



Universidad Carlos III

Arquitectura de Datos

Curso 2024-25

## Práctica 2

Resultados del desarrollo de la migración de la base de datos a  
Cassandra

**Ingeniería Informática, Cuarto curso**

Adrián Fernández Galán (NIA: 100472182, e-mail: 100472182@alumnos.uc3m.es)

César López Mantecón (NIA: 100472092, e-mail: 100472092@alumnos.uc3m.es)

Manuel Gómez-Plana Rodríguez (NIA: 100472310, e-mail: 100472310@alumnos.uc3m.es)

**Prof . Lourdes Moreno López**

**Grupo: 81**

# Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Prueba de las queries</b>	<b>2</b>
2.1. Query de las funciones operativas . . . . .	2
2.2. Querys del análisis estadístico 1 . . . . .	2
2.3. Querys del análisis estadístico 2 . . . . .	3
2.4. Querys del análisis estadístico 3 . . . . .	4
<b>3. Conclusiones</b>	<b>5</b>

## 1. Introducción

En este documento se recoge los resultados obtenidos durante el desarrollo de la práctica 2 de la asignatura *Arquitectura de Datos*. En esta práctica se tratará de completar una migración de una base de datos desde **MongoDB** a **Cassandra**. Además, se computarán nuevas tablas con el fin de permitir el análisis estadístico, aprovechando las cualidades de **Cassandra** para el análisis de datos gracias a su capacidad para la consulta masiva de datos de una misma columna.

En el diseño de consultas se tratará de aprovechar al máximo las capacidades de **Cassandra** en la lectura y escritura, dejando a la aplicación otra clase de operaciones. De esta forma, ambos sistemas trabajarán en conjunto, garantizando la eficiencia de la aplicación.

## 2. Prueba de las queries

Para probar que las tablas de cassandra funcionan, es necesario comprobar que las queries explicadas con anterioridad funcionan de acuerdo a lo exigido por los casos de uso en el enunciado. Así, se ejecutarán las queries y se mostraran algunas filas que sirvan como ejemplo de salida.

### 2.1. Query de las funciones operativas

Para los resