Hadoop & Kerberos

DevFest2016

Sobre nosotros

Sobre el taller

Preparar un servidor para instalar

Desplegar un cluster de un solo nodo [pseudo distribuido]

Kerberizar el cluster que hemos desplegado

Hablar sobre medidas de seguridad en Hadoop

Entorno para Laboratorio

Maquina Virtual con Centos - Instalación con imagen mínima!

- 4GB RAM
- Networking en modo brigde
- 20GB HDD Thin provisioned

Laboratorio Uno!

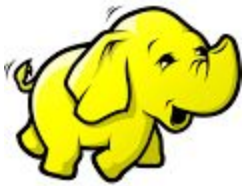
HADOOP

Hadoop: Creado por Doug Cutting, inspirado por Google File System y MapReduce.

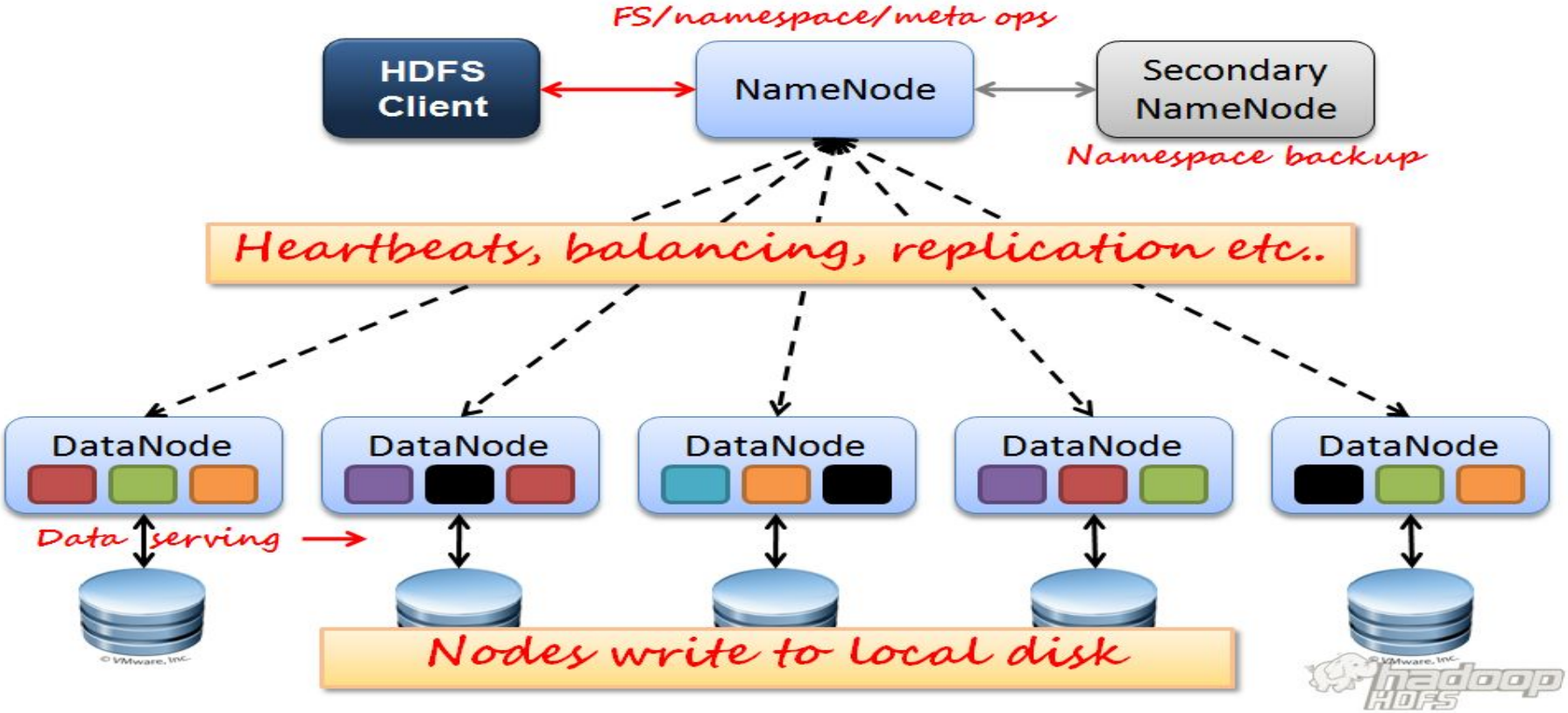
Almacenamiento y procesamiento de grandes volúmenes de datos. Sobre “comodity hardware”.
Tolerancia a errores incluida en el diseño.
Compuesto por:

