



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Introducción al Big Data

Presentación



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Antonio Cachuán

@antoniocachuan



Big Data
Data Engineer @ **COE**
Center of Excellence

Comuniquémonos



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería



<https://bit.ly/2HyVT9o>

Presentación



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Este curso ayuda a

- Conocer el Big Data y sus oportunidades.
- Entender las áreas involucradas en las cuales uno pueda especializarse y desarrollar una carrera.
- Enriquecer el conocimiento con nuestras experiencias e intercambiar ideas y opiniones.

Presentación



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Conozcámonos

- ✓ Nombre
- ✓ ¿Qué viste en el curso pasado que (sí/no) te gustó?
- ✓ ¿Qué esperas del curso?
- ✓ Otros comentarios

Presentación



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Este curso ayuda a

- Conocer el Big Data y sus oportunidades.
- Entender las áreas involucradas en las cuales uno pueda especializarse y desarrollar una carrera.
- Enriquecer el conocimiento con nuestras experiencias e intercambiar ideas y opiniones.

Metodología



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Clases y ponencias (la teoría es importante)

Lecturas

Videos

Debates

Ejemplos prácticos

Evaluación



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Participación/Exposición (70%)

Examen (30%)

Agenda



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

- Qué es Big Data
- Conceptos y Terminología
- Aspectos Generales de Big Data
- Fuentes de Datos por Tipos de Datos
- Criterios para la adopción del Big Data
- Ejercicios Prácticos

Todo gira por los datos



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

La recopilación, almacenamiento y análisis de datos se han hecho desde siempre

- ☐ Ciencia
- ☐ Ciencias Sociales y Humanidades
- ☐ Medicina
- ☐ Comercios y logística

Todo gira por los datos



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

[https://www.youtube.com/watch?v=0Q3sRSUYm](https://www.youtube.com/watch?v=0Q3sRSUYmys)
[ys](#)

*Big Data will impact every part of your life -
TEDxFultonStreet*

Todo gira por los datos

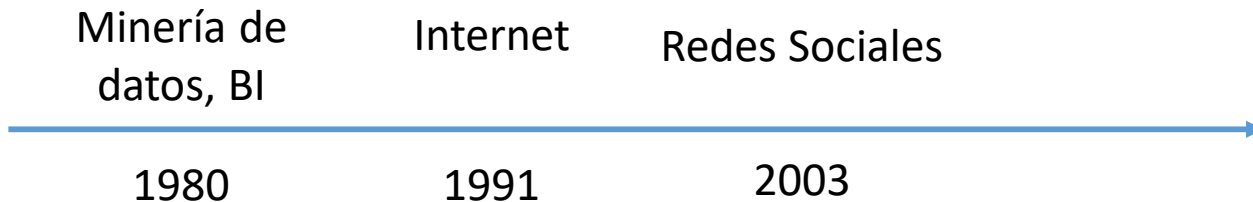


CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

La mayoría de datos están digitalizados

*Menos del 10% de los datos en el 2007 eran analógicos.



Todo gira por los datos



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

El 90% de todos los datos se crearon en los últimos 3 años

3ª etapa de Internet		Quién aporta información a internet		Ventaja competitiva
1991-2003 La tecnología cara y compleja	Web	Internet corporativa Empresas	Productos	Optimización de los procesos Logística
2003-2012 La tecnología baja de precio y de complejidad	Web 2.0 Smartphone 2007	Personas	Opiniones	Reputación
2012- La tecnología sigue bajando de precio y de complejidad	IoT	Objetos	Sensores Digitalización del mundo físico	Datos



Al flujo digital que es internet se van incorporando cada vez mas actores

[ECV, 2016]

Big Data



No existe una definición fija

“Big data” se refiere a todo lo relacionado con la captura, almacenamiento, análisis y visualización de conjunto de datos caracterizados por su **Volumen**, **Velocidad** y **Variedad** [ECV, 21016]



Conceptos y Terminología



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Volumen

La cantidad de datos crece exponencialmente

Las empresas manejan volúmenes en alrededor de TB

En el 2020 se tendrá 35ZB

Terabytes

Petabytes

Exabytes

Zettabytes

Conceptos y Terminología



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Velocidad

Hoy en día se genera datos rápidamente

What happens in an
INTERNET MINUTE?



Conceptos y Terminología



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Velocidad

Y muchos datos necesitan ser procesados rápidamente



Safety Check

Connect with friends and loved ones during a crisis.



Conceptos y Terminología



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Variedad

No solo tenemos los datos tradicionales, se han sumado diferentes estructuras.

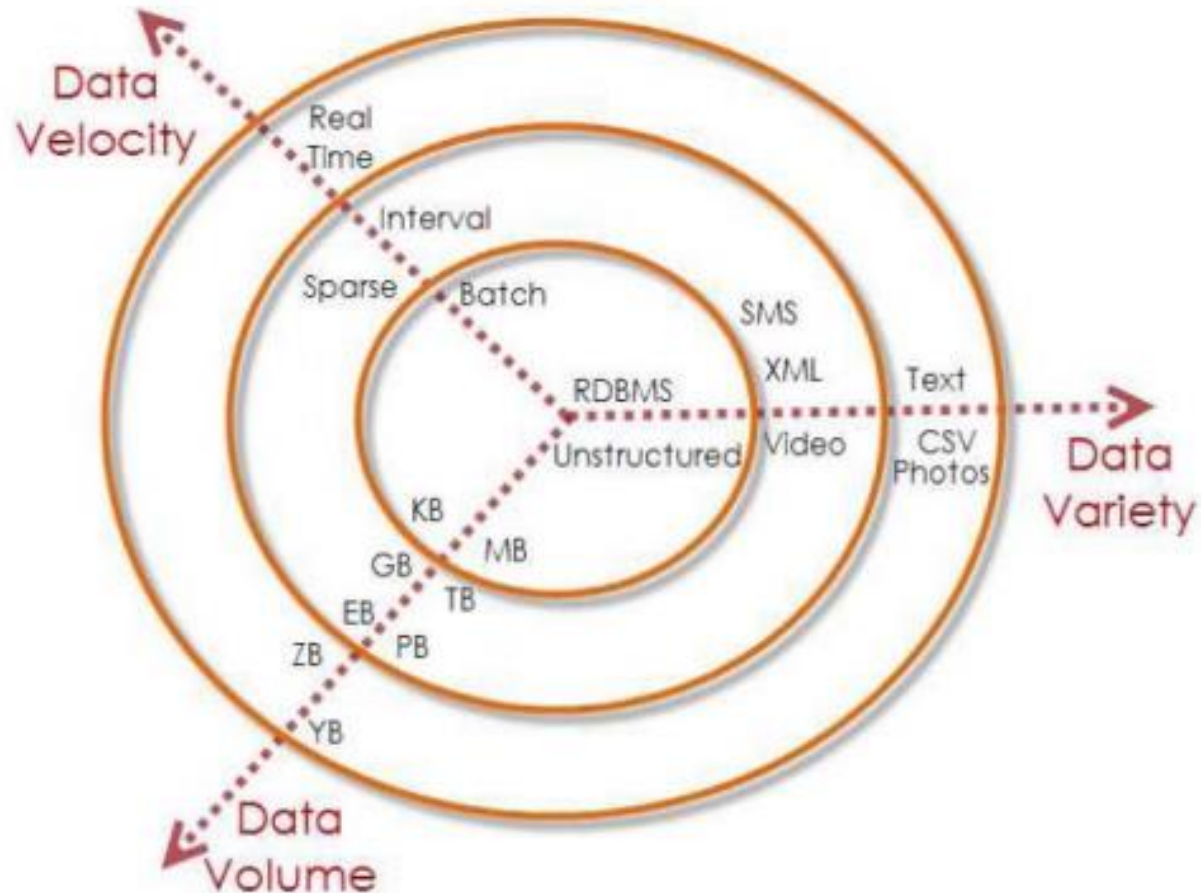


Aspectos Generales



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería



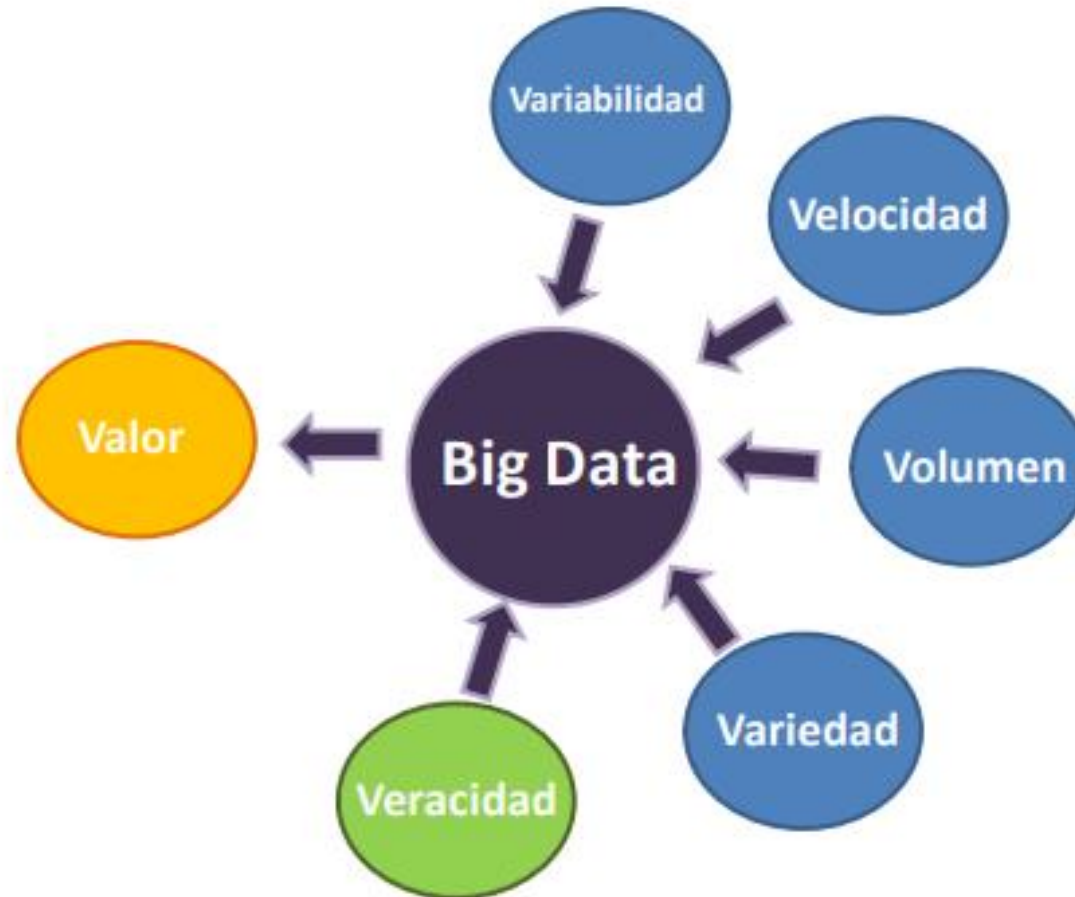
[ECV, 2016]

Aspectos Generales



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería



Aspectos Generales



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Veracidad

- “No todo lo que brilla es data útil”
- Es necesario evaluar la veracidad de la información




Google

All News Videos Images Shopping More Search tools

About 58,800,000 results (0.80 seconds)

In the news



Donald Trump Pivots to Economic and Domestic Policy Posts
