Azure Databricks

Big data analytics and AI with Apache Spark

Insights from all data and AI solution with Azure Databricks. Set up environment in minutes in minutes, autoscale and collaborate. Databricks supports Python, Scala, R, Java and SQL. In addition, data science frameworks, such as, TensorFlow, PyTorch, and Scikit-learn.

* Fast, optimized Apache Spark environment;
* Interactive workspace with built-in support for popular tools, languages, and frameworks;
* machine learning on big data with native Azure Machine Learning integration;
* High-performance modern data warehousing with Synapse

**Data Engineering Light**

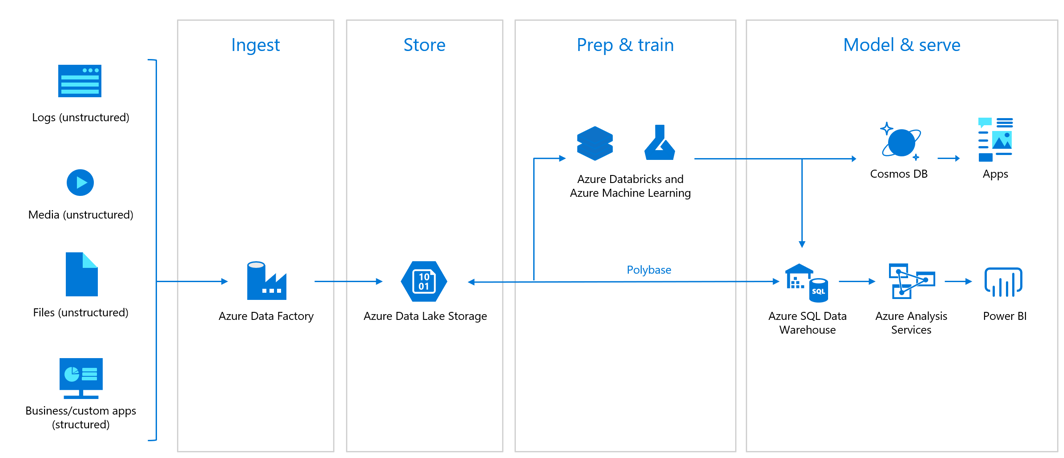
A new low-priced workload called Data Engineering Light that enables customers to run batch applications on managed Apache Spark. It is meant for simple, non-critical workloads that don’t need the performance, autoscaling, and Data Analytics workloads.

**Managed MLflow**

MLflow is an open source framework for managing the machine learning lifecycle. You can access it natively from the Azure Databricks environment and leverage Azure Active Directory for authentication. Track experiments by automatically recording parameters, results, code, and data; package machine learning code and dependencies locally in a reproducible project format and execute remotely on a Databricks cluster.

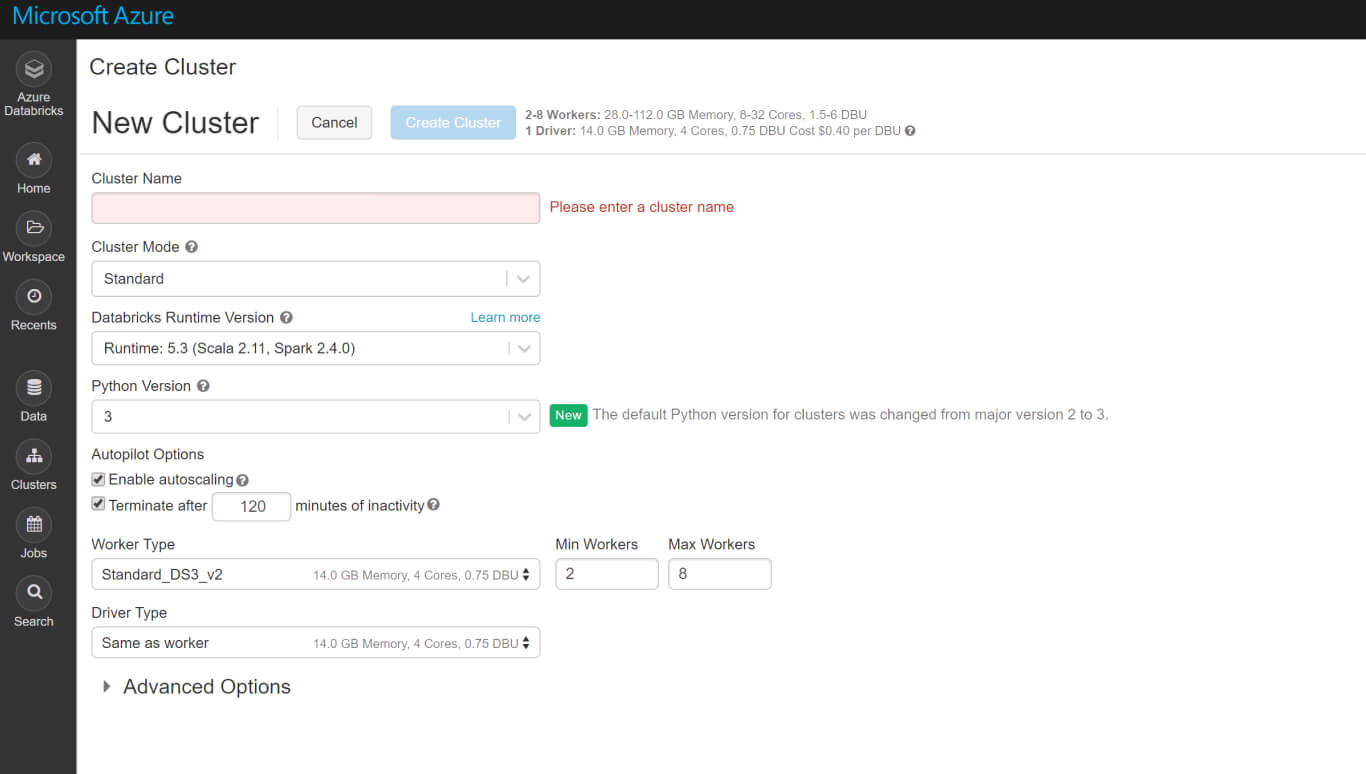
**Azure Machine Learning compliments Azure Databricks**

