Microsoft

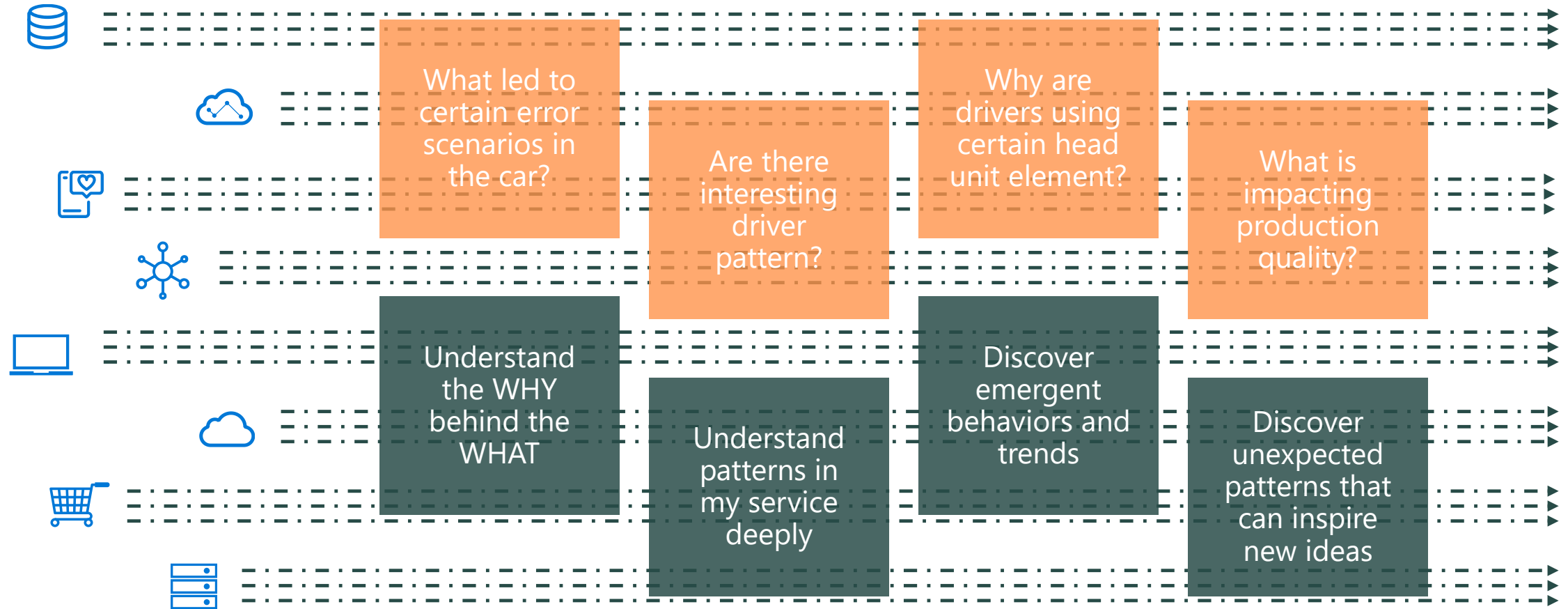# BMW ADX intro

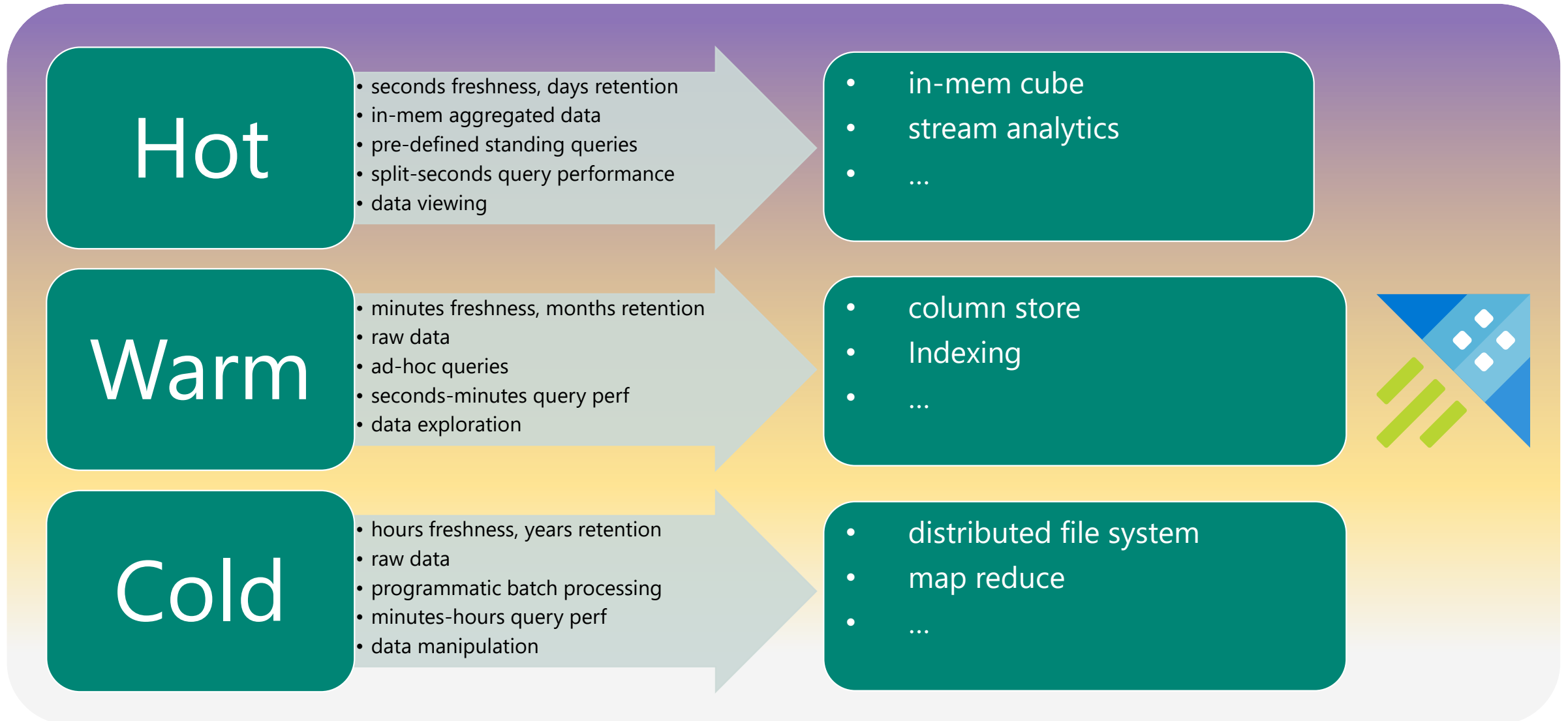Henning Rauch (30 min)
Adi Eldar (60 min)

Take action on details that differentiate your business

| What led to certain error scenarios in the car? | Are there interesting driver pattern? | Why are drivers using certain head unit element? | What is impacting production quality? |

| Understand the WHY behind the WHAT | Understand patterns in my service deeply | Discover emergent behaviors and trends | Discover unexpected patterns that can inspire new ideas |

New insight develops when you explore your data freely, unconstrained.

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
Fast and fully managed data analytics platform

Fully managed for efficiency

Focus on insights, not the infra-structure for fast time to value
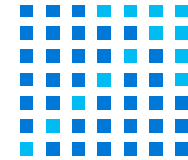
No infrastructure to manage; provision the service, choose the SKU for your workload, and create database.

