

Capítulo 2: El Big Data y la ciencia de datos

Capítulo 3: Exploración de Big Data

Capítulo 4: Hadoop



División de Alta Tecnología

3

Exploración del Big Data

Los datos son la nueva ciencia.
El Big Data son las respuestas. (Pat Gelsinger)

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.




Objetivos

Al finalizar el capítulo, el alumno logrará:

- Comprender los mecanismos de la importancia del Big Data para los clientes.
- Operatividad y consideraciones de implementación del Big Data.

3 - 2

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Agenda

- La vista 360 del Cliente
- Análisis, Diseño y Operación del Big Data
- Procesamiento del Big Data
- Seguridad en el Big Data

3 - 3

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



La vista 360 del cliente

FACTORES QUE INFLUYEN EN EL COMPORTAMIENTO DE LOS CONSUMIDORES



3 - 4

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Análisis en Big Data Conocimiento del negocio

- Definición de los objetivos del negocio
- Identificar los orígenes de los datos
- Entregables
 - Documento marco
 - Matriz de orígenes de datos
 - Diccionario de datos



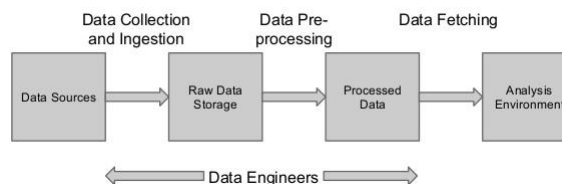
3 - 7

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Análisis en Big Data Adquisición y comprensión de los datos

- Carga de datos
- Exploración de los datos
- Configuración de una canalización de los datos
- Entregables
 - Informe de calidad de datos
 - Arquitectura de solución de datos



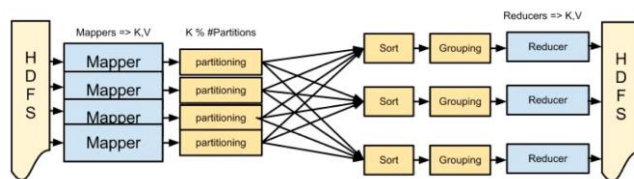
3 - 8

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Operación del Big Data

- Despliegue de la solución
- Pruebas del método procesamiento de Big Data
- Pruebas funcionales para consultas y explotación de datos
- Evaluar seguridad de la solución



The MapReduce Pipeline

A mapper receives (Key, Value) & outputs (Key, Value)
 A reducer receives (Key, Iterable[Value]) and outputs (Key, Value)
 Partitioning / Sorting / Grouping provides the Iterable[Value] & Scaling

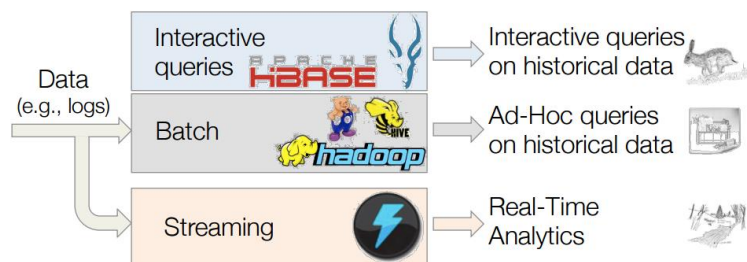
3 - 11

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Procesamiento en Big Data

- Procesamiento por lotes (Batch Processing)
- Procesamiento en tiempo real (Streaming Processing)
- Procesamiento Híbrido (Arquitectura Lambda)

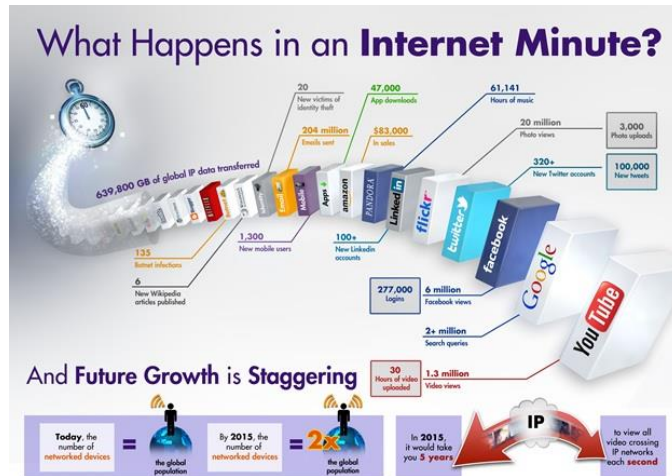


3 - 12

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Procesamiento en Big Data

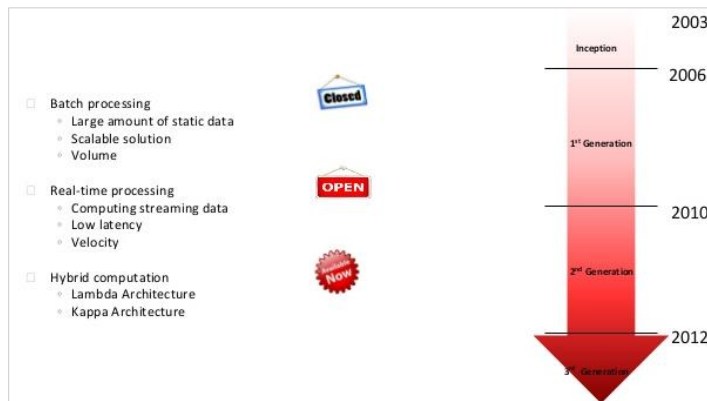


3 - 13

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Procesamiento en Big Data Timeline

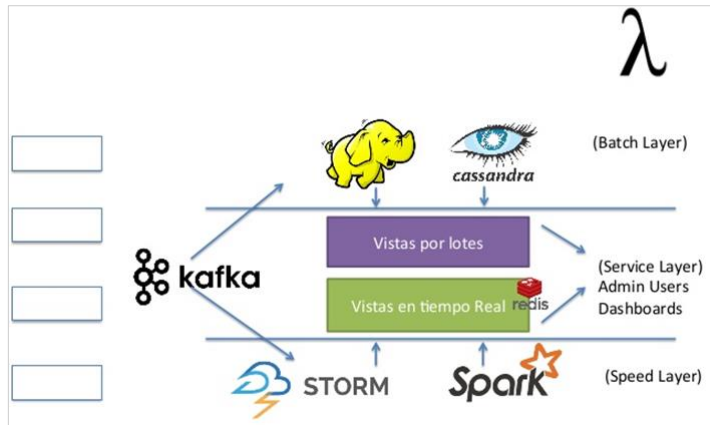


3 - 14

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Procesamiento en Big Data Arquitectura Lambda

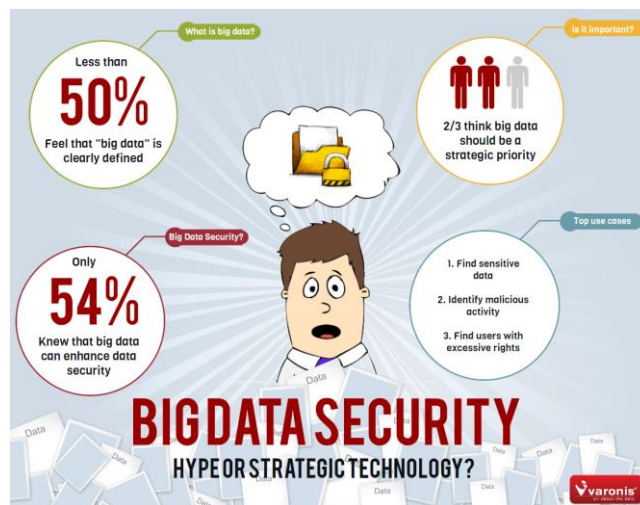


3 - 15

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Seguridad en el Big Data

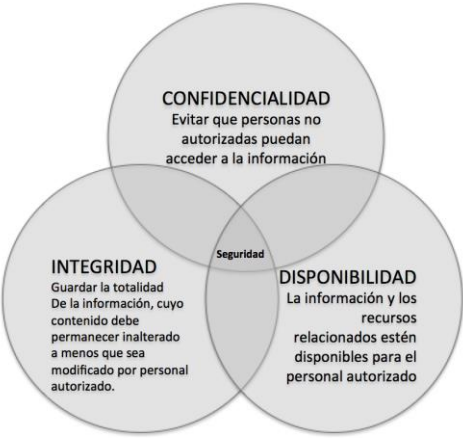


3 - 16

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Seguridad en el Big Data



3 - 17

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Seguridad en el Big Data

Los 14 capítulos de la Norma ISO 27002

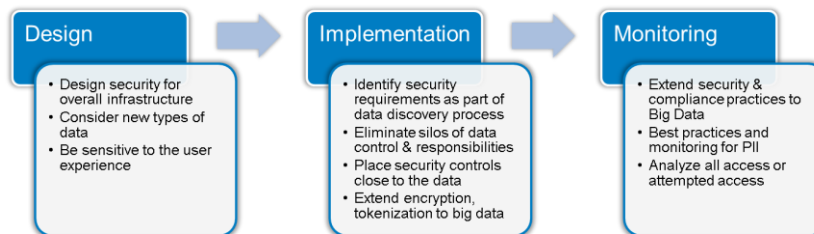


3 - 18

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Seguridad en el Big Data



3 - 19

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Tarea N° 3: Curso Big Data

Al finalizar la tarea, el alumno logrará:

- Aprender los conceptos y la terminología básica de Big Data.

3 - 20

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Laboratorio N° 3.1: Metodología de proyectos de Big Data

Al finalizar la tarea, el alumno logrará:

- Aprender como aplicar la metodología para proyectos de Big Data.

3 - 21

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Lecturas adicionales

Se sugiere revisar los siguientes enlaces para profundizar en los conceptos tratados en el presente capítulo:

- Construyendo Customer 360 con Big Data
- Tipos de Implementación de Big Data
- Seguridad y Privacidad en Big Data

3 - 22

Copyright © Todos los Derechos Reservados - Cibertec Perú SAC.



Resumen

- En este capítulo, hemos aprendido la importancia de utilizar la vista 360 del cliente.
- Además, hemos hecho una revisión general de los procesos de análisis, diseño y operación del Big Data.
- Finalmente, aprendimos de los aspectos del procesamiento y seguridad de Big Data.

