

Sesión 11 Jenkins

big data (Cloudera)

cloudera-introduction.pdf

- MapReduce → Spark
 - MapReduce : repartir la carga entre X servidores que no tengan colisiones y meterlo en un lugar común
 - Hive : SQL to MapReduce
 - Pig : lenguaje para no SQL (pero que se le parece tanto...)
 - kafka : cluster
 - impala : consumo intensivo RAM, accede al FS directamente (sin pasar por HDFS)
 - HDFS : Hadoop Distributed FS (Java) → lento
 - SOLR : indexador
 - Kite : ¿api?
 - Resource Management
 - YARN : Yet Another Resource Negotiation
 - sistema de colas
 - Filesystem: HDFS
 - Relational: kudu
 - Security
 - Sentry : kerberos
 - RecordService: ¿registro a qye se ha accedido?
 - NoSQL: HBase (el perdedor contra Casandra)
 - Storage
 - Batch
 - sqoop
 - Real-Time
 - Kafka
- Hadoop Cluster
 - NameNode (de 2 a 3, este último para Journal)
 - zookeeper : service discovery
 - secondaryNameNode : helper
 - Resource manager (YARN)
 - DataNode
 - Just a Band of Disc
 - RAID1 para disco OS
 - hacen el trabajo
 - HDFS
 - sistema raid
 - bloques de 128MG
 - replicación (3 dataNode)
 - reglas de afinidad de la replicación
 - sueltas el archivo de Gb o Tb

Jenkins

Last update:

22/03/2019

08:46

info:cursos:pue:devops:sesion11 <https://miguelangel.torresegea.es/wiki/info:cursos:pue:devops:sesion11?rev=1553269617>

From:

<https://miguelangel.torresegea.es/wiki/> - miguel angel torres egea

Permanent link:

<https://miguelangel.torresegea.es/wiki/info:cursos:pue:devops:sesion11?rev=1553269617>

Last update: **22/03/2019 08:46**