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

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

© Microsoft Corporation
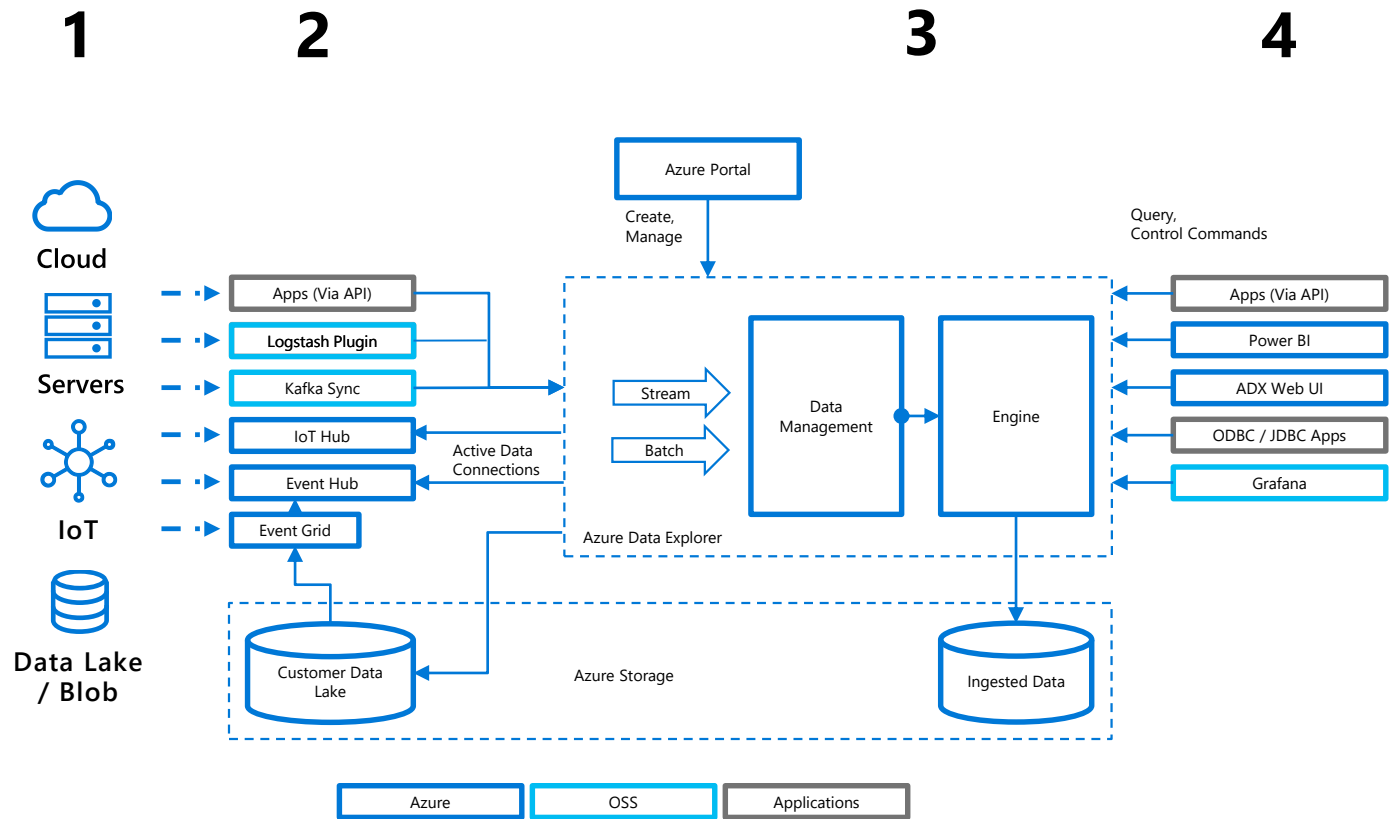
Simple provisioning

