

Features of Azure for Analytics and Data Science



By Michael Kightley

The Azure cloud platform provides powerful, yet easy to use cloud-based tools for data transformation, data analytics and data science. These tools offer a lot of flexibility to Azure developers and cater to a variety of skill levels. This blog provides a brief run through of some of the features of the Azure tools for data analysts and data scientists.

For further information, or if you have more questions you'd like to discuss, please feel free to contact us at <u>contact@inflexionanalytics.com</u> and we'd be happy to assist you.

Azure Options for Data Analytics

Azure Synapse Analytics

Azure Synapse Analytics is the next generation of the Azure SQL Data Warehouse. It lets users load many data sources from relational and non-relational databases at a single given instance. These databases can reside locally or in the Azure cloud. The data is unified, processed, and analyzed using SQL. The <u>Azure Synapse</u> Studio acts as a workspace for data analysis and AI tasks.

Azure Databricks

Azure Databricks is an analytical service tool based on Apache Spark. Large datasets can be processed quickly. Databricks supports several languages such as Java, Python, Scala, SQL, as well as libraries such as PyTorch and TensorFlow. Spark data can be integrated with any of these languages and frameworks.

Microsoft Partner

www.inflexionanalytics.com

contact@inflexionanalytics.com

?7 Stratford Place London W1C 1AY

<u>Databricks</u> also offers integration with Azure Machine Learning giving an Azure developer access to hundreds of pre-determined machine learning algorithms. It minimises the complexity of setting up a Spark data center locally through auto-termination and auto-scaling.

Azure Data Factory

Azure Data Factory is an ETL (Extract Transform Load) service, used for processing structured data at scale. An ETL process extracts the data from various data sources, cleans and transforms the data and converts it into a format that is suitable for analysis. <u>Data Factory</u> helps users build ETL flows using a visual editor without code.

There are over 90 built-in connectors to common data sources, including BigQuery, S3 and many others. Data can also be effortlessly copied from Data Factory to Azure File Storage.

Azure Stream Analytics

Azure Stream Analytics is a real-time analytics service. It provides you an end-to-end bridge for streaming processes based on serverless technology. A data analytics pipeline is defined for streaming the data and data processes are defined through SQL syntax. The processing can scale up dynamically depending on the throughput and volume of the data. It also offers built-in recovery and machine learning capabilities.

<u>Azure Stream Analytics</u> lets you add Power BI as an output, and this allows Azure developers to visualise those data streams in real-time in the Power BI Service.

Data Lake Analytics

Data Lake Analytics is an on-demand analytics job service that helps in simplifying big data. It dynamically provisions resources and the system can automatically scale up or down as required. Using Azure Data Lake Analytics, you can perform data transformations using a variety of languages such as Python C#.Net, and SQL as well as others. <u>Data Lake Analytics</u> connects to other Azure data sources like Azure Data Lake Storage and performs data analytics on-the-go. An advantage is that it is a cost-effective solution for running big data workloads.

Azure Analysis Services

<u>Azure Analysis Services</u> can fetch data from multiple sources and build a single semantic model for processing. This model can help you develop high-end business intelligence service solutions with security and reduced delivery time. Analysis Services is highly scalable, and it is possible to import existing tabular models or SQL tables into the system.

Azure Tools for Data Science

Azure Machine Learning Studio

<u>Azure Machine Learning</u> Studio is a cloud-based drag and drop, collaborative platform where users of varying skill levels can build, deploy and test machine learning solutions, either using a no-code designer or built-in Jupyter notebooks for a code-first experience. It provides automated machine learning which aids both professional and non-professional data scientists to build ML models rapidly. Pre-configured machine learning algorithms and management modules of Azure data warehouses are also available, meaning Azure ML Studio ideal for any data scientist looking to efficiently research a machine learning models' performance.

Azure ML Studio offers built-in integrations with other Azure services such as Data Bricks, Data Lake Storage etc. It also supports open-source frameworks and languages like MLFlow, Kubeflow, PyTorch, TensorFlow, Python and R.



Azure Cognitive Services

<u>Azure Cognitive Services</u> offers a selection of pre-built ML and AI models. It can be used by any developer and doesn't require machine-learning expertise. Some of the features of cognitive services are:

- Image processing algorithms that can identify, index and caption your images.
- Speech recognition algorithms that can convert audio into readable and searchable text. Integrate real-time speech translation into your apps. Allows for voice verification based on audio.
- Mapping of complex data for semantic search and smart recommendations.
- Allowing apps to process natural language with pre-built commands to understand the user's needs. It also can detect and translate more than 60 supported languages.
- Access to billions of web pages, images through a single call by adding Search API's to your apps. Enables safe, ad-free search and advanced features like video search on your app.

Data Science Virtual Machine

<u>Azure Data Science Virtual Machine</u> (DSVM) is a virtual machine with pre-configured data science tools. ML solutions can be developed in a pre-designed environment. DSVM can be the ideal environment for data scientists to learn and compare machine learning tools as it is a complete development setting for ML on the Azure platform. Some of the data science tools in DSVM are data platforms, ML and AI tools, data visualization tools and development tools.

The DSVM environment can significantly reduce time to install, troubleshoot and manage data science frameworks. You can use it to evaluate or learn new data science tools. For more information on the full list of tools included with DSVM, <u>click here</u>.



www.inflexionanalytics.com

contact@inflexionanalytics.com

Inflexion Analytics © 2020