Microsoft Azure

Azure Data Explorer
Fast and highly scalable interactive analytics

Start exploring your data

4. Februar 2020

Hans-Peter Bareiner - habarein@microsoft.com

Azure Data Explorer
Fast and fully managed data analytics service

Fully managed for efficiency

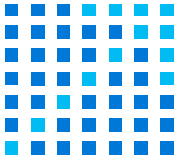
Focus on insights, not the infra-structure for fast time to value

No infrastructure to manage; provision the service, choose the SKU for your workload, and create database.

Optimized for streaming data

Get near-instant insights from fast-flowing data

Scale linearly up to **200 MB per second per node** with highly performant, low latency ingestion**.**



Designed for data exploration

Run ad-hoc queries using the intuitive query language

Returns results from **1 Billion records < 1 second** without modifying the data or metadata

Multi-temperature data processing paths

Hot

- seconds freshness, days retention
- in-mem aggregated data
- pre-defined standing queries
- split-seconds query performance
- data viewing

- in-mem cube
- stream analytics
- …

Warm

- minutes freshness, months retention
- raw data
- ad-hoc queries
- seconds-minutes query perf
- data exploration

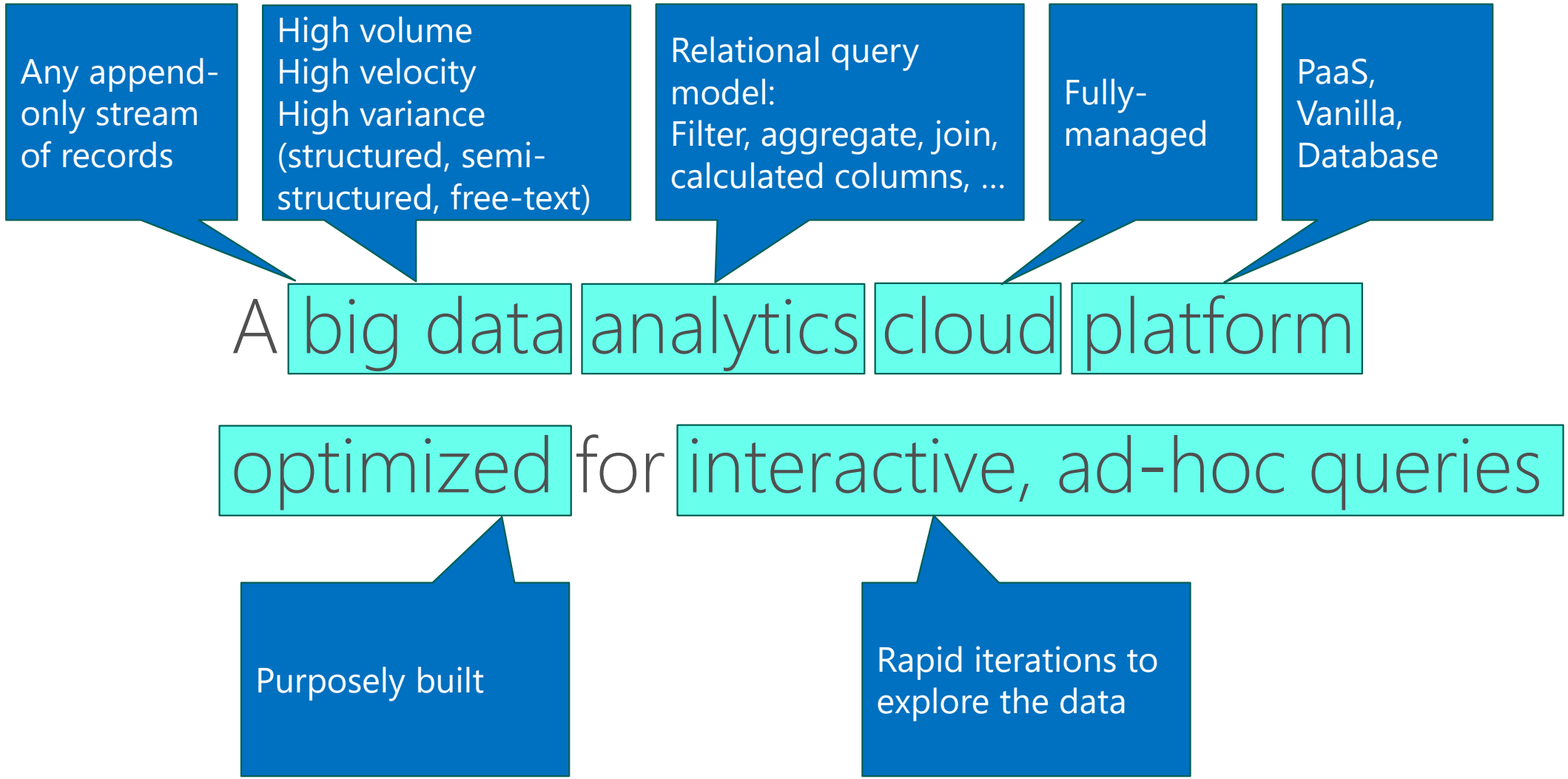
- column store
- Indexing
- …

Cold

- hours freshness, years retention
- raw data
- programmatic batch processing
- minutes-hours query perf
- data manipulation

- distributed file system
- map reduce
- …

Azure Data Explorer
In a sentence



Proven Technology

In production since 2015 for internal Microsoft workload, GA since Feb 2019.

Battle tested for Microsoft internal workload

The platform for analytical solutions (SaaS)

Azure Monitor

Security Products

IoT

Gamin g

Available as PaaS

#PoweredByADX

Common Use Cases

Time Series, Logs, Events, Transactions Data analytics platform

- Consolidate and correlate your logs and events data across on-prem, cloud, 3rd party
- Replace legacy log search solutions - save cost, infra, and index management overhead
- Accelerate your AI Ops journey (pattern recognition, anomaly detection, forecasting, etc.)
- IoT Analytics on Time Series and Telemetry data
- Replace HBase and Time Series databases

Build Analytical SaaS Solutions

- Build multi-tenant or single tenant SaaS analytics solution for Time Series, Logs, Events, Transactions, and security data

