



Cloud based Data Lake - modern Data Platform

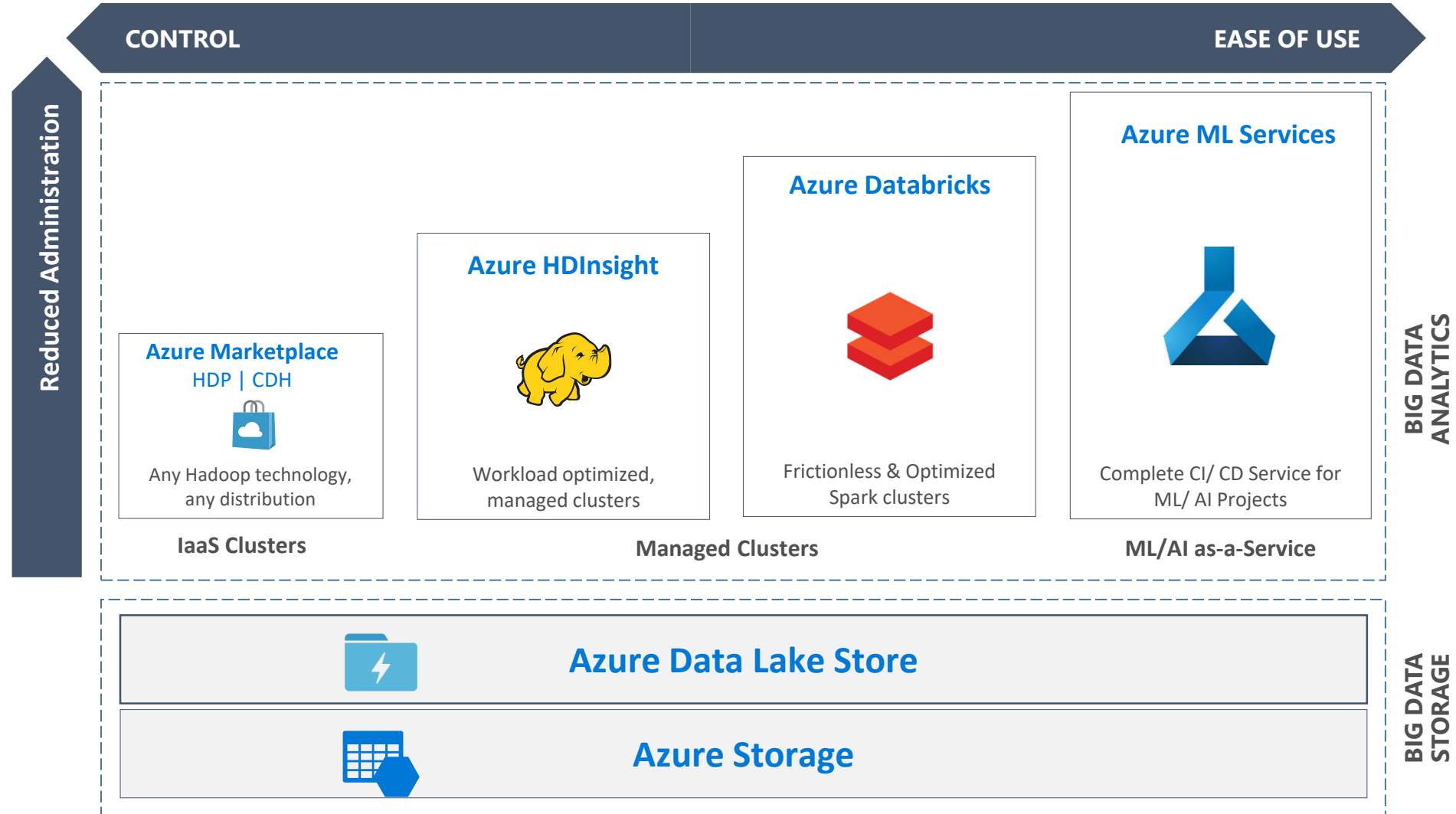
Guido Jacobs

Cloud Solution Architect – Big Data/ Data & AI

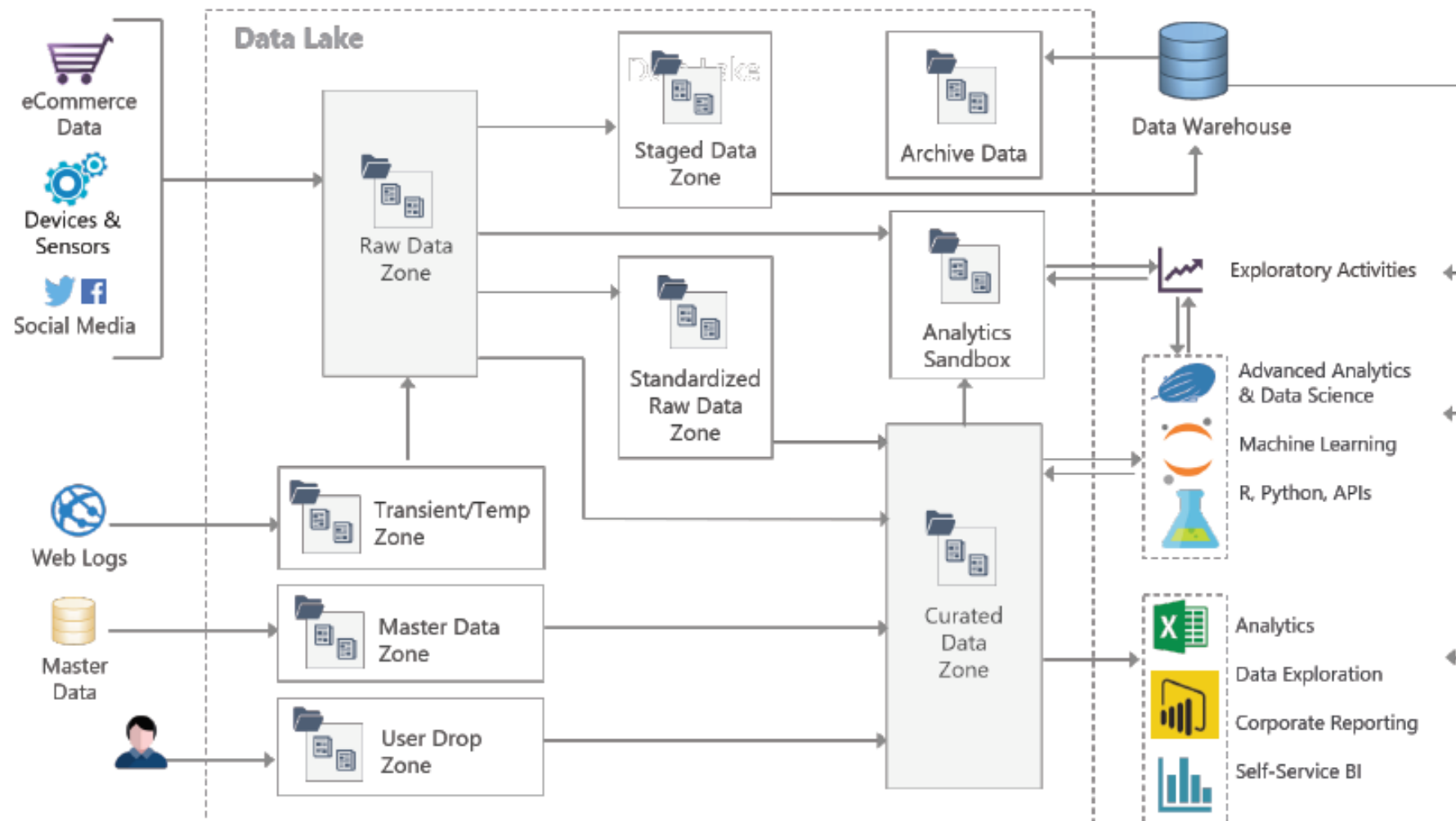
Guido.Jacobs@microsoft.com

Microsoft Deutschland GmbH

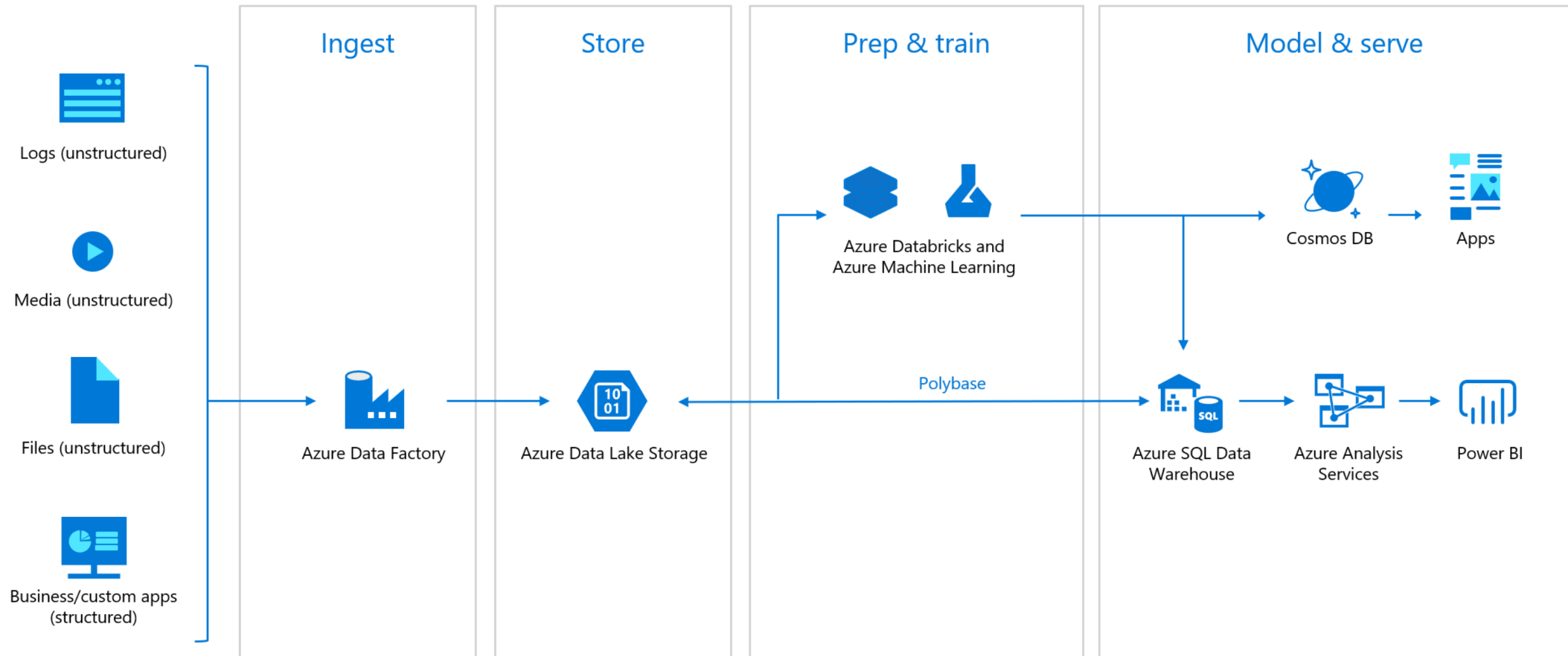
Cloud based Data Lake Platform



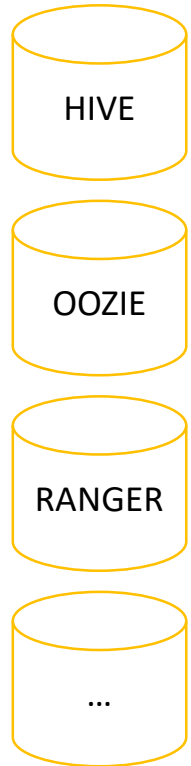
Zone-Concept of a Data Lake



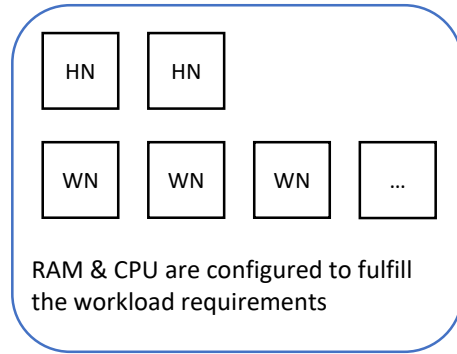
Cloud-based Data Lake „in action“



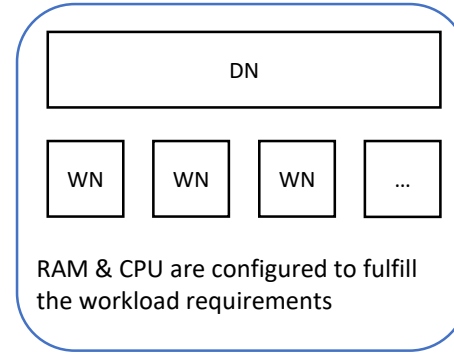
Shared Meta-Data



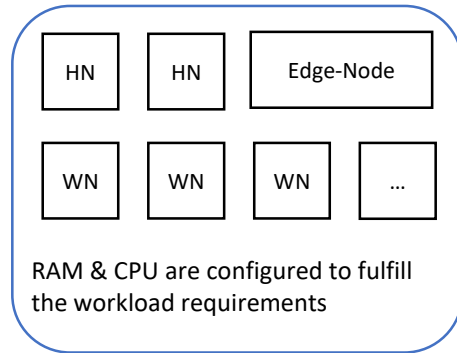
HDInsight Cluster
Type: Spark



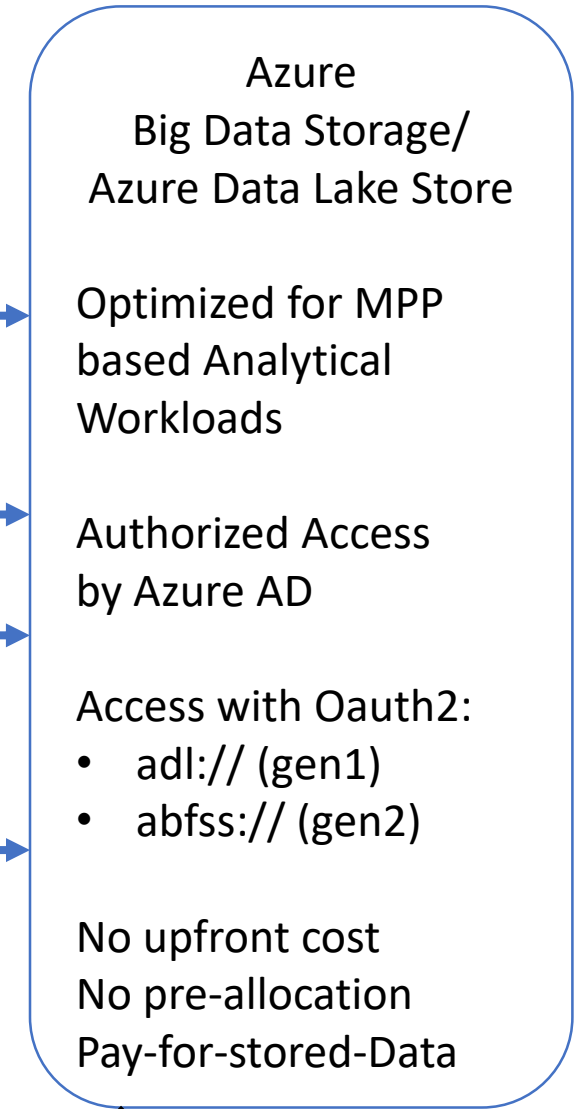
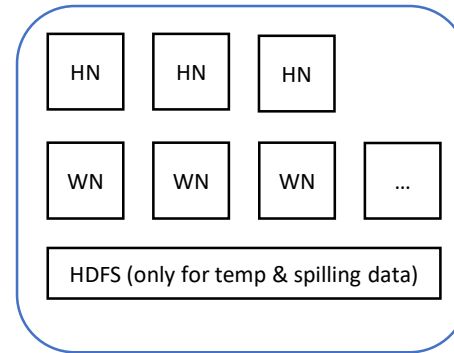
Databricks Cluster
Type: Spark



HDInsight Cluster
Type: HIVE LLAP



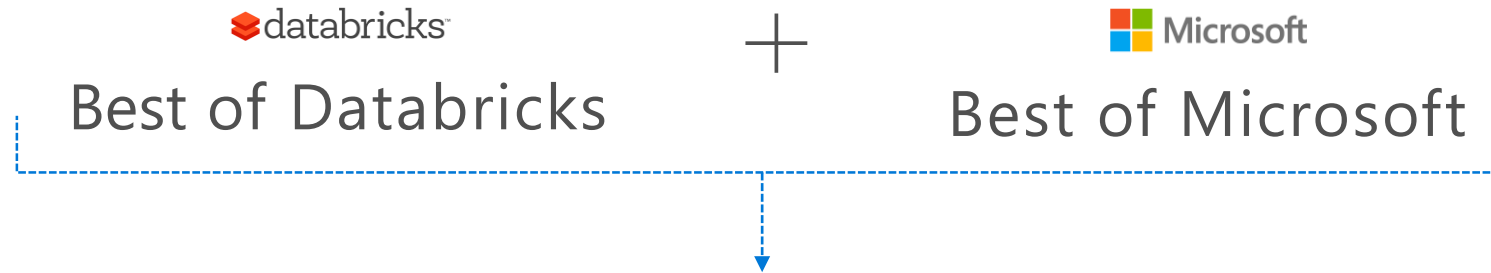
HDP (IaaS)
Type: Cloudbreak



Azure Databricks

What is Azure Databricks?

A fast, easy and collaborative Apache® Spark™ based analytics platform optimized for Azure



