

---

# Event Hubs como Sistema de Tópicos en Azure

---

BIG DATA ACADEMY

# Event Hubs

Es el servicio propietario de Azure para gestión de tópicos basados en el estándar Kafka

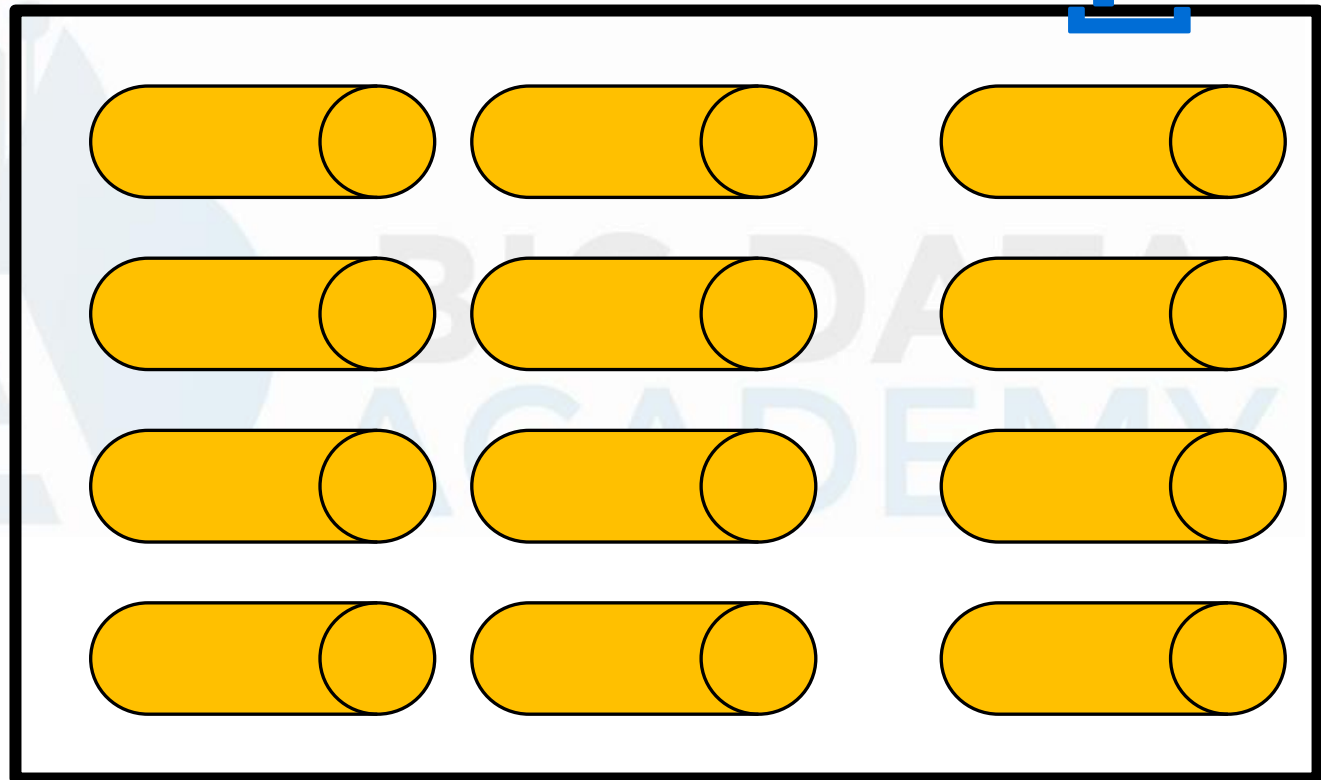


## Event Hubs

Es un servicio serverless, es decir no tendremos que administrar la infraestructura de Kafka ni de ZooKeeper

# Namespaces para agrupar los tópicos

Así como una base de datos agrupa tablas, en Event Hubs un namespace agrupa uno o varios tópicos.



# Unidades de Procesamiento

En Azure, una unidad de procesamiento puede procesar hasta 1000 registros por segundo



Por ejemplo, si se esperan 4500 registros por segundo, necesitaríamos 5 unidades de procesamiento

# Directiva de limpieza

Como los registros se almacenan en memoria RAM, y la memoria RAM es volátil, **Azure los elimina luego de cierto tiempo** que debemos configurar.

El tiempo de eliminación puede variar desde 1 hora hasta 90 días

Como los procesos de tiempo real están orientados a procesarlos en algunos segundos o minutos, **el tiempo mínimo recomendado es de 1 día (para desarrollo) y 7 días (para producción)**

```
{"ID_PERSONA": "1", "ID_EMPRESA": "6", "MONTO": "1935"}
```

```
{"ID_PERSONA": "3", "ID_EMPRESA": "1", "MONTO": "2993"}
```

```
{"ID_PERSONA": "7", "ID_EMPRESA": "8", "MONTO": "7956"}
```

...

```
{"ID_PERSONA": "9", "ID_EMPRESA": "2", "MONTO": "1540"}
```



Event Hubs

# Planes tarifarios

En Event Hubs, existen tres planes tarifarios

## Básico

- **Retención de mensajes:** Hasta 1 día
- **Costo:** 11.16\$ por cada unidad de procesamiento
- **Recomendado:** Para entornos de **desarrollo**

## Estándar

- **Retención de mensajes:** Hasta 7 días
- **Costo:** 22.32\$ por cada unidad de procesamiento
- **Recomendado:** Para entornos de **producción**

## Premium

- **Retención de mensajes:** Hasta 90 días
- **Costo:** 764.09\$ por cada unidad de procesamiento
- **Recomendado:** Para entornos **críticos** (p.e., para almacenamiento de transacciones bancarias)

# Procesamiento Real-Time con PySpark en Azure