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

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

logstash

kafka

SDKs and APIs

Microsoft .NET   Java   python   +R :)   node JS   REST API

Intuitive querying
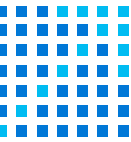
Simple and powerful

- Rich relational 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
- SQL
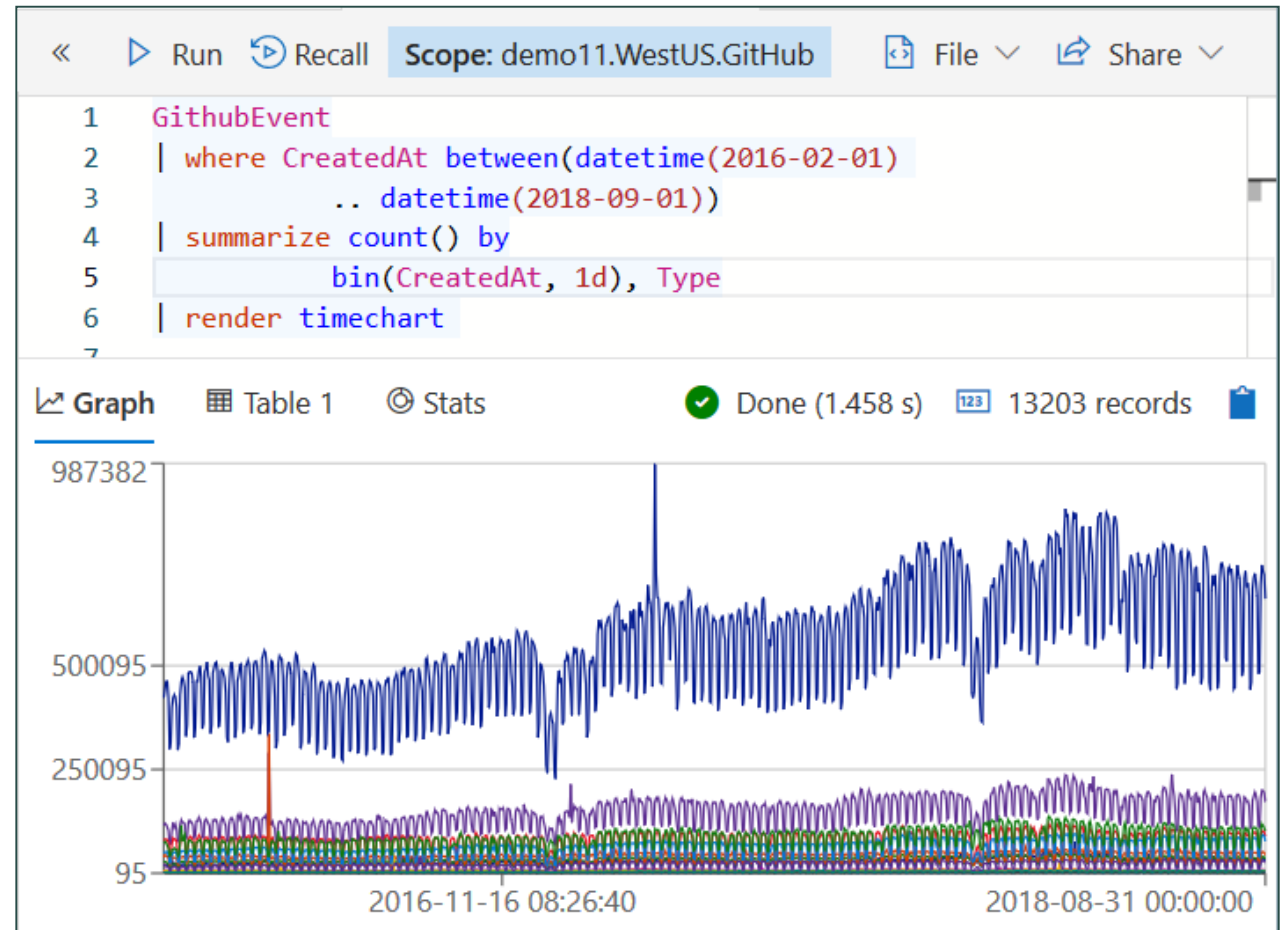
Typical scenarios

Log Analytics Platform

Fast, fresh, cost efficient, indexed access to all activity records, events and time series data
Platform for internal applications and services

Multitenant Analytics Solutions as a Service

Customer facing analytical solution in a variety of domains
Complementing the main product or the product itself

Advanced Analytics

Power data science and machine learning workloads
Boost data science team productivity

The platform for analytical Solutions(aaS)

Azure Monitor
- Log Analytics
- Application Insights

Security Products
- Windows Defender
- Azure Security Center
- Azure Sentinel

IoT
- Time Series Insights
- Azure IoT Central
- Microsoft Connected Vehicle Platform

Gaming
- PlayFab

#PoweredByADX

The data lake at your fingertips

Cached, indexed access to the data lake

- 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

Continuous export to the data lake

- Data ingested via other channels
- Automatically saved to a data lake folder
- Parquet, CSV

Ad hoc query over raw data (New)

- 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

- Auto Clustering
- Anomaly detection
- Regressions
- Forecasting
- Series shape detection

Distributed Custom Code Execution

- Distributed Python execution
- Push code near the data
- User defined functions
- Stored functions

Spark Integration

- Connector for heavy duty model training
- Upload model or data into ADX for ongoing model scoring