New York Times - 40 mins ago
Mr. Trump returned to New York on Sunday night as his transition team looked toward filling ...

Trump's White House transition also aims at party unity, mending fences
Fox News - 12 hours ago

How Christie lost his spot on the Trump transition team
Yahoo - 16 hours ago

[More news for trump transition](#)

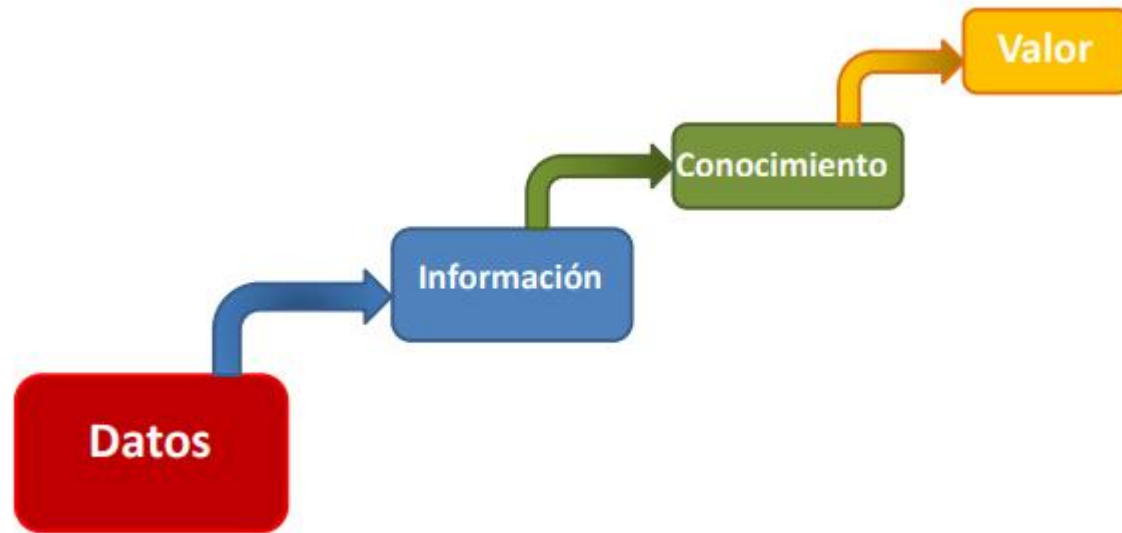
Trump cranks up speed of transition - POLITICO
www.politico.com/story/2016/11/trump-transition-twitter-231684
1 day ago - After days of chaos in his transition effort, Trump is now moving full-speed ahead on the massive task of preparing to lead the executive branch.

Aspectos Generales



Valor

Las herramientas tradicionales no soportan un conjunto de datos con las 3V



Aspectos Generales

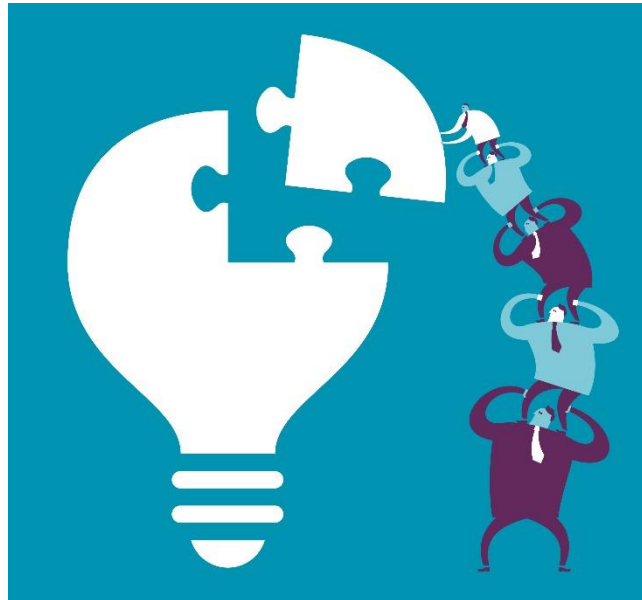


CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Valor

Necesita nuevas formas (herramientas, tecnologías y algoritmos) para gestionar los datos y extraer valor.



Aspectos Generales



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Big Data no pretende

Eliminar las herramientas de BI tradicionales

Eliminar las BBDD tradicionales

Busca cubrir los agujeros que las maneras de trabajar anteriores no cubren

Repasemos



Pensemos en un caso con las 3Vs del Big Data

Caso: Carros inteligentes Tesla

Volumen: Existen más de 100 mil autos generando datos que viajan a los servidores de Tesla.

Velocidad: A cada segundo los carros generan información de los caminos, de accidentes, los cuales necesitan transmitidos rápidamente.

Variedad: Además accede a la información de tu calendario para hacerte llegar temprano a las reuniones.

Ejercicios



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

¡Ahora te toca! – Nota participación

Imagina un caso parecido y contéstalo en el aula virtual.

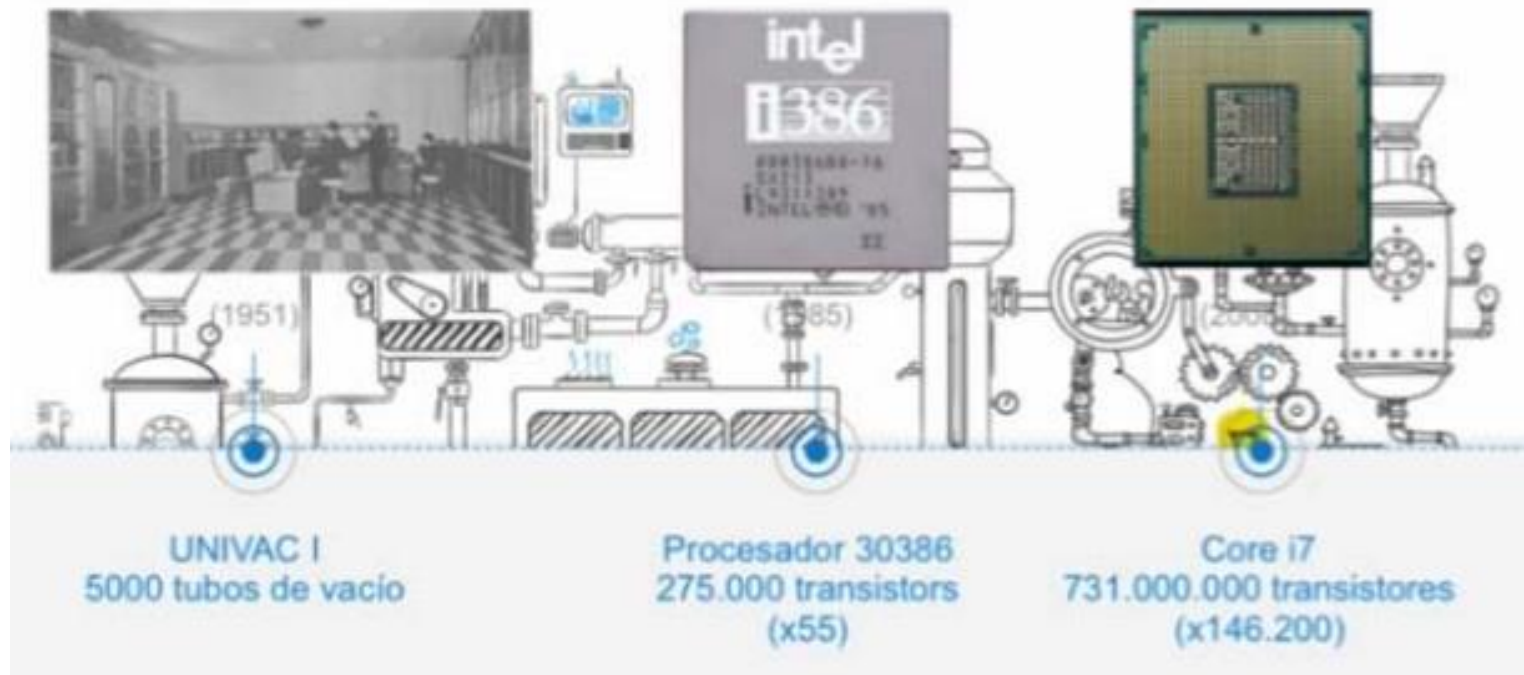
Ecosistema Big Data



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Debemos recordar que todo este avance se logró por la evolución de la tecnología



Ecosistema Big Data



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Este cambio necesita de nuevos perfiles

Data Engineer

Data Scientist

Data Quality

Data Architect

Data Expert

DevOps

Hadoop Admin

Cloud Architect

Data Analyst

Chief Data Officer

Ejercicios



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

¡Ahora te toca! – Nota participación

¿Cómo estos perfiles ayudarían al caso de Big Data que planteaste?

Incluye al menos tres perfiles.

Fuentes de datos por tipo de dato

Estructurado



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Base de datos relacionales: MySQL, Oracle, SQL Server, etc.

Mainframe: cpy, occure, redefine.

Excel

Fuentes de datos por tipo de dato

Semi Estructurado

JSON: APIs Facebook, Twitter, Google.

Web Scrapping



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Fuentes de datos por tipo de dato



No Estructurado

Correos: horas más se trabaja, tópicos conversación

Audios: conversaciones call center.

Imágenes: detectar tendencia en ropa en instagram, reconocer colores, formas, objetos (TC)

Videos: reconocer personas

Calendarios: tópicos reunión

Criterios de Adopción Big Data



Negocio

Se tiene una necesidad de brindar algo mejor a los clientes.

Se quiere enriquecer el conocimiento que tenemos de los clientes.

Todo esto se acompaña de un cambio en la cultura de trabajo.

Criterios de Adopción Big Data



Tecnológico

Se tiene una necesidad de ampliar la capacidad operativa.

Se quiere optimizar el costo del tamaño de los data centers.

Todo esto se acompaña de un cambio en la cultura de trabajo.

Ejercicio Práctico



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

¡Esto es Data!

Su equipo quiere proponer un proyecto a la institución* y para iniciarlo el área de negocio le pide un detalle y justificación para el uso del Big Data.

*Salud, Banca, Educación, Agricultura,
Transporte, Telecomunicaciones.

Ejercicio Práctico



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

¡Esto es Data! – Consideraciones

El negocio no sabe qué es Big Data y por qué lo necesitan.

No conoce el resumen del proyecto

No sabe que perfiles necesitan

No sabe que tipo de datos necesitarían

En 30 minutos el equipo elabora la presentación a exponer en 3 minutos.

¡Premio!



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Preguntas

Bibliografía



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

<http://blog.cloudera.com/>

<https://code.facebook.com/posts/>

<https://es.hortonworks.com/blog/>

<https://medium.com/netflix-techblog>

<https://labs.spotify.com/>

<https://eng.uber.com/>

Actualicémonos @ CTIC



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Por cada equipo elijan un artículo relacionada a Big Data para compartirlo la próxima clase y publicarlo en Medium.





CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Gracias =)



CTIC UNI

Centro de Tecnologías de Información y Comunicaciones
Universidad Nacional de Ingeniería

Introducción al Big Data