POWER BI

DAX OVER VIEW AND FAQS

MSBI
IS, AS, RS & MDS

POWER BI
SERVER, DESKTOP & DAX

DAX few important points for understanding

The below is simple understanding. For more detals go through

"DAX_BASICS_1.pdf in other documents folder"

What is DAX?

DAX: Data Analysis Expressions

This is a **formula / functional langage** (not a programming language).

Why do we require:

To work with Power Pivot [2010], Tabular Model[2012], and
Power BI [2013]

What it does?

Help us to create new measures, columns, tables which are suitable for

a) Modeling
b) Analysis

Note: It can transform data with few operations
[Data Conversion, mathematical calculations, formations etc...]

In which year introduced?

2009 End

What is the Inspration:

Excel functions and MDX[Multi Dimensional Expressions] functions are inspiration.

What Excatly DAX Contain?

DAX is a Formula Language / Functional Language, which contain multiple elements like other languages

a) **Identifier :** Name it, what kind of chars / digits / special you can take, how do you recognize a name etc...

b) **Data Type:** Integer, String, Date, Boolean, BLOB etc...

c) **Operators:** Arithemetic [+, -...], Concatination [+,||], Logical [and or not] etc....

d) **Parameter naming convention**

e) **Functions** [9 kinds of functions available]

DateTime

Mathematical

Parent Child etc...

These are of two types

a) Functions without arguments [parameters]

Rand() –which will show you random value between 0 and 1

b) Functions with arguments [parameters]

Sum(columnname) –It will perform total value

Where do we write DAX?

a) PowerPivot: Expression Bar or Any Cell
b) Tabular Model: Expression Bar or Any Cell or On Cube Database [SSMS]
c) Power BI: Expression Bar / DAX Studio / SSMS

Are we going to use any tools?

DAX is a **formula language**, we can also use like **query language.**

There are two types of tools

a) SSMS: Sql Server Management Studio [Native and from MicroSoft]

b) DAX Studio [DAX Org]

DAX availability?

In Two ways

a) Expressions

b) Querying

What is Parameter in DAX?

Parameters are called as arguments in other languages.

It may be an expression / table / ties/ column etc...

What is Context, how it impacts?

Context is the filter used to evaluate a DAX function.

There are multiple contexts.

a)Row
b) Column
c) Multi-row
d) Filter

etc...
Ex:
You created a DAX function with SUM(DiscountFee), but the value changes based on the filter / row / column / multi-row selection.

What are the important terminologies need to remember to work with DAX?

a) Single value return: **New measure**

b) Multiple values return in a single column: **New column**

c) Multiple columns returns : **New Table**

d) Taking Microsoft defined calculated formulas: **New Quick Measure**

Where do we write DAX?

a) Power Pivot in Excel

b) SSDT -Sql Server Data Tools (Visual Studio)

c) SSMS--Sql Server Management Studio

d) Power BI Desktop (Cloud/ On-premises)

e) DAX Studio

Explain object referencing in DAX?

Tablename

'Tablename'

'Table name'

Tablename[Columnname]

'Table name'[columnname]

What kind of statements we write in DAX queries frequently?

a) **Evaluate :** like Select statement

b) **DEFINE:** Like Scope / CTE (with in the query it will be executed)

c) **Var:** Intermediate value storage holder

How do we work with variables in DAX?

By defining and using

```
Define
        Var     totalval=sum(...)
If
(totalval>1000000,'Good Business', 'Poor Business')
```

What is quick measure in Power BI?

Quick measure creates a measure from the selection of a measure in the given list of measures [these are given by Microsoft]

Advanatage:

Reduces time to create

Note:
Microsoft Observed many orgnizations and defied few measures by default, those are quick measures or add Business Intelligence measures

Real-time working perspective important functions

Date Functions
Time Intelligence
Parent Child Functions
Text Functions
Logical Functions
Calculate, Format, Filter, SUM/AVG/MAX/MIN/COUNT [X or without X], CROSSFILTER, USERELATIONHIP, RANK, TOP, ROW, VALUES, EVALUATE, VAR, DEFINE, SUMMARIZE, GROUP BY, SUMMARIZE Columns,ALL, ALLSELECTED, SELECTEDVALUE etc.

DAX FAQS

How do we establish relationships and Directions using DAX without establishing in the Model?

USERELATIONSHIP and CROSSFILTER will help you to establish at DAX level

How do we establish relationship between two tables at runtime using DAX without establishing in the Model?

USERELATIONSHIP

How do we specify Directions (single / both) between two tables at run time using DAX without establishing in the Model?

