**Data Science**

Data science is working with large amounts of data to reveal patterns and gain insights. Data science involves many steps which can be facilitated using Microsoft Azure including storage, processing and pipeline management of data. Microsoft Azure supports a number of popular big data frameworks including Spark, Map Reduce, and Kafka.

**Big Data Storage: Azure Data Lake Store**

You need a place to store your data. Azure Data Lake Store is an enterprise-wide hyper-scale repository for big data analytic workloads. Azure Data Lake enables you to capture data of any size, type, and ingestion speed in one single place for operational and exploratory analytics. Azure Data Lake Store is the equivalent to HDFS in Hadoop. It provides Petabyte storage with open integration for any processing framework that utilizes open source HDFS. Azure Data Lake Store offers additional capabilities not provided by HDFS including: data encryption as rest, role based security, and integration with a full-fledged key vault.

[Learn More](https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-overview)  
[Get Started](https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-get-started-portal)

**Big Data Jobs: Azure Data Lake Analytics**

Scaling job processing for big data is a challenge. You can focus on writing, running, and managing jobs rather than on operating distributed infrastructure. With Azure Data Lake Analytics, instead of deploying, configuring, and tuning hardware, you write queries to transform your data and extract valuable insights. You only pay for your job when it is running, making it cost-effective. It uses U-SQL, a language that unifies the benefits of SQL with the expressive power of user code to analyze data in the store and across SQL Servers in Azure, Azure SQL Database, and Azure SQL Data Warehouse.

[Learn More](https://docs.microsoft.com/en-us/azure/data-lake-analytics/data-lake-analytics-overview)

[Get Started](https://docs.microsoft.com/en-us/azure/data-lake-analytics/data-lake-analytics-get-started-portal)

**Machine Learning: Azure Machine Learning**

Azure Machine Learning is a cloud predictive analytics service that makes it possible to quickly create and deploy predictive models as analytics solutions. Azure Machine Learning Studio provides a drag and drop interface for building predictive analytics solutions. Azure Machine Learning also provides a fully managed service you can use to deploy your predictive models as ready-to-consume web services.

[Learn More](https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-what-is-machine-learning)

[Get Started](https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-create-experiment)

**Spark/Hadoop: HDInsight**

Hadoop is possible on the Microsoft Azure platform with Azure HDInsight. HDInsight is a full Hadoop Distribution based on the Horton Platform (HDP) which supports major Hadoop components including but not limited to HBase, YARN, MapReduce, and Spark. Microsoft Azure makes the setup and use of the Hadoop clusters straightforward. With HDInsight, your data is stored in Azure Storage, so you can safely delete a cluster when it is not in use.

[Learn More](https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-introduction)[Get Started](https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-linux-tutorial-get-started)

**Event/Data Distribution: Event Hub**

Real time big data processing involves stream processing. Azure Event Hubs provides a mechanism, similar to Apache Kafka, for a publish/subscribe model to consume stream events and distribute to multiple systems. Azure Event Hubs is a highly scalable data streaming platform and event ingestion service capable of receiving and processing millions of events per second.

[Learn More](https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-what-is-event-hubs)

[Get Started](https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-create)

**Stream Processing: Stream Analytics/Event data processing/IoT**

Azure IoT services including Azure IoT Hub, Azure IoT Suite and Azure IoT Developer Center provide the capabilities to process device data from various devices. The Azure stream processing functionality provides several pre-packaged input sources and sinks and allows data transformation using a SQL like expression language.

[Learn More](https://docs.microsoft.com/en-us/azure/iot-hub/iot-hub-what-is-azure-iot)

[Get Started](https://docs.microsoft.com/en-us/azure/iot-hub/iot-hub-get-started)

**Data Movement: Azure Data Factory**

Source data comes from multiple locations including the cloud, SQL databases and online transaction systems. Azure Data Factory is a service that allows you to script the movement of data from source to destination, processing and publishing of job results. Azure Data Factory can be used to produce reference data on a regular time frame for use in streaming or batch processing jobs on HDInsight for data enrichment.

[Learn More](https://docs.microsoft.com/en-us/azure/data-factory/data-factory-introduction)

[Get Started](https://docs.microsoft.com/en-us/azure/data-factory/data-factory-copy-data-from-azure-blob-storage-to-sql-database)

**Data Discovery: Azure Data Catalog**

Discovery of organizational information often relies on an individual's knowledge of the datasets in the enterprise. Azure Data Catalog makes source data discoverable. With Data Catalog, any user (analyst, data scientist, or developer) can discover, understand, and consume data sources. Data Catalog includes a crowdsourcing model of metadata and annotations. It is a single, central place for all of an organization's users to contribute their knowledge and build a community and culture of data.

[Learn More](https://docs.microsoft.com/en-us/azure/data-catalog/data-catalog-what-is-data-catalog)

[Get Started](https://docs.microsoft.com/en-us/azure/data-catalog/data-catalog-get-started)