

Taller — Comparación de arquitecturas Big Data

Propósito: analizar, comparar y documentar las principales **arquitecturas Big Data** (Hadoop, Spark, NoSQL, arquitecturas en la nube) evaluando sus componentes, ventajas, limitaciones y casos de uso reales.

Escenario guía (ejemplo): *SmartRetail*, una empresa de comercio electrónico que procesa grandes volúmenes de transacciones, reseñas de clientes y datos de IoT de sensores de inventario en tiempo real.

1) Entregable principal — Documento comparativo

Qué es: un documento técnico que describe y compara al menos 3 arquitecturas Big Data. Incluye **diagramas**, **tablas comparativas** y **ejemplos aplicados**.

1.1 Introducción

- **Propósito** — ¿Por qué es importante comparar arquitecturas Big Data?
- **Alcance** — Plataformas a comparar: Hadoop, Spark, y una arquitectura en la nube (ej. AWS o Azure).
- **Glosario** — Definir términos: HDFS, YARN, RDD, ETL, etc.

1.2 Descripción de arquitecturas

- **Hadoop** — HDFS, YARN, MapReduce.
- **Spark** — Procesamiento en memoria, APIs en Scala/Python, librerías (MLlib, SparkSQL).

- **Arquitecturas en la nube** — Servicios gestionados como AWS EMR, GCP BigQuery, Azure Synapse.

1.3 Tabla comparativa

Aspecto	Hadoop	Spark	Nube (ej. AWS)
Procesamiento	Batch con MapReduce	En memoria, batch y streaming	Batch y streaming gestionado
Escalabilidad	Horizontal manual	Horizontal con cluster manager	Automática, según demanda
Facilidad de uso	Alta complejidad	APIs más amigables	Interfaces simplificadas
Costos	Infraestructura propia	Cluster dedicado	Paga por uso
Casos de uso	Procesos batch grandes	Análisis interactivo, ML	Analítica empresarial en tiempo real

1.4 Requerimientos de implementación

- Hardware necesario.
- Configuración de clústeres.
- Dependencias de software.

Importante: justificar con métricas simples (tiempo de ejecución, costo estimado, complejidad de gestión).

2) Actividades del taller

1. Investigar y documentar 3 arquitecturas Big Data.
2. Elaborar una **tabla comparativa** como la sección 1.3.
3. Redactar un **caso de uso aplicado** en un sector específico (salud, retail, banca, ciudades inteligentes).

4. Crear un **diagrama arquitectónico** en PlantUML, [Draw.io](https://draw.io) o Lucidchart.
5. Concluir con una reflexión crítica: ¿qué arquitectura sería más adecuada para *SmartRetail* y por qué?

3) Instrucciones paso a paso para la entrega

1. **Selecciona un escenario** (ej. e-commerce, salud, banca).
2. Redacta la **introducción** con propósito, alcance y glosario.
3. Describe cada arquitectura con componentes principales.
4. Construye la **tabla comparativa**.
5. Incluye al menos **un caso de uso detallado**.
6. Diseña un **diagrama** que represente la arquitectura elegida.
7. Concluye con un **análisis crítico**.

4) Entregables

taller-arquitecturas-bigdata/

└ DocumentoComparativo	# Análisis detallado
└ TablaComparativa	# Tabla de comparación
└ CasoDeUso	# Caso práctico aplicado
└ DiagramaArquitectura	# Esquema visual
└ ReflexionCritica	# Conclusión del taller