Big Data & Data Warehousing

Microsoft's big data and data warehousing solutions handle all types of data, end-to-end: streaming, collection, processing, storage, and analytics

Data, data, everywhere.

Data is expanding ten-fold every five years with over 85% of the increase coming from new sources outside the traditional relational data warehouse. Data sources include mobile, social, videos, sensors, devices, RFID, web logs, advanced analytics, and click streams. Microsoft's big data & data warehousing offerings are able to process and query data wherever it might live.

Leader

Gartner has named Microsoft a leader in vision and ability to execute for their 2016 Magic Quadrant for Data Warehouse and Management Solutions for



Store any data of any size and speed

- Relational data and non-relational data
- Real-time data
- Data of any size from terabytes to petabytes
- Dynamically scales to match your business priorities

Process and query any data, anywhere

- Distributed query federates and joins heterogeneous data sources from on-premises and cloud, structured and unstructured
- Create managed data pipelines to orchestrate data transformation

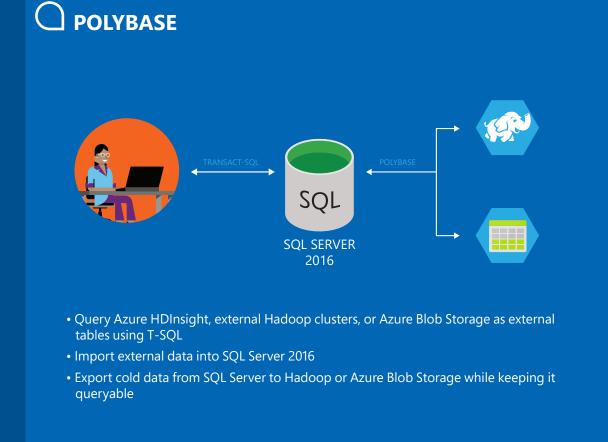
On-premises and in the cloud

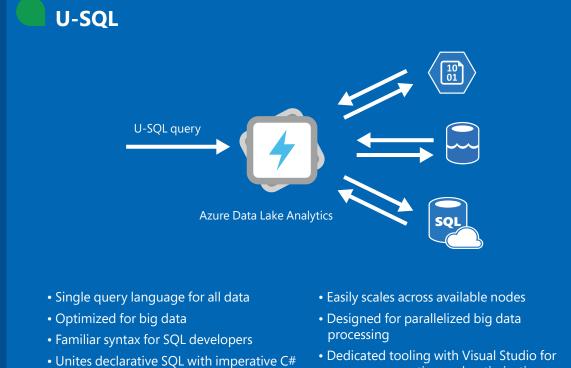
- On-premises options with enterprise software, reference architectures and appliances
- Cloud with virtual machines (laaS) and managed services (PaaS)
- Hybrid deployments across on-premises and cloud

Choice of tools and strong ecosystem

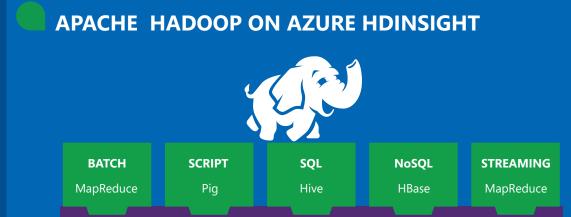
- Managed and supported open source tools and Microsoft services for all functions
- Designed for seamless interaction between tools and services
- Strong partner ecosystem to integrate and extend the solution







easy query creation and optimization



· Works across structured, semi-structured,

and unstructured with federated query

• Hadoop as a service on Azure: cost-effective, elastic & flexible • Works on Azure Storage or Data Lake Store • Customize clusters to run other Hadoop open-source projects • Crunch all data—structured, semi-structured & unstructured • Scale elastically on demand • Develop in your favorite language