 Designed in collaboration with the founders of Apache Spark

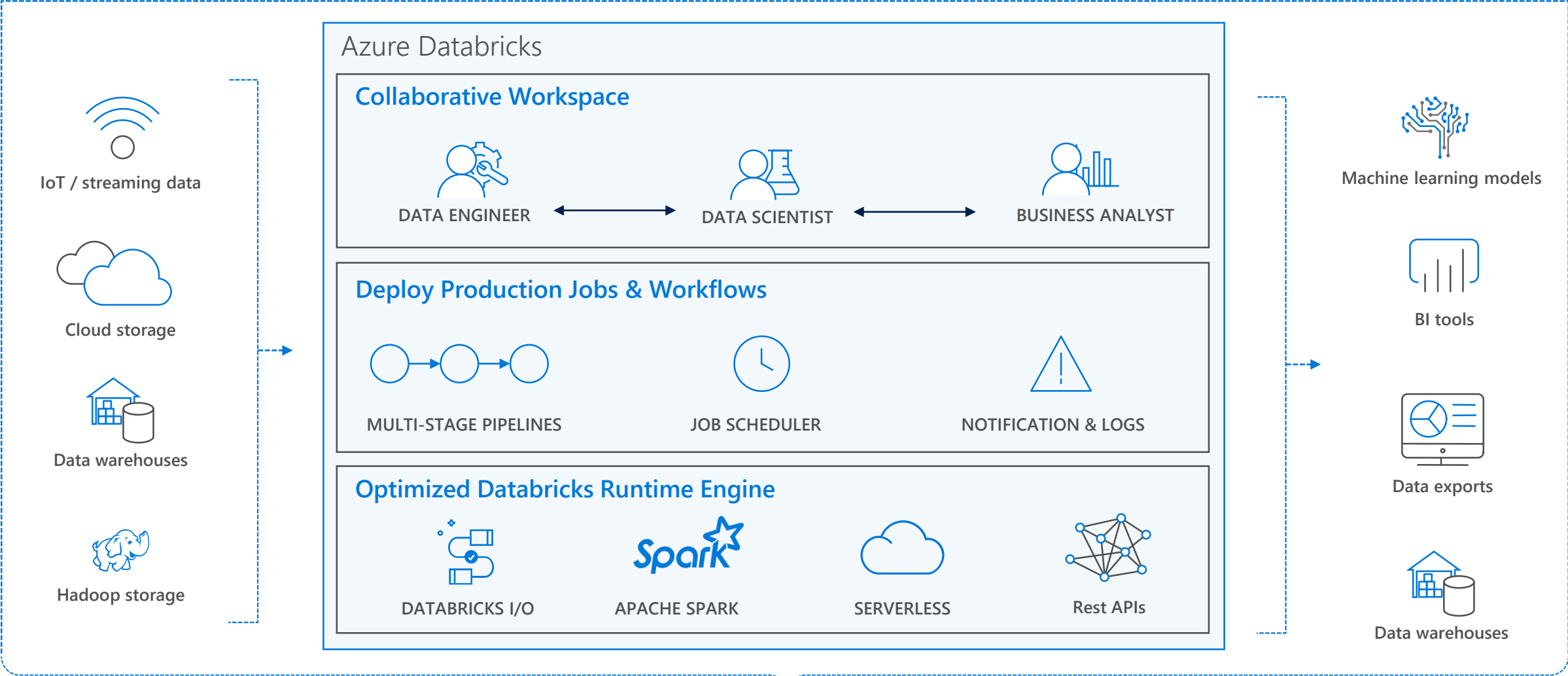
 One-click set up; streamlined workflows

 Interactive workspace that enables collaboration between data scientists, data engineers, and business analysts.

 Native integration with Azure services (Power BI, SQL DW, Cosmos DB, Blob Storage)

 Enterprise-grade Azure security (Active Directory integration, compliance, enterprise-grade SLAs)

