

Microsoft Azure & Databricks = Cloud-Scale Spark Power

Colby Ford



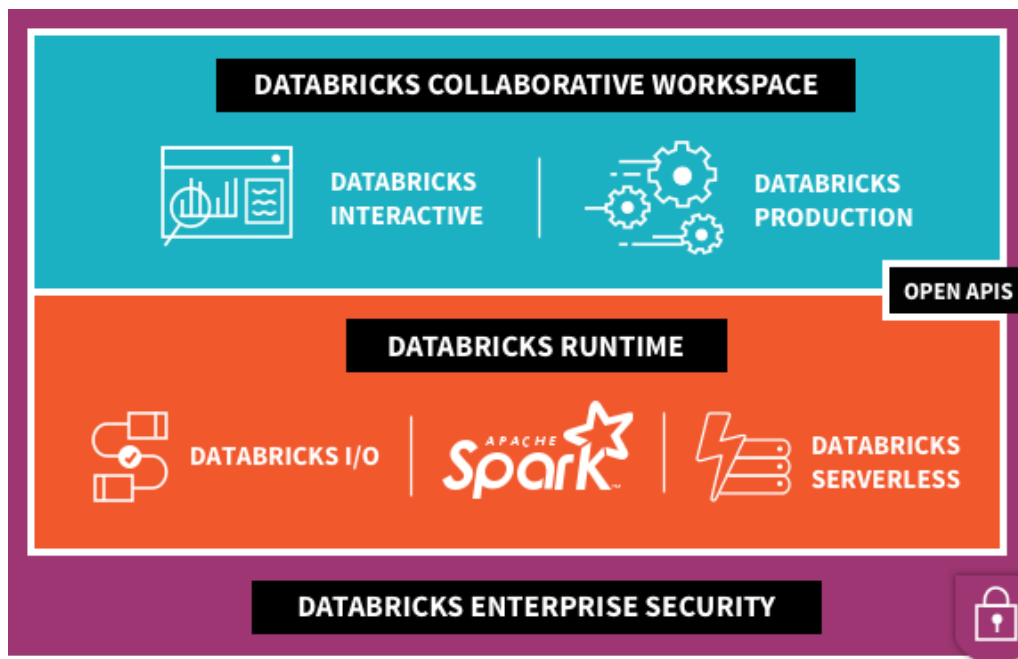
Recently, Microsoft and [Databricks](#) made an exciting announcement around their partnership that will soon result in a cloud-based, managed Spark service on [Azure](#). Currently, some select customers are allowed into a "private preview" mode of the service, and over the next few weeks, a "gated public preview" will ensue for around 150 clients. In January 2018, the service will be available for everyone to try. While the full details are not known about the partnership or full features of the platform, here is how Azure Databricks will likely enhance your Big Data capabilities in the cloud.

What is Databricks?

Databricks is a company that was started by the team that originally created Spark at UC Berkeley. They have created a [Unified Analytics Platform](#) that aims to be the single system for everything from analytics workflows to Spark integration to security.

Databricks boasts various benefits of their Unified Analytics Platform such as:

- **UNIFY ANALYTICS WITH APACHE SPARK** - Eliminate the need for disparate tools.
- **STREAMLINE ANALYTIC WORKFLOWS** - Reduce deployment time to minutes.
- **INCREASE PRODUCTIVITY OF DATA SCIENCE TEAMS** - With Databricks, they'll be 5x more productive.
- **REDUCE RISK** - Enable innovation with out-of-the-box enterprise security and compliance.



[[Source](#)]

Apache® Spark™ on Databricks is said to have a 5x performance gain over that of the open-source version.

Looking at the [Databricks' Feature Comparison](#) page, there are quite a few features that could likely make it into the Azure version in the near future.

CLOUD OPTIMIZATION: <ul style="list-style-type: none">• Tuned Apache® Spark™ clusters• High availability for Spark Streaming• Built-in file system COST MANAGEMENT: <ul style="list-style-type: none">• Autoscaling Apache Spark clusters• Multi-user cluster sharing BUILT-IN EXPLORATION TOOLS: <ul style="list-style-type: none">• Notebooks with real-time collaboration + revision history• Publish notebooks as production dashboards	BUILT-IN PRODUCTION TOOLS: <ul style="list-style-type: none">• Spark job monitoring alerts• One-click deployment from notebooks to Spark Jobs• APIs to build workflows in notebooks SECURITY: <ul style="list-style-type: none">• Access control for clusters and notebooks• Permission-based job and workflow execution• Authenticated SQL server
--	--

Expect a Familiar Azure Experience

Azure already has a managed Hadoop™ offering known as HDInsight. You can spin up a custom HDInsight cluster with your specifications from the Portal. Then, you pay for the time that you have your cluster running. Support for HDInsight is provided by the Microsoft Azure support team.

As for [Azure Databricks](#), the experience will be very similar. Simply spin up an Azure Databricks cluster directly from the Portal and Azure will do the setup work for you. No licensing is required other than your Azure subscription.

For support, Microsoft and Databricks will have a seamless system for users to get help with their individual needs. Since the service is within Azure, you will go through Microsoft for support, which will now be fully integrated with the Databricks expert support team.

Want to learn more about how you can take advantage of this exciting announcement at your

organization? [Contact BlueGranite!](#)



About The Author

Dr. Colby Ford is the Principal of Life Sciences at BlueGranite. Coming from a background in mathematics, data science, and computational biology, he combines this expertise to architect scalable solutions in the Azure cloud. Using R, Python, and Spark, he puts AI and bioinformatics to work to provide valuable insights from data. Outside of BlueGranite, Colby is an avid researcher in infectious diseases and human genomics. Check out Colby's website at www.colbyford.com.