



Azure Data Explorer

Timeseries Database Functionality Whitepaper

Doc Author Rolf Tesmer
Cloud Architect - Azure | Data | AI
([LinkedIn](#)) ([Blog](#))

Published August 2021

Doc Version 1.00



Introduction and Key Objectives



2

- This whitepaper outlines 29 key functionality areas expected for typical timeseries data processing and analytics applications, and how these are met via the core capabilities provided within Azure Data Explorer (ADX) – the massive scale PaaS timeseries database hosted in the Azure cloud.
- Information presented in this whitepaper provides a reference view on core ADX “out-of-the-box” functionality (as at writing) against a set of common timeseries analytics and applications requirements. All information provided has been referenced back to official MS Docs pages.
- **NOTE - This paper lists key functionality only, and does not provide performance benchmarks, or any form of industry product comparison.**

Requirements Summary



3

Timeseries Database Functionality – Key Requirements

• Statistical Functions	• Linear Engine Scalability
• Down Sampling (e.g CQ)	• Encryption (in transit) - Transport Layer Security (TLS)
• Time-Stamped Measurement	• Authentication and Authorization
• Data Summarization	• Encryption (at rest)
• Data Lifecycle Management	• Backup / Restore
• Data Back-Filling	• High Availability (HA) and Fault Tolerance (resiliency)
• Support for Advanced Data Queries	• Rolling Upgrades and Restarts
• Ingestion Source; Internal to Azure	• Migration Tools
• Ingestion Source; External to Azure	• Integrate with BI Visualization Tools / Dashboard Platforms
• Data Concurrency and Consistency	• Integrate with External Tools / Frameworks
• Data Compression	• Management Tools – Administration
• Metadata Tagging	• Management Tools – Deployment Automation
• Reference Data	• Monitoring / Audit / Logs
• High Ingest Rate (write)	• Maintenance (SLA)
• High Query Rate (read)	

Timeseries Requirements / Feature Review

Core Functionality



Requirement	ADX Support / Response
1. Statistical Functions	<p>Supported.</p> <ul style="list-style-type: none"> • ADX supports a large set of statistical aggregation functions that support timeseries queries / data analysis summarize operator - Azure Data Explorer Microsoft Docs
2. Down Sampling (e.g CQ)	<p>Supported.</p> <ul style="list-style-type: none"> • ADX supports a down sampling function that can be applied on timeseries data https://docs.microsoft.com/en-us/azure/data-explorer/kusto/functions-library/series-downsample-fl • ADX also supports Materialized Views (MV) which can be used for downsampling/aggregations – and supports back-filling and auto updates of deltas. Materialized views - Azure Data Explorer Microsoft Docs • ADX also supports a large set of time series functions including regressions, forecasting, anomaly detection and many more Time Series Analysis - Azure Data Explorer Microsoft Docs
3. Time-Stamped Measurement	<p>Supported.</p> <ul style="list-style-type: none"> • The Kusto Engine (v3) is ADX's next generation storage and query engine. It's designed to provide unparalleled performance for ingesting and querying telemetry, logs, and time series data. It includes new optimized storage format and indexes, and uses advanced data statistics query optimizations to create an optimal query plan and just-in-time compiled query execution. The engine also has improved disk cache resulting in unparalleled query performance Azure Data Explorer Kusto EngineV3 Microsoft Docs
4. Data Summarization	<ul style="list-style-type: none"> • ADX also has lots of native features including timecharts, summarize-by, make-series operators, bin(), and more to measure change over time. ADX performs well in these areas as ADX/Kusto indexes all columns.
5. Data Lifecycle Management	<p>Supported.</p> <ul style="list-style-type: none"> • ADX supports many data “policies” which can be applied at the database and table level to automatically apply rules and manage data over time Kusto retention policy controls how data is removed - Azure Data Explorer Microsoft Docs and Auto delete policy - Azure Data Explorer Microsoft Docs • The ADX engine supports a hot/cold cache and continuous export, which can be controlled by the customer, and assist ADX to manage costs vs performance Cache policy (hot and cold cache) - Azure Data Explorer Microsoft Docs and Continuous data export - Azure Data Explorer Microsoft Docs