Azure Databricks



Enhance Productivity

Build on secure & trusted cloud

Scale without limits

Time Series Analytics with ADX

Azure Data Explorer Übersicht

1. Unterstützung vieler Dateiformate und Strukturen

Strukturiert (Zahlenbasiert), Semi-Strukturiert (JSON/XML) und Frei-Text

2. Batch oder Streaming ingestion

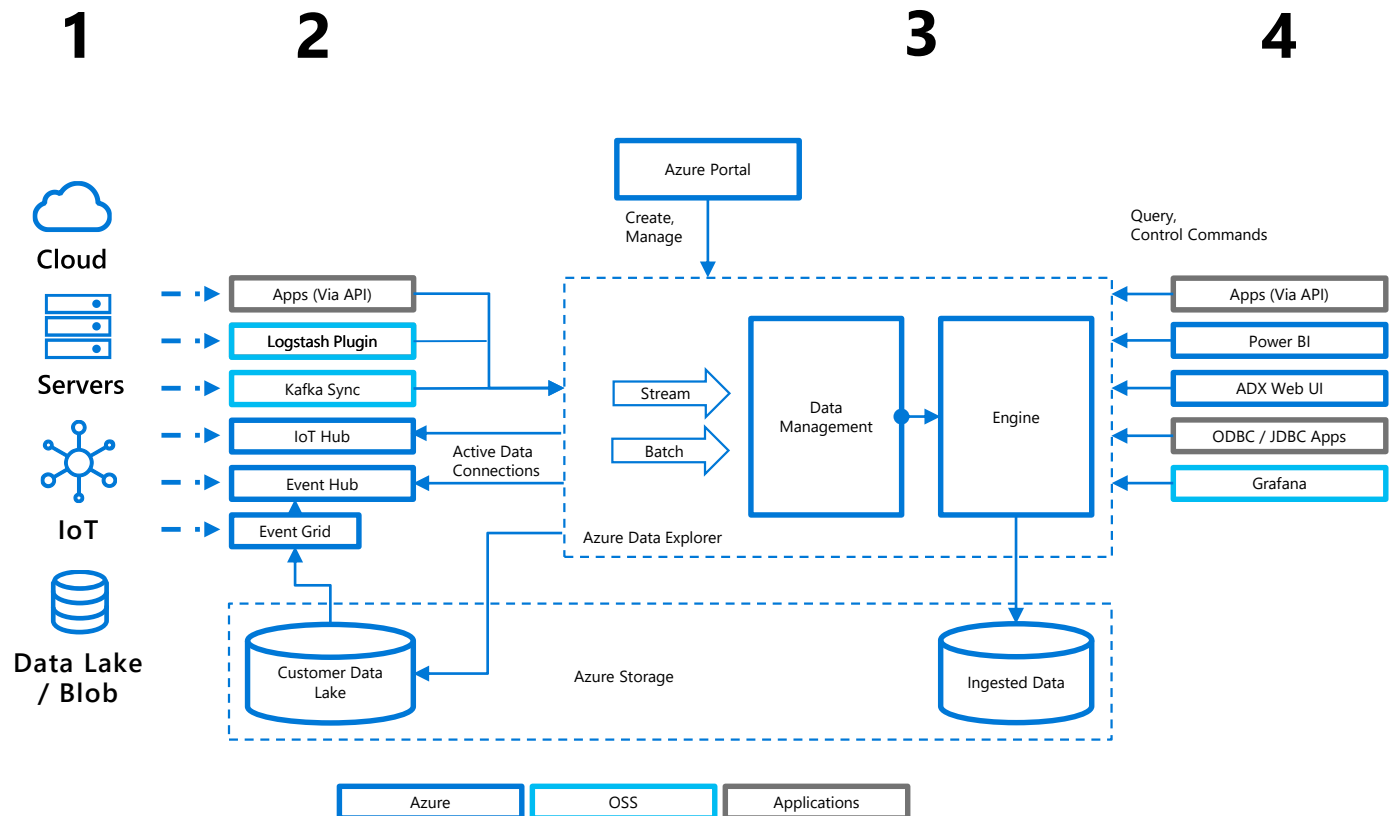
Nutzung von „managed“ Ingestion Pipelines oder automatisierte „Pull Ingestion“

3. Trennung von Compute und Storage

- Flexible Skalierung
- Langfristige Speicherung in Azure Blob Storage
- „Low-latency“ Caching auf den Compute Ressourcen

4. Verschiedene Standard Schnittstellen für den Zugriff

Out-of-the box Tools und Schnittstellen oder mit Hilfe von APIs/SDKs frei definierbar