ML enables data scientists of all skill levels to identify suitable algorithms and hyperparameters faster. Enables DevOps for machine learning for simplified management, monitoring, and updating of machine learning models; deploy models from the cloud and the edge; a central registry for experiments, machine learning pipelines, and models at the enterprise level.

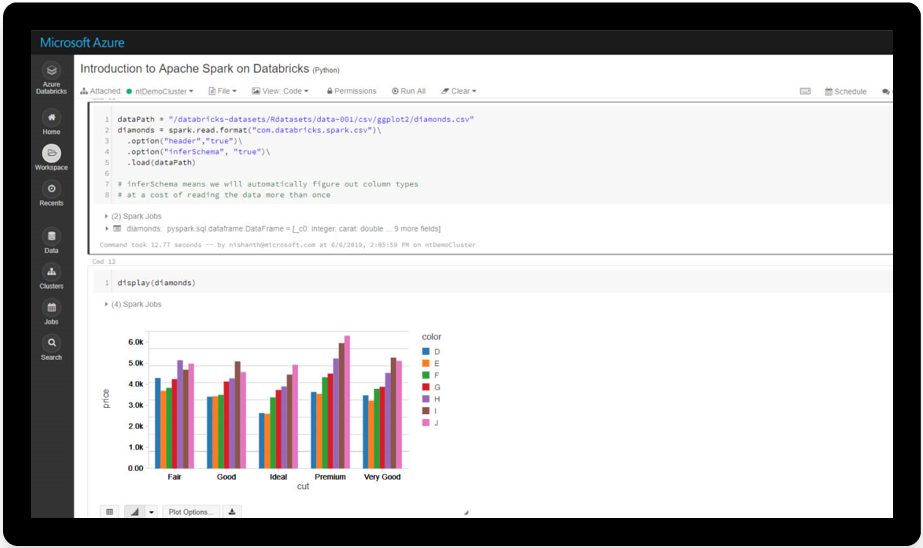


*Azure Databricks at a glance*

## Start quickly with an optimized Apache Spark environment

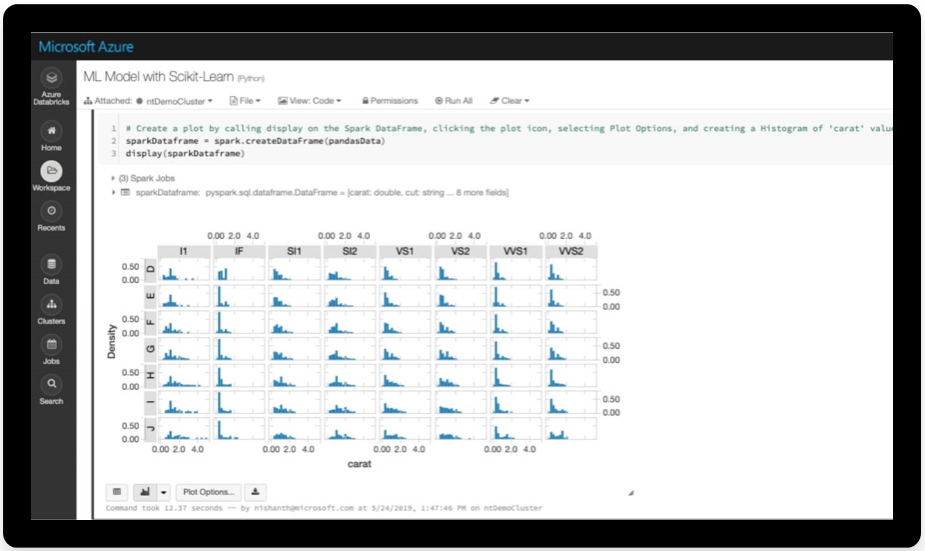
Azure Databricks provides the latest versions of Apache Spark to seamlessly integrate with open source libraries. Set up clusters quickly in a fully managed Apache Spark environment with global scale and availability. Optimized clusters are set up, configured, and fine-tuned to ensure reliability and performance with no need for monitoring. Autoscaling and auto-termination improves the total cost of ownership (TCO).