CROSSFILTER

Difference between Summarize and Group By?

GROUBY

The GROUPBY function is similar to the SUMMARIZE function.
GROUPBY does not do an implicit
CALCULATE for any extension columns that it adds.
GROUPBY attempts to reuse the data that has been grouped making it highly performant.
Group isn't placed into the filter context.
Start with the specified table (and all related tables in the "to-one" direction)
CURRENTGROUP supports AverageX, CountAX, CountX, GeoMeanX, MaxX,MinX, ProductX, StDevX.S, StDevX.P, SumX, VarX.S, VarX.P.
No such option
No such option

Differences between DATES (MTD/YTD/QTD) AND TOTAL (MTD/YTD/QTD)?

DATES (MTD/YTD/QTD)--return DATES for the function.
For totals, again we need to take aggregate function.

TOTAL (MTD/YTD/QTD)--Total Value (SUM) returned for the function

Differences between First Date(or Lastdate) and First Non Blank Date (or Last NonBlank Date)?

First Date--First date of business
Last Date--Last date of business

First Non Blank Date-- First business value started date
Last Non Blank Date-- Last business value date

Differences between IFERROR and isERROR?

IFERROR does not require conditional validation, if it is error, then handle automatically with second argument action. Whereas iserror returns TRUE / False, so separate conditional validation (like if clause) is required.

Differences between PARALLEL PERIOD and SAME PERIOD LAST YEAR?

Parallel period is super set and give you parallel date with any interval
Sameperiodlastyear will give you only one interval past data.

What are the options PARALLEL PERIOD support?

YEAR, QUARTER, MONTH
Note: Week and Day not possible

What is the purpose of PARALLEL PERIOD?

Numorous number of intervals you want to go back / move forward, then it is required.

Can we use PARALLEL PERIOD is the replacement for other functions?

Yes, we can use it for the replacement of PREVIOUS AND NEXT FUNCTIONS

Differences between ALL and ALLEXCEPT?

All will ignore filters and consider all values.

AllExcept will consider all values except the specified column values.

Difference between ALL and ALL SELECTED?

All will ignore filters and consider all values.

AllSelected will keep all explicit filters and ignore filter section values.
This is helpful to findout visual aggregations.

Differences between Related and Related Table?

Related will get required value based on model relationship.

Related Table will get set of rows based on the table data filtered or sliced in the current context.

Note: Both functions require a proper data model relationship

Differences between Calculate and Calculate Table?

Calculate perform calculations, whereas calculate table return set of rows based on condition

Differences between IN, CONTAINS and CONTAINSROW?

IN--Multiple values against single column

CONTAINS--Single value against single column [can be multiple columns and multiple values]

CONTAINSROW--Multiple values against multiple columns **in a row**

Differences between Related and LookupValue?

Related will use model join and gets result column value. Whereas Lookup will give you result column value after manual condition match (manual join).

Related works at row level for each cell value, where as Lookup returns a single value.

When do we return TRUE or FALSE values apart from Informational Functions?

When manually we write condition and eveluate, if condition satisfied return TRUE, otherwise FALSE.

Differeneces between SUM, SUMX, AVG, AVGX, MIN, MINX, MAX, MAXX, PRODUCT,PRODUCTX etc..?

SUM--column wise operation, multiple rows passed to it. Only one argument it will take.

SUMx--Row wise operation, conditional expression used here. Two arguments it will take.

Differeneces between SUMx and Calculate?

SUM X always take same table [single] fields in the first and second argument.

Calculate can take multiple table arguments. It uses internal model join condition to perform calculation.

Differences between Generate and GenerateSeries?

Generate will do cross join of two tables, whereas GenerateSeries will generate sequence of values.
(similar to identity topic in SQL server)

Differences between Generate and Crossjoin?

Crossjoin perform cartesian product of two resultsets, whereas Generate passes each value from first resultset to second resultset and then perform operation to provide cartesian result.

Crossjoin simply map and give you, generate will **evaluate** (secondset), map and give you.

Differences between COUNT, COUNTA, COUNTAX, COUNTX?

The COUNT function counts the **number of cells in a column that contain numbers.**

The COUNTA function counts the number of cells in a column that are not empty. It counts not just rows that contain numeric values, but also rows that contain **nonblank values, including text, dates, and logical values.**

The COUNTAX function counts nonblank results when evaluating the result of an **expression** over a table.

COUNTX counts the number of rows that contain a number or an expression that evaluates to a number, when evaluating an expression over a table.

Differences between MAX and MAXA?

MAX--Numeric values MAXA: Numerics + Logical values+ Blanks

Differences between MIN and MINA?

MIN--Numric values MINA: Numerics + Logical values

Difference between Values, Row, and DataTable?

Values:Single column values retrieval

Row: Single row with multiple columns values

Datatable: Help us to create table with data

Difference between RANDOM and SAMPLE?

Random: Gives random value between the range specified
SAMPLE: Returns the specified sample number of rows

Differnece between Concatinate and CombineValues?

Combinevalues: Combine based on delimiter

Concatinate: Concatinate the specified values (no delimiter option)

Difference between Count and CountRows?

Count works on column wise and count rows works on row wise.

I want to findout number of string values or date values in a column?

COUNTA

How do we implement SubSrting functionality in DAX?

Using MID (middle chracters, it will bring from the specified point to the length of characters)

Differences between FIND and Search?

Search:
Returns the number of the character at which a specific character or text string is first found, reading left to right. **Search is case-insensitive and accent sensitive.**
=search("te","VINAYTECH",2) --RESULT IS 6

Find:
Returns the starting position of one text string within another text string. FIND is case-sensitive.
=search("te","VINAYTECH",2) --FAILS

Differences between REPLACE and SUBSTITUTE?

Substitute:
Replaces existing text with new text in a text string. **Multiple times** it can subsititute.
=SUBSTITUTE([Product Code], "NW", "PA")

REPLACE replaces part of a text string, based on the number of characters you specify, with a different text string. **One time replacement** based on the positions mentioned.
=REPLACE('DIMCOURSE'[COURSE DESC], 3,9,"VINAYTECH")

Differences between Values and Value?

Value--consider **string as value**
=Value("3")

Values--Return **set of column values**
=values(dimcourse[courseid])

How do we establish bidirectional cross filtering in DAX dynamically?

Using CROSSFILTER and specifying both option

Difference between earlier and Earliest?

Earliest has one additional level of recursion

Difference between RANKX and RAN.EQ?

RankX for column values based ranking, Rank.EQ for comparing one column values with others and taking a rank value.

How do we get earliest value and comparing with other values like correlated sub query?

Earlier function

Can we create a table in DAX with columns and data types?

DATA TABLE function

Difference between CROSS JOIN and GENERATE?

Cross join works and Generate close to each other, but Generate support dat context, where as cross join does not support context.

Difference between Today and Now?

Today will show you current date and time as 12 AM, where as Now will show you current date and current time.

Difference between Calendar and Calendar Auto?

Calendar shows the calendar for the given dates, whereas calendarauto generate calendar based on model dates

How do we format Date / Numeric values?

Using FORMAT function

How do we format currency values?

CURRENCY

How do we identify user and manager hierarchy using DAX?

Using Path function

I want to identify a user existence in a hierarchy?

PATHCONTAINS

How do we perform division of two values in DAX?

DIVIDE

Differences between IF and SWITCH in DAX?

IF is for single condition evaluation, whereas SWITCH for multiple conditions evaluation

When do we use VAR / Variable?

Intermediate data holding to passing to the next expression to implement operations.

What is Evaluate?

DAX Query language statement to bring data

What is ADDCOLUMNS?

Adding columns to the first dataset argument

Differences between TOP N and SAMPLE N?

Top N perform operation (desc / asc) and then give you top values, whereas sample does not perform any operation and produce random sample n rows

How do we verify a particular column filtered or not?

ISFILTERED

How do we show user selected value?

SELECTEDVALUE

How do we get multiple values of user selected?

ALLSELECTED / VALUES

What exactly PREVIOUS YEAR / QUARTER /MONTH / DAY will do?

Complete last YEAR / QUARTER / MONTH / DAY data and does not apply any fiters.

What exactly NEXT YEAR / QUARTER /MONTH / DAY will do?

Complete NEXT YEAR / QUARTER / MONTH / DAY data and does not apply any fiters.

What exactly SAMPERIODLASTYEAR will do?

Exactly last year / quarter / month / combination data and apply filters.

What exactly STARTOF YEAR / QUARTER /MONTH will do?

In the given context Starting of Year / Quarter / Month, no filters applied

What exactly OPENINGBALANCE YEAR / QUARTER /MONTH / DAY will do?

Gets context start date value, if it is not available last closing value

What exactly CLOSINGBALANCE YEAR / QUARTER /MONTH / DAY will do?

