



Microsoft  
**AI+Tour**  
LATAM





BAM! Llevando Machine Learning al siguiente nivel con Azure Databricks

# TRACK #2: Científico de Datos

Para escuchar esta charla, por favor sintonizar el siguiente canal en sus auriculares

**Canal #2** Audio original

**Canal #4** Traducción inglés/español



Microsoft  
**AI+Tour**  
Argentina

# BAM! Llevando Machine Learning al siguiente nivel con Azure Databricks

Facundo Santiago

Arquitecto de Soluciones

Analítica Avanzada e Inteligencia Artificial

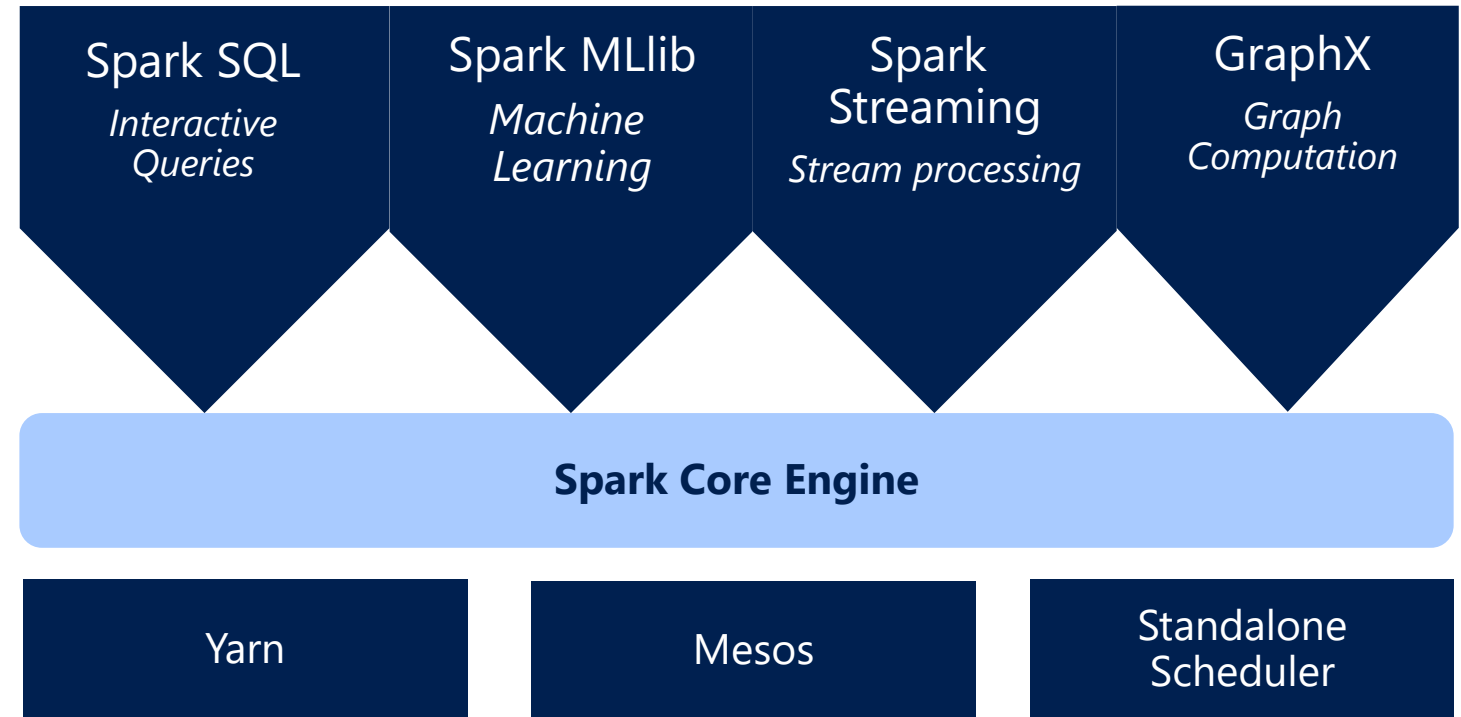
[Medium.com/@santiagof](https://medium.com/@santiagof)



# Azure Databricks for Machine Learning

## Apache Spark

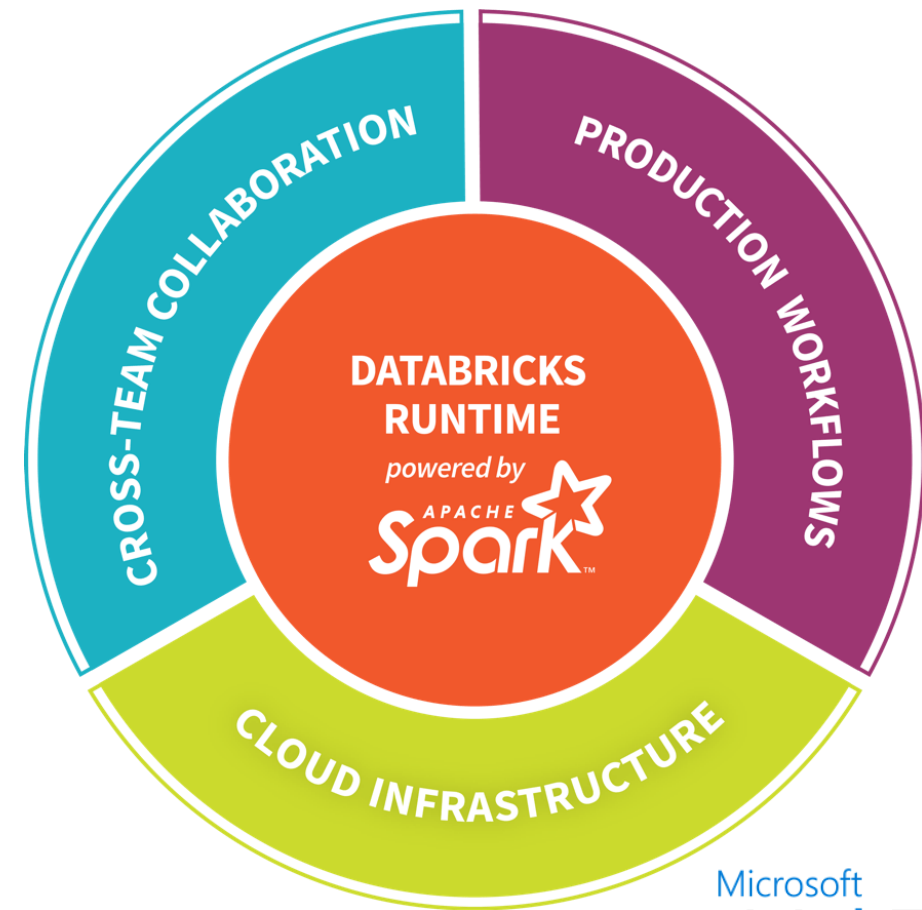
Motor de procesamiento de datos de código abierto unificado y paralelo



# Azure Databricks for Machine Learning

## Databricks

- Founded in late 2013
- By the creators of Apache Spark, original team from UC Berkeley AMPLab
- Largest code contributor code to Apache Spark
- Level 2/3 support partnership with
  - Hortonworks
  - MapR
  - DataStax
- Provides [certifications](#) such as Databricks Certified Application, Databricks Certified Distribution and Databricks Certified Developer
- Main Product: The [Unified Analytics Platform](#)
- In Oct 2017, introduced [Databricks Delta](#).



# Azure Databricks for Machine Learning

## Databricks integrado en Azure

- Azure Portal
- Azure Storage Services
- Azure Active Directory
- Azure SQL DW and Azure Cosmos DB
- Apache Kafka for HDInsight
- Azure Billing
- Azure Power BI





# ¿Y Machine Learning?



# Azure Databricks

## Librerías de ML y DL (hay un par más)



Microsoft Machine Learning  
for Apache Spark





# Azure Databricks

# Machine Learning and Deep Learning



<b>Clasificación y regresión</b>	<ul style="list-style-type: none"> <li>Modelos lineales (SVMs, regresión logística, regresión lineal)</li> <li>Naïve Bayes</li> <li>Árboles de decisión</li> <li>Ensamblados de árboles (bosque aleatorio, árboles impulsados por gradiente)</li> <li>Regresión isotónica</li> </ul>
<b>Agrupación en clústers</b>	<ul style="list-style-type: none"> <li>Medios K y streaming de medios K</li> <li>Mezcla gaussiana</li> <li>Agrupación en clústers de interacción de potencia (PIC)</li> <li>Asignación de Dirichlet latente (LDA)</li> </ul>
<b>Filtrado colaborativo</b>	<ul style="list-style-type: none"> <li>Alternar mínimos cuadrados (ALS)</li> </ul>
<b>Reducción de dimensionalidad</b>	<ul style="list-style-type: none"> <li>SVD</li> <li>PCA</li> </ul>
<b>Minería de patrones frecuentes</b>	<ul style="list-style-type: none"> <li>Crecimiento FP</li> <li>Reglas de asociación</li> </ul>
<b>Estadísticas básicas</b>	<ul style="list-style-type: none"> <li>Estadísticas de resumen</li> <li>Correlaciones</li> <li>Muestreo estratificado</li> <li>Pruebas de hipótesis</li> <li>Generación de datos aleatorios</li> </ul>



