



MÓDULO 1

Introducción a Tecnologías Big Data

Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Contenidos

- Introducción a Big Data
- ¿Qué es Big Data?
- Escenarios y retos

Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.



Introducción a Big Data

Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.



Introducción a Tecnologías Big Data

Introducción a Big Data



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.



Introducción a Tecnologías Big Data

Introducción a Big Data



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Introducción a Big Data

Data
is the new
OIL

We need to find it,
extract it, refine it,
distribute it and
monetize it.

David Buckingham



Impacto en la UE
206.000 M€ en 2020
3,7M de empleos en 2018

Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Introducción a Big Data

Data
is the new
OIL

We need to find it,
extract it, refine it,
distribute it and
monetize it.

David Buckingham



Sembrar
Regar
Cuidar
Recolectar

Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

¿Cuáles son las fuentes de los datos?

1. Datos de procesos internos
2. Comercio electrónico
3. Telecomunicaciones
4. Redes sociales
5. Internet de las Cosas

En 2020, 30mil millones de dispositivos estarán conectados a internet



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Introducción a Big Data

Data is growing at a 40 percent compound annual rate, reaching nearly 45 ZB by 2020

Data in zettabytes (ZB)



Source: Oracle, 2012

Humanity Passes 1 Zettabyte Mark in 2010

A zettabyte is 1,000,000,000,000,000,000 bytes (that's 21 zeroes for those counting), or one trillion gigabytes. That's enough data to fill 75 billion 16-gigabyte-sized iPads.



Graphic by Karl Tate
Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Introducción a Big Data

Gartner

El volumen de datos generados cada año crece de forma **exponencial** en todos los sectores

IBM

El **90% de la información** existente ha sido generada en los **dos últimos años**

COMPUTERWORLD

El **80%** de la información es **desestructurada**

FORRESTER

Las empresas sólo **aprovechan** en torno al **5%** de la información generada

Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

De los datos al conocimiento

1. Data
2. Anonymous Data
3. Aggregated Data
4. Insights



My name is John (27) and go to London on Mondays at 7 AM



I am (27) and go to London on Mondays at 7 AM

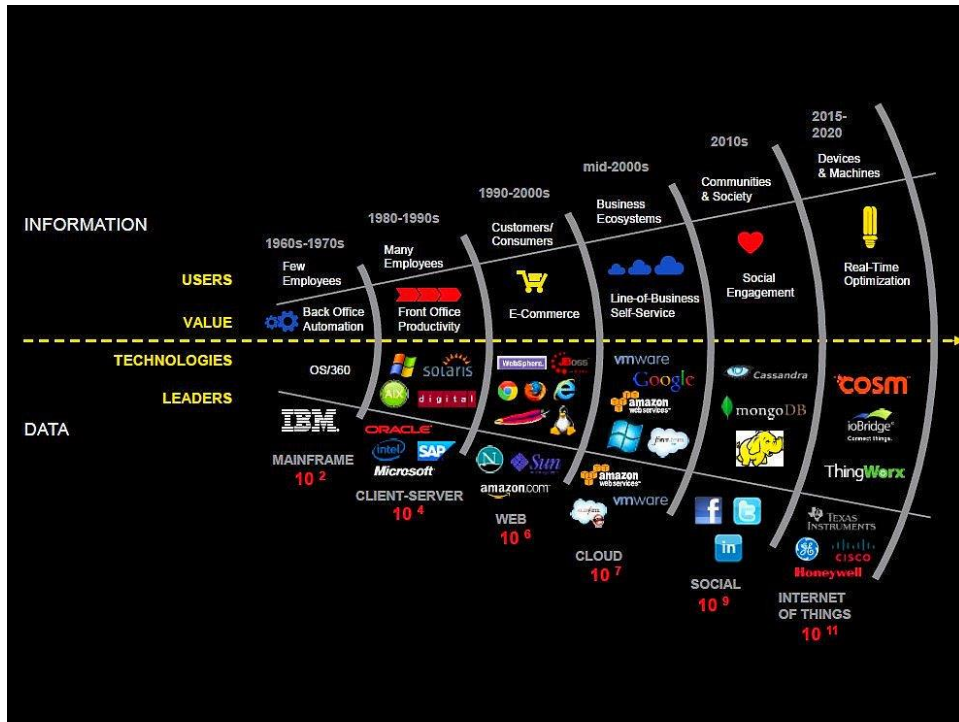


We are 100 people under 30 and go to London on Mondays at 7 AM



27% of people under 30 go to London on Mondays at 7 AM

Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.



Introducción a Big Data

¿Hemos asistido a la revolución?

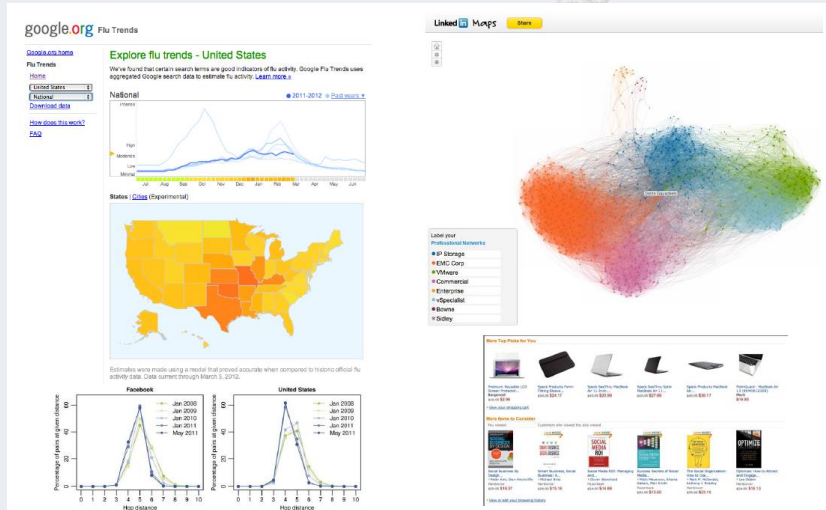
- Almacenar datos no es suficiente
- Extraer valor de los datos es la clave
- Información = Ventaja competitiva
 - 20% del tiempo se invierte en buscar datos
 - 61% de los ejecutivos quieren acceso rápido a datos
 - 80% de las decisiones se toman en base a datos
- Las empresas necesitan:
 - Conocer y mejorar sus procesos
 - Descubrir tendencias
 - Evaluar el impacto y el riesgo
 - Dirigirse al target (personalizar)
 - Apoyo en la toma de decisiones
 - Anticiparse o predecir

El negocio de datos



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Introducción a Big Data



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Introducción a Big Data

¿Alrededor de los datos sólo hay negocio?

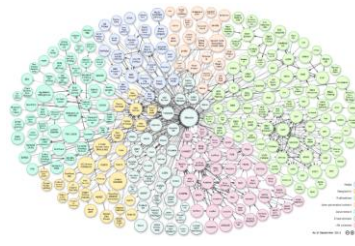
Open Data

Filosofía de datos abiertos que nació con la necesidad de transparencia de las instituciones públicas



Linked Data

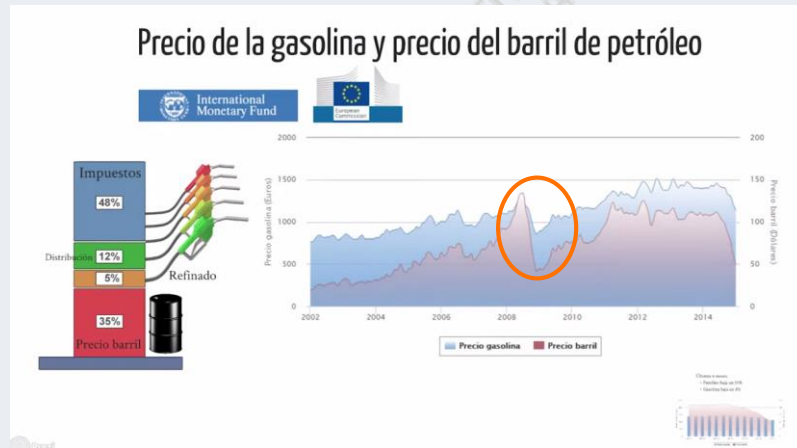
Define estándares para poder intercambiar información entre organizaciones



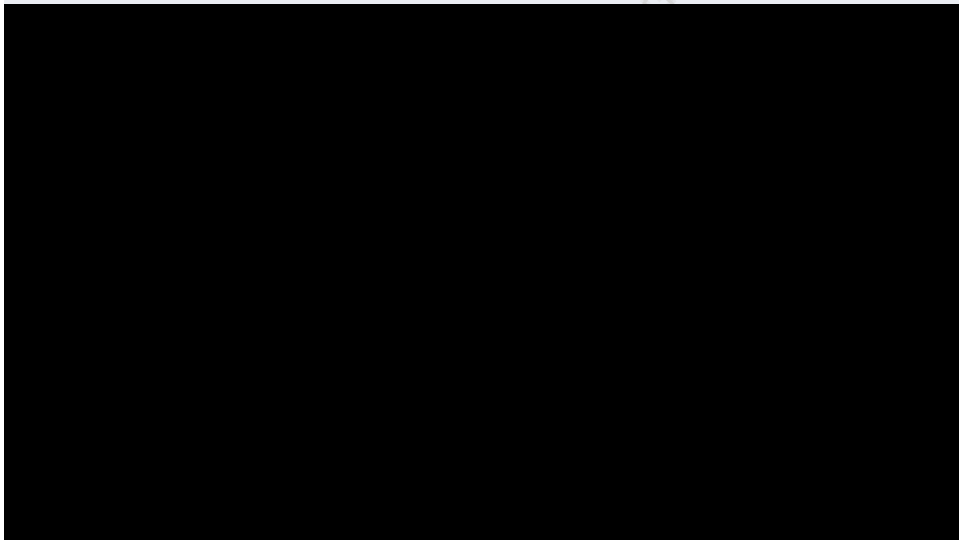
Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Introducción a Big Data

(Big + Open) Data



Introducción a Big Data



Introducción a Big Data

SPRING SPREE
spending patterns in spain during easter 2011

Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Introducción a Big Data

Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Introducción a Big Data



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Introducción a Big Data



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.



¿Qué es Big Data?

Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

¿Qué es Big Data?

Definición de Big Data

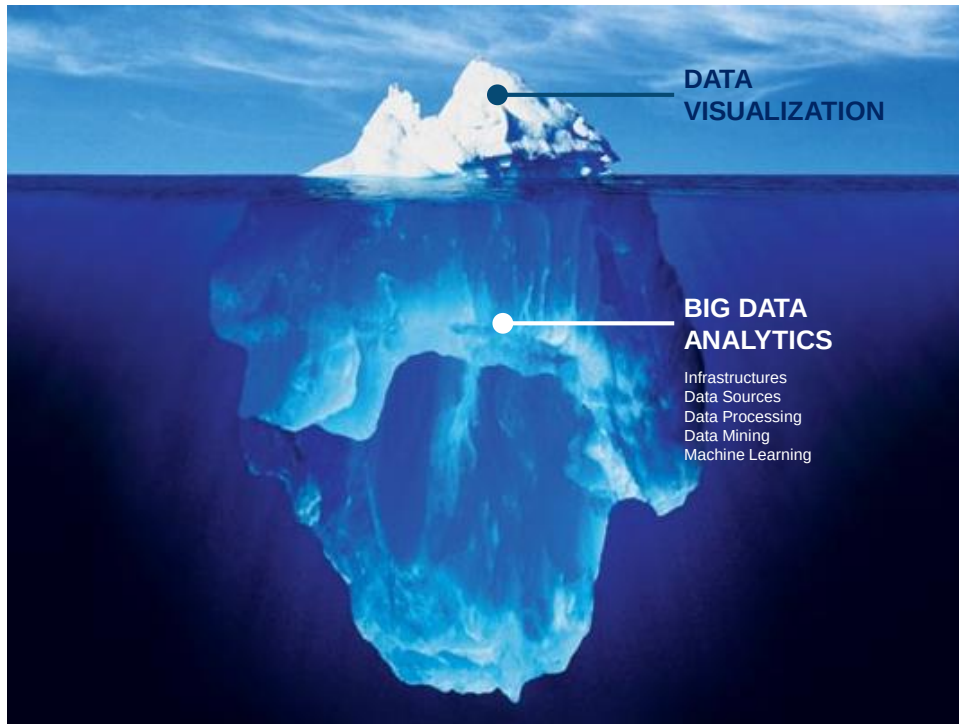
Término que hace referencia a una **cantidad de datos tal que supera la capacidad del software habitual** para ser capturados, gestionados y procesados en un tiempo razonable