Gets context closing date value

Difference between measure and quick measure?
Measure created by the user and quick measure generated from selection
Which function is useful to identify the number of student names in DimStudent?
CountA
Which function is useful to identify the maximum value in the given list(34, 'Ramana' Null, 400, 500)?
MAXA
Which function is usful to identify Online mode students Discount Fee?
SUMX
What exactly MAXA and MinA will do when we have FALSE and TRUE values in the column?
False considered as 0 and True considered as 1. So which ever is highest, MAXA will display. Whichever is lowerest MINA will display. Ex: a)If the column contains all values as False: MaxA and MinA: 0 b) If the column contains all values as True: MAXA and MinA :1 c) If the column contains mix of values (False and True), then MAXA: 1   MinA: 0

DAX OPTIMIZATION STEPS

Improve Power BI Performance by Optimizing DAX

DAX Ganeric

Power BI performance issues are often a result of sub-optimal Data Analysis Expressions (DAX) language. DAX is a collection of functions, operators, and constants that you can use in a formula, or expression, to calculate and return one or more values. Stated more simply, DAX helps you create new information from data already in your model.

DAX Performance Issues

Optimized DAX provides deep insights, shedding new light on complex data. Inefficient DAX slows processing,

a) Clogs premium capacity

b) Increases wait time

c) Hinders refreshes

d) Report load time.

Few DAX Performance Tips

We have experience delivering rich insights with DAX in hundreds of enterprise-grade Power BI implementations. Use these 20 DAX best practices to ensure your dashboard users get the most out of their data.

1. Clear the DAX cache before optimizing DAX

- Caches result from internal VertiPaq queries.
- Clear your cache from within DAX Studio.
- Resetting the cache lets you measure effective performance gain.

2. Format your code

- Use the DAX Formatter.
- Formatted code is easier to read and maintain.

3. Do not change BLANK values to zeros or other strings

- It is common practice to replace blanks with zeros or other strings.
- However, Power BI *automatically* filters all rows with blank values. When viewing results from tables with large amounts of data, this limits the result set and prevents poor performance.
- If the blanks are replaced, Power BI does not filter the unwanted rows, negatively affecting performance.

4. Use the DISTINCT() and VALUES() functions consistently

- Power BI adds a blank value to a column if it finds a referential integrity violation.
- For a direct query, Power BI adds a blank value to the columns because it cannot check for violations.
- The DISTINCT() and VALUES() functions differ:
  - DISTINCT(): Does not return a blank that is added due to an integrity violation. The DISTINCT() function includes a blank only if it is part of the original data.
  - VALUES(): Includes any blank that Power BI adds due to referential integrity violations.
- Maintain consistent usage of the DISTINCT() and VALUES() functions throughout the entire report.
- Power BI recommends using the VALUES() function if blank values are not an issue.

5. Use ISBLANK() instead of =Blank() check

- Use the built-in function ISBLANK() to check for any blank values instead of using the comparison operator = Blank().

6. Use = 0 instead of checking for ISBLANK() || = 0

- The BLANK value in Power BI is associated with the base value of a column’s data type
- The BLANK value corresponds to zero for integers, “(empty string)” for string columns, and “1–1–1900” for date fields.
- ISBLANK() || = 0 enacts two checks: ISBLANK() and comparing with zero.
- Use = 0, which internally performs both checks.
- To perform only the check for zero, use the IN operator.

7. Use SELECTEDVALUE() instead of HASONEVALUE()

- It is common practice to use HASONEVALUE() to check if there is only one value present in a column after applying slicers and filters. You also have to use the VALUES(ColumnName) DAX function to retrieve that single value.
- SELECTEDVALUE() performs the above steps internally. It automatically retrieves the single value if there is one and returns a blank if there are multiple values available.

8. Use SELECTEDVALUE() instead of VALUES()

- The VALUES() function returns an error if it encounters multiple values. Normally, users address the error using Error functions, which negatively affects performance.
- Instead of using VALUES(), use SELECTEDVALUE(). The SELECTEDVALUE() function is safer and returns a blank if it encounters multiple values.

9. Use variables instead of repeating measures inside the IF branch

- Incorrect DAX:
  Ratio = IF([Total Rows] > 10, SUM(Revenue) /[Total Rows], 0)
  - Because measures are calculated continuously, the [Total Rows] expression is calculated twice: first for the condition check, then for the true condition expression.
- Correct DAX:
  VAR totalRows = [Total Rows]; Ratio = IF(totalRows > 10, SUM(Revenue) / totalRows,0)
- Instead of calculating the same expression multiple times, you can store the resulting measure value in a variable.
- You can use a variable reference wherever required. The same variable process applies to all instances where you call the same measure.
- Variables can help you avoid repetitive functions.
- Note: Be aware that variables are actually constants.

