Admin.
↳ Refer phone for photo.

↳ Publishing apps: We can select reports & dashboards to publish as an app so large groups of people, both internal & external to our organization, can view them.

↳ Deployment pipelines: Allows us to manage the lifecycle of our organization's content (ALM) by developing + testing in Power BI before it's consumed.

Development → Test → Production.

→ This process creates workspaces automatically for all three stages.

↳ Data lineage: Refers to the flow of data from a data source to a report and dashboard.

Datasets → Reports → Dashboards.

↳ Incremental Refreshing:

→ The process of refreshing only the modifying data & appending it to the constant data instead of refreshing the entire dataset. This greatly reduces the processing + refreshing time.

→ (Eg) Macro datasets such as population level.

→ Incremental Refreshing can be done only if the source data supports Query folding.

→ Sources that support Query Folding:

* Relational DB * Odata * Microsoft Exchange * Azure Active Direc.

→ Sources that don't support QF:

* Flat Files, Azure blob storage, web page data.

→ Incremental refresh only works with a Date/Time Column.

↳ Large Dataset Storage Format: Is used for datasets over the 10 GB refresh limit in service.

↳ Endorsement: It is a way to flag content that's ready to use. Any content owner or member with write permissions can endorse content.

→ Certification means that the content meets the organizations quality standards & can be regarded as reliable, authoritative and ready to use.

↳ Sensitivity labels: Provides a simple way to classify and safeguard sensitive content by labeling reports, dashboards, datasets etc.

Quiz-7:

① RLS rules can be applied to users with Viewer permission level.