**Shared workspace and common languages**



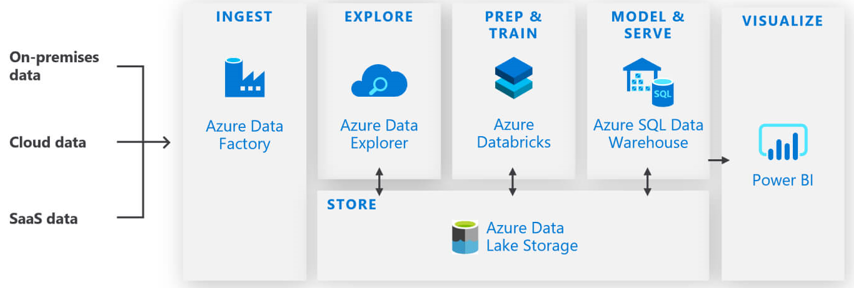
Collaborate on shared projects. Language choices are Python, Scala, R, and SQL. Get easy version control of notebooks with GitHub and Azure DevOps.

**Quick machine learning on big data**



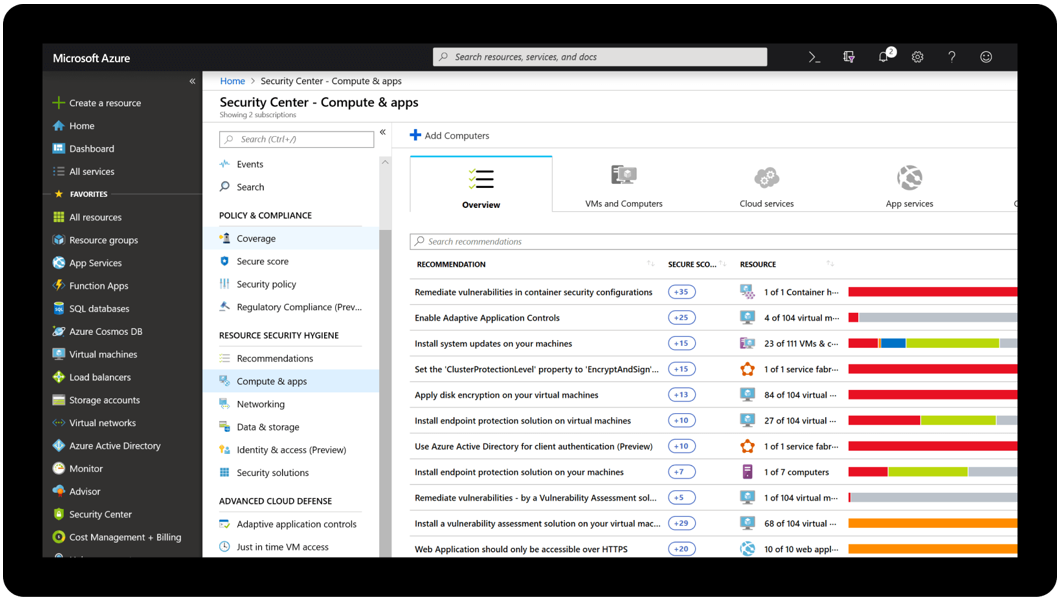
Advanced automated machine learning capabilities using the integrated Azure Machine Learning to quickly identify suitable algorithms and hyperparameters. Simplify management, monitoring, and updating of machine learning models deployed from the cloud to the edge. And a central registry for experiments, machine learning pipelines, and models.

**Modern data warehousing**



Data warehouse in the cloud for unmatched performance and scalability. Combine data at any scale, and get insights through analytical dashboards and operational reports. Automate data movement using Azure Data Factory, load data into ADLS Gen 2, transform and cleanse using Azure Databricks.

**Security and compliance**



Native integration with Azure Active Directory for role-based access control. Secure architectures without compromising on compliance using configurable virtual networks. Fine-grained user permissions for Azure Databricks notebooks, clusters, jobs, and data.

[Delta Lake](https://delta.io/) is an [open source storage layer](https://github.com/delta-io/delta) that brings reliability to data lakes. Delta Lake provides **ACID transactions, scalable metadata handling, and unifies streaming and batch data processing**. Delta Lake runs on top of your existing data lake and is fully compatible with Apache Spark APIs.

Delta Lake on Azure Databricks allows you to configure Delta Lake based on your workload patterns and provides optimized layouts and indexes for fast interactive queries.