Analytic sandboxes

- Short term adhoc exploratory analysis

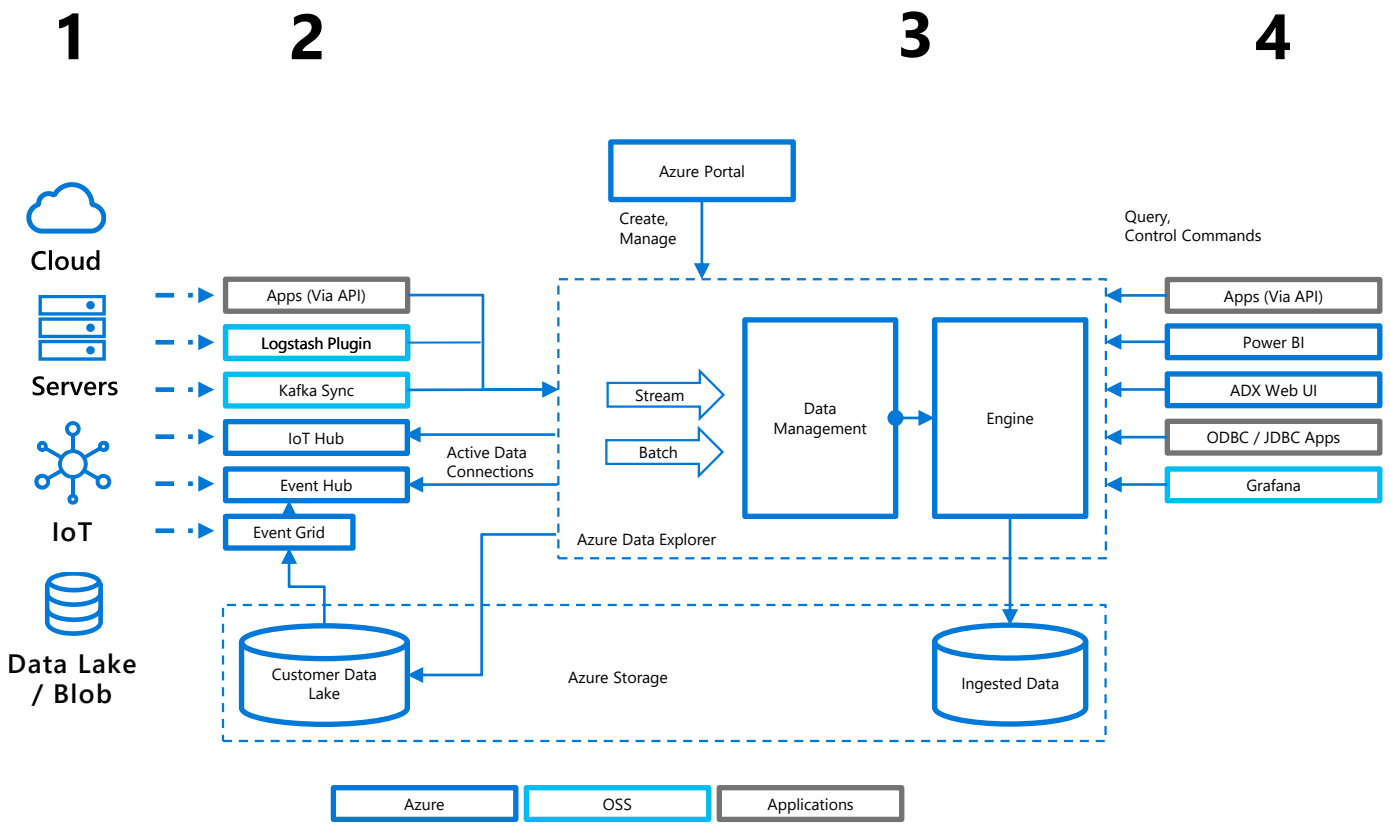
Azure Data Explorer overview

1. **Capability for many data types, formats, and sources**
   Structured (numbers), semi-structured (JSON\XML), and free text

2. **Batch or streaming ingestion**
   Use managed ingestion pipeline or queue a request for pull ingestion

3. **Compute and storage isolation**
   - Independent scale out / scale in
   - Persistent data in Azure Blob Storage
   - Caching for low-latency on compute

4. **Multiple options to support data consumption**
   Use out-of-the box tools and connectors or use APIs/SDKs for custom solution



Simple provisioning

Fully managed for efficiency

Easy provisioning

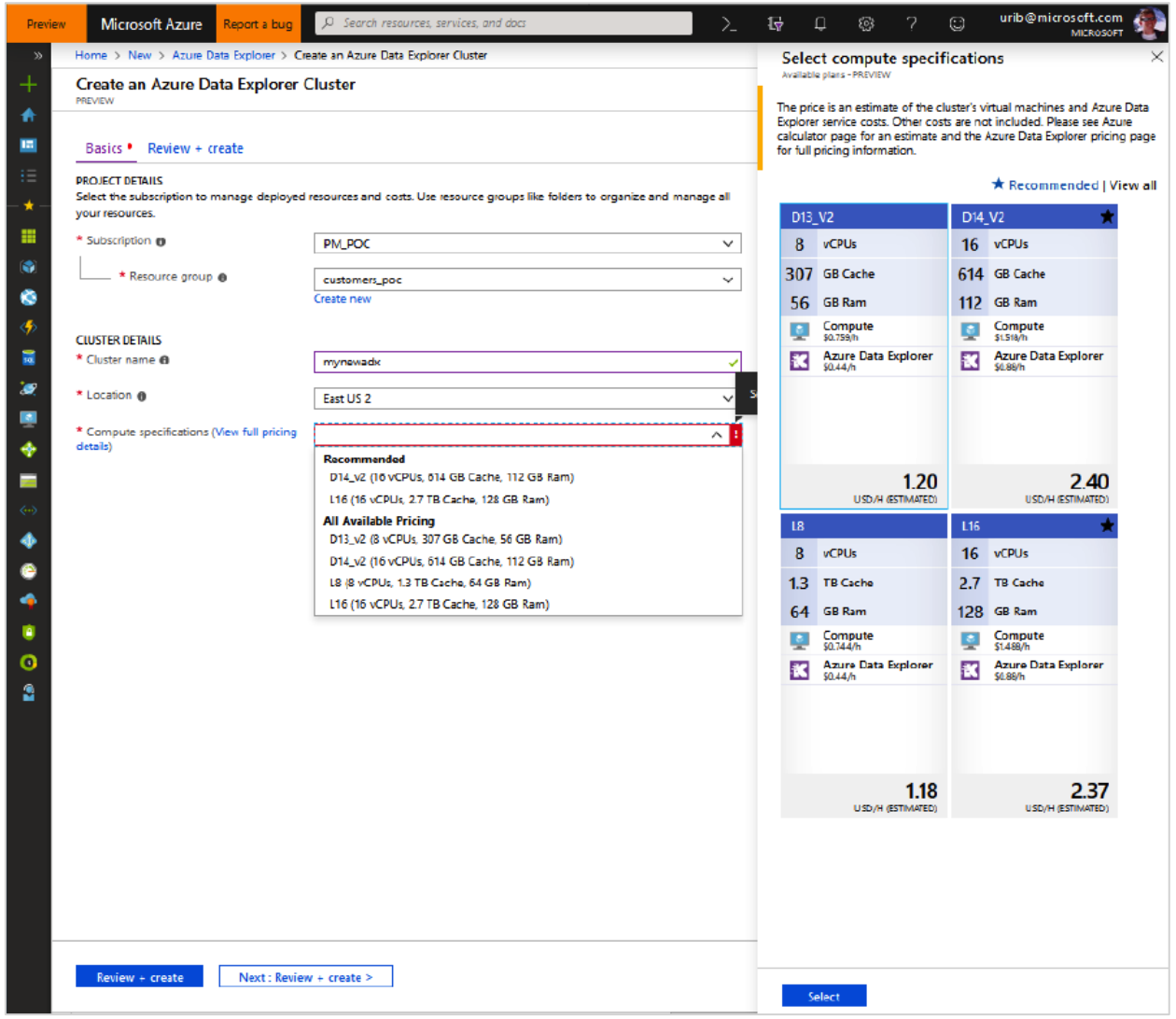
- No infrastructure to manage: Azure PaaS
- Use Azure Portal, APIs, or PowerShell to provision
- Storage Optimize/Compute Optimize SKUs
- Flexible data caching and retention options at database and table level

Rapid elasticity

- Buy only what you need
- Scale out/in manually or use autoscale
- Dedicated resources

Maintenance-free

- All columns are compressed and indexed during ingestion
- No index maintenance required



Fast ingestion

Optimized for streaming data

Easy input from multiple data sources

Multiple data sources

- Managed ingestion (e.g. Event Hub, IoT Hub) or programmatic ingestion (e.g. connectors, SDKs)

Multiple formats

- Tabular formats: CSV, TSV, PSV, SCSV
- JSON (line-separated, multiline), Avro
- ZIP and GZIP compression (for Batch)

Versatile ingestion

- Use batch or streaming ingestion

Easy input from multiple formats

- Tabular formats: CSV, TSV, PSV, SCSV
- JSON (line-separated, multiline), Avro
- ZIP and GZIP compression (for Batch)

Instant integration with simple transforms

- Reshape the data with update policies (Database Ingest Triggers)

