pragsis Bidoop
Tecnología e Innovación

BIG DATA & SEGURIDAD
UN MATRIMONIO DE FUTURO

PRESENTACIÓN

pragsis Bidoop
Tecnología e Innovación

ANTONIO GONZÁLEZ CASTRO

IT SECURITY DIRECTOR EN PRAGSIS TECHNOLOGIES

agcastro@pragsis.com

antoniogonzalezcastro.es

@agonzaca

linkedin.com/in/agonzaca

CEH Certified Ethical Hacker™ | CHFI Computer Hacking Forensic INVESTIGATOR™ | ECSA EC-Council Certified Security Analyst™ | LPT Licensed Penetration Tester™

ISO/IEC 27001:2005 | ITIL® IT Service Management | CISM | cloudera hadoop

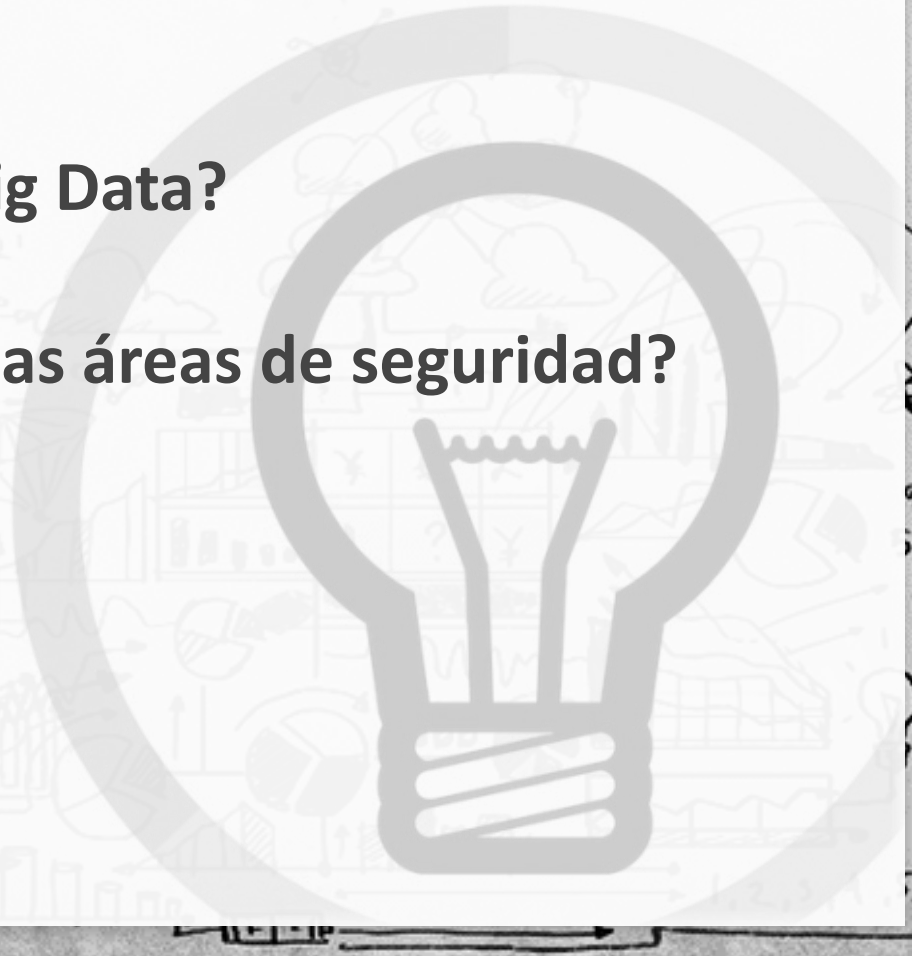
CONTENIDO

pragsis Bidoop
Tecnología e Innovación

¿Qué es esto del Big Data?

¿Son realmente seguras las plataformas Big Data?

¿Podemos aprovechar esta tecnología en las áreas de seguridad?



¿QUÉ ES ESTO DEL BIG DATA?

pragsis Bidoop
Tecnología e Innovación

NO ES SOLO TECNOLOGÍA…

Conjunto de **procesos, tecnologías y modelos de negocio** que **están basados en datos** y en **capturar el valor** que los propios datos encierran.

¿Qué lo hace diferente e innovador?

VOLUMEN

VARIEDAD

VELOCIDAD

CRECIMIENTO DE LOS DATOS EN EL MUNDO REAL

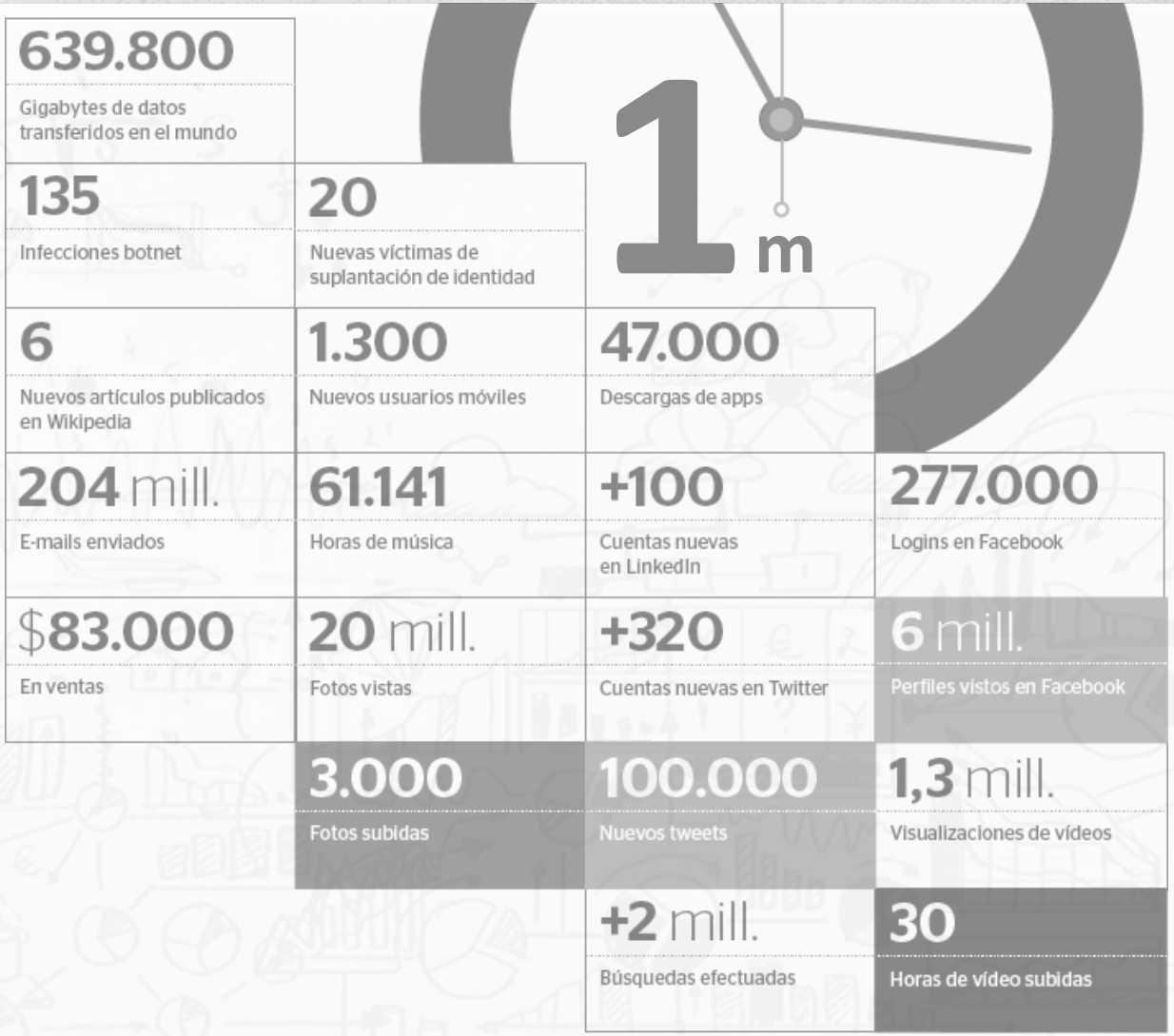
pragsis Bidoop
Tecnología e Innovación

A DÍA DE HOY

El número de dispositivos en red, equivale a la población mundial.

x2

EN EL AÑO 2016



Fuente: Intel: What Happens in an Internet Minute?

TIPOS DE DATOS

pragsis Bidoop
Tecnología e Innovación

SMART DATA

Datos referentes al negocio (online/offline). En este grupo podemos encontrar cifras de ventas, estrategia de negocio, datos sobre los clientes, etc. Todo aquello que este relacionado con los objetivos de la empresa.

INDENTITY DATA

Datos referentes que nos permiten identificar a nuestros clientes actuales, así como datos de sus gustos, historial de compras, interacciones con nuestros contenidos, etc.

OPEN DATA

Agrupa al resto de datos externos a la empresa y que son accesibles por todo el mundo.

¿UNA MODA?

pragsis Bidoop
Tecnología e Innovación

BIG DATA YA ES UNA REALIDAD

Las empresas ya están empezando a analizar lo que realmente les interesa: **SUS CLIENTES.**

Empresas como Facebook, Google y Amazon han llegado al éxito gracias a esta tecnología.

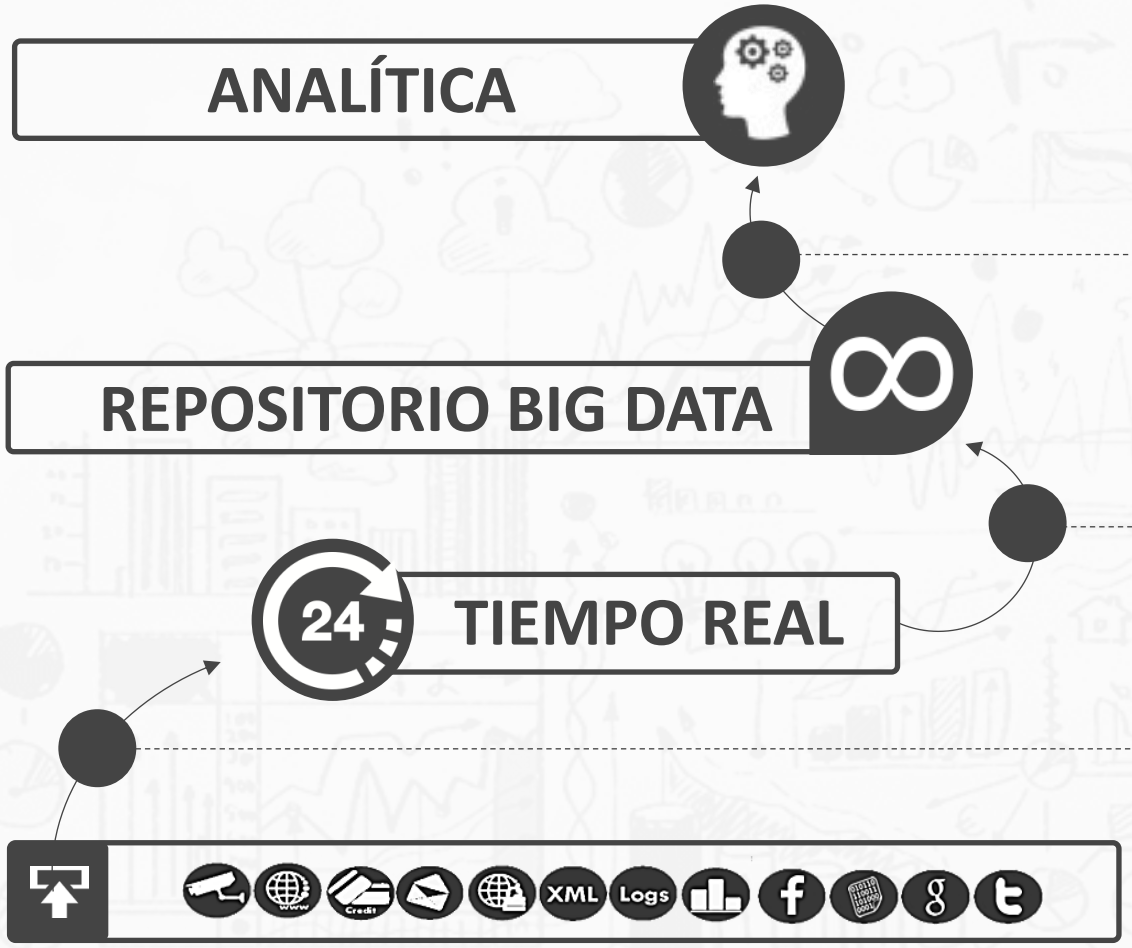
Ya se esta poniendo un gran foco en desarrollar aquellos productos que el mercado demanda.

DETECTAR EL FRAUDE

ALGUNAS EMPRESAS QUE UTILIZAN BIG DATA

pragsis Bidoop
Tecnología e Innovación

CAPAS FUNCIONALES DE INTEGRACIÓN



- Modelos analíticos.
- Motor de reglas.

PROCESAMIENTO BATCH

- Datos en brutos.
- Datos transformados.

ALMACENAMIENTO

- Motor de reglas.

ETL

- Datos estructurados.
- Datos no estructurados.

TECNOLOGÍA BIG DATA, HADOOP

OPEN SOURCE (PROYECTO APACHE)

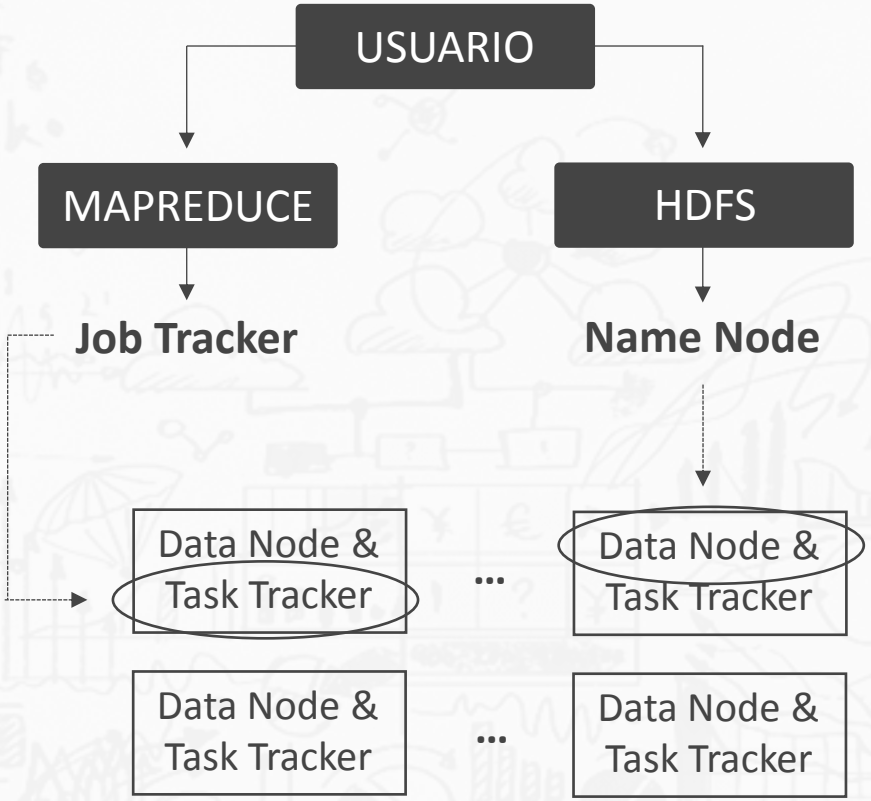
- Almacenar y procesar gran cantidad de datos.
- Implementado en JAVA.
- Posibilidad de desarrollar en otros lenguajes.
- Económico, rápido y eficiente.

HDFS

- Sistema de archivos distribuido.
- Los datos se replican en varias máquinas.

MapReduce

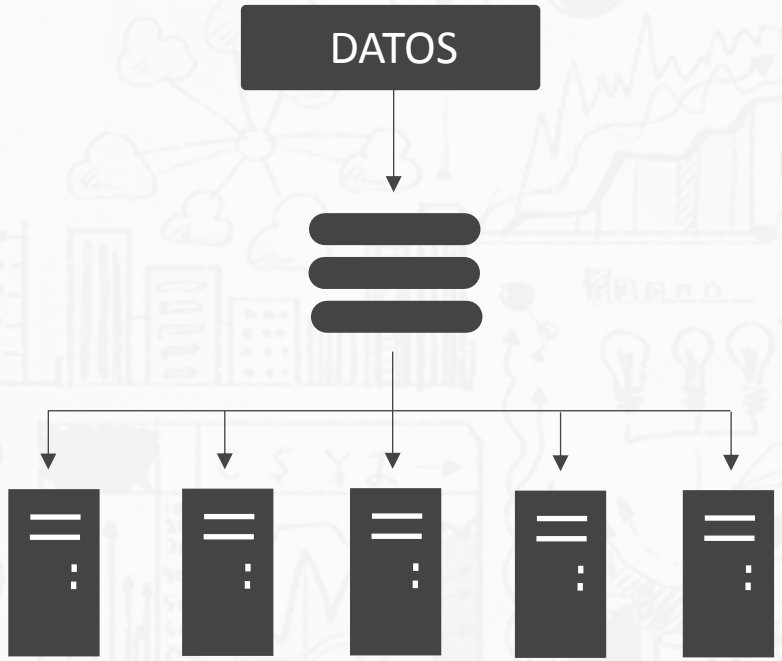
- Procesamiento por lotes.
- Consulta de datos sobre HDFS.
- MapReduce + HDFS = Localidad.



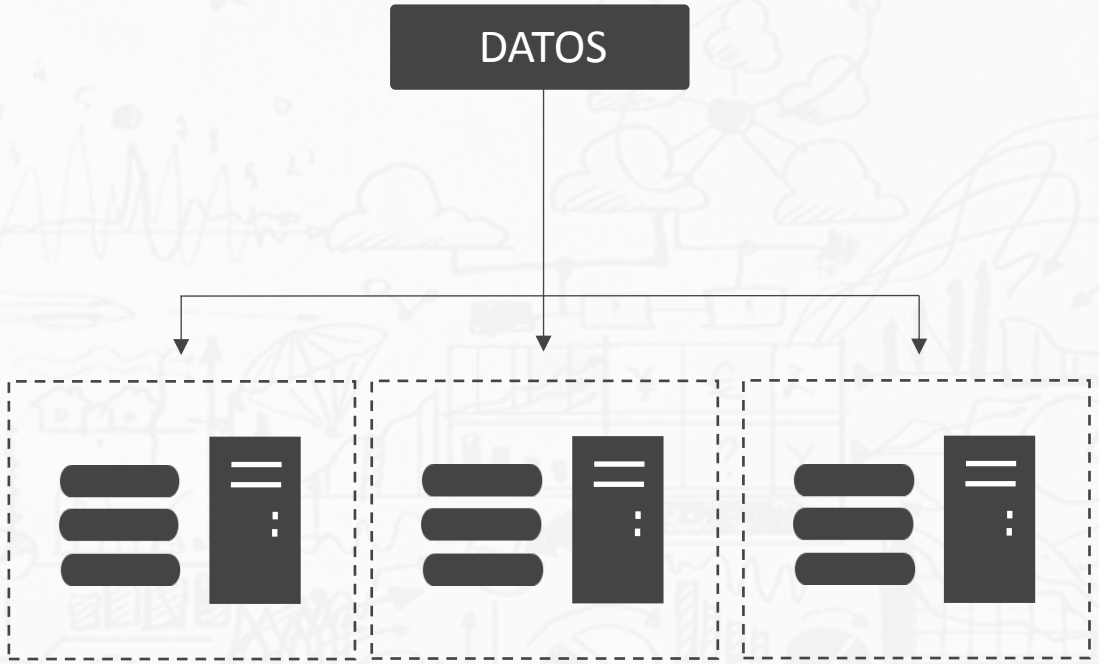
¿POR QUÉ HADOOP FUNCIONA?

pragsis Bidoop
Tecnología e Innovación

ESCENARIO TRADICIONAL



ESCENARIO HADOOP



COMPONENTES DE HADOOP

pragsis Bidoop
Tecnología e Innovación

CORE
HADOOP

HBase | Flume | Sqoop | Mahout

Hive | Pig | Impala | …

ECOSISTEMA
HADOOP

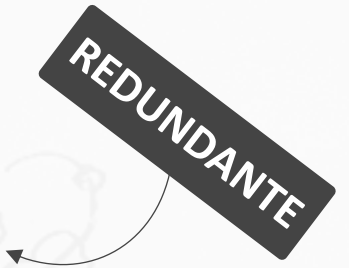
EL HARDWARE Y EL CÁLCULO DE CAPACIDADES

pragsis Bidoop
Tecnología e Innovación

NODOS ESCLAVOS

- Procesadores: 2x6 core 2.9 GHz.
- Memoria RAM: 48-96 GB.
- Red: 10 GB.
- Disco Duro: 12x3 TB (NO-RAID)

NODOS MAESTROS



- Carrier-class.
- Dos tarjetas de red.
- Disco Duro en RAID.
- Dos fuentes de alimentación.

POR NODO = ESPACIO EN DISCO / 4

SI MUY BIEN, ¿PERO ESTOS DATOS ESTAN SEGUROS?

pragsis Bidoop
Tecnología e Innovación

NO! Actualmente existen varios problemas de seguridad

AUTENTICACIÓN

AUTORIZACIÓN

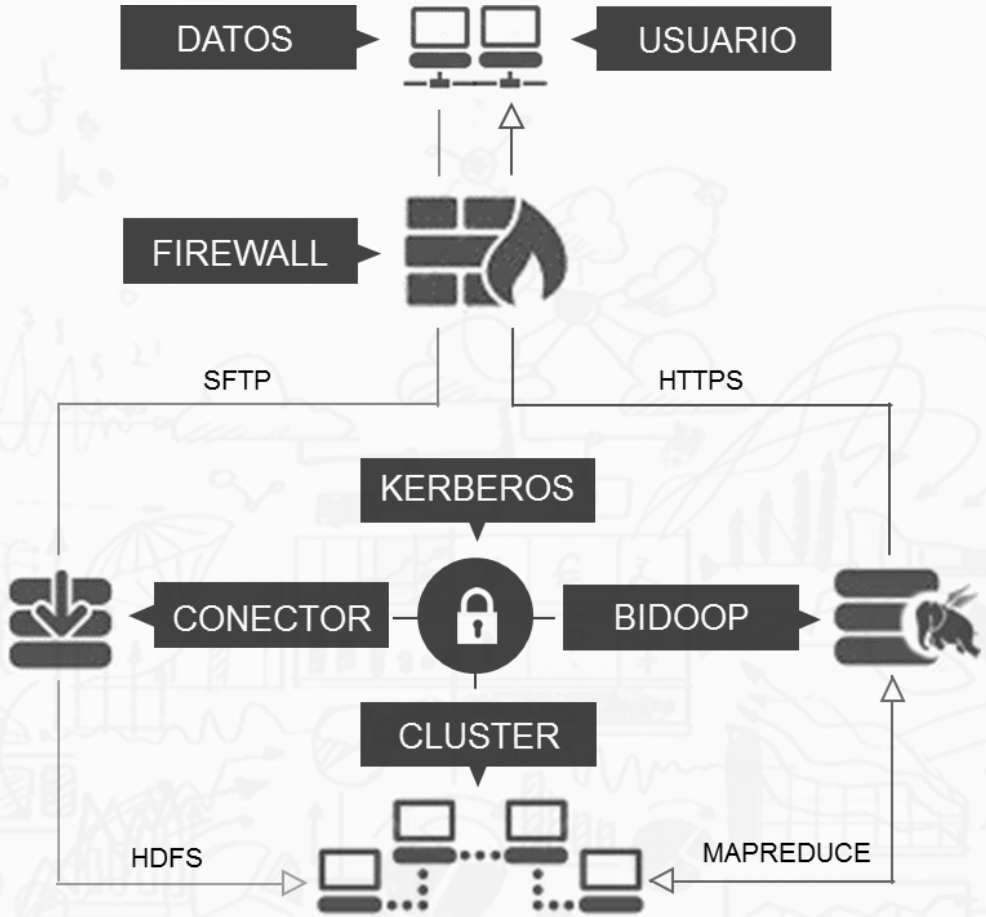
CIFRADO

EJECUCIÓN DE CÓDIGO

TODO TIENE SOLUCIÓN, ARQUITECTURA SEGURA

pragsis Bidoop
Tecnología e Innovación

- CONTROL DE ACCESOS
- CIFRADO DEL TÁFICO DE RED
- ENMASCARADO DE INFORMACIÓN
- ARQUITECTURA RED AISLADA



SEGURIDAD SOBRE HADOOP AL DETALLE

pragsis Bidoop
Tecnología e Innovación

AUTENTICACIÓN

- **Kerberos RPC (SASL / GSSAPI)** autenticar a los usuarios.
- **Consolas web HTTP (HTTP SPNEGO)** conexiones HTTP.
- **Tokens de delegación** después de la autenticación para evitar carga.

AUTORIZACIÓN

- **Autorización a datos en HDFS** a través del NameNode basado en el control de acceso (ACL) de los usuarios y grupos.
- **Bloque de Tokens (HMAC-SHA1)** control de acceso a los bloques de datos.

CIFRADO

- Conexiones SASL **(Kerberos y Autenticación RPC).**
- Consolas Web y Operaciones MapReduce **(SSL).**
- HDFS **Soluciones comerciales.**

EJECUCIÓN CÓDIGO

- **Se soluciona con los pasos indicados en autorización (Tokens).**

MENSAJE PARA LOS SECURITY RESEARCHER

pragsis Bidoop
Tecnología e Innovación

security@hadoop.apache.org

MMM! ¿SI LO UTILIZAMOS PARA SEGURIDAD?

pragsis Bidoop
Tecnología e Innovación

- CENTRALIZACIÓN DE EVENTOS
- DETECCIÓN DEL FRAUDE
- ANÁLISIS FORENSE
- CIBERVIGILANCIA
- DETECCIÓN DE AMENAZAS Y ATAQUES

CASO DE USO, DETECCIÓN DE FRAUDE

pragsis Bidoop
Tecnología e Innovación

DATOS

- **TRANSACIONES** → ID CLIENTE / IP ORIGEN / FECHA / TARJETA / CANTIDAD
- **PERFIL DE CLIENTE** → ID CLIENTE / MEDIA CONSUMO / POSICION HABITUAL
- **CIBERVIGILANCIA** → USUARIOS / TARJETAS

ESCENARIO

CUANDO **QUIEN** **DONDE** **EL QUE**

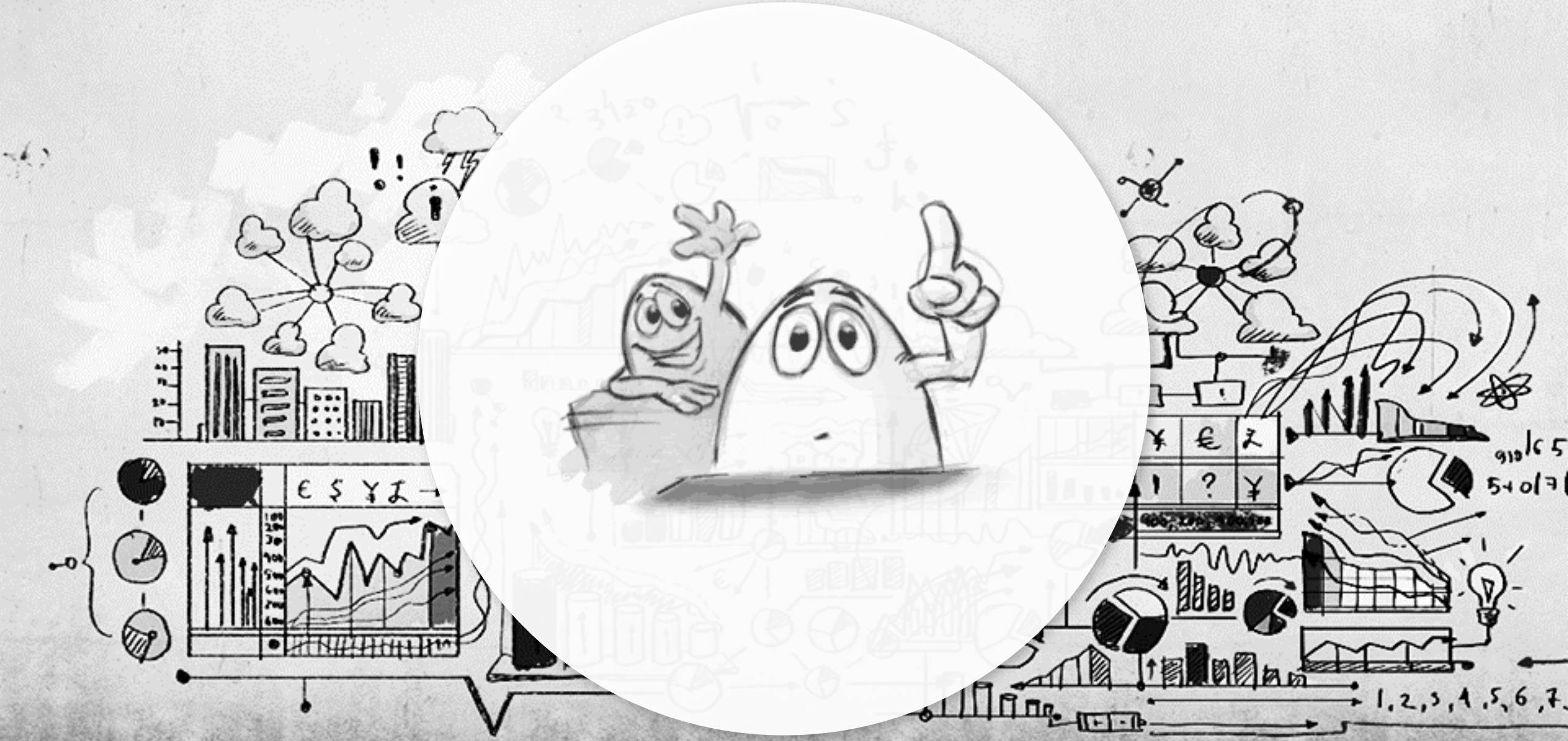
[13/06/2014 10:13:52] [5598CM23 1234567890123457] [80.26.83.175] [150]

[13/06/2014 10:43:12] [5598CM23 1234567890123457] [148.245.38.39] [45]

[12/06/2014 09:20:35] [5598CM23 1234567890123457] [195.60.81.64] [4900]

¿PREGUNTAS?

pragsis Bidoop
Tecnología e Innovación



Fin de la presentación, haga clic para cerrar. Será si quiero!

MUCHAS GRACIAS!

[X] CERRAR