

Curso

Herramientas de Computación en la Nube

Clase 4
Febrero 14 de 2026

Almacenamiento de Objetos y Arquitecturas tipo S3

Guía de estudio

Profesor
Alvaro Mauricio Montenegro Díaz, Ph.D.
Universidad de la Sabana

Febrero 2026

1. Propósito Académico de la Lección

Esta lección no trata sobre Amazon como proveedor. Tampoco trata sobre aprender comandos S3.

El objetivo es comprender:

- el **modelo arquitectónico del almacenamiento de objetos**,
- la **separación radical entre almacenamiento y cómputo**,
- las implicaciones de consistencia fuerte,
- la semántica de objetos inmutables,
- y el rol fundacional de S3 en el diseño moderno de sistemas de datos.

“Estudiar S3 no como un producto, sino como un modelo arquitectónico.”

Ese es el eje conceptual de la clase.

2. Competencias que el Estudiante Debe Desarrollar

Al finalizar la lección, el estudiante debe ser capaz de:

1. Diferenciar almacenamiento de objetos de filesystem y base de datos.
2. Explicar por qué no existen directorios reales en S3.
3. Analizar las consecuencias arquitectónicas del uso de prefijos.
4. Comprender la diferencia entre durabilidad y disponibilidad.
5. Explicar por qué S3 no es OLTP.
6. Argumentar por qué el almacenamiento y el cómputo deben desacoplarse.
7. Diseñar una organización lógica tipo data lake usando buckets.

3. Conceptos Fundamentales que Debe Dominar

A. Modelo Conceptual

- Object Storage
- Objeto (payload + metadatos + key)
- Bucket
- Key
- Prefijo
- Inmutabilidad lógica

B. Semántica y Consistencia

- Strong read-after-write consistency
- No transacciones multiobjeto
- Diseño idempotente
- Flujos reproducibles

C. Durabilidad vs Disponibilidad

Debe poder explicar:

- Por qué 11 nueves no es marketing sino diseño
- Cómo se logra (replicación, erasure coding, autosanación)
- Por qué durabilidad \neq disponibilidad

D. Economía del almacenamiento

- Clases de almacenamiento
- Lifecycle policies
- Control de costo por diseño

E. Seguridad y Gobernanza

Debe entender:

- IAM
- cifrado en reposo
- versionado
- retención legal (WORM)

F. Arquitectura de Datos Moderna

- Data Lake
- Lakehouse
- Separación almacenamiento–cómputo

- Motores externos (Spark, Trino, DuckDB)
- Storage is cheap, compute is ephemeral

4. Preguntas de Autoevaluación Técnica

Nivel Conceptual

1. ¿Por qué en S3 no existen directorios reales?
2. ¿Qué implica que mover un objeto sea copiar + borrar?
3. ¿Por qué S3 no soporta transacciones multiobjeto?
4. ¿Cuál es la diferencia conceptual entre durabilidad y disponibilidad?
5. ¿Por qué renombrar es una operación costosa?

Nivel Arquitectónico

6. ¿En qué casos NO deberías usar S3?
7. ¿Por qué el diseño de prefijos es una decisión arquitectónica?
8. ¿Qué significa desacoplar almacenamiento y cómputo?
9. ¿Por qué S3 no implementa joins?
10. ¿Qué ventajas aporta esta simplicidad para escalar?

5. Ejercicio de Diseño Arquitectónico

Escenario

Una empresa necesita:

- almacenar logs crudos,
- procesarlos diariamente,
- conservar histórico por 7 años,
- reducir costos a largo plazo.

Preguntas:

1. ¿Cómo organizarías los buckets?
2. ¿Qué prefijos usarías?
3. ¿Qué lifecycle policies aplicarías?
4. ¿Qué clases de almacenamiento utilizarías?

5. ¿Dónde viviría el cómputo?

6. Errores Conceptuales que Deben Evitarse

Un estudiante no ha comprendido la lección si:

- Cree que S3 es un filesystem.
- Cree que existen carpetas reales.
- Cree que es adecuado para OLTP.
- Confunde durabilidad con disponibilidad.
- Cree que el query ocurre dentro del almacenamiento.

7. Nivel de Profundización Esperado

Esta lección exige comprensión estructural, no operativa.

El estudiante debe ser capaz de:

- Diseñar un mini data lake conceptual.
- Explicar el impacto de inmutabilidad.
- Justificar separación storage/compute.
- Evaluar trade-offs frente a bases de datos.

No basta con repetir definiciones.

8. Reflexión Final de la Lección

El almacenamiento de objetos redefine la arquitectura de sistemas.

Al eliminar:

- jerarquía física,
- locking,
- transacciones complejas,

y conservar:

- durabilidad extrema,
- simplicidad semántica,
- API uniforme,

S3 se convierte en la base sobre la cual se construyen:

- data lakes,
- pipelines analíticos,
- sistemas de ML,
- arquitecturas lakehouse modernas.

Comprender este modelo no es aprender una tecnología. Es comprender la lógica estructural de la nube contemporánea.