

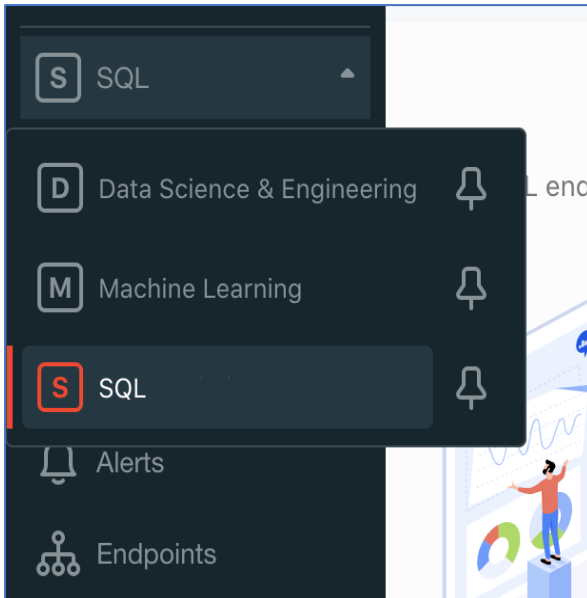


Azure Databricks

Azure Databricks Nedir?

Databricks, Apache Spark ile çalışmak için otomatik cluster yönetimi ve iPython tarzı notebooklar sağlayan web tabanlı, işbirliğine dayalı bir veri bilimi, veri mühendisliği ve veri analizi platformudur. Azure Databricks, bulut için optimize edilmiş uçtan-uca yönetilen bir Apache Spark platformu olan Databricks'in gücünü Microsoft Azure platformunun kurumsal ölçeği ve güvenliğiyle birleştirerek, büyük ölçekli Spark iş yüklerini çalıştırmayı kolaylaştırır.

Azure Databricks birçok Azure servisi ile sıkıca entegredir. Örneğin; Azure Data Lake Storage, Azure Data Factory, Azure Machine Learning, Azure Synapse Analytics, Power BI vb. Bu sayede tüm verilerinizi basitçe bir veri gölünde depolayabilir ve tüm analiz veya yapay zeka iş yüklerinizi birleştirebilirsiniz.



Azure Databricks yoğun veri içeren uygulamalar geliştirmek için 3 ortam sunar:

Databricks SQL veri göllerinde SQL sorguları çalıştırmak, birden çok görselleştirme türü oluşturmak ve panolar oluşturup paylaşmak isteyen analistler için kullanımı kolay bir platform sağlar .

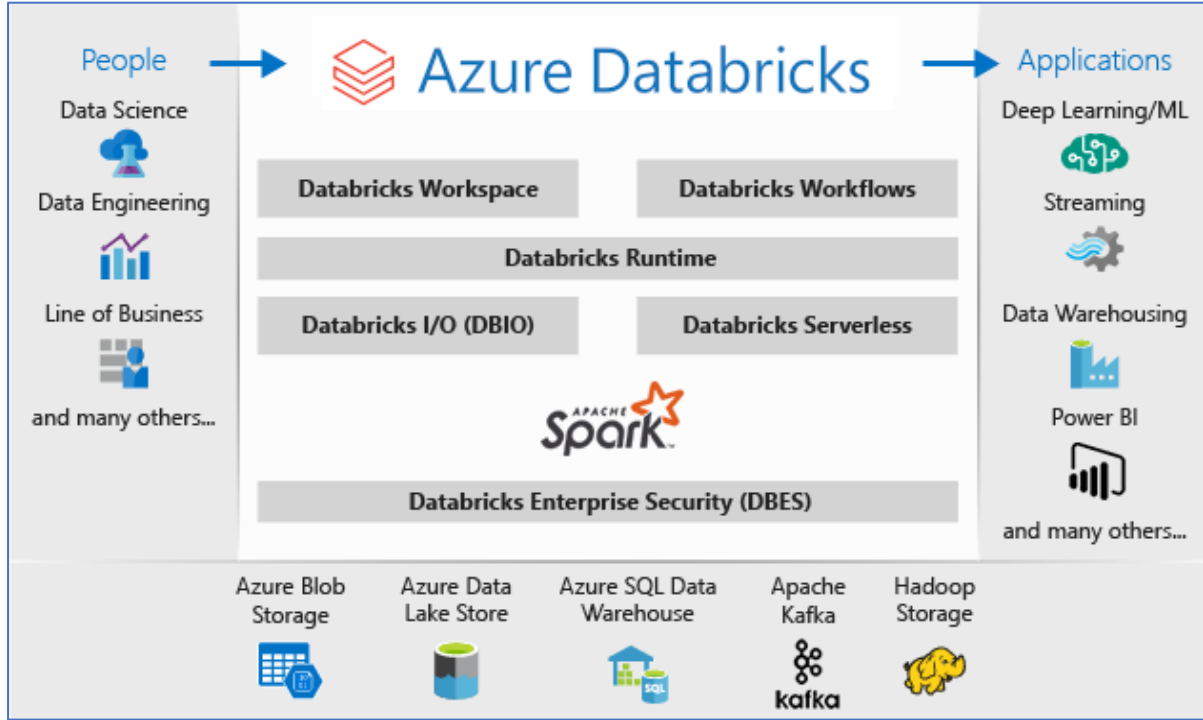
Databricks Data Science & Engineering, veri mühendisleri, veri bilimcileri ve makine öğrenmesi mühendislerinin işbirliği yapabileceği bir çalışma alanı sağlar. Bir big data pipeline'ında, veriler(ham veya

