Frequently Asked Questions

Kusto Query Language

Table of Contents

Order of where clause matters? No optimizer?

I am sure there are optimizer to choose most efficient data access method but most likely not as smart as SQL Server where it can re-arrange the order of filter predicates. From Microsoft documentation:

```
Traces
| where Timestamp > ago(1h)
    and Source == "MyCluster"
    and ActivityId == SubActivityId
```

> *Records that are no older than 1 hour, and come from the Source called "MyCluster", and have two columns of the same value.*
>
> *Notice that we put the comparison between two columns last, as it can't utilize the index and forces a scan.*

Can we have indexes on these fields?

We cannot create any indexes manually. Once a data store is created, it is sharded, indexed and immutable for many reasons.

> *Azure Data Explorer has a unique inverted index design. In the default case, all string and dynamic (JSON-like) columns are indexed. If the cardinality of the column is high, meaning that the number of unique values of the column approaches the number of records, then the engine defaults to creating an inverted term index with two "twists". The index is kept at the shard level so multiple data shards can be ingested in parallel by multiple Compute nodes, and is low granularity so instead of holding per-record hit/miss information for each term, we only keep this information per block of about 1,000 records. A low granularity index is still efficient in skipping rarely occurring terms, such as correlation IDs, and is small enough so it's more efficient to generate and load. Of course, if the index indicates a hit, the block of records must still be scanned to determine which of the individual records matches the predicate, but in most cases this combination results in faster (potentially much faster) performance.*

- Having low granularity, and therefore small, indexes also makes it possible to continuously optimize how data shards are stored in the background. Data shards that are small are merged together as a background activity, improving compression and indexing. For example, because the data they contain comes in continuously and we want to keep query latency small. Beyond a certain size, the storage artifacts holding the data itself stop getting merged, and the engine just merges the indexes, which are usually small enough so that merging them results in improved query performance.-- Azure Data Explorer Technology 101

Can you export queries into Power BI?

Yes, you can export queries into Power BI. Microsoft documentation does not explicitly explain this for Azure Monitoring (Log Analytics Workspace) but you can follow instructions form here and it worked in my testing.

Can you do a search with "like" on the left or right?

Yes, you can do that using hasprefix and hassuffix operator. See 'search operator' documentation for details.

Is == case insensitive by default?

'==' is case sensitive and is an exact match. Good reference 'String operators'

Do you have to use pipes for each filter, or can you combine into one pipe with all filters?

You can combine filters that are using same operator. Otherwise you have separate pipes.

These two queries (taken from Robert Cain's Pluralsight course) will produce same result set.

```
// Combining and's and or's
Perf
| where TimeGenerated  >= ago(1h)
    and (CounterName == "Bytes Received/sec"
         or
         CounterName  == "% Processor Time"
        )
    and CounterValue > 0

// Stackable where operators
Perf
| where TimeGenerated  >= ago(1h)
| where (CounterName == "Bytes Received/sec"
         or
         CounterName  == "% Processor Time"
        )
| where CounterValue > 0
```

If you want to add another operator other than search you must use a pipeline instead of using 'and'.

```
Perf
| where TimeGenerated  >= ago(1h)
    and (CounterName == "Bytes Received/sec"
         or
         CounterName  == "% Processor Time"
        )
    and CounterValue > 0
| take 5
```

Does piping impact the performance of the query? For example, use all pipes is faster or slower than using only one pipe?

Yes, it will impact performance. It is creating a filtered data set each time passing via a new pipeline.

Is there a way to take top x instead of random (with take)?

Yes, by using top operator. Returns the first N records sorted by the specified columns.

Does the position of count matter in relation to the where clause?

You can only have one count at the end which returns number of records in the input record set.

Can I rename columns?

Yes, there are few options to do that.
Project: Select the columns to include, rename or drop, and insert new computed columns.
project-rename: Renames columns in the result output.
Extend operator can be used to rename a column but not recommended. Because index might not be used in some complex scenarios. *If the goal is to rename a column, use the project-rename operator instead.*

Is the backtick character a break key, like in PowerShell?

backtick (`) will give a syntax error. Line break will not cause issues like it does in PowerShell

Is the innerunique relevant only to left side? Or right as well?

Yes, only on left side. Only one row from the left side is matched for each value of the on key. The output contains a row for each match of this row with rows from the right. Details about innerunique and other types of joins join operator.

Can you do cross workspace queries?

Yes, you can do cross workspace queries. Details Perform cross-resource log queries in Azure Monitor.

How long can data be kept in Log Analytics?

With Pay-as-you-go model default retention is 31 days. You can increase up to 730 days. Details 'Manage usage and costs with Azure Monitor Logs'.

Can you throttle what flows into workspace?

There are metrics/events that get collected by default and you cannot turn those off. You can add custom collections. Few examples Custom metrics in Azure Monitor

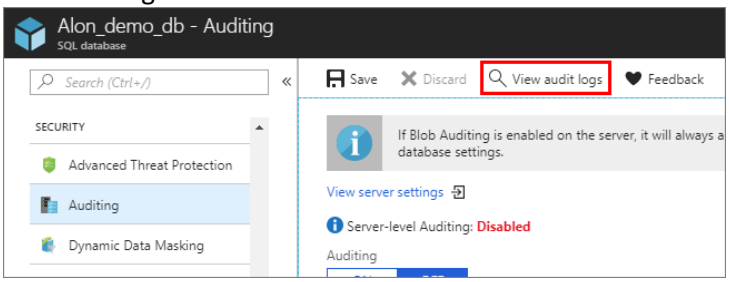
You can throttle retention by data types which is a new feature announced in October 2019.

- Retention by data type
- Azure Monitor Log Analytics retention is now configurable by data type

How do you read the data archived in a storage account?

There are few options when it comes to reading audit logs from an Azure storage account. Details Analyze audit logs and reports .

- Azure Storage Explorer
- Use the Azure portal. Open the relevant database. At the top of the database's Auditing page, click View audit logs.



- Use the system function sys.fn_get_audit_file (T-SQL) to return the audit log data in tabular format. For more information on using this function, see sys.fn_get_audit_file.
- Use SQL Server Management studio
- Use Power BI. You can view and analyze audit log data in Power BI. For more information and to access a downloadable template, see Analyze audit log data in Power BI.
- View blob auditing logs programmatically, Query Extended Events Files by using PowerShell.

There seems to be an inferred time filter when using the UI which changes to "Query Specified Time Window". Does this execute early in the pipeline, before other operations?

I do not know for sure as I cannot see the execution plan. I will update this document if I get a definitive answer to this question.
In my test using the demo portal (https://aka.ms/LADemo) I constantly get better execution time when the TimeGenerated predicate is used as the first compare to last.

```
Event
| where TimeGenerated between (ago(365d) .. startofmonth(now()))
| where EventLog == "Operations Manager"
| where Source == "NPMD Agent"
| where SourceSystem == "OpsManager"
| where EventLevel == 4
| where EventLevelName == "Information"
| where EventID == 101
| where Type == "Event"
```

```
Event
| where EventLog == "Operations Manager"
| where Source == "NPMD Agent"
| where SourceSystem == "OpsManager"
| where EventLevel == 4
| where EventLevelName == "Information"
| where EventID == 101
| where Type == "Event"
| where TimeGenerated between ( ago(365d) .. startofmonth(now()) )
```

Does 'and' short-circuit?

I do not know for sure as I cannot see the execution plan. I will update this document if I get a definitive answer to this question.