

Introdução ao Databricks



databricks™

Making Big Data Simple



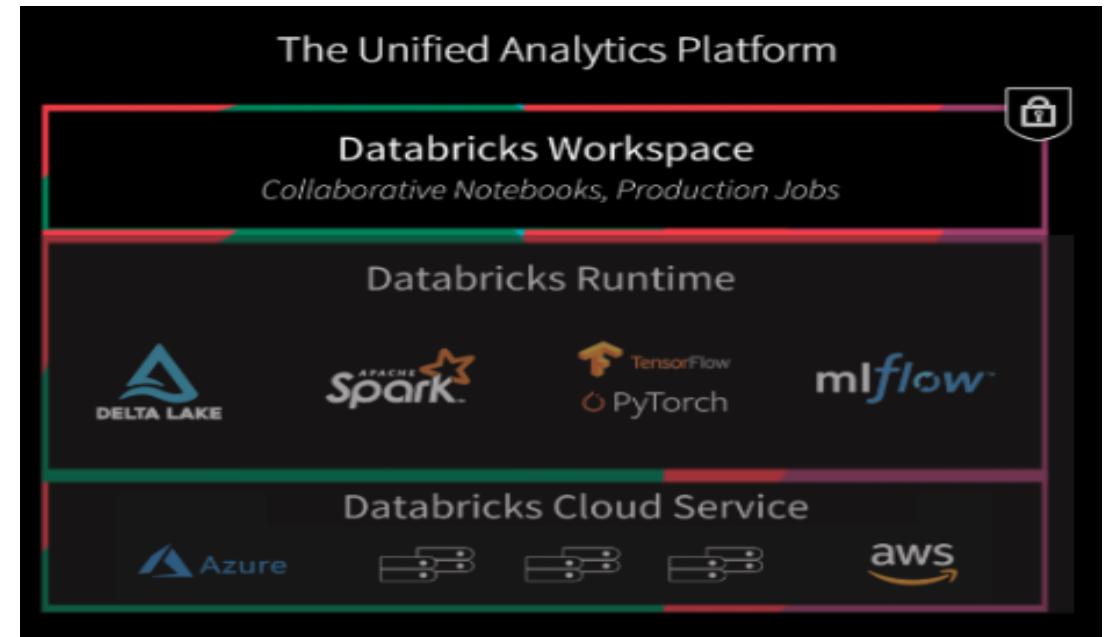
<What is Databricks?/>

Unifying Data Science and Engineering

Databricks Unified Analytics Platform, from the original creators of Apache Spark™, unifies data science and engineering across the Machine Learning lifecycle from data preparation, to experimentation and deployment of ML applications.

Data Engineering

Speed up the preparation of high quality data, essential for best-in-class ML applications, at scale.



Data Science

Collaboratively explore large datasets, build models iteratively and deploy across multiple platforms.

<Why use Databricks?/>

Databricks Just-in-Time Data Platform

Powered by Apache® Spark™

Integrated Workspace

Notebooks

Dashboards

BI Tools

Qlik

+ a b l e a u

Your Custom Spark Apps

Production Jobs

Orchestrated Apache® Spark™ in the Cloud

Open Source  +  databricks™ Managed Services

Your Storage



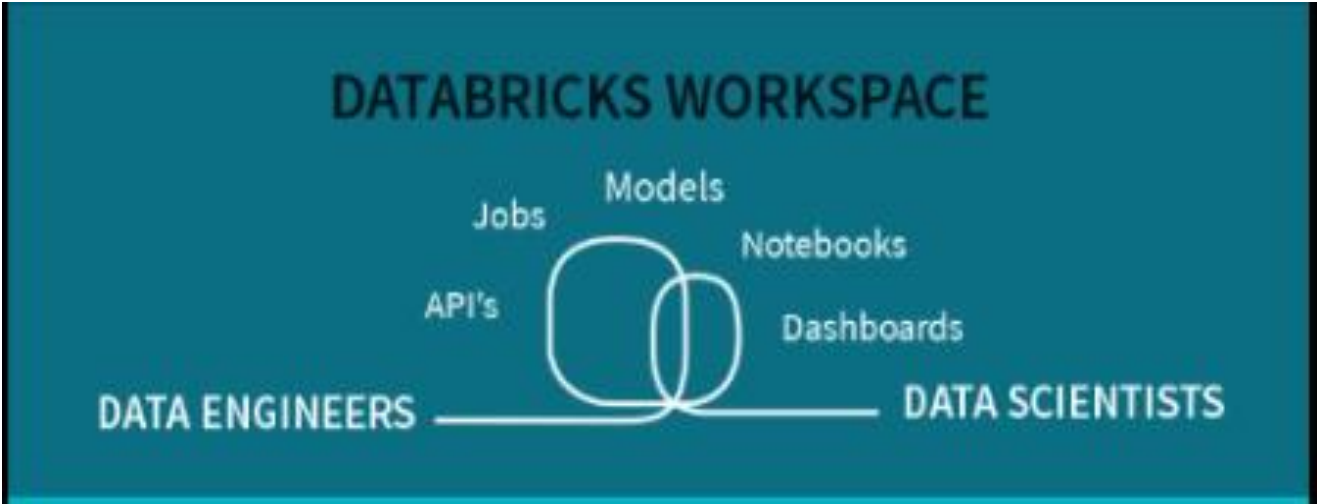
Cloud Storage | Data Warehouses | Data Lakes



<Databricks Workspace/>

Local onde se armazena as pastas, notebooks, bibliotecas e dashboards.

- Notebooks**
Blocos de códigos com Integração com o GitHub (Revision History)
- Ecosystema integrado**
Conectividade com diversas ferramentas, facilidade na importação de bibliotecas e frameworks.
- Workspace Access Control**
Controle de acesso a nível de Notebooks e Pastas.



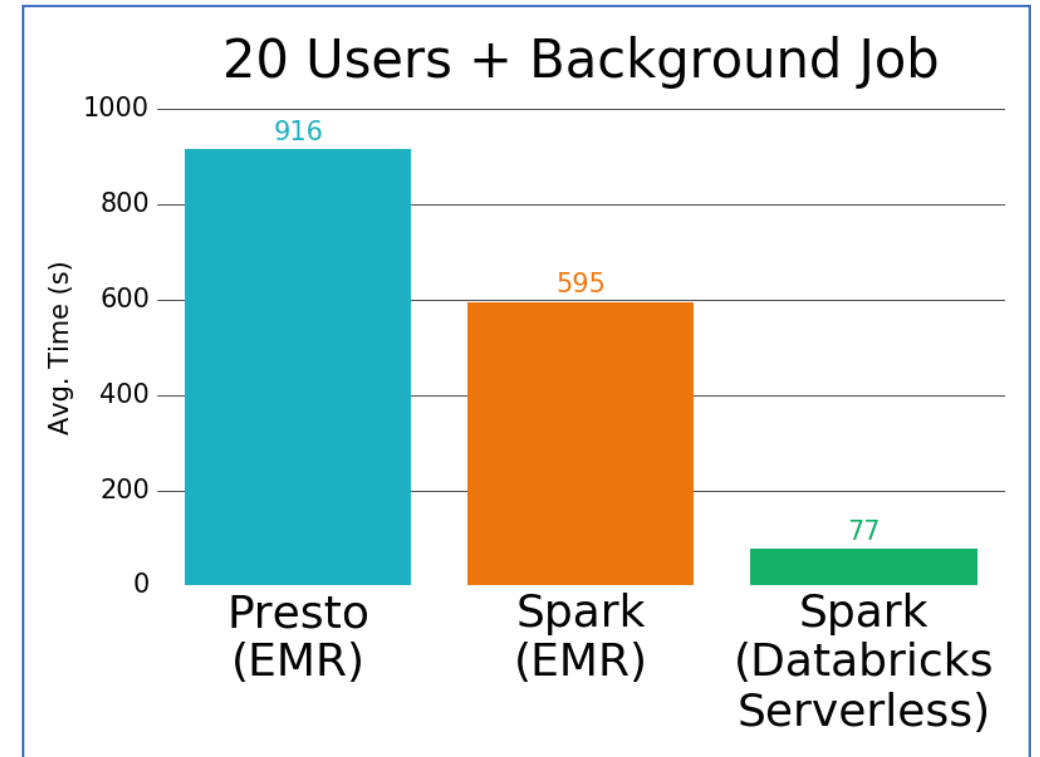
Notebook permissions

Ability	No Permissions	Read	Run	Edit	Manage
View cells		x	x	x	x
Comment		x	x	x	x
Attach and detach notebooks			x	x	x
Run commands			x	x	x
Edit cells				x	x
Change permissions					x

<Databricks Runtime/>

Otimização e simplificação dos processos de Spark

- **Machine Learning at Massive Scale**
Aplicação de Machine Learning em dados de grande escala tanto em processos Batch como Streaming.
- **Easy Scheduling**
Facilidade no agendamento de tarefas, disparo de alertas e criação de pipelines
- **Quick release cycle**
Acesso rápido às novas features e correções de bugs
- **Concorrência / Jobs Complexos**
Performance mais rápida para casos de concorrência / Jobs Complexos sendo processados simultaneamente



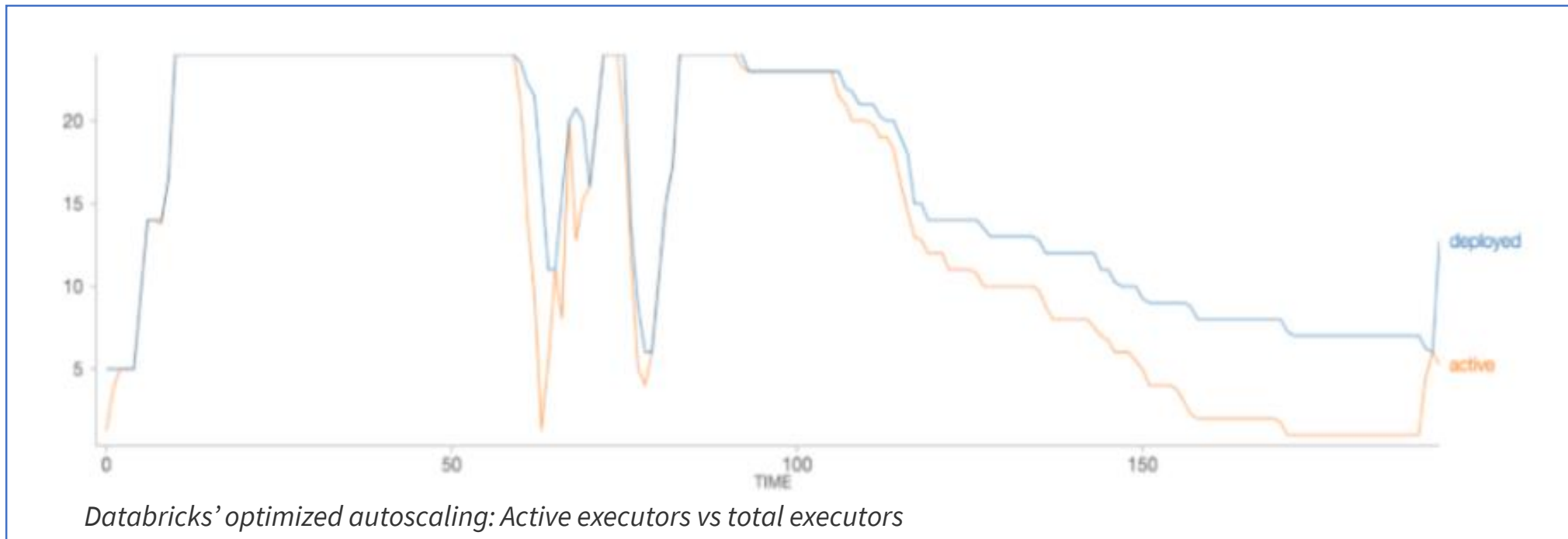
<Databricks Cloud Service/>

Redução da necessidade de infraestrutura / DevOps simplificado

Auto-config: Configuração automática do Spark para as aplicações e distribuição dos recursos do cluster.

Auto-scaling: Adequa o número de executores à necessidade da aplicação de acordo com o range fornecido pelo usuário.

Logging and Monitoring: Acesso ao Spark UI e Spark Driver Logs



<So Azure Databricks is.../>

Plataforma de Análise baseada no Apache Spark
otimizada para a plataforma de serviços de nuvem do Microsoft Azure