Managed services


Azure Event Grid


Azure Event Hub


Azure IoT Hub

Azure Blob

Azure Data Lake

Connectors/Plugins

SDKs and APIs

Intuitive querying

Designed for data exploration

Simple and powerful

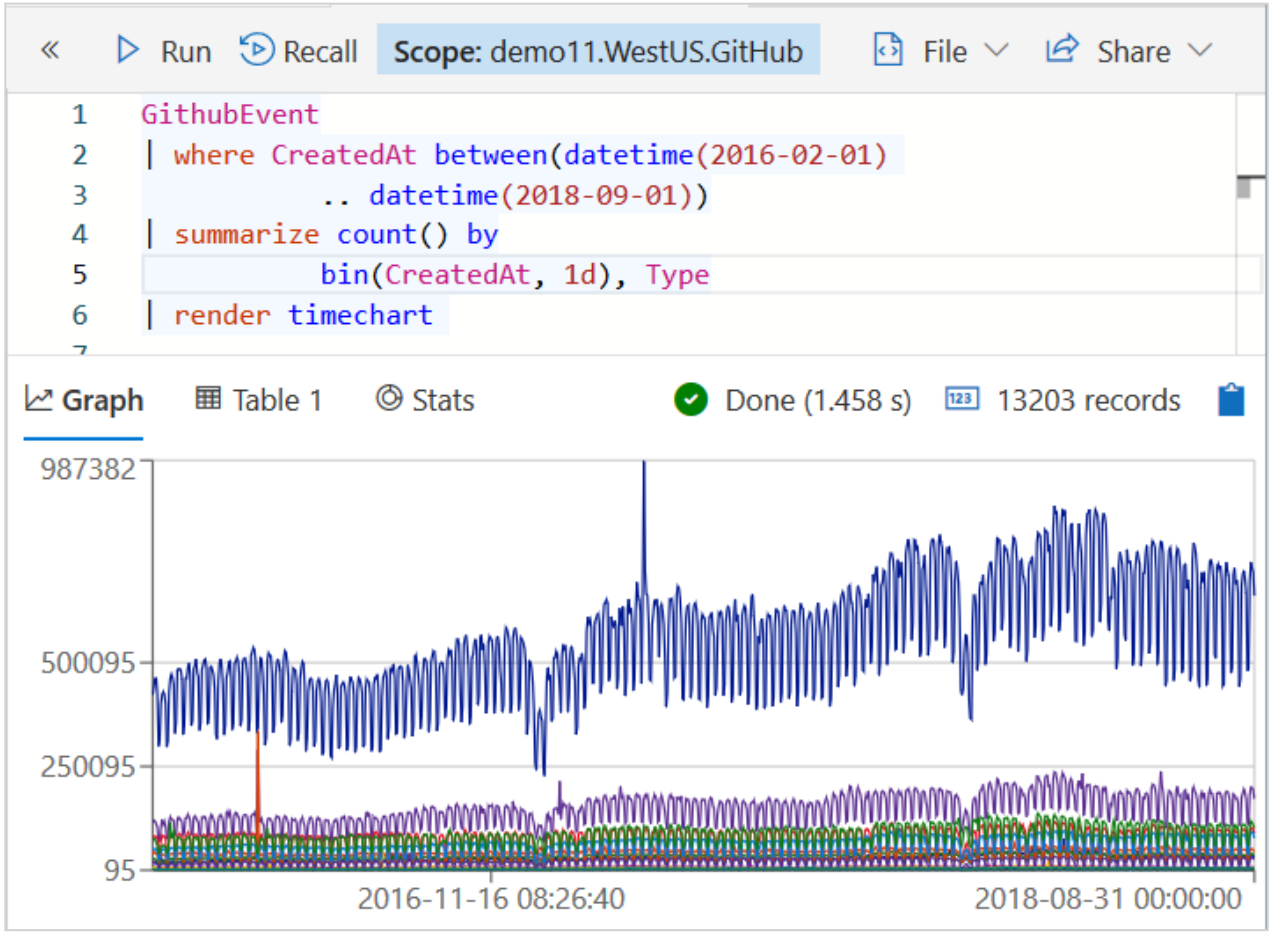
- Rich rational query language (filter, aggregate, join, calculated columns, and more)
- Built-in full-text search, time series, user analytics, and machine learning operators
- Out-of-the box visualization (render)
- Easy-to-use syntax + Microsoft IntelliSense
- Highly recognizable hierarchical schema entities