- HDFS - Hadoop Distributed Filesystem
- YARN - Yet Another Resource Negociator
- MapReduce - Sistema de procesamiento en paralelo basado en YARN
- Librerías comunes

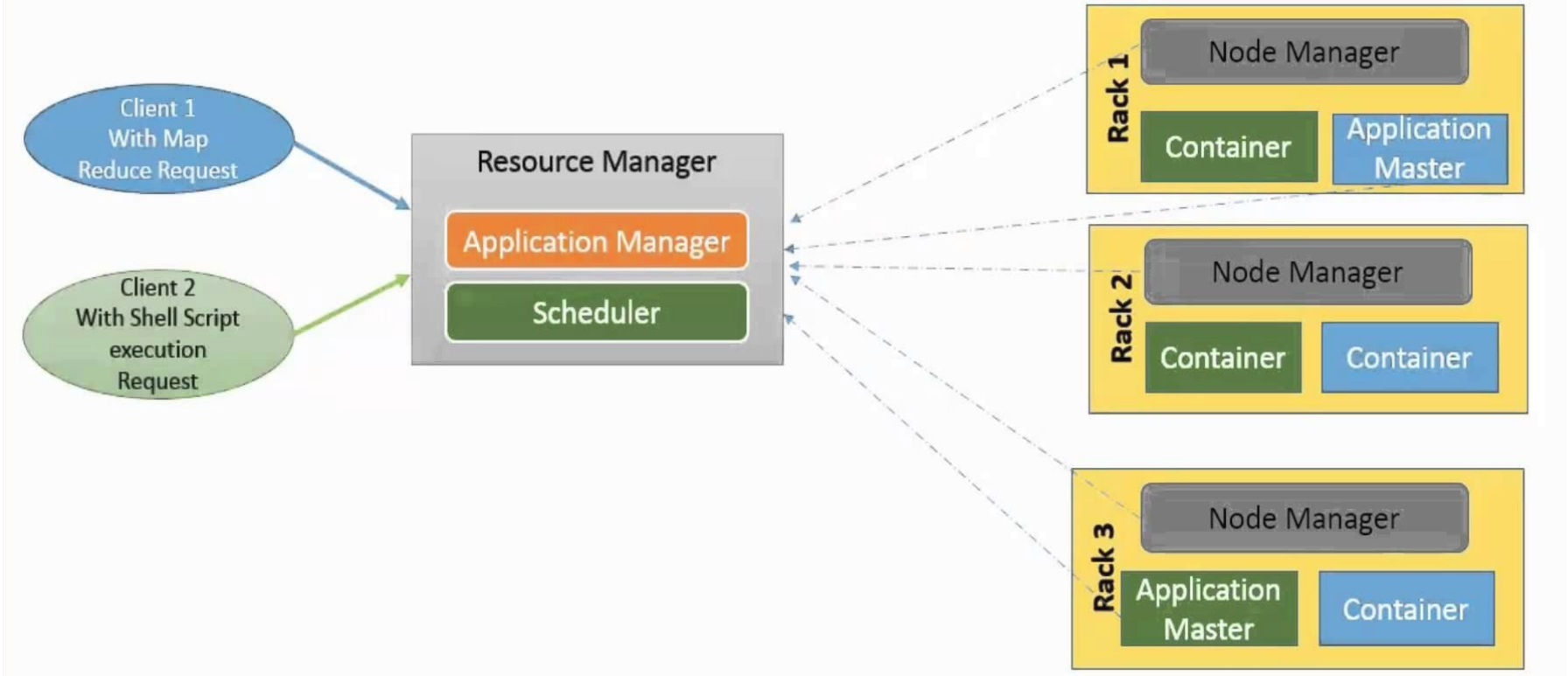
Ecosistemas Hadoop: conjunto de software relacionado con el proyecto Hadoop.