Timeseries Requirements / Feature Review

Core Functionality



Requirement	ADX Support / Response
6. Data Back-Filling	<p>Supported.</p> <ul style="list-style-type: none"> • ADX supports a data “policy” to set the incoming data with the timestamp of the ingestion time (as measured at the ADX cluster). Thus when data arrives it is automatically set with the ingestion datetime. IngestionTime policy - Azure Data Explorer Microsoft Docs • In “backfill” scenarios, data can arrive late and therefore setting the ingestion time at the cluster is not valid, and must be over-ridden. ADX supports many methods to ingest data, which also support to set the ingestion datetime as per fields on the incoming record, and not the ADX cluster. • Example for data ingested from Azure Storage Ingest from storage using Event Grid subscription - Azure Data Explorer Microsoft Docs
7. Support for Advanced Data Queries	<p>Supported.</p> <ul style="list-style-type: none"> • ADX supports the “Kusto” query language, which is a very mature data language used by several Azure services over many years. It's based on relational database management systems, and supports entities such as databases, tables, and columns. Complex analytical queries are made using the Kusto Query Language, including calculated columns, searching and filtering on rows, group by-aggregates, join functions. Getting started with Kusto Microsoft Docs • ADX can also be queried using Python via ADX “Plugins” Python plugin - Azure Data Explorer Microsoft Docs, or via using Azure Notebooks Notebooks with Kqlmagic (Kusto Query Language) in Azure Data Studio - Azure Data Studio Microsoft Docs
8. Ingestion Source; Internal to Azure	<p>Supported.</p> <ul style="list-style-type: none"> • ADX supports many native ingestion points within Azure, including hub services hosted (Azure EventHub, Azure IoT Hub, Event Grid), and Azure Storage. • ADX supports 3rd party ingestion points such as Logstash, Kafka, and Spark. • Data can be ingested via a stream or in batch. Azure Data Explorer data ingestion overview Microsoft Docs
9. Ingestion Source; External to Azure	<p>Supported (with Additional Services)</p> <ul style="list-style-type: none"> • Currently ADX does not support ingesting data from streaming services hosted on 3rd party locations, such as on-prem, AWS or GCP • Simple technical options exist to redirect streams from those locations and onto Azure for ingestion into ADX. • <i>For example - AWS;</i> An AWS Lambda Function can be created which triggers from messages landing in AWS Kinesis. The Function would pickup and route the message from AWS into an Azure EventHub / IoT Hub hosted in Azure, where its then natively ingested directly in ADX.

Timeseries Requirements / Feature Review

Core Functionality



Requirement	ADX Support / Response
10. Data Concurrency and Consistency	<p>Supported.</p> <ul style="list-style-type: none">• ADX supports high end user query concurrency across a deployed cluster, and also manages data consistency via a Strong/Weak consistency model set by the client query. Optimize for high concurrency with Azure Data Explorer Microsoft Docs and Query consistency - Azure Data Explorer Microsoft Docs• ADX also publishes a Best Practices list and QRG for development of queries.• Query best practices - Azure Data Explorer Microsoft Docs and KQL quick reference Microsoft Docs
11. Data Compression	<p>Supported.</p> <ul style="list-style-type: none">• ADX supports multiple ingestion formats and 2 primary compression types (Gzip, Zip) which can be natively ingested in ADX with no transformation required to the source data. Data formats supported by Azure Data Explorer for ingestion. Microsoft Docs• Once landed, the ADX engine applies a built-in columnar compression algorithm against all data in the engine. This optimizes the use of data in all parts of the service.
12. Metadata Tagging	<p>Supported.</p> <ul style="list-style-type: none">• ADX provides ability to apply one or many “tags” to the data ingested. This is performed at ingestion time at the “extent” level, and can be used to efficiently describe properties that are common to all data in an extent.• Extents (data shards) - Azure Data Explorer Microsoft Docs
13. Reference Data	<p>Supported.</p> <ul style="list-style-type: none">• ADX databases store ingested data in Tables. These can be organised in any way desired including acting as a “fact” or “dimension”. Dimensions are essentially reference tables that can be used as filters against Fact tables to limit rows based on a query.• Fact and dimension tables - Azure Data Explorer Microsoft Docs

Timeseries Requirements / Feature Review

Performance



<u>Requirement</u>	<u>ADX Support / Response</u>
1. High Ingest Rate (write)	Supported. <ul style="list-style-type: none">The Kusto Engine (v3) is ADX's next generation storage and query engine. It's designed to provide unparalleled performance for ingesting and querying telemetry, logs, and time series data. It includes new optimized storage format and indexes, and uses advanced data statistics query optimizations to create an optimal query plan and just-in-time compiled query execution. The engine also has improved disk cache resulting in unparalleled query performance, which has a sustained ingestion benchmark at 200 MB/sec (or 703 GB/hour) a single 16-core low-cost standard Azure VM (ie 1x D14 SKU node) . Azure Data Explorer Kusto EngineV3 Microsoft Docs
2. High Query Rate (read)	<ul style="list-style-type: none">The various ADX ingestion methods are outlined here. Azure Data Explorer data ingestion overview Microsoft Docs
3. Linear Engine Scalability	Supported. <ul style="list-style-type: none">The ADX engine architecture has been built with scalability in mind. As such supports methods for both manual and/or auto scale-up or scale-out.<ul style="list-style-type: none">Manual Horizontal Cluster Scaling. Manage cluster horizontal scaling (scale out) to match demand in Azure Data Explorer Microsoft DocsManual Vertical Cluster Scaling. Manage cluster vertical scaling (scale up) to match demand in Azure Data Explorer Microsoft DocsAuto-Scaling Clusters (preview). How to save ADX cost with the new Predictive Autoscale - Microsoft Tech CommunityAlso, when ADX is scaled, the engine has been designed to specifically provide linear scaling.Azure Data Explorer Kusto EngineV3 Microsoft Docs