10. Use (a-b)/b along with variables instead of a/b — 1 or a/b*100–100

- It is common practice to use a/b — 1 to calculate a ratio and avoid duplicate measure calculations.
- However, you can achieve the same performance by using variables and using (a-b)/b to calculate the ratio.
- Here’s why you should use (a-b)/b: If both a and b are blank values, then (a-b)/b returns a blank value and Power BI will filter the values out. a/b — 1 would return -1 as the result because both a and b are integers.

11. Stop using IFERROR() and ISERROR()

- The IFERROR() and ISERROR() functions were widely used in Excel when applying the FIND() and SEARCH() functions. They were necessary because FIND() and SEARCH() returned errors if the query did not obtain the required result.
- The IFERROR() and ISERROR() functions force the Power BI engine to perform a step-by-step execution of each row to check for errors. There is currently no method to directly state which row returned the error.
- The FIND() and SEARCH() DAX functions provide an extra parameter that the query can pass. The parameter is returned if the search string is not present.
- The FIND() and SEARCH() DAX functions check if more than one value is returned. They also ensure nothing is divided by zero. You can avoid using the FIND() and SEARCH() DAX functions by using the correct DAX functions such as DIVIDE() and SELECTEDVALUE(). The DIVIDE() and SELECTEDVALUE() functions perform the error check internally and return the expected results.
- You can always use DAX expressions in such a way that they never return an error.

12. Use DIVIDE() instead of /

- / raises an exception if the denominator is zero.
- The DIVIDE() function internally performs a check to validate whether the denominator is zero. If it is, it returns the value specified in a third (extra) parameter.
- For “invalid denominator” cases, use the IF condition while using the “/” operator.
- Note: If you are certain the denominator value is not zero, it is better to use the “/” operator without an IF check. The DIVIDE() function will always perform an IF check internally.

13. Do not use scalar variables in SUMMARIZE()

- The SUMMARIZE() function is traditionally used to group columns and return resulting aggregations.
- However, the SUMMARIZECOLUMNS() function is newer and more optimized. Use it instead.
- Only use SUMMARIZE() for grouped elements of a table without any associated measures or aggregations. For example: `SUMMARIZE(Table, Column1, Column2)`

14. Use KEEPFILTERS() instead of FILTER(T)

- The FILTER function overrides any existing set of filters on a column applied via slicers.
- The KEEPFILTER function does not override the existing set of filters. Instead, it uses the intersection of values present in both, thus maintaining the current context. Use it when you want to maintain any filters applied by slicers or at a report level while performing calculations.

15. Use FILTER(all(ColumnName)) instead of FILTER(values()) or FILTER(T)

- To calculate measures independent of any filters applied to a column, combine the All(ColumnName) function with the FILTER function instead of using Table or VALUE(). For example: CALCULATE([Total Sales], FILTER(ALL(Products[Color]), Color = 'Red'))
- Use ALL along with the FILTER function if there is no need to keep the current context.
- Directly applying filters using expressions instead of the FILTER function behaves in the same way as mentioned above. This method internally translates using the ALL function in the filter. For example: CALCULATE([Total Sales], FILTER(ALL(Products[Color]), Color = 'Red'))
- It is always better to apply filters to the desired column than to the whole table for the sake of scalability.

16. Avoid using the AddColumns() function inside measure expressions

- Measures are calculated iteratively by default.
- If measure definitions use iterative functions such as AddColumns(), Power BI creates nested iterations, which negatively affects report performance.

17. Use the correct data types based on column values

- If there are only two distinct values in a column, check if it can be converted to a Boolean data type (true/false).
- This speeds up processing when you have large number of rows.

18. Use COUNTROWS instead of COUNT:

- Use the COUNT function to count column values, or we can use the COUNTROWS function to count table rows. Both functions will achieve the same result, providing that the counted column contains no

BLANKs. Sales Orders = COUNT(Sales [OrderDate]) Sales Orders = COUNTROWS(Sales)

- There are three reasons why the second measure definition is better:
  - It's more efficient, and will perform better.
    - It doesn't consider BLANKs contained in any column of the table.
    - The formula intention is clearer and self-describing.

19. Use SEARCH() with the last parameter

- The SEARCH() DAX function accepts the last parameter as the value that the query must return if the search string is not found.
- You should always use the SEARCH() function instead of using Error functions along with SEARCH().

20. ALL vs. ALLExcept

- ALLEXCEPT() behaves exactly like ALL(), VALUES() as long as the "exempted" columns are columns on the pivot.
- ALLEXCEPT() does NOT preserve pivot context on columns that are not on the pivot.
- Use ALL() instead of ALLEXCEPT() when using VALUES()