

## Base de Datos II (IC4302) – Semestre 1, 2023

### Resumen #2 – Data Warehousing on AWS

David Suárez Acosta – 2020038304

---

#### Introducing Amazon Redshift

Amazon Redshift es una tecnología de almacenamiento de datos en la nube que es rápida, "fully managed" y de tamaño de datos petabyte, que soluciona problemas que existían en el pasado como lo era la creación de análisis y reportes para los usuarios de las empresas.

Esta tecnología tiene motores de almacenamiento de datos en paralelo y se ofrece más espacio de almacenamiento por menos precio.

#### Modern Analytics and data warehousing architecture

*Data warehouse:* son optimizadas para operaciones de lectura de volúmenes grandes de información.

*OLTP databases:* son optimizadas para escritura continua y altos volúmenes de operaciones pequeñas de lectura.

AWS analytics services permite la conversión rápida de información que va desde almacenamiento en la nube hasta lagos de datos sin server.

#### *Analytics architecture*

Un analytics pipeline está conformado por:

1. Collect data
2. Store data
3. Process data
4. Analyze and visualize data

#### Data warehouse technology options

##### *Row-oriented databases*

Estas bases de datos almacenan filas completas en un bloque físico, esto permite un alto rendimiento en las operaciones de lectura.

##### *Column-oriented databases*

Estas bases de datos organizan las columnas en bloques físicos, esto permite que sean más eficientes para consultas de solo lectura.

#### Amazon redshift deep dive

##### *Integración con lago de datos*

Amazon Redshift tiene una feature llamada Redshift Spectrum que facilita tanto consultas como escritura de datos en un lago de datos.

### *Rendimiento*

Amazon Redshift es rápido, tiene alto rendimiento y es flexible, esto se debe a su hardware de alto rendimiento, tecnología "AQUA", almacenamiento eficiente, views materializadas, la maximización del rendimiento gracias a machine learning y caching resultante.

### *Durabilidad y disponibilidad*

Para aumentar la durabilidad y disponibilidad, esta tecnología detecta y reemplaza nodos fallidos de un cluster automáticamente.

### *Elasticidad y escalabilidad*

Con la feature de "Elastic resize", se puede alterar el tamaño de un cluster fácilmente con solo añadir o eliminar nodos, mientras que con la feature "Concurrency scaling" se aguantan usuarios de concurrencia ilimitados así como consultas de concurrencia.

## **Operations**

### *Amazon Redshift Advisor*

Ayuda a mejorar el rendimiento y reducir los costos de los clusters.

### *Interfaces*

Amazon Redshift tiene drivers JDBC y ODBC personalizados que permite el uso de clientes SQL.