Intuitive Abfrage

Leistungsfähig und simple in der Anwendung

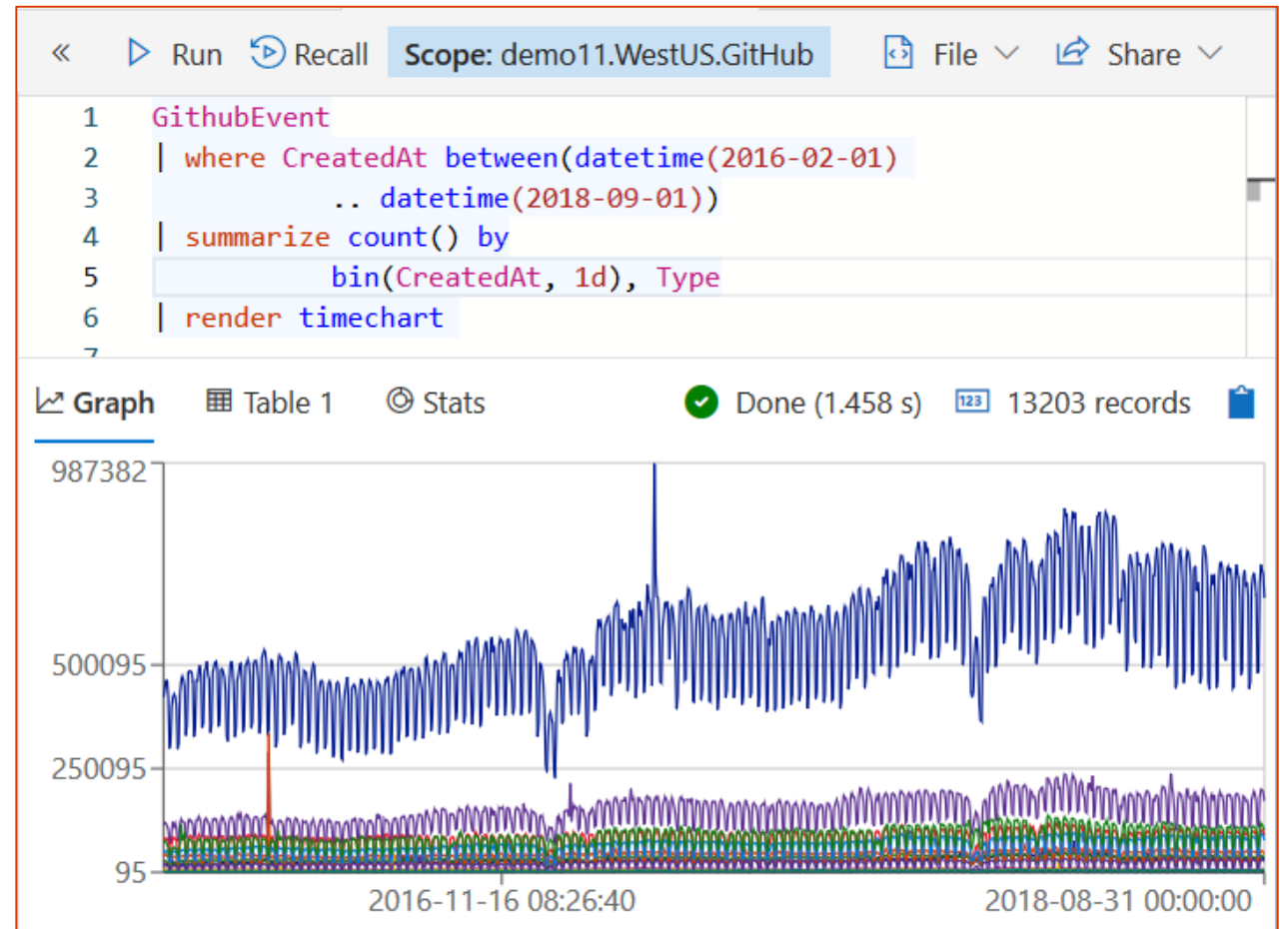
- Umfangreiche Querysprache (Filter, Aggregate, Joins, Calculated columns, ...)
- Built-in Full-text Search, Time Series, User Analytics, und Machine Learning Operatoren
- Out-of-the box Visualisierungs-Möglichkeiten
- Easy-to-use syntax + Microsoft IntelliSense

Übergreifend Einsetzbar

- Optimiert für Analysen über strukturierte, semi-strukturierte und unstrukturierte Daten zeitgleich

Flexible Erweiterbar

- In-line Python
- SQL





© 2016 Microsoft Corporation. All rights reserved.