Timeseries Requirements / Feature Review

Security



<u>Requirement</u>	<u>ADX Support / Response</u>
1. Encryption (in transit) - Transport Layer Security (TLS)	<p>Supported.</p> <ul style="list-style-type: none">• ADX supports many security features inbuilt into the service, including isolation into a virtual network, Identity, Role-Based access controls, disk/storage encryption (including BYO Key). Secure Azure Data Explorer clusters in Azure Microsoft Docs• Microsoft also publish a full set of security recommendations for ADX deployments, which includes TLS configurations.• Azure security baseline for Azure Data Explorer Microsoft Docs
2. Authentication and Authorization	<p>Supported.</p> <ul style="list-style-type: none">• A high level overview of the various methods of Authentication and Authorization support is outlined here.• Kusto Access Control Overview - Azure Data Explorer Microsoft Docs• ADX supports Azure Active Directory (AAD) authentication into the service/database/tables. In order to connect to ADX, all users/applications require an account managed within AAD. Kusto Azure Active Directory (AAD) authentication - Azure Data Explorer Microsoft Docs• Once authenticated, Azure Data Explorer uses a role-based authorization control (RBAC) model, where principals are ascribed to one or more security roles. Authorization succeeds if one of the principal's roles is authorized.• Role-based Authorization in Kusto - Azure Data Explorer Microsoft Docs
3. Encryption (at rest)	<p>Supported.</p> <ul style="list-style-type: none">• ADX supports many security features inbuilt into the service, including isolation into a virtual network, Identity, and Role-Based access controls.• ADX also specifically provides disk/storage encryption (including BYO Key). Secure Azure Data Explorer clusters in Azure Microsoft Docs• Microsoft also publish a full set of security recommendations for ADX deployments, which includes TLS configurations. Azure security baseline for Azure Data Explorer Microsoft Docs

Timeseries Requirements / Feature Review

Availability, Resilience, Recovery



<u>Requirement</u>	<u>ADX Support / Response</u>
1. Backup / Restore	<p>Supported. (Automated and managed by Microsoft)</p> <ul style="list-style-type: none">• As the service is based on PaaS, Microsoft manages the back-end infrastructure, including service availability, data recovery (including accidental deletes initiated by the customer), database backups and recovery.• Internal backup/restore services are not exposed to the customer for manual operations, however depending on the requirement ADX offers several tools and functions to provide capabilities in this space.• SyncKusto Tool can copy/synchronise an ADX database schema from one cluster to another. This is schema only, not data. microsoft/synckusto: Synchronize database schemas between a Kusto cluster and the local file system (github.com)• Continuous Export Feature can be used to export data to Azure storage, which can then be later ingested to a target cluster as/when needed or can be queried through ADX External Tables. Continuous data export - Azure Data Explorer Microsoft Docs
2. High Availability (HA) and Fault Tolerance (resiliency)	<p>Supported. (Automated and managed by Microsoft, and additional options for Customer specific configuration)</p> <ul style="list-style-type: none">• By default the ADX engine provides for a 99.9% SLA for all ADX deployments. SLA for Azure Data Explorer Microsoft Azure• The ADX team have assessed the service against many common failure scenarios, and accounted for how the service can be configured to provide availability in that scenario, or how the service could be recovered should that scenario occur (including human errors, high availability, and multiple disaster recovery configurations). Azure Data Explorer and business continuity disaster recovery Microsoft Docs• In addition the ADX engine can be specifically configured (by the customer) to provide for enhanced availability/recovery, including deployment across regions, to support highly available workloads. Create business continuity and disaster recovery solutions with Azure Data Explorer Microsoft Docs
3. Rolling Upgrades and Restarts	<p>Supported. (Automated and managed by Microsoft)</p> <ul style="list-style-type: none">• By default the ADX engine provides for a 99.9% SLA for all ADX deployments. SLA for Azure Data Explorer Microsoft Azure• The ADX service is a “PaaS” based service hosted in the Azure cloud, and managed by Microsoft. As such patching, minor upgrades etc are all managed by the Microsoft service with no impact to the customer.

