

# Unidad 1

## Introducción

BIG DATA

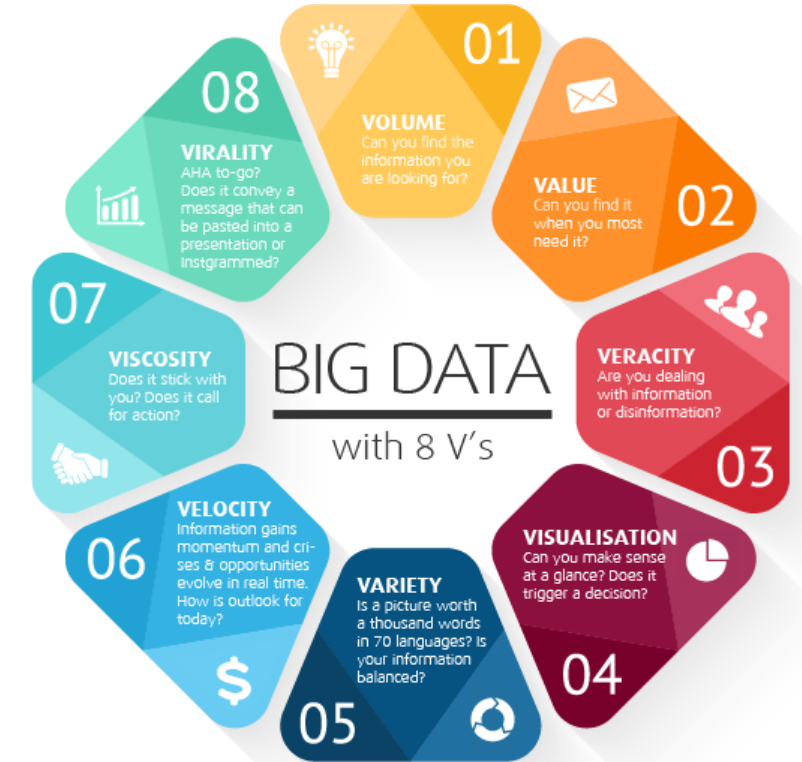
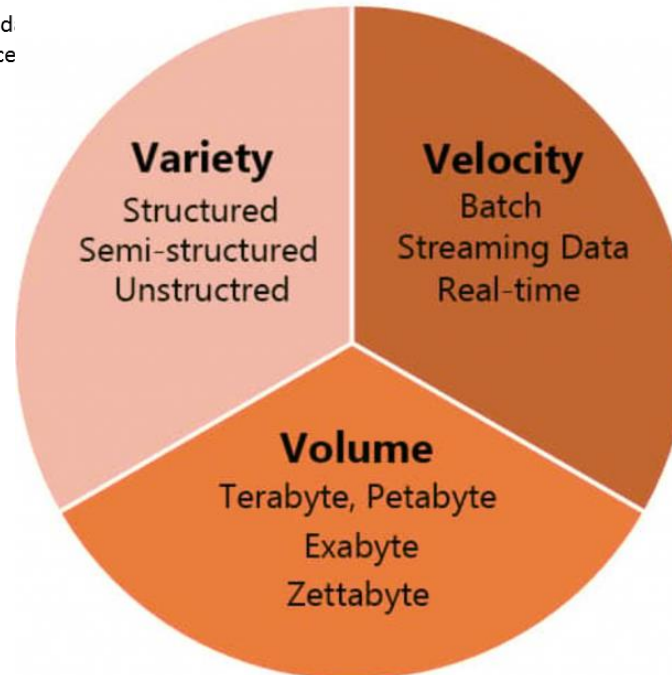
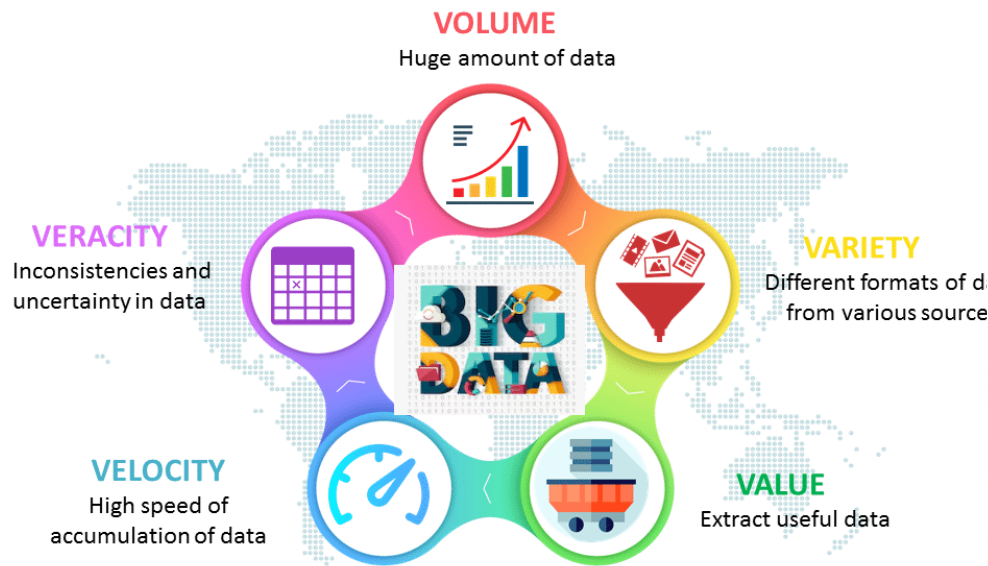


## ¿Qué es BigData?

- **The with book of.. FUJITSU;** Is about quickly deriving business value from a range of new and emerging data sources, including social media data, location data generated by smartphones and other roaming devices, public information available online and data from sensors embedded in cars, buildings and other objects and much more besides.
- **Big Data for DUMMIES;** The challenge today is how companies can make sense of the intersection of all these different types of data. When you are dealing with so much information in so many different forms, it is impossible to think about data management in traditional ways. Although we have always had a lot of data, the difference today is that significantly more of it exists, and it varies in type and timeliness. Organizations are also finding more ways to make use of this information than ever before. Therefore, you have to think about managing data differently.
- **Wikipedia;** Es un término que hace referencia al concepto relativo a conjuntos de datos tan grandes y complejos como para que hagan falta aplicaciones informáticas no tradicionales de procesamiento de datos para tratarlos adecuadamente. Por ende, los procedimientos usados para encontrar patrones repetitivos dentro de esos datos son más sofisticados y requieren software especializado.
- **Gartner;** Son datos que contienen una mayor **variedad** y que se presentan en **volúmenes** crecientes y a una **velocidad** superior. De aproximadamente 2001 (y que continúa siendo la definición de referencia).



## Características



## Característica

- **Valor**
- Visualización
- Viabilidad
- **Veracidad**
- **Variedad**
- **Velocidad**
- **Volumen**
- Vulnerabilidad

## Significado

- Datos que se transforman en información
- Legibles y accesibles para encontrar patrones y claves ocultas en el tema a investigar
- Uso eficaz del gran volumen de datos
- Grado de fiabilidad de la información recibida
- Formas, tipos y fuentes
- Rapidez en la que son creados, almacenados y procesados
- Cantidades masivas de datos que se almacenan
- Mas datos, mas riesgo



## Tipos

- Estructurados
  - Longitud definida
  - Cada campo o columna está asociado a un tipo de dato
  - Cada tipo de dato tiene un tamaño fijo y determinado
  - Normalmente vistos a semejanza de una hoja de cálculo con renglones, columnas y cada una de ellas identificadas por nombre
- No estructurados
  - No tienen un formato específico
  - No tiene una estructura
- Semi-estructurados
  - Tiene algo de estructura



## Tipos; Comparación

PROPERTIES	STRUCTURED DATA	SEMI-STRUCTURED DATA	UNSTRUCTURED DATA
Technology	Basado en tablas en base de datos relacionales	Basado en XML/RDF	Basado en catacteres binarios y texto
Transaction management	Grandes transacciones y diversas técnicas de concurrencia	Las transacciones son adaptadas debido a un manejador no maduro	No existe manejo de transacciones ni concurrencia
Version management	Diversas versiones de datos en registros y tablas	Diversas versions sobre tablas y graficas es posible	Diversas versiones pero solo sobe toda la BD
Flexibility	Basado en una estructura tipo schema	Es mas flexible que datos estructurados pero menos flexible que los no estructurados	Muy flexible y no existe el concepto de schemas
Scalability	Es muy difícil actualizar un schema de base de datos	Actualización es más fácil que los datos no estructurados	Muy escalable
Robustness	Muy robusto	Nueva tecnología, no muy conocida	—
Query performance	SQL permite joins complejos	Sentencias en diferentes nodos de forma anónima es posible	Solo sentencias bajo texto es posible

ATA



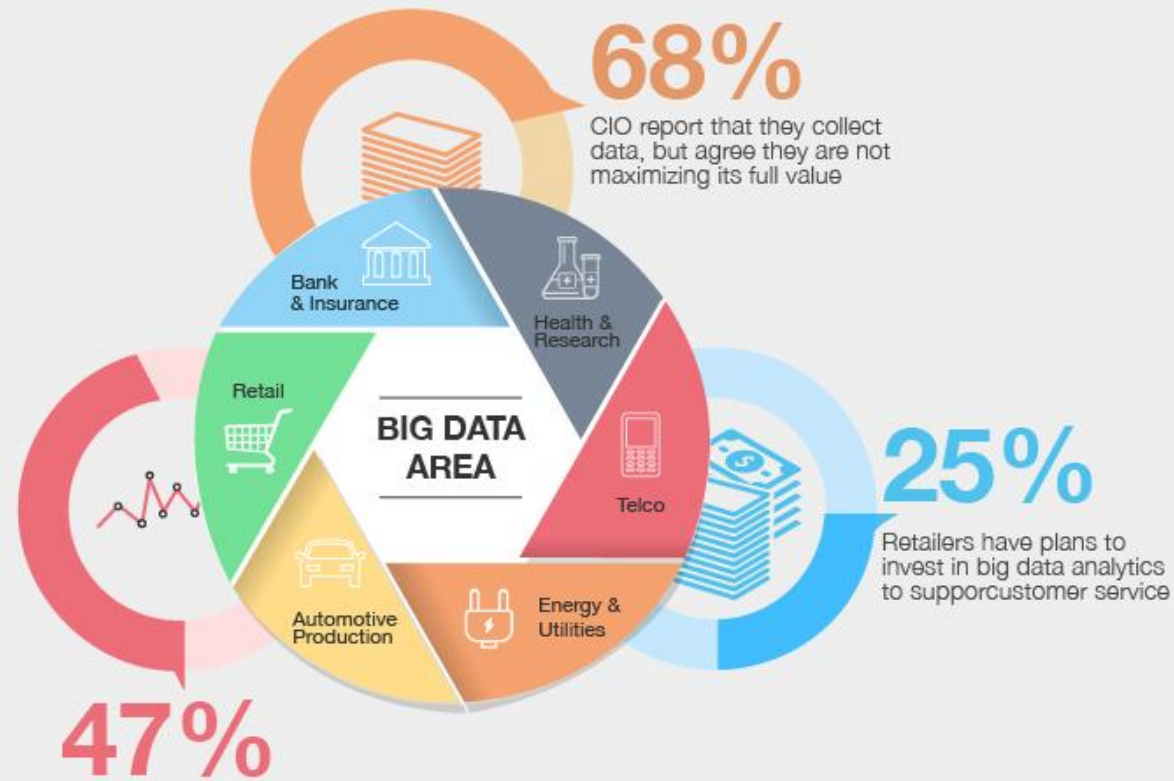


## Fuentes de Información

- Estructurados
  - Machine generated
    - Sensores
    - Server logs
    - Puntos de venta
    - Financieros
  - Human
    - Capturas de información
    - Click
    - Juegos
- No estructurados
  - Texto en lenguaje natural
  - Películas
  - Audios
  - Imágenes
- Semi-estructurados
  - Open Data
  - Redes Sociales
  - Sensores
- Internas; dentro de la empresa
  - Finanzas
  - Mercadeo
  - Comercial
  - Producción
  - Planificación estratégica
- Externas; fuera de la empresa
  - Tio de cambio
  - Bolsa se valores
  - Bancos de datos



## BIG DATA & ANALYTICS



CIO having invested in cross-channel analytics for deriving multichannel insights and enhancing the company's ability to reach consumers effectively

 motivitylabs

BIG DATA



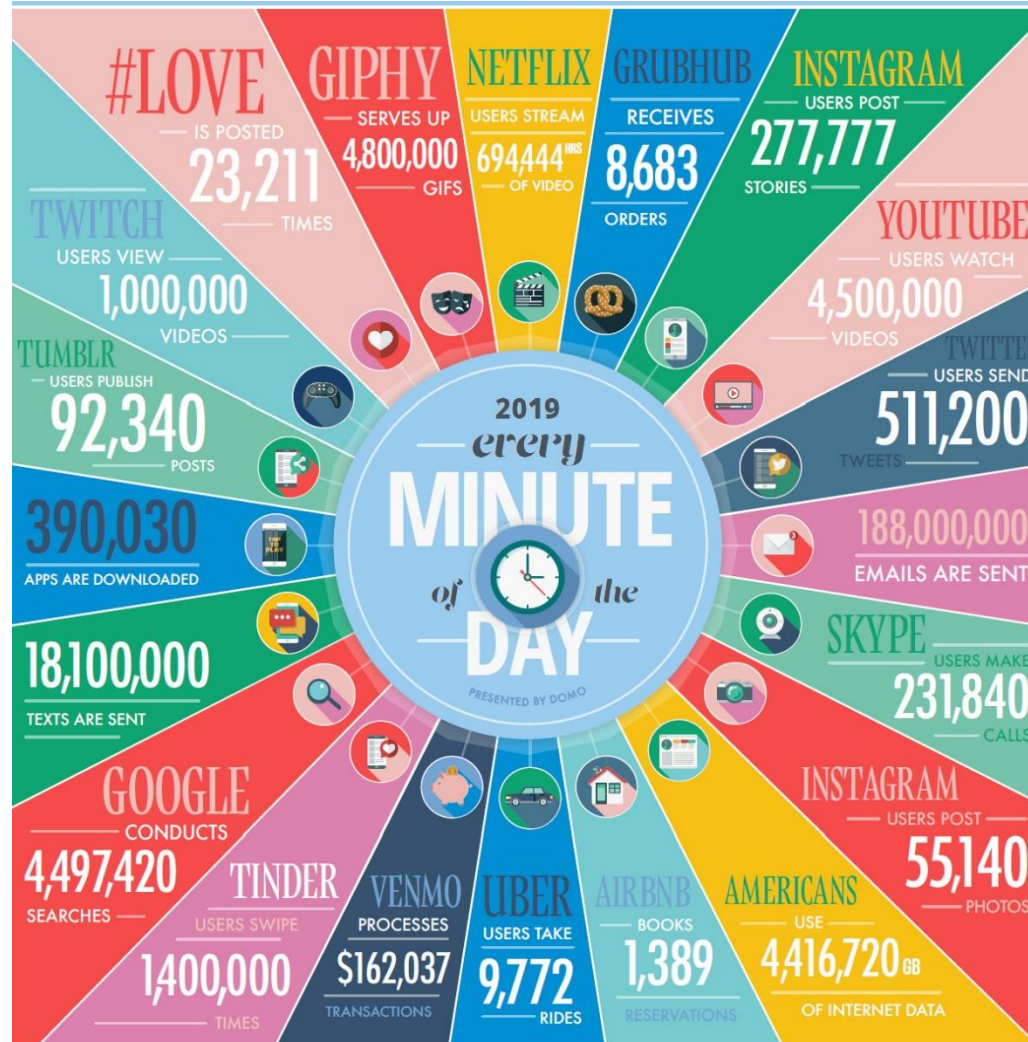




TECNOLÓGICO  
NACIONAL DE MÉXICO

# Instituto Tecnológico de Saltillo

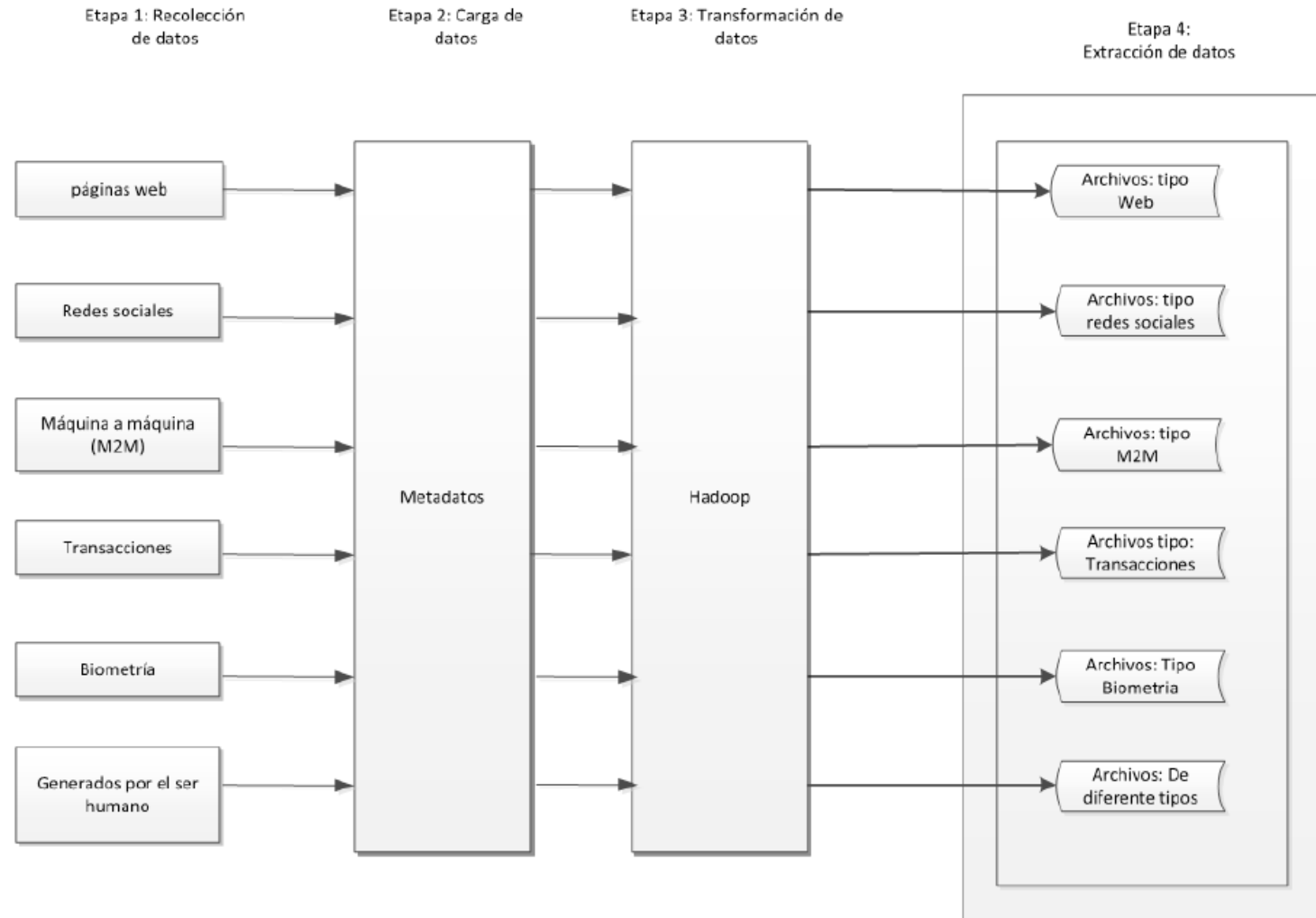
## Ingeniería en Sistemas Computacionales



BIG DATA



## Arquitectura de Big Data propuesta por Krishnan

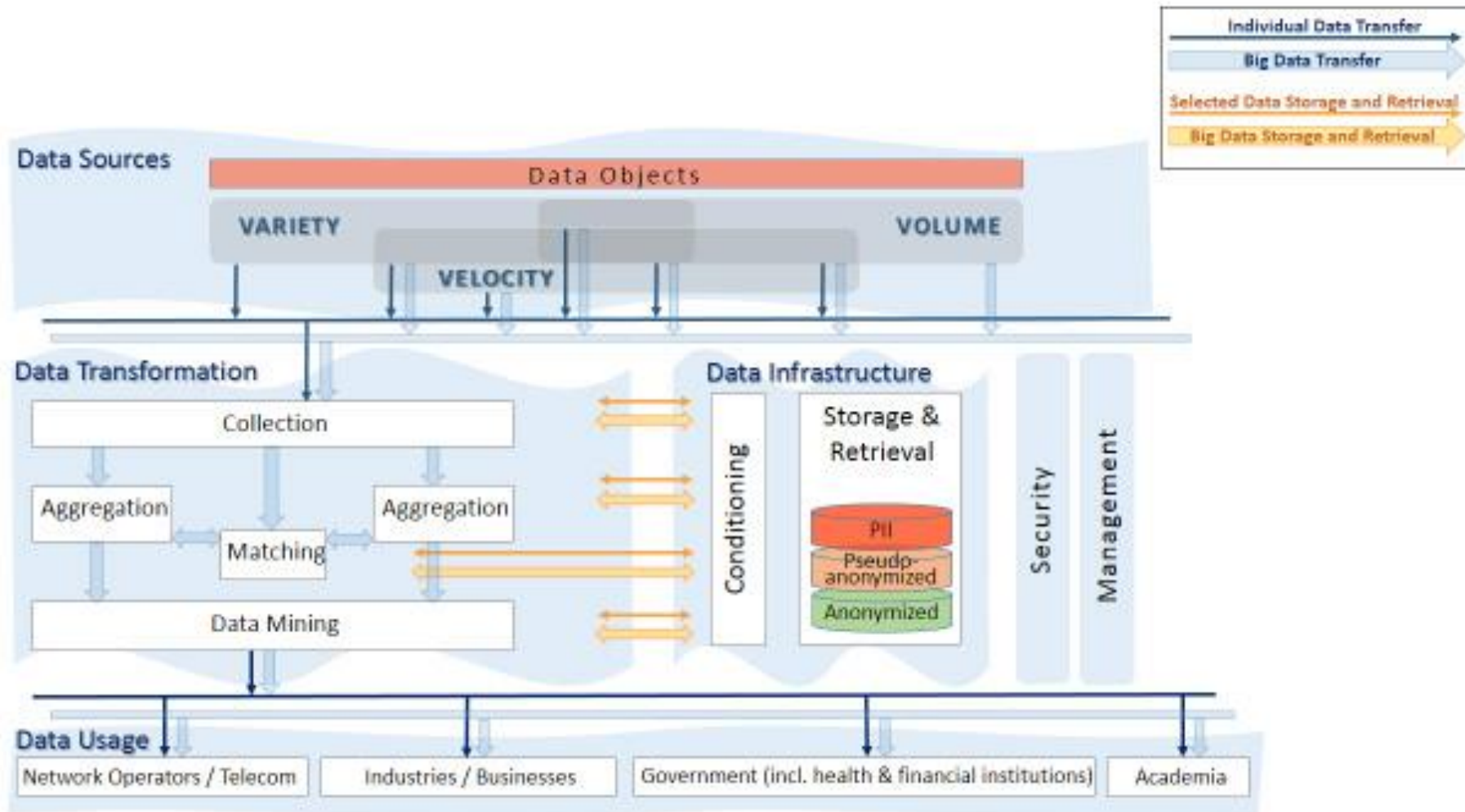


## Arquitectura de Big Data propuesta por Bob Marcus

- A. Fuentes de datos externos Suministra las entradas de datos externas y la producción de los componentes internos de Big Data
- B. Secuencia y procesamiento ETL.
  - Filtrar y transformar los flujos de datos provenientes de los recursos externos.
  - El procesamiento de datos que se realiza es el llamado “en movimiento” entre los almacenes de datos
- C. Fundación altamente escalable, base al grandes volúmenes de datos
  - El primero, a nivel de la infraestructura existe con el fin de poder atender el almacenamiento y procesamiento
  - El segundo se refiere a los almacenes de datos.
  - El tercero refiere a escalonar el procesamiento buscando aprovechar las ventajas de los almacenes de datos distribuidos escalables.
- D. Bases de datos operacionales y de analítica.
  - Bases de datos analíticas. Toma los datos procesados y escalonados de la sección anterior. Son altamente optimizadas para sola lectura.
  - Bases de datos operacionales. Mantienen una excelente operación en lectura y escritura en general de forma eficiente.
  - En memoria. Se refieren a los datos ubicados en memorias cachés.
- E. Análisis e interfaces de bases de datos.
  - Análisis de interfaces de procesos en lotes. Se refiere al tipo de interfaz usada para el procesamiento de datos que provienen en lotes o Batch.
  - Análisis de interfaces interactivas.
  - Análisis de interfaces en tiempo real.
- F. Aplicaciones e interfaces de usuario; Aplicaciones e interfaces de usuario, las cuales no deben ser algoritmos complejos, al usar grandes cantidades de datos distribuidos.
- G. Servicios de Apoyo; Diseñar, desarrollar e implementar herramientas, seguridad, gestión de procesos y recursos de datos



## Arquitectura de Big Data propuesta por Microsoft





## El ciclo del Big data



BIG DATA



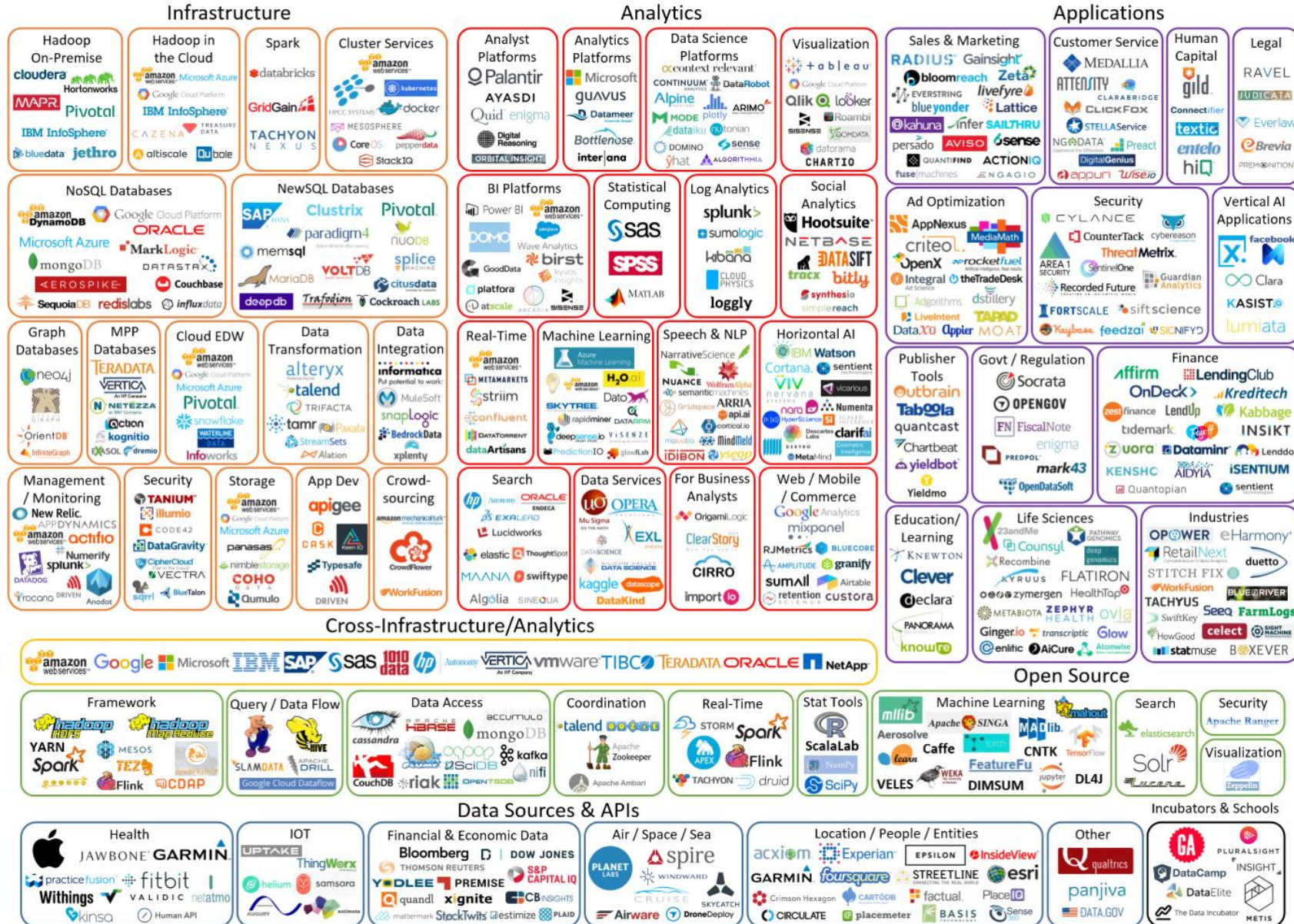




TECNOLOGICO  
NACIONAL DE MEXICO

# Instituto Tecnológico de Saltillo

## Ingeniería en Sistemas Computacionales



BIG DATA

## Tareas y Laboratorios

- Instalo manejador de base de datos NOSQL (MongoDB) y crear 3 estructuras de datos.
- Investigar y exponer un caso de estudio donde el diseño se planteó en base a un modelo NOSQL pudiendo ser diseñado con un modelo relacional o viceversa.
- Modelar una base de datos con el entorno de desarrollo diferente a MongoDB.