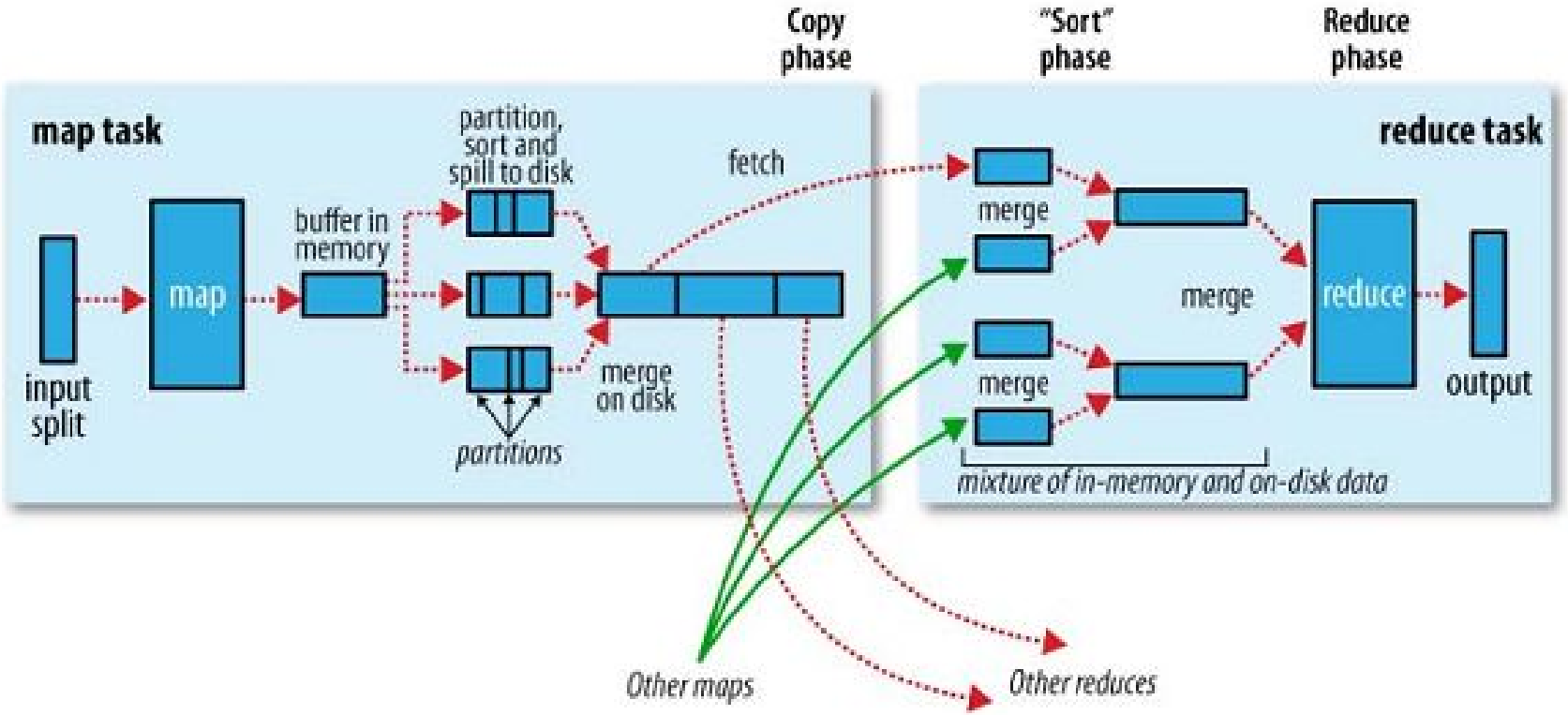
HADOOP - HDFS


FS/namespace/meta ops
HDFS Client
NameNode
Secondary NameNode
Namespace backup
Heartbeats, balancing, replication etc..
DataNode
DataNode
DataNode
DataNode
DataNode
Data serving
Nodes write to local disk
© VMware, Inc.
hadoop HDFS