Fuente: Wikipedia



¿Estamos hablando de gestionar
VOLUMEN de datos?

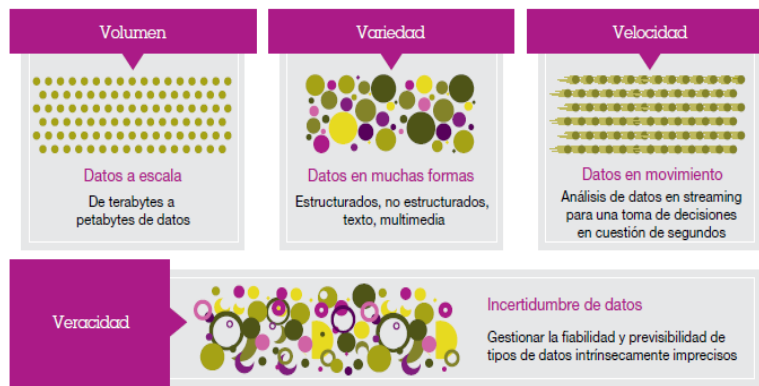
Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.



¿Qué es Big Data?

Características de Big Data

Las 4 Vs



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

¿Qué es Big Data?

Volumen, Variedad, Velocidad, Veracidad

- Tecnologías actuales no resuelven el problema
- Ha sido necesario crear técnicas y tecnologías para:
 - Bases de datos escalables
 - Sistemas de archivos distribuidos
 - Tratamiento masivo
- El estándar es  **hadoop**
- Han proliferando soluciones **cloud (IaaS)** para dar respuesta a las necesidades de elasticidad

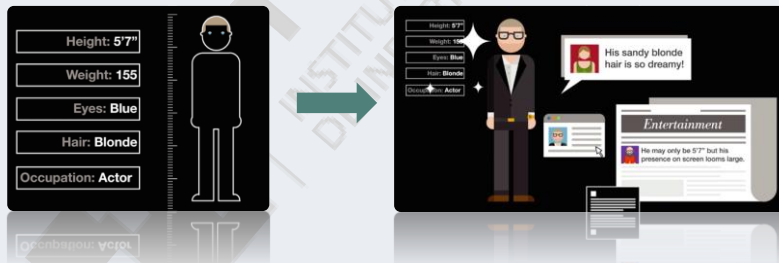


Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

¿Qué es Big Data?

Volumen, Variedad, Velocidad, Veracidad

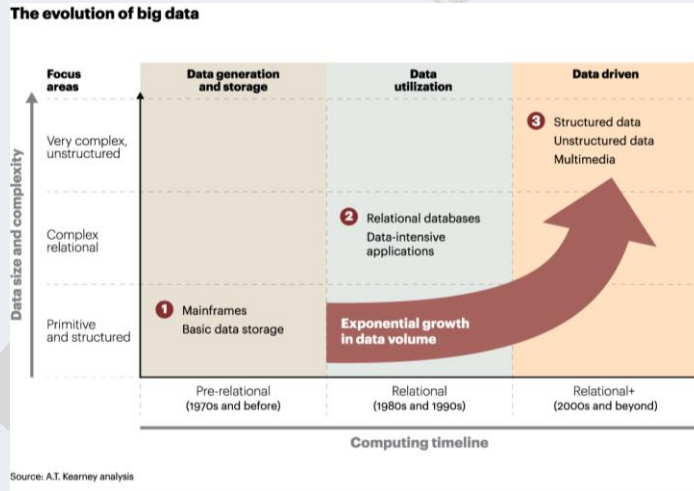
Hoy en día, 80% datos no estructurados: imágenes, vídeos, voz, documentos de texto completos...



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

¿Qué es Big Data?

Volumen, Variedad, Velocidad, Veracidad



¿Qué es Big Data?

Volumen, Variedad, Velocidad, Veracidad

Tipos de datos complejos:

- **Estructurados:** BBDD relacionales, hojas de cálculo
- **No estructurados:** texto, multimedia, logs, emails, documentos, ...
- **Semiestructurados:** xml, html, json, ...

Aparición de nuevas tecnologías especializadas:

- NoSQL



Cassandra



mongoDB



elastic

- NewSQL o Bases de Datos Distribuidas



NUODB



MEMSQL

¿Qué es Big Data?

Volumen, Variedad, Velocidad, Veracidad

Son necesarias tecnologías capaces de:

- **Recoger y transformar** datos generados a gran velocidad (Streaming)
- **Procesar y analizar** de forma rápida (Real-Time)

Ejemplos:

- Detección de **fraudes** en transacciones bancarias
- Análisis de **riesgos**
- Detección de **anomalías** en altas frecuencias
- **Mensajes virales** en redes sociales (twitter)
- Interacciones en **juegos online** (MMOG)
- Recogida de **datos en sensores** (IoT)
- ...

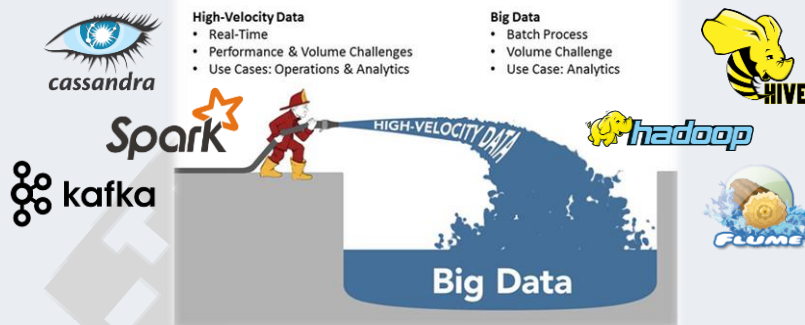


Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

¿Qué es Big Data?

Volumen, Variedad, Velocidad, Veracidad

Comparing High-Velocity Data & Big Data



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

¿Qué es Big Data?

Volumen, Variedad, Velocidad, Veracidad

- Establecer criterios que nos permitan confiar en los datos
- Incluye:
 - Autenticidad (blockchain)
 - Disponibilidad
 - Fiabilidad
 - Precisión
 - Confiabilidad (reputación)
- Es necesario identificar los datos que aporten valor
- Los datos obsoletos se almacenan (coste elevado de identificación)

Ley de los datos de Parkinson

"Los datos se expanden para llenar el espacio disponible"



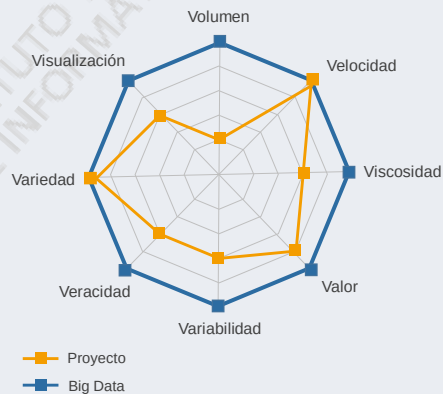
Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

¿Qué es Big Data?

Características de un proyecto Big Data

Actualmente están apareciendo otras características:

- **Variabilidad:** significado de los datos cambia constantemente (lenguaje natural)
- **Visualización:** representación de los datos
- **Valor:** toma de decisiones
- ¿Algunas más? Viscosidad, Viralidad ...



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.



Escenarios y Retos

Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Escenarios y Retos

Data-Driven Decision Making (DDD)

Aquellas empresas que tienen un proceso de toma de decisiones basado en los datos obtienen un **incremento medio de la productividad en un 3%**

MIT, 2010



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Escenarios y retos

Business Intelligence vs Data Analytics

- BI tradicional se centra en "lo que ha pasado" y un enfoque "causal"
- Data Analytics se centra en "lo que va a pasar" y un enfoque "correlacional".

Data Analytics vs Big Data Analytics

- En DA usamos un "conjunto de datos limitado" y aplicamos técnicas de reducción
- En BDA "no hay límite" y por tanto el resultado arroja un error menor (el 99%)

Big Data vs Big Data Analytics

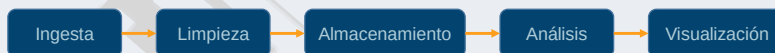
- Del almacén y gestión de los datos al análisis de los mismos
- De los **datos** a la **información**, de información a **conocimiento**.

Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Escenarios y Retos

Procesos involucrados

- **Ingesta:** ¿Podemos recibir / recoger y organizar los datos?
- **Limpieza:** ¿Se puede limpiar y transformar los datos recogidos?
- **Almacenamiento:** ¿Se puede almacenar y acceder a una gran cantidad de datos?
- **Análisis:** ¿Se puede enriquecer y analizar estos datos en tiempo razonable?
- **Visualización:** ¿Podemos servir los datos a un aplicativo?

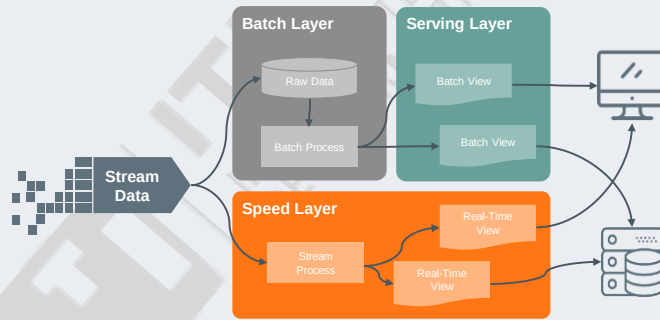


Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Escenarios y Retos

Escenarios de actuación

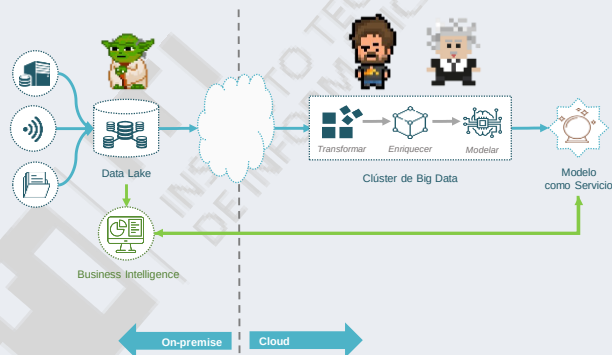
Las tecnologías Big Data se especializan en los siguientes grupos:



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Escenarios y Retos

Arquitectura de referencia



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

Servicios de Análisis de Datos Avanzados

sada@iti.es
formacion@iti.es

+34 963 879 958

Follow us



Copyright © Instituto Tecnológico de Informática (ITI). Prohibida la reproducción total o parcial sin permiso del ITI.

 **ITI**
INSTITUTO TECNOLÓGICO
DE INFORMÁTICA

Introducción a Tecnologías Big Data