


A complex network graph with numerous blue and white nodes connected by thin lines, serving as a background for the title.

# Big Data analytics con Apache **Spark**

*Live Demo de Spark + Cassandra + Zeppelin*

# Sobre mí



 @Neuw84  
@IK4\_IKERLAN



Ángel Conde Manjón

DevOps y DataEngineer en  
resolviendo los desafíos de la Industria 4.0.  
relacionados con Big Data.



*Mantenimiento  
Predictivo*

*Monitorización  
en Tiempo Real*

*Ingesta de  
Datos*

*Analíticas en  
RT/Batch*

*Plataformas  
Digitales*

# Colaboración en datahack...

---



# ... como escuela de formación Big Data



MÓDULO 1 : DATA  
DISCOVERY



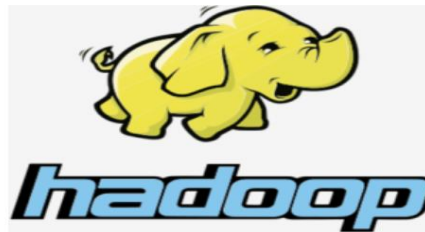
MÓDULO 2 :  
MACHINE LEARNING  
CON PYTHON



MÓDULO 3:  
MACHINE LEARNING  
CON SAS



MÓDULO 4: BASES  
DE DATOS NOSQL



MÓDULO 5: HADOOP  
BASICS



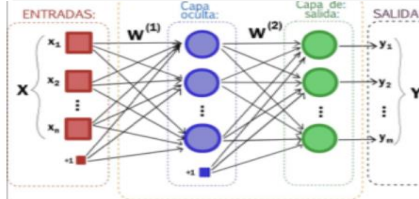
MÓDULO 6:  
ECOSISTEMA  
HADOOP



MÓDULO 7: APACHE  
SPARK



MÓDULO 8:  
VISUALIZACIÓN



MÓDULO 9:  
ALGORITMOS  
AVANZADOS

# Live Demo de Spark y Zeppelin

---



# Analítica



Analíticas Big data en memoria.

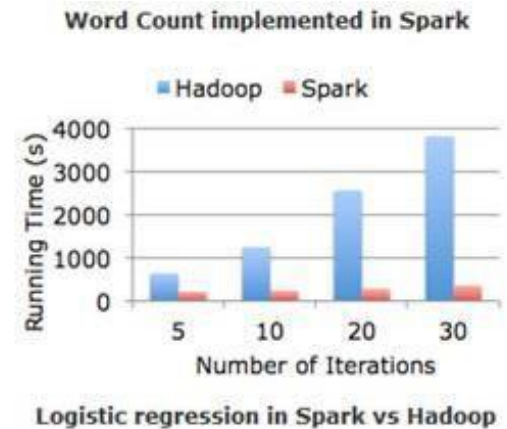
Soporte SQL.

Aprendizaje automático.

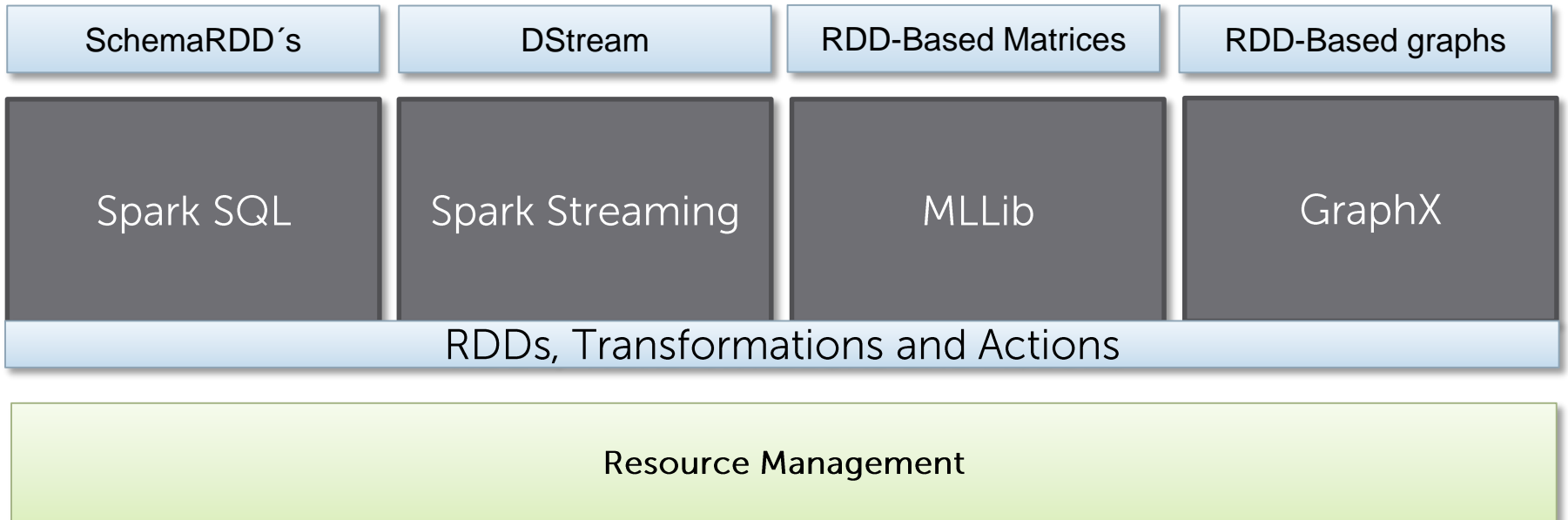
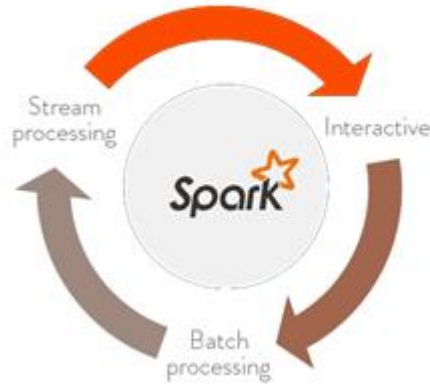
Procesamiento en tiempo real.

Grafos.

Scala/Java/Python/R



# Arquitectura



# Almacenamiento

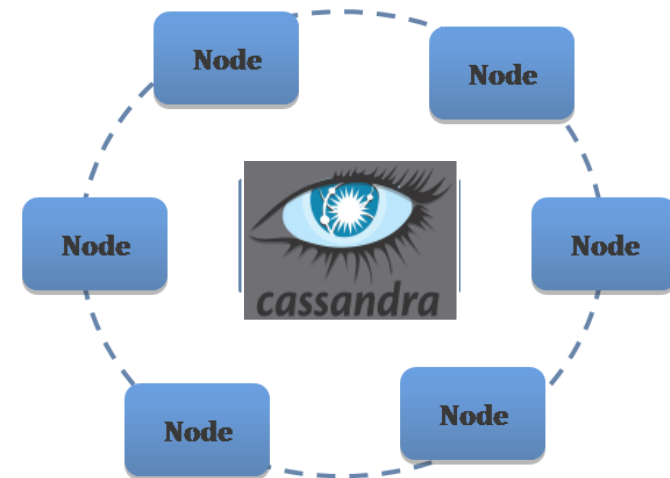
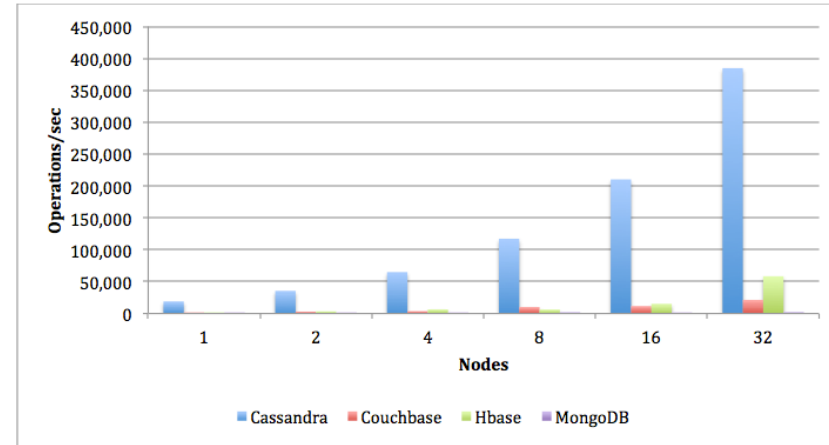


Escalabilidad horizontal.

Arquitectura “masterless”.

AP en el teorema CAP.

Conectores para analíticas (Spark).





# Web Notebook



Analítica de datos interactiva.

Scala, Python R.

Soporte nativo para Spark.

Visualizaciones interactivas.



The background of the slide is a complex network graph. It consists of numerous nodes, represented by circles of varying sizes and shades of blue and grey, connected by a dense web of thin, light blue lines. The graph is centered and fills most of the slide area.

# Gracias por tu atención

<https://github.com/Neuw84/euskal2k17>

[aconde@lkerlan.es](mailto:aconde@lkerlan.es)