HADOOP - YARN


Client 1
With Map
Reduce Request
Client 2
With Shell Script
execution
Request
Resource Manager
Application Manager
Scheduler
Rack 1
Node Manager
Container
Application
Master
Rack 2
Node Manager
Container
Container
Rack 3
Node Manager
Application
Master
Container

HADOOP - MapReduce



Laboratorio Dos!

Kerberos

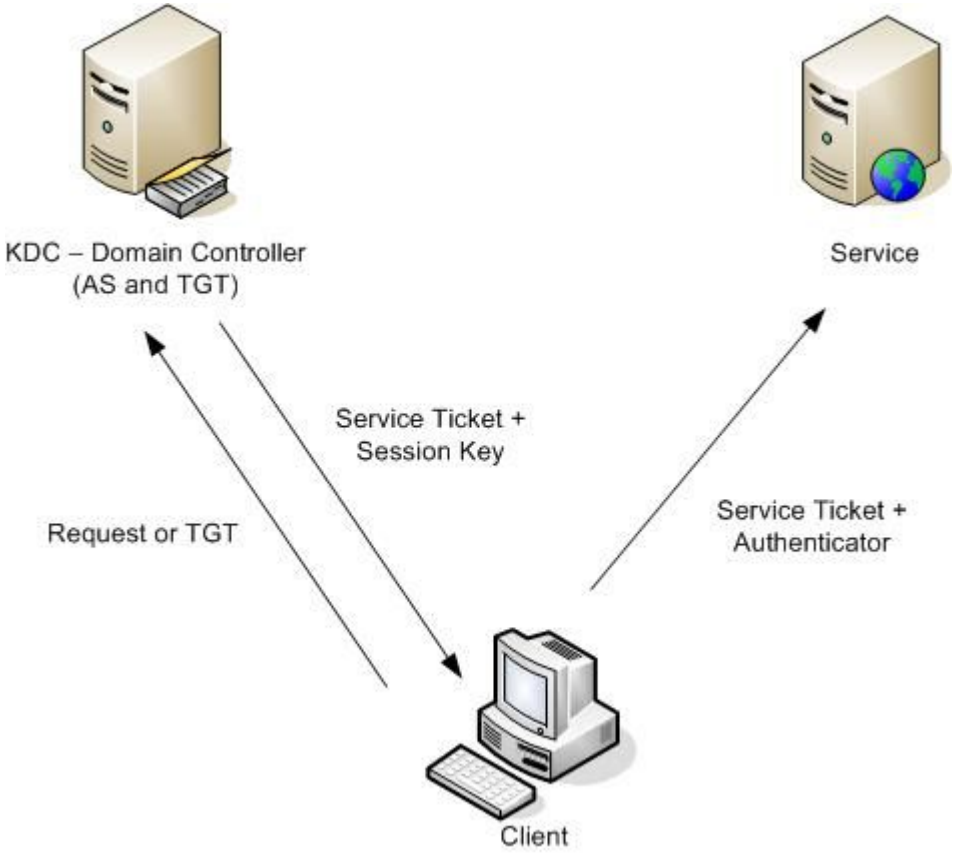
Autenticación en sistemas distribuidos

Kerberos

kinit <identidad>

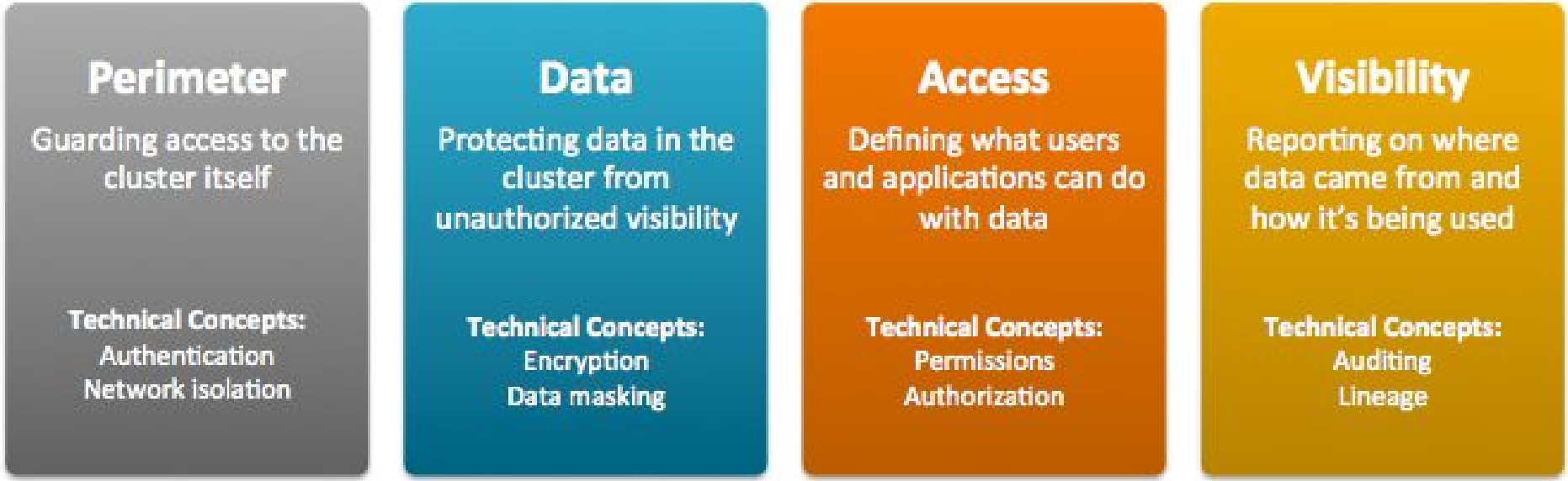
klist -ef

kdestroy



Laboratorio Tres!

Seguridad en Hadoop



Perímetro - Autenticación

Proveedor de Autenticación:

- MIT Kerberos o compatible

Proveedor de usuarios:

- Sistema operativo / Integración mediante NSS [LDAP, NIS, etc ...]
- LDAP

Gestores de Identidad:

- FreeIPA
- Active Directory

Perímetro - Seguridad en Red

Defensa perimetral

- filtrado a nivel de red (access list / firewalls)
- filtrado a nivel de host (iptables)
- uso de gateway o maquina de acceso

Datos - Cifrado

Cifrado de datos en tránsito

- Habilitar soporte TLS para HTTP
- Habilitar cifrado
  - RPC
  - Bloques

Cifrado de datos en reposo

- Hadoop Key Management Server

Datos - Protección de datos

Cloudera - Recordservice

- Nuevo servicio y protocolo - Requiere cambios
- Enmascaramiento y filtrado - columnas y filas

Hortonworks - Ranger

- Enmascaramiento y filtrado - columnas y filas
- Solo en Hive

Acceso - Permisos

Permisos

- Permisos Unix
- ACLs

Acceso - Autorización

Autorización de uso

- ACL a nivel de servicio

Cloudera

- Sentry

Hortonworks

- Ranger

Apache
Ranger

Visibilidad - Auditoría

Cloudera

- Cloudera Navigator

Hortonworks

- Ranger

Apache
Ranger

Acceso - Linaje

Cloudera

- Cloudera Navigator

Hortonworks

- Atlas