Comprehensive

- Built for querying over structured, semi-structured and unstructured data simultaneously

Extensible

- In-line Python and R
- **SQL**



The data lake at your fingertips

Cached, indexed access to the data lake	Continuous export to the data lake	Ad hoc query over raw data (New)
• Ad hoc query over raw data with full indexing • Automatic ingestion from selected lake repositories • Sensing new data as it enters the lake • ADLS G2 and Azure Blob storage	• Data ingested via other channels • Automatically saved to a data lake folder • Parquet, CSV	• External tables • Ad hoc query data in its natural format in the lake • Leverage data partitioning to optimize query time • Join across indexed and natural lake data

Advanced Analytics - Machine Learning with Azure Data Explorer

Out of the box	Distributed Custom Code Execution	Spark Integration	Tools
• Auto Clustering • Anomaly detection • Regressions • Forecasting • Series shape detection	• Distributed Python and R execution • Push code near the data • User defined functions • Stored functions	• Native Spark connector for heavy duty model training • Upload model or data into ADX for ongoing model scoring	• Jupyter Integration with KQL Magic • Python, Java SDKs

Enterprise Ready

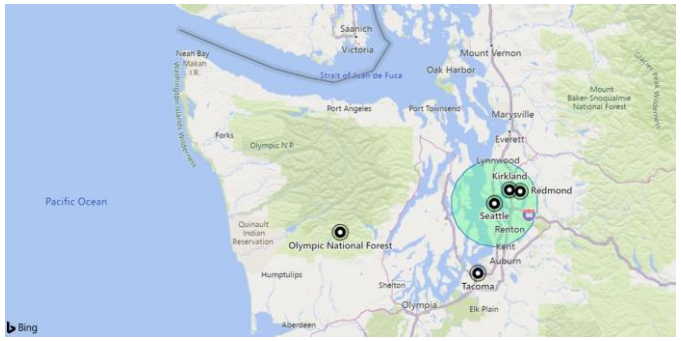
Mission critical

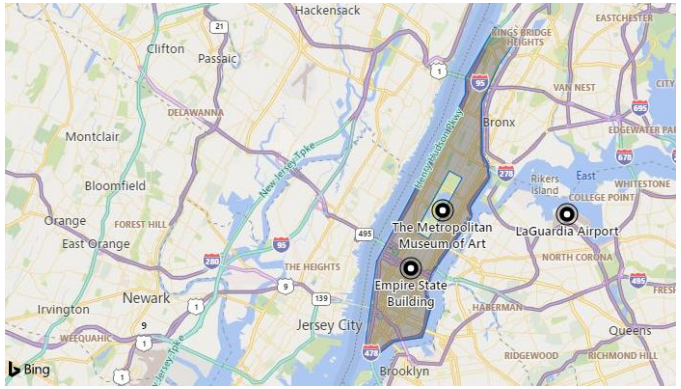
- ✓ Azure Active Directory Integration
- ✓ Role based authorization
- ✓ Virtual Network Support
- ✓ Encryption at Rest
- ✓ Encryption in Transit (https)
- ✓ Bring Your Own Keys
- ✓ Availability Zones
- ✓ Auto Scale-Up/In
- ✓ Globally available
- ✓ CI/CD Integration
- ✓ Automated provisioning
- ✓ Monitoring

New geospatial features in ADX

1. Geohash support
   - Transformation from coordinates to geohashes and back
   - Use-case: Summarization by geographical buckets, store locations based on a single column
2. Distance
   - Calculate the distance between two points
3. Contains
   - Check whether a point is in a given circle
   - Next: support for lines & polygons (i.e. check whether a point/line/polygon is in a polygon)
4. Next: Intersection
   - Support for lines & polygons

Geospatial coordinates are interpreted as represented per the WGS-84 reference system.





Thank you

Next steps

1. Visit the Azure Data Explorer product page to learn more
2. Access documentation, quick starts, and tutorials
3. Find pricing information for Azure Data Explorer
4. Get started with Azure Data Explorer now