



Programa AWS Big Data – Analytics Specialist



“Big Data, el nuevo petróleo”

Reflexión de la noche



Coméntanos....



Sesión 2: No Sql - Streaming

Sql vs No SQL

	Traditional RDBMS	NoSQL
Data model	Fixed schema	Flexible schema
DB Transaction	Full ACID compliance	*No ACID compliance
Performance	Optimized for storage	Optimized for compute
Scale	Scale vertically	Scale horizontally

Tipos de NoSQL Databases

- ❏ Columnar data store
- ❏ Document data store
- ❏ Graph data store
- ❏ In-memory key-value store

NoSQL – Columnar Data Source

- Column-oriented storage (blocks)
- Improved analytics performance
- Fast query execution

Examples:

- Apache Cassandra
- Apache Hbase

SSN	Email	Age	City
12031582	abc@email.com	42	Boston
97281858	xyz@email.com	22	Los Angeles
85293827	pqr@email.com	29	Chicago

Block 1

12031582|97281858|85293827

Block 2

abc@email.com|xyz@email.com|pqr@email.com

Block 3

42|22|29

NoSQL – Document Data Source

- Stores semi-structured data (JSON)
- Flexible data model
- Nested data access
- E.g., MongoDB, Couchbase, Cassandra, MarkLogic

```
{
  "person_id" : "2132152"
  "SSN" : "12031582",
  "email" : abc@email.com,
  "age" : "42",
  "City" : "Boston",
  "Addresses" : [
    "700 Second St",
    "5 Lamp St"
  ],
  "medals_won":{
    "Running" : "5",
    "Tennis": "1"
  }
}
```

JSON

NoSQL- Document Data Source

Amazon DynamoDB



- **Documento NoSQL totalmente gestionado y con almacén de valores (key-value)**

Tablas no tienen esquema fijo.

- **Performance más alto**

Latencia de milisegundos de un solo dígito

Se ejecuta en unidades de estado sólido

- **Transacciones**

Incluye soporte para ACID transacciones.

- **Bajo demanda**

El modo de capacidad bajo demanda acomoda las cargas de trabajo a medida que aumentan o disminuyen

- Atomicidad: cada transacción es ejecución "todo o nada".
- Consistencia: todos los datos tienen que ser válidos
- Aislamiento: cada transacción es independiente de otra.
- Durabilidad: que una vez que se confirmó una transacción (commit), los datos quedaran persistidos

NoSQL- Document Data Source

Amazon DynamoDB – Data Types

Scalar

- Number, string, binary, Boolean, and null

```
FavoriteThings: ["Books", "Cars",  
3.14159]
```

Document

List

- Ordered collection of values
- No restrictions on data types stored inside a list

```
{  
  Day: "Friday",  
  ThingsToDo: 5,  
  ItemsOnMyDesk: [ "Pens", "Stick Notes" ]  
}
```

Map

- Unordered collection of name-value pairs

Sets

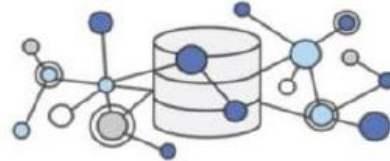
- Values in a set must be of the same type
- Values must be unique

```
["Red", "Green", "Blue"]
```

```
[24, 923, 2]
```

Amazon Neptune: Graph Database

- Build and run applications that work with highly connected datasets.
- Use open and popular graph query languages to execute powerful queries that are easy to write and perform well on connected data.
- Optimized for storing billions of relationships and querying the graph with milliseconds latency.
- Use cases:
 - Fraud detection
 - Knowledge graphs
 - Drug discovery
 - Network security



MUCHA TEORIA
A PRACTICAR



Gracias