yapılandırılmış) Azure Data Factory ile toplu olarak veya Apache Kafka, Event Hub ya da IoT Hub ile neredeyse gerçek zamanlı akış olarak Azure'a alınabilir. Bu veriler, uzun süreli kalıcı depolama için veri gölüne gelir, Azure Blob Storage veya Azure Data Lake Storage kalıcı depolama için kullanılabilir. Birden çok veri kaynağından veri okumak ve bu verileri Spark'ın gücünden faydalanarak analiz etmek için Azure Databricks'i kullanabilirsiniz.

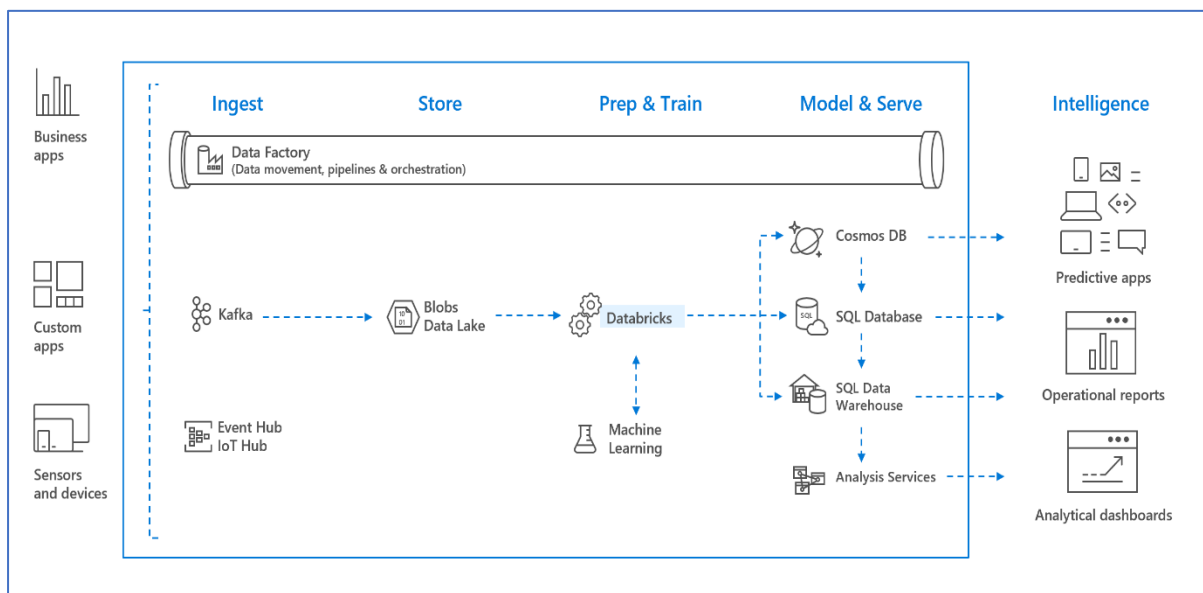
Databricks Machine Learning, experiment takibi, model eğitimi, özellik geliştirme ve yönetimi ile özellik ve model sunumu için yönetilen hizmetler içeren, uçtan-uca entegre bir makine öğrenimi ortamıdır.

Databricks Data Science & Engineering

Bu yazıda Databricks Data Science & Engineering (kısaca “Workspace” olarak da isimlendirilir) ortamına odaklanacağız. Databricks Data Science & Engineering Apache Spark’a dayalı bir analiz platformudur. Tek-tıkla kurulum, modern iş akışları ve etkileşimli bir çalışma ortamı sunmak için Azure ile entegre edilmiştir.



Analiz iş akışınızın bir parçası olarak Azure Blob Storage, Azure Data Lake Storage, Azure Cosmos DB veya Azure SQL Data Warehouse gibi birçok veri kaynağından veri okumak için Azure Databricks’i kullanabilir ve Spark ile verilerinizden ileri düzeyde bilgiler elde edebilirsiniz.



Azure Databricks'te Apache Spark

Emirhan Gürel's Cluster

☐ Multi node ☒ Single node

Access mode [?](#) Single user access [?](#)

Single user [?](#) Emirhan Systems [?](#)

Performance

Databricks runtime version [?](#)

Runtime: 10.4 LTS (Scala 2.12, Spark 3.2.1) [?](#)

☐ Use Photon Acceleration [?](#)

Node type [?](#)

Standard_DS3_v2 14 GB Memory, 4 Cores [?](#)

☒ Terminate after 120 minutes of inactivity [?](#)

Azure Databricks, Apache Spark'ın özelliklerini temel alır. Saniyeler içinde clusterlar oluşturabilir, dinamik olarak oluşturduğunuz clusterları otomatik ölçeklendirebilir veya takımlar arasında kaynak paylaşımı yapabilir, REST API'larını çağırarak clusterları programlı olarak kullanabilirsiniz.

Databricks Runtime

Databricks Runtime Apache Spark üzerinde oluşturulmuştur ve Azure platformu için inşa edilmiştir. Altyapı karmaşıklığını tamamen ortadan kaldırır. Azure Databricks üretim işlerinin performansını önemseyen veri mühendisleri için I/O katmanında ve işleme katmanında çeşitli optimizasyonlar yapar ve bu sayede daha hızlı ve performanslı bir Spark motoru sağlar.

Databricks runtime version [?](#)

Runtime: 10.4 LTS (Scala 2.12, Spark 3.2.1) [?](#)

Standard	>	11.1	Scala 2.12, Spark 3.3.0
ML	>	11.0	Scala 2.12, Spark 3.3.0
		10.5	Scala 2.12, Spark 3.2.1
		10.4 LTS	Scala 2.12, Spark 3.2.1
		9.1 LTS	Scala 2.12, Spark 3.1.2
		7.3 LTS	Scala 2.12, Spark 3.0.1

Workspace

Microsoft Azure | Databricks

databricks

Data Science & Engi... [?](#)

Create

Workspace

Workspace

Users [?](#)

emirhan.gurel [?](#)

emirhan.gurel [?](#)

Trash [?](#)

11-Delta-Lake-Architecture [?](#)

Home [?](#)

Databricks Data Science & Engineering, işbirlikçi ve entegre bir ortam aracılığıyla, Spark'ta veri keşfetme, prototip oluşturma ve veriye dayalı uygulamaları çalıştırma sürecini oldukça kolaylaştırır. İlerlemelerinizi R, Python, Scala ve SQL notebookları ile belgeleyebilir, verileri birkaç tıklamayla Matplotlib, ggplot veya d3 gibi araçlarla görselleştirebilir, dinamik raporlar oluşturmak için panolar kullanabilirsiniz.

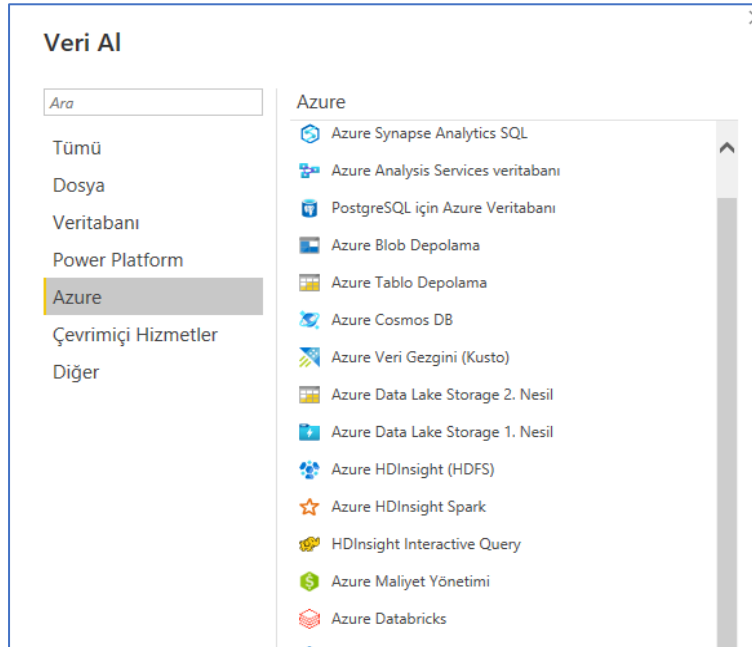
Kurumsal Güvenlik

Azure Databricks, kurumsal düzeyde Azure güvenliği sağlar. İşinizi ve verilerinizi korumak için Azure Active Directory entegrasyonu, rol tabanlı kontroller ve SLA'lar içerir.

Azure AD entegrasyonu, Azure Databricks kullanarak eksiksiz, Azure tabanlı çözümler çalıştırmanıza olanak tanır. Rol tabanlı erişim notebooklar, clusterlar, işler ve veriler için ayrıntılı kullanıcı izinleri ayarlamanıza izin verir..

Azure Servisleri ve Power BI ile Entegrasyon

Databricks Data Science & Engineering Azure veri tabanları ve depolamaları ile derin bir şekilde entegredir. Synapse Analytics, Cosmos DB, Data Lake Store ve Blob Storage Databricks'in entegre olduğu başlıca Azure servisleri arasındadır.



Power BI ile olan zengin entegrasyon sayesinde, Databricks Data Science & Engineering verilerinizden elde ettiğiniz çıkarımları etkili, hızlı ve kolay bir şekilde paylaşmanıza olanak tanır. Ayrıca diğer BI yazılımlarını da kullanabilirsiniz.

Azure Databricks ilginizi çektiyse veya analiz işleriniz için Azure Databricks kullanmayı düşünüyorsanız daha detaylı bilgi için Azure Databricks dökümanlarına [göz atabilirsiniz.](#)