**Name -Monika Gariya**

**Email-** [**monikagariya2023@gmail.com**](mailto:monikagariya2023@gmail.com)

**Data Engineering Batch 1**

**Date – 12-02-2024**

**Topic - Introduction to Microsoft Azure**

## ****What is Azure?****

[Azure is Microsoft’s](https://www.geeksforgeeks.org/microsoft-azure/) cloud platform, just like Google has its Google Cloud and Amazon has its Amazon Web Service or AWS.000. Generally, it is a platform through which we can use Microsoft’s resources. For example, to set up a huge server, we will require huge investment, effort, physical space, and so on. In such situations, Microsoft Azure comes to our rescue. It will provide us with virtual machines, fast processing of data, analytical and monitoring tools, and so on to make our work simpler. The pricing of Azure is also simpler and cost-effective. Popularly termed as “Pay As You Go”, which means how much you use, pay only for that.

## How Does Microsoft Azure Work?

It is a private and public cloud platform that helps developers and IT professionals to build deploy and manage the application. It uses the technology known as virtualization. Virtualization separates the tight coupling between the hardware and the operating system using an abstraction layer called a hypervisor. Hypervisor emulates all the functions of a computer in the virtual machine, it can run multiple virtual machines at the same time and each virtual machine can run any operating system such as Windows or Linux. Azure takes this virtualization technique and repeats it on a massive scale in the data center owned by Microsoft. Each data center has many racks filled with servers and each server includes a hypervisor to run multiple virtual machines. The network switch provides connectivity to all those servers.

Azure will provide the Microsoft Azure is a cloud computing platform which offers

* Infrastructure as a service (IaaS).
* Platform as a service (PaaS).
* Software as a service (SaaS).

### **Infrastructure as a service (IaaS)**

Virtual machines, storage, and networking will come under the category of infrastructure as a service but the users have to do manually the build and deploy of the applications. Azure will support a wide range of operating systems because of its Hyper-hypervisor.

### **Platform as a service (PaaS)**

Azure app service, Azure functions, and logic apps are some services that are offered by Azure under the platform as a service. This service will provide autoscaling and load balancing and also there will be a pre-configured environment for the application.

### **Software as a service (SaaS)**

Office 365, Dynamics 365, and Azure Active Directory are some of the services provided by Microsoft Azure under Software as a Service (SaaS) the complete application will be managed by the Microsoft azure including deploying, scaling and load balancing.

**What Is Microsoft Azure Used For?**

Following are the some the use cases that Microsoft Azure Used.

**1)Deployment Of applications:**You can develop and deploy the application in the azure cloud by using the service called Azure App Service and Azure Functions after deploying the applications end users can access it.

**2)Identity and Access Management:**The application and data which is deployed and stored in the Microsoft Azure can be secured with the help of Identity and Access Management. It’s commonly used for single sign-on, multi-factor authentication, and identity governance.

**3)Data Storage and Databases:**You can store the data in Microsoft azure in service like blob storage for unstructured data, table storage for NoSQL data, file storage, and Azure SQL Database for relational databases. The service can be scaled depending on the amount of data we are getting.

**4)DevOps and Continuous Integration/Continuous Deployment (CI/CD):**Azure DevOps will provide some tools like ncluding version control, build automation, release management, and application monitoring.

**Databricks in Azure**

Azure Databricks is a data analytics platform optimized for the Microsoft Azure cloud services platform. Azure Databricks offers three environments:

* Databricks SQL
* Databricks data science and engineering
* Databricks machine learning

### **Databricks SQL**

Databricks SQL provides a user-friendly platform. This helps analysts, who work on SQL queries, to run queries on [**Azure Data Lake**](https://intellipaat.com/blog/what-is-azure-data-lake/), create multiple virtualizations, and build and share dashboards.

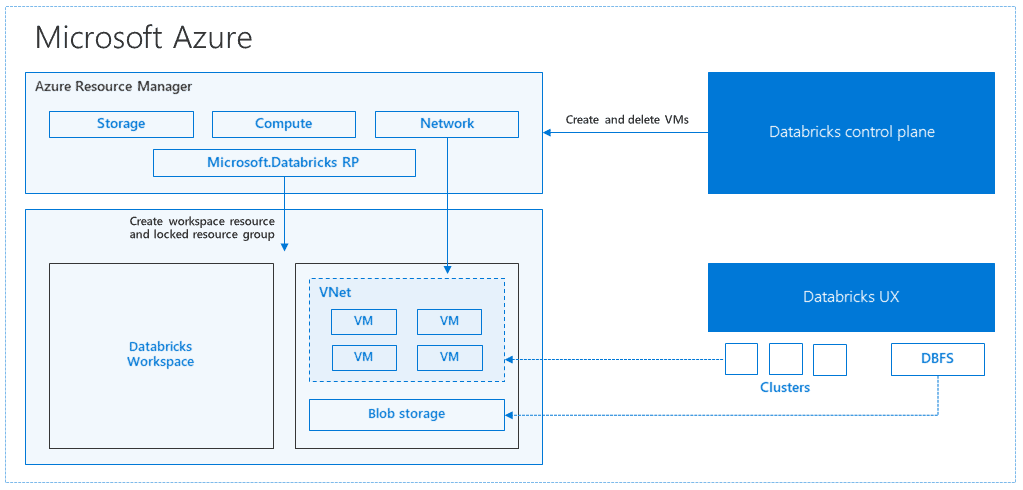
### **Databricks Data Science and Engineering**

Databricks data science and engineering provide an interactive working environment for data engineers, data scientists, and machine learning engineers. The two ways to send data through the big data pipeline are:

* Ingest into Azure through [Azure Data Factory](https://intellipaat.com/blog/what-is-azure-data-factory/) in batches
* Stream real-time by using Apache Kafka, Event Hubs, or IoT Hub

# **Databricks architecture**

**Azure Databricks** is an easy, fast, and collaborative Apache spark-based data analytics platform for the Microsoft Azure cloud services platform. It accelerates innovation by bringing data science, [data engineering](https://k21academy.com/microsoft-azure/data-engineer/data-engineering-on-microsoft-azure/) and business together. It makes the process of data analytics more productive more secure more scalable and optimized for Azure.



1)When we launch a cluster via Databricks, a “Databricks appliance” is deployed as an Azure resource in our subscription.

2)Then we specify the types of VMs to use and how many, but Databricks handle all other elements.

3)A managed resource group is deployed into the subscription that we populate with a VNet, a storage account, and a security group.

4)Once these services are ready, we will control the Databricks cluster over the Databricks UI.

Data bricks Azure Workspace is an analytics platform based on Apache Spark.

For the big data pipeline, the data is ingested into Azure using [Azure Data Factory](https://k21academy.com/microsoft-azure/azure-data-factory/). This data lands in a data lake and for analytics, we use Databricks to read data from multiple data sources and turn it into breakthrough insights.

**Why Azure Databricks ?**

**1)Optimized Environment**

* Databricks Azure was optimized automatically from the ground up for cost-efficiency and performance in the cloud.
* Auto-scaling and auto-termination of Spark clusters, no doubt it minimizes costs automatically.
* Optimizations including indexing, caching, and advanced query optimization, which can enhance performance by as much as 10-100x over conventional Apache Spark deployments in the cloud.

**2)** **Persistent collaboration**

* Notebooks on Databricks are live and easy to share, with real-time teamwork.
* Dashboards allow business users to call a current job with new parameters.
* Databricks integrates closely with PowerBI for hand-on visualization.

**3) Simple to use**

* Azure Databricks comes with notebooks that let you run machine learning algorithms, connect to common data sources, and learn the basics of Apache Spark to get started rapidly.
* It also a unified debugging environment features to let you analyze the progress of your Spark jobs from under interactive notebooks, and powerful tools to examine past jobs.
* No need to install common analytics libraries, such as the Python and R data science stacks, which are preinstalled.

**DATABRICKS CLUSTER**

A cluster in Databricks is a group of virtual machines that are configured with Spark/PySpark and has a combination of computation resources and configuration on which your application can run. In a simple way, the cluster executes all of your databricks code.

Workloads that the Databricks cluster can run are ETL pipelines, Machine Learning models, Streaming, Batch analytics, and ad-hoc analytics.

How to make a cluster in azure databricks