Timeseries Requirements / Feature Review

Integration Points



<u>Requirement</u>	<u>ADX Support / Response</u>
1. Migration Tools	<p>Supported. (Automated and managed by Microsoft)</p> <ul style="list-style-type: none">• The ADX service is a “PaaS” based service hosted in the Azure cloud, and managed by Microsoft. As such patching, minor upgrades etc are all managed by the Microsoft service with no impact to the customer.• For ADX internal engine upgrades; occasionally the ADX service may offer a significant version upgrade, such as the recent Gen2 to Gen3 ADX engine change (March 2021). For such upgrades the ADX Support Team offer all customers a Microsoft performed/supported database migration pathway that is initiated by the customer.• For migration to ADX from other engines (ie historical data migration); then ADX provides several tools to perform this. The LightIngest CLI tool provides one of the fastest methods. Azure Data Explorer data ingestion overview Microsoft Docs
2. Integrate with BI Visualization Tools / Dashboard Platforms	<p>Supported.</p> <ul style="list-style-type: none">• ADX supports native dashboarding within the ADX query tools (Data Explorer Dashboards), as well as tight integrations into external BI tools, including Power BI, Grafana, Tableau, Qlik, etc.• Azure Data Explorer data visualization Microsoft Docs
3. Integrate with External Tools / Frameworks	<p>Supported.</p> <ul style="list-style-type: none">• ADX supports native integration with many Azure services and 3rd party products, including data ingestion points, query tools, visualization/BI tools, Notebooks (ie Jupyter, etc), orchestration tooling, data sharing and source control.• Azure Data Explorer tools and integrations overview - Azure Data Explorer Microsoft Docs• ADX also supports a full suite of client development libraries, CLI tools, Powershell, and REST API.• Client libraries overview - Azure Data Explorer Microsoft Docs

Timeseries Requirements / Feature Review

Administration



<u>Requirement</u>	<u>ADX Support / Response</u>
1. Management Tools – Administration	<p>Supported.</p> <ul style="list-style-type: none">• ADX can be deployed and managed via the Azure Portal like any service in Azure.• ADX also offers several tools to perform simple administration on the ADX Cluster, as well as execute and monitor KQL queries, including Kusto Explorer and the Web UI. Azure Data Explorer tools and integrations overview - Azure Data Explorer Microsoft Docs• Quickstart: Query data in Azure Data Explorer Web UI Microsoft Docs
2. Management Tools – Deployment Automation	<p>Supported.</p> <ul style="list-style-type: none">• ADX can also be deployed via Infrastructure as code (IaC) for Devops processes including using ADX CLI, API's, Sync Kusto and Delta Kusto. Many 3rd party IaC solution support ADX, including Terraform.• Azure DevOps task for Azure Data Explorer Microsoft Docs and azurearm_kusto_cluster Resources hashicorp/azurearm Terraform Registry• Kusto CLI - Azure Data Explorer Microsoft Docs and Kusto Delta – Azure Data Explorer Github
3. Monitoring / Audit / Logs	<p>Supported.</p> <ul style="list-style-type: none">• ADX has built in metrics collection and visualization for all cluster, database, and query statistics and operations. ADX metrics/logs are (by default) tightly integrated into Azure Monitor (Log Analytics) Service for automated data collection, aggregation, storage and analysis.• Customers can use ADX native dashboards, or Azure Monitor / Log Analytics tooling to monitor and alert to platform issues.• Monitor Azure Data Explorer performance, health & usage with metrics Microsoft Docs• Azure Data Explorer Insights (ADX Insights preview) - Azure Monitor Microsoft Docs
4. Maintenance (SLA)	<p>Supported. (Automated and managed by Microsoft)</p> <ul style="list-style-type: none">• The ADX service hosted in Azure provides for multiple layers of redundancy built into the core components leveraged to provide the service• By default the ADX engine provides for a 99.9% SLA for all ADX deployments. SLA for Azure Data Explorer Microsoft Azure• The SLA definition includes total available cluster uptime, which takes into account time for PaaS based maintenance activities, and automatic Microsoft owned service restoration should the service become unavailable at any time

Appendix

Microsoft Docs Refs | Links



12

1. **ADX - MS Docs Ref**

[Data Explorer - Analytics Service | Microsoft Azure](#)

2. **ADX - Pricing Details**

[Azure Data Explorer | Microsoft Azure](#)

3. **ADX – Azure Reference Architectures / Patterns**

[Browse Azure Architecture - Azure Architecture Center | Microsoft Docs](#)