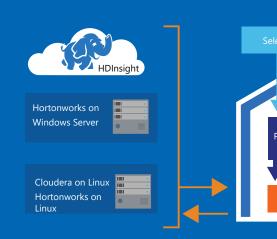
AZURE SQL DATA WAREHOUSE



Big data & data warehousing from beginning to end

- Enterprise-class cloud data warehouse with T-SOL • Dynamically scale and pause in seconds • Queries integrate relational data with data in Azure Blob Storage
- Deploy in seconds • Automatic backup to Azure Blob Storage

ANALYTICS PLATFORM SYSTEM



- Scale-out, massively parallel processing system supporting integrated data warehouse
- Easy to deploy, ships to your datacenter with hardware and software pre-installed and
- Queries across relational and non-relational data by leveraging PolyBase • Offers the lowest price per terabyte for large data warehouse workloads

AZURE SQL DATABASE DOCUMENT DB



- Create pools of elastic databases to manage performance and cost
- Develop scalable SaaS applications • Enterprise grade security Work within your preferred

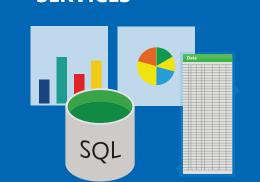
 Natively supports JSON and JavaScript • Schema-agnostic documents, automatically indexed Supports SQL queries

• Share dashboards and reports

Bring together data from a variety • Transform and model your data • Create linked, interactive visuals

POWER BI

SQL SERVER REPORTING SERVICES



• Show charts, maps, and KPIs Programming features for • Interact with your data anywhere on

MICROSOFT R SERVER SQL SERVER





- Flexible and agile with exceptional performance & enterprise support • Use ScaleR to compute large data sets No memory constraints



AZURE MACHINE LEARNING



AZURE DATA FACTORY

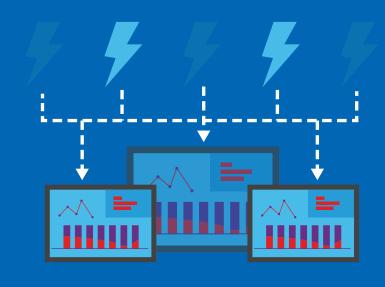
for business needs



- Uncover patterns hidden in data • Apply statistical methods to solve any
- Get started in minutes with drag & Leverage familiar R and Python

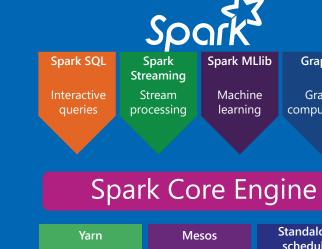
Compose and orchestrate data movement and processing at scale Visualize data orchestration Connect to on-premises and cloud Data workflow scheduling

AZURE STREAM ANALYTICS



 Real-time insights from devices and sensors • Enable rapid development with SQL based syntax Achieve mission-critical reliability and scale • Integrate directly with Power BI to publish real-time data

APACHE SPARK ON AZURE HDINSIGHT



= orchestration

Batch processing Real-time processing Stream processing

Machine learning

✓ Interactive SQL

 Open source processing framework for data analytics Parallel data processing persists data in-memory, on disk • Suited to ETL, batch, interactive queries Real-time processing for real-time scenarios



• Enterprise scale & performance:

MICROSOFT R SERVER ON AZURE HDINSIGHT

Scales from workstations to large Growing portfolio of parallelized Runs R functions in parallel • Secure, scalable R deployment &

- Write once, deploy anywhere: standalone R server Linux: RedHat and SuSE Cloudera, MapR

AZURE DATA LAKE

STORE

- · No limits to scale—architected for cloud scale and performance High frequency, low latency,
- HDFS for the cloud · Optimized for massive throughput • Stores data in native format
- Enterprise ready: secure, manageable & reliable



AZURE DATA LAKE

ANALYTICS

- On-demand job service built on YARN Pay for what you use • Use U-SQL—familiar, easily extensible • Develop faster and debug smarter with
- Query any data store with federated Enterprise ready: access control &

Search for big data & data warehouse on msdn.microsoft.com