

Actividad 2: Spark Streaming y Kafka

Parte 2

- 1) Iniciamos el flujo de transmisión de datos y ejecutamos la sentencia SQL para observar como tenemos el dataframe inicialmente.

```
SSH en el navegador SUBIR ARCHIVO DESCARGAR ARCHIVO
Linux ipmd-act-cluster-m 5.10.0-0.deb10.16-cloud-amd64 #1 SMP Debian 5.10.127-2~bpo10+1 (2022-07-28) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
araceli.ruiz2002@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-topics.sh --zookeeper localhost:2181 --create
--replication-factor 1 --partitions 1 --topic retrasos
Created topic retrasos.
araceli.ruiz2002@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-topics.sh --zookeeper localhost:2181 --list
retrasos
araceli.ruiz2002@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-console-producer.sh --broker-list ipmd-act-clu
ster-w-0:9092 --topic retrasos
>
+----+-----+
|dest|retraso_medio|
+----+-----+
+----+-----+
```

- 2) Enviamos el primer mensaje, y vemos como cambia el dataframe

```
SSH en el navegador SUBIR ARCHIVO DESCARGAR ARCHIVO
Linux ipmd-act-cluster-m 5.10.0-0.deb10.16-cloud-amd64 #1 SMP Debian 5.10.127-2~bpo10+1 (2022-07-28) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
araceli.ruiz2002@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-topics.sh --zookeeper localhost:2181 --create
--replication-factor 1 --partitions 1 --topic retrasos
Created topic retrasos.
araceli.ruiz2002@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-topics.sh --zookeeper localhost:2181 --list
retrasos
araceli.ruiz2002@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-console-producer.sh --broker-list ipmd-act-clu
ster-w-0:9092 --topic retrasos
>{"dest": "GRX", "arr_delay": 2.6}
>
+----+-----+
|dest|retraso_medio|
+----+-----+
| GRX|          2.6|
+----+-----+
```

- 3) Enviamos el segundo mensaje

```
SSH en el navegador SUBIR ARCHIVO DESCARGAR ARCHIVO
Linux ipmd-act-cluster-m 5.10.0-0.deb10.16-cloud-amd64 #1 SMP Debian 5.10.127-2~bpo10+1 (2022-07-28) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
araceli.ruiz2002@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-topics.sh --zookeeper localhost:2181 --create --replica
tion-factor 1 --partitions 1 --topic retrasos
Created topic retrasos.
araceli.ruiz2002@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-topics.sh --zookeeper localhost:2181 --list
retrasos
araceli.ruiz2002@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-console-producer.sh --broker-list ipmd-act-clu
ster-w-0:9092 --topic retrasos
>{"dest": "GRX", "arr_delay": 2.6}
>{"dest": "MAD", "arr_delay": 5.4}
>
```

```

+----+-----+
|dest|retraso_medio|
+----+-----+
| MAD|          5.4|
| GRX|          2.6|
+----+-----+

```

4) Enviamos el tercer mensaje



SSH en el navegador

Linux ipmd-act-cluster-m 5.10.0-0.deb10.16-cloud-amd64 #1 SMP Debian 5.10.127-2-bpo10+1 (2022-07-28) x86_64

The programs included with the Debian GNU/Linux system are free software; the exact distribution terms for each program are described in the individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent permitted by applicable law.

```

araceli@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-topics.sh --zookeeper localhost:2181 --create --replication-factor 1 --partitions 1 --topic retrasos
Created topic retrasos.
araceli@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-topics.sh --zookeeper localhost:2181 --list
retrasos
araceli@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-console-producer.sh --broker-list ipmd-act-cluster-w-0:9092 --topic retrasos
>{"dest": "GRX", "arr_delay": 2.6}
>{"dest": "MAD", "arr_delay": 5.4}
>{"dest": "GRX", "arr_delay": 1.5}
>

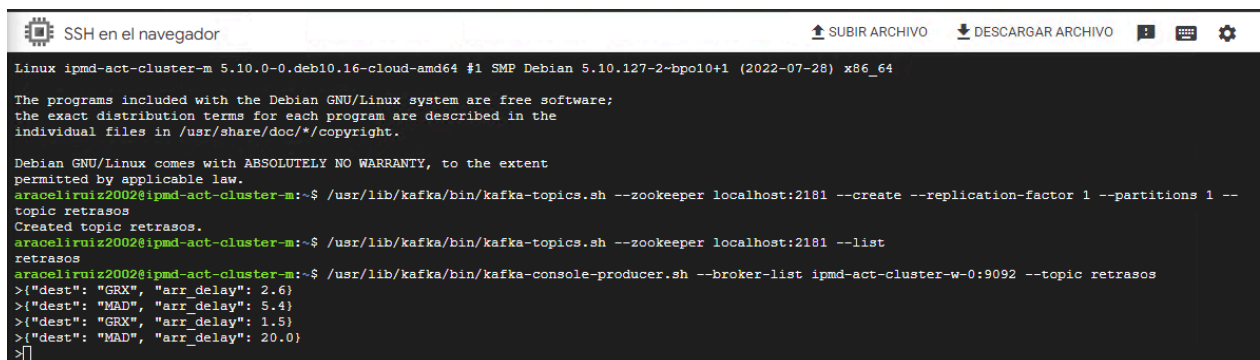
```

```

+----+-----+
|dest|retraso_medio|
+----+-----+
| MAD|          5.4|
| GRX|          2.05|
+----+-----+

```

a) Enviamos el cuarto mensaje



SSH en el navegador

Linux ipmd-act-cluster-m 5.10.0-0.deb10.16-cloud-amd64 #1 SMP Debian 5.10.127-2-bpo10+1 (2022-07-28) x86_64

The programs included with the Debian GNU/Linux system are free software; the exact distribution terms for each program are described in the individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent permitted by applicable law.

```

araceli@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-topics.sh --zookeeper localhost:2181 --create --replication-factor 1 --partitions 1 --topic retrasos
Created topic retrasos.
araceli@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-topics.sh --zookeeper localhost:2181 --list
retrasos
araceli@ipmd-act-cluster-m:~$ /usr/lib/kafka/bin/kafka-console-producer.sh --broker-list ipmd-act-cluster-w-0:9092 --topic retrasos
>{"dest": "GRX", "arr_delay": 2.6}
>{"dest": "MAD", "arr_delay": 5.4}
>{"dest": "GRX", "arr_delay": 1.5}
>{"dest": "MAD", "arr_delay": 20.0}
>

```

```

+----+-----+
|dest|retraso_medio|
+----+-----+
| MAD|         12.7|
| GRX|          2.05|
+----+-----+

```