Tools

- Jupyter Integration with KQL Magic
- Python, Java SDKs

Truck telemetry analytics

Truck telemetry analytics - architecture
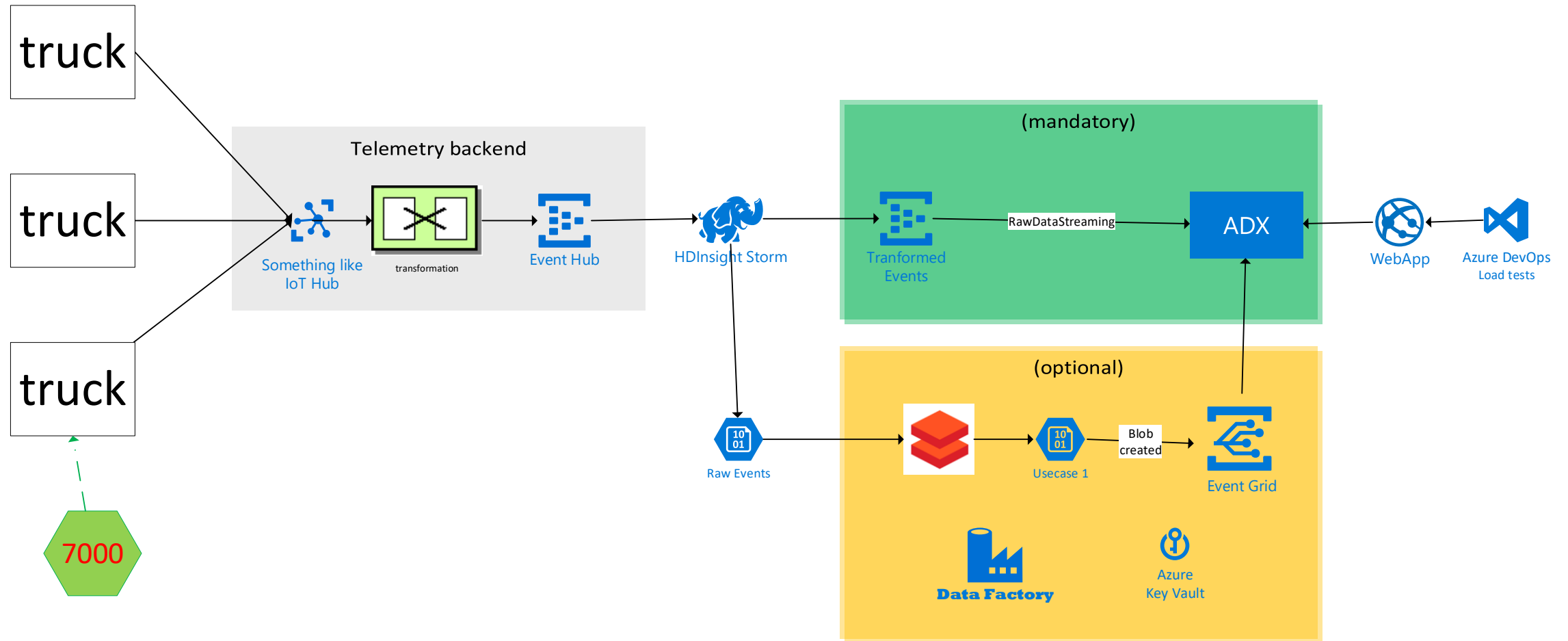
Truck telemetry analytics - data

```
{
  "VIN": "XXX",
  "MSG_TYPE": "SIGNAL",
  "TSC": [
    {
      "E": 1550478143711,
      "V": 2
    },
    {
      "E": 1550478189310,
      "V": 2
    },
    {
      "E": 1550478169077,
      "V": 2
    },
    {
      "E": 1550478166098,
      "V": 1
    },
    {
      "E": 1550478187382,
      "V": 1
    }
  ],
  "CGV": [
    {
      "E": 1550478140127,
      "V": 6
    },
    {
      "E": 1550478183965,
      "V": 0
    },
```

Vin	TS	TSC	CGV	...	...
xxx	2019-02-20 15:13:20	2	6		
yyy	2019-02-20 15:13:20	3	5		

```
{
  "StartTimestamp": "1558275110000",
  "SpacialCount": "120",
  "EndTimestamp": "1558275229000",
  "TriggerType": "NA",
  "SpacialSum": "9.242128906250073",
  "Sequence": "402",
  "Trigger": "analytics",
  "Version": "3",
  "Source": "ctpSource",
  "Signal": "Accel_Y_Cval",
  "Max": "1.4604609375000006",
  "Pts": "1558393275917",
  "Rts": "1558393275856",
  "Min": "-1.5297734374999994",
  "Vin": "3AKJHHDR9LSLP8810",
  "T0s": "0.2104609375000006",
  "T110s": "0.2700312500000006",
  "T101s": "0.2602656250000006",
  "T116s": "0.3701289062500006",
  "T108s": "0.2700312500000006",
  ...
```

Truck telemetry analytics - result

· Reimplement current dashboard functions using Azure Data Explorer

· Insert data continuously (100 raw event inserts per second)

ID	eature	PoC median response (ms) 60 concurrent users	Old system median response (ms)
1	Dashboard map	968	2 concurrent users = 13,700 10 concurrent users = 96,000
2	Crash events	386	10 concurrent users = 324 60 concurrent users = 957
3	Fuel Efficiency	179	10 concurrent users = 2,145 60 concurrent users = 7,527

Truck telemetry analytics - conclusion

Higher performance:

