Azure can help with large, complex, and tabular structured and unstructured data forms coming from devices, services, and applications. Azure offers a more sophisticated level of processing beyond traditional data warehousing.

**DATA ANALYTICS**

Azure has a comprehensive set of services to ingest, store, and analyze all types and scales of data ranging from; spanning table, file, streaming, and other data types.

The Azure platform provides tools across the Data Analytics lifecycle. This allows you to:

- Ingest data into Azure using robust services for batch ingestion, or real-time ingestion so that events are captured as they are being generated from your devices and services.

- Store structured or unstructured data globally at virtually unlimited scale.

- Train and prepare the data in data stores to derive insights and create predictive/prescriptive models on data using Machine Learning and Deep Learning techniques. These capabilities can extend to real-time processing of streaming or log data. You can even leverage artificial intelligence with machine learning and cognitive services for automated machine analysis.
• Serve and publish analyzed data to an Operational or Analytical store to help with visualizing as part of reports and dashboards. Analyzed data can also be leveraged directly to your apps, while ensuring access is secure and performance meets required service levels.

AZURE DATA MOVEMENT CAPABILITIES

The first step in data analysis is connecting disparate datasets from multiple sources and ingesting them into Azure. Data may originate in a data center, in cloud services, or span both.

Azure Data Factory is the primary service for batch ingestion of data. Azure Data Factory is an ingestion, orchestration, and scheduling service that determines what happens when certain events occur and determines which engines to use to optimally analyze and process your data. This service allows you to create sophisticated data pipelines from ingestion of data to processing, storing, and making it available for your end users and apps to tap into.

There are other Data movement capabilities on Azure too. If you have a massive one-time upload, the Azure Import Export Service manages the bulk loading of large data sets into Azure Blob storage and Azure Files by shipping drives to an Azure data center.

If you have structured data, the Azure Data Migration Service migrates data from on-premises structured databases directly into Azure, maintaining the same relational structures leveraged by your current apps.

Azure also has engines for ingesting real-time data streams. These engines can ingest data at a fast pace and cater to processing needs in the future.

Azure Event Hubs enables large scale telemetry and event ingestion with durable buffering and low latency from millions of devices and events. Azure IOT Hub is a device-to-cloud telemetry data service to track and understand the state of your devices and assets.

For custom operations to perform and in order to scale out ingestion engines with custom logic, Azure also supports the open source Apache Kafka in HDInsight as a managed high-throughput, low-latency service for real-time data.

Azure Command Line Interface (CLI) allows you to programmatically target and ingest multiple data formats into Azure. If you’re a developer, APIs can be called using the Azure Software Development Kit (SDK) to bring in your data.

STORING DATA

As you plan for the process of data ingestion, it is also important to plan for where and how the data will be stored in Azure. Azure Blob Storage is a managed service that can store massive datasets, regardless of their structure, or the lack of it, and keep it ready for analysis, including video, images, scientific datasets, and more.

For demanding analytical throughput requirements, or large file sizes that need to be optimized for analysis, using Azure Data Lake Store is optimal. Azure Data Lake Store allows you to analyze both structured and unstructured data with a very high throughput, generally desired by analytics engines. It can store trillions of files and a single file can be larger than one petabyte in size.

Azure SQL DB is used for operational and transactional data in structured or relational form. Azure SQL DB is an Azure service that works similarly to SQL Server, so the management and scaling of your host infrastructure is secure. Existing database apps can also be in hosted Windows- or Linux-based virtual machines.

For analytical data that has been aggregated over the years, Azure SQL Data Warehouse provides an elastic, petabyte scale service, which allows you to dynamically scale data either on-premises or in Azure.

Azure Cosmos DB is a turnkey globally distributed NoSQL DB service that allows you to bring in data that is schema-agnostic. Azure Cosmos DB uses key-value, graph, and document data together, with multiple consistency levels to cater to specific app requirements.

Whatever the need, Azure has an optimal store for you. Interestingly, all these stores integrate seamlessly to the analytics engines as sources of data.
TRAINING AND PREPARING DATA

With data stored in Azure, there are many analytics options for training and preparing your data, spanning from super-scalable and involved approaches, to data engineering, through automated machine analytics on serverless infrastructures.

Azure Databricks is an optimized Apache Spark-based analytics cluster service. It offers the best of Spark with collaborative notebooks and enterprise features. Azure Databricks leverages integrated Azure Active Directory and native connectors to other Azure data services. Azure Databricks is a hub of Spark-based analytics including Batch, Streaming or Machine Learning.

HDLInsight is a managed cluster service for a variety of Open Source Big Data analytics workloads. HDInsight helps to clean, curate, process and transform data in addition to scaling machine learning workloads. Using HDInsight, you can create scale-out clusters for Hadoop, Spark, Hive, HBase, Storm, and Microsoft R Server without monitoring and administering the underlying infrastructure.

For scale-out compute engines similar to traditional SQL infrastructure, Data Lake Analytics allows you to develop and run large scale, parallel data transformation and processing programs in U-SQL over petabytes of data from your Data Lake. Data Lake Analytics can leverage the familiarity and extensibility of U-SQL to scale machine learning models to work against massive amounts of data from R or Python. Most importantly, it is a serverless environment that allows you to request and leverage compute resources on a per query basis, which makes scaling and parallel execution easy for the maintenance of large clusters of data.

Azure also has engines for processing real-time data streams. To analyze data logged in real time from devices, sensors, and more, Azure Stream Analytics offers a powerful event processing engine. Azure Stream Analytics together with Event Hubs, allows you to ingest millions of events to find patterns, detect anomalies, create power dashboards, or automate event-driven actions. With the simplicity and familiarity of a SQL like language, real-time streams are processed efficiently.

Azure HDInsight and Azure Databricks also allow you to leverage streaming capabilities within the scale-out processing engines, like Structured Streaming with Spark.

For advanced analytics, Azure Machine Learning and Microsoft Machine Learning Server provide infrastructure and tools to efficiently build intelligent apps and services by analyzing data, creating high quality data models, and by training and orchestrating machine learning.

In addition to these tools, scale-out cluster technologies like Azure Databricks, also allows for scalable machine learning with Spark ML and Deep Learning libraries like CNTK and TensorFlow. Cognitive Services; a first level AI service, provides pre-built intelligent services for vision, speech, text, understanding and interpreting.

SERVING DATA

After you have analyzed and derived insights from your data, you want to effectively Serve this enriched data to your end-users.

Within Azure, the best destination for all analyzed data is Azure SQL Data warehouse. In Azure SQL Data warehouse, you can combine new insights with historical trends and drive a targeted conversation by maintaining one version of data for your org. Azure SQL Data warehouse integrates well with Business Intelligence tools and supports seamless connectivity to analytics tools and services. Azure Analysis Services and Power BI provide powerful options to find and share further data insights.

Analyzed data that contains insights valuable to end-consumers can be populated into Operational stores like Azure SQL DB and Azure Cosmos DB, so that web and app experiences can be augmented by those insights.

Using Azure Functions, you can pipe data directly to your apps with Azure Platform tools for developers, including Visual Studio, Azure Machine Learning workbench or custom server-less apps and services.
AZURE ACTIVE DIRECTORY

With Azure, you can ensure that data is consumed securely by intended users and groups, while network performance SLAs and privacy requirements are met using Azure Express Route. Azure key management services allows you to hold the keys to your data once it’s in the cloud with Azure.

Data Analytics

Demo Topics

INGEST

Your data may originate in your data center, in cloud services, or span both. If you need to do a large one-time data upload, you can ship your disk drives to an Azure data center. You can then use the Azure Import Export Service to manage the bulk loading of large datasets into Azure Blob storage and Azure Files. If you have structured data, the Azure Data Migration Service migrates data from your on-premises structured databases directly into Azure, maintaining the same relational structures leveraged by your current apps. To ingest data from multiple on-premises and cloud sources, use Azure Data Factory, a globally-deployed data movement service in the cloud. Azure Event Hubs enables large scale telemetry and event ingestion with durable buffering and low latency from millions of devices and events. Azure IOT Hub is a device-to-cloud telemetry data service to track and understand the state of your devices and assets.

STORE

Azure Blob Storage can store massive datasets, irrespective of their structure, or the lack of it, and keep it ready for analysis; including video, images, scientific datasets, and more. If you have a particularly demanding analytical throughput requirement, or you have humongous file sizes that you need optimized for analysis, you want to use Azure Data Lake Store. For operational and transactional data in structured or relational form, you can use Azure SQL DB.

Azure Cosmos DB provides native support for NoSQL choices, offers multiple well-defined consistency models, guarantees single-digit-millisecond latencies at the 99th percentile, and guarantees high availability with multi-homing capabilities and low latencies anywhere in the world. Whatever the need, Azure has an optimal store for you. Interestingly, all these stores integrate seamlessly to the analytics engines as sources of data.

TRAIN AND PREPARE

With your data stored in Azure, there are many analytics options for training and preparing your data - spanning from super-scalable and involved approaches to Data Engineering through to automated machine analytics on serverless infrastructure. Azure Databricks is an optimized Apache Spark-based analytics cluster service. HDInsight is a managed cluster service for a variety of Open Source Big Data analytics workloads. It helps you clean, curate, process and transform your data in addition to scaling your machine learning workloads. For scale-out compute engines similar to traditional SQL infrastructure, Data Lake Analytics lets you develop and run large scale, parallel data transformation and processing programs in U-SQL over petabytes of data from your Data Lake. Beyond this, we’ve also built engines for processing real-time data streams as well as a number of first level AI services called Cognitive Services, providing pre-built intelligent services for vision, speech, text, understanding and interpreting.
**Azure Learning Paths**

Azure Solution Architect

**Microsoft Mechanics**

An introduction to Azure Analysis Services

**Hands-on Labs**

**Power BI Embedded**

Flight Delay analysis using Cosmos DB, Spark cluster in HDInsight and Power BI

**Serve and Publish**

Visualization tools like Power BI provide powerful options to find and share further data insights among your data scientists, developers, and users. You can store data in a data warehouse and pipe it directly to your apps with Microsoft Azure Developer Tools, including Visual Studio, Azure Machine Learning Workbench, and custom serverless apps and services using Azure Functions. Within Azure the best destination for all this analyzed data is Azure SQL Data warehouse, where you can now combine new insights with historical trends and drive a targeted conversation by maintaining one version of data for your org. If the analyzed data contains insights valuable to end-consumers, these can be populated into the Operational stores like Azure SQL DB and Azure Cosmos DB, so that web and app experiences can be augmented by those insights.

**Additional Resources**

Explore services in Azure that can help you analyze your structured and unstructured data. You can learn more with these useful resources: