



UTNMDP
Regional Mar del Plata

BIG DATA

BASE DE DATOS II - TECNICATURA EN
PROGRAMACIÓN

Di Meglio Melina,
Fernandez Brenda,
Castagnino Lucia
Docente: Eduardo Mónaco

Resumen: Big Data - Base de Datos II

1. ¿Qué es Big Data?

Big Data se refiere al tratamiento de grandes volúmenes de datos que, por su tamaño, complejidad y velocidad de generación, no pueden ser gestionados con tecnologías tradicionales. Va más allá del volumen, ya que implica nuevas formas de capturar, almacenar, procesar, analizar y visualizar datos para transformarlos en conocimiento útil y aplicable.

- Se genera a partir de diversas fuentes: redes sociales, sensores, cámaras, teléfonos, transacciones, etc.
- Tiene impacto en todos los niveles: individual, organizacional y estatal.
- Permite pasar del análisis histórico al análisis predictivo y prescriptivo, generando decisiones automáticas o semi-automáticas.

2. Las 5 V's del Big Data y sus extensiones

Las 5 V principales:

1. Volumen: magnitud de los datos generados (terabytes, petabytes).
2. Velocidad: rapidez con la que se generan, transmiten y deben procesarse.
3. Variiedad: diversidad de formatos (texto, video, audio, sensores, etc.).
4. Veracidad: confiabilidad y consistencia de los datos.
5. Valor: utilidad estratégica que se puede obtener del análisis de datos.

Extensiones propuestas:

- Viabilidad: capacidad técnica y económica para implementar soluciones Big Data.
- Visualización: herramientas para representar gráficamente los resultados.
- Variabilidad: los datos pueden cambiar de forma o significado según el contexto.
- Vulnerabilidad (*extra de la presentación*): alude a los riesgos de seguridad y privacidad.

3. Tipos de datos en Big Data

1. Estructurados: organizados (bases de datos relacionales, hojas de cálculo).
2. No estructurados: sin formato fijo (videos, audios, publicaciones, logs).
3. Semiestructurados: contienen etiquetas que organizan la información (JSON, XML, correos electrónicos).

Cada tipo requiere tecnologías distintas:

- Estructurados → Data warehouses
- No estructurados → Data lakes o bases orientadas a documentos
- Semiestructurados → Bases NoSQL con esquemas flexibles

4. Tecnologías y arquitecturas para Big Data

Bases de datos NoSQL:

- Clave-valor: Redis
- Columnas: HBase, BigTable
- Documentos: MongoDB, CouchDB

Tecnologías clave:

- Hadoop: procesamiento distribuido con HDFS + MapReduce.
- Apache Spark: análisis en memoria, ideal para tiempo real.
- Data Lakes: almacenan datos en bruto, de múltiples fuentes, sin transformarlos.
- Bases en la nube: escalables y accesibles (Google BigQuery, Amazon Redshift).

5. Desafíos en la gestión del Big Data

1. Calidad de los datos: ruido, duplicados, inconsistencias.
2. Seguridad y privacidad: el uso de datos sensibles requiere cifrado, autenticación, control de accesos.
3. Escalabilidad: los sistemas deben crecer horizontalmente (más nodos).

6. Desafíos éticos y sociales

- ¿Sabemos qué datos compartimos y quién los usa?
- ¿Existe reciprocidad entre lo que entregamos y lo que recibimos?
- Discriminación algorítmica: decisiones basadas en sesgos de los datos.
- Brechas de acceso: no todos tienen las mismas capacidades para acceder, analizar o monetizar datos.

Se requieren marcos legales sólidos, educación digital y regulación estatal y ética.

7. Análisis de datos y toma de decisiones

Tipos de análisis:

Tipo	Pregunta que responde	Herramientas
Descriptivo	¿Qué pasó?	Informes, dashboards
Diagnóstico	¿Por qué pasó?	Estadísticas, correlaciones
Predictivo	¿Qué pasará?	Machine learning
Prescriptivo	¿Qué debería hacerse?	Optimización, IA

Herramientas:

Lenguajes de programación:

- Python: el más popular en Big Data. Se usa para análisis de datos, machine learning, automatización y visualización (con librerías como Pandas, NumPy, Scikit-learn, etc.).
- Java: muy usado en sistemas grandes y en frameworks como Hadoop y Apache Spark.
- R: ideal para análisis estadístico y visualización de datos. Muy usado por científicos de datos.

Frameworks y Herramientas

- Apache Hadoop: permite procesar grandes volúmenes de datos distribuidos.
- Apache Spark: más rápido que Hadoop, permite procesamiento en tiempo real y batch.

Lenguajes para Bases de Datos

- SQL: lenguaje estándar para consultar y manejar bases de datos relacionales. Aún fundamental en entornos Big Data.

Plataformas en la Nube

- AWS (Amazon Web Services) y Azure (Microsoft): ofrecen servicios específicos para almacenar, procesar y analizar grandes volúmenes de datos.

Por ejemplo: Amazon Redshift, EMR, S3, Azure Synapse, etc.

8. Aplicaciones reales y casos de uso

- Salud: medicina personalizada, monitoreo en tiempo real, dispositivos wearables.

- Educación: Learning Analytics, predicción de abandono, planes personalizados.
- Gobierno: portales de datos abiertos, urbanismo inteligente, participación ciudadana.
- Industria: mantenimiento predictivo, reducción del consumo energético (ej. General Electric).
- Comercio y marketing: personalización de ofertas, análisis de comportamiento de clientes.
- Finanzas: detección de fraudes, gestión de riesgos.

9. Ventaja competitiva y cambio organizacional

- Big Data como activo estratégico: mejora procesos y crea nuevos modelos de negocio.
- Organizaciones basadas en datos: decisiones rápidas, precisas y en tiempo real.
- Cultura de datos: promueve transparencia, colaboración y empoderamiento de los empleados.
- Sustitución de KPIs tradicionales por paneles de monitoreo en tiempo real.

10. Impacto del Big Data en la sociedad

- Transformación del sistema de salud (atención continua vs. puntual).
- Mejora en transparencia y eficiencia de la gestión pública.
- Personalización educativa.
- Desarrollo sostenible y ciudades inteligentes.

Riesgo: que solo unos pocos concentren el poder de los datos, generando nuevas formas de desigualdad.

11. Conclusiones y proyección futura

Big Data no es solo una tendencia tecnológica, sino una transformación estructural de cómo funciona el mundo. Su combinación con:

- Inteligencia Artificial
- Machine Learning
- Cloud Computing

...hará que los sistemas sean más eficientes, pero exigirá más ética, transparencia y responsabilidad social.

Frase clave:

Los datos no resuelven problemas; las decisiones que tomamos a partir de ellos son las que marcan la diferencia.