- Response time: Up to 100 times faster
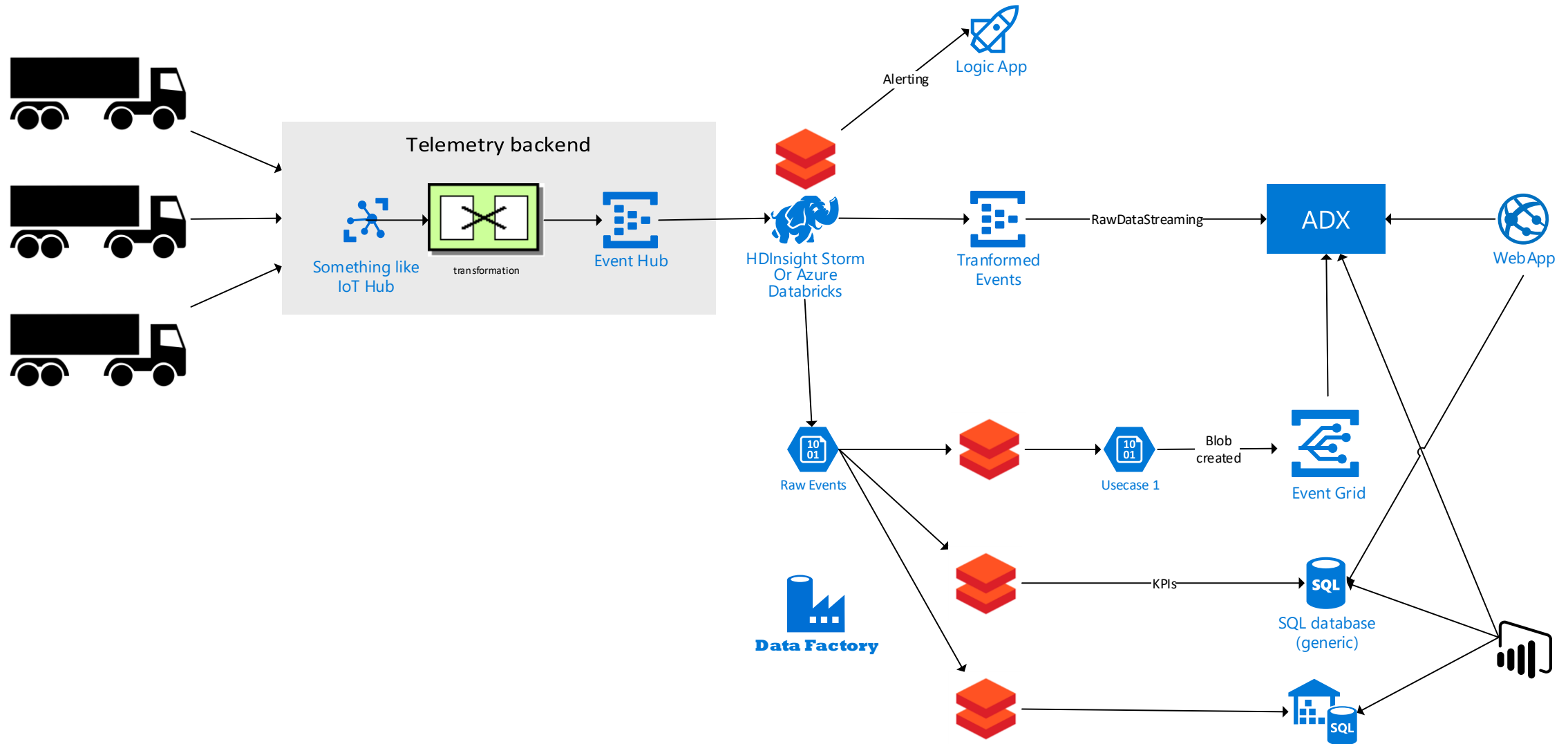
Costs are <u>lower</u>

- Cost per vehicle: At least 18.73% (up to 80%) less
- Fully PaaS service

Code is <u>simpler</u>

- Fewer lines of code
- Query language is well-suited to IoT telemetry analysis

Truck telemetry analytics – bigger scope

Azure Data Explorer customers

BÜHLER

ECOLAB®

snelstart

esri

INMOBI™

ZOOMD
KNOW MORE, DO MORE.

DocuSign®

"In terms of usability it is beyond compare any other database query language we have used so far."

Emilian Ertel
Senior Key Expert

SIEMENS
Healthineers

"The solution was so simple that we were up and running in a week, ingesting and analyzing 17 TB of data per day."

Ariel Pisetzky
Vice President of IT

Taboola™