Hands-on

Microsoft  
AI+Tour  
LATAM



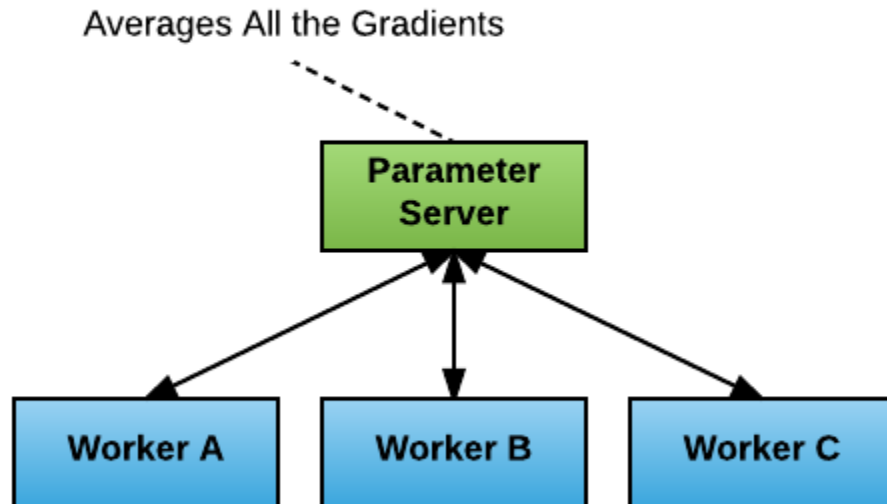
1	16:40
2	17:30
3	23:00
4	09:00
5	

23 5499  
46 6308  
23° 12' 56" S  
46° 38' 20" O

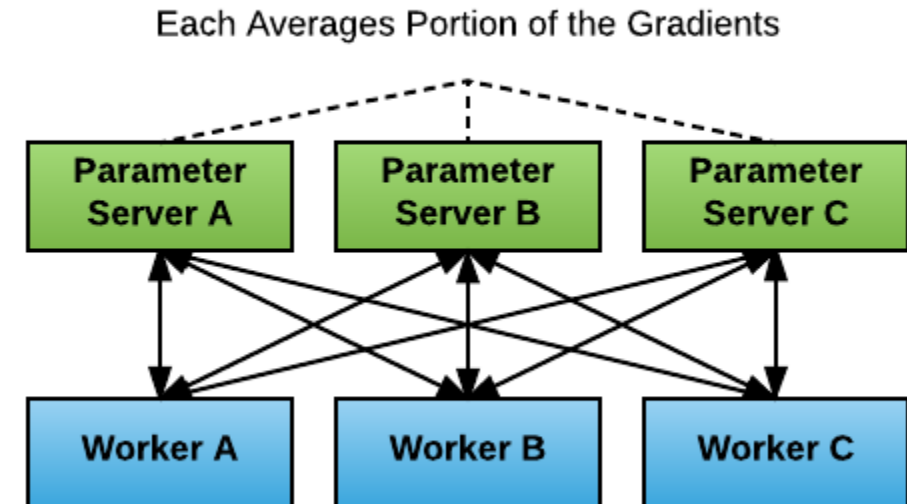
1	16:40
2	17:30
3	23:00
4	09:00
5	



# Azure Databricks Machine Learning and Deep Learning



or





Hands-on

Microsoft  
AI+Tour  
LATAM



1	16:40
2	---
3	17:30
4	23:00
5	09:00

23 5499  
46 6308  
23° 12' 56" S  
46° 38' 20" O



# Azure Databricks

## Machine Learning and Deep Learning



Microsoft Machine Learning  
for Apache Spark



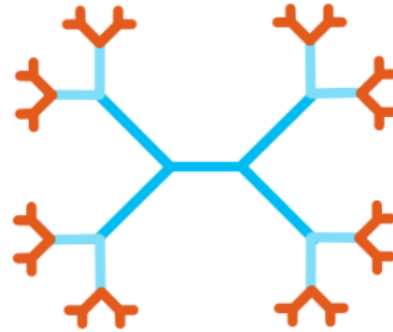
Vowpal Wabbit on Spark

Fast, Sparse, and Effective  
Text Analytics



The Cognitive Services on  
Spark

Leverage the Microsoft  
Cognitive Services at  
Unprecedented Scales in  
your existing SparkML  
pipelines



LightGBM on Spark

Train Gradient Boosted  
Machines with LightGBM



Spark Serving

Serve any Spark  
Computation as a Web  
Service with Sub-  
Millisecond Latency



Hands-on

Microsoft  
AI+Tour  
LATAM



1	16:40
2	17:30
3	23:00
4	09:00
5	

23 5499  
46 6308  
23° 12' 56" S  
46° 38' 20" O



Muchas gracias  
¿Preguntas?





¿Quieres tu LinkedIn Badge  
de Científico de Datos  
Participando en el  
AI+ Tour **Argentina** ?

Conseguilo ahora en:  
**<https://aka.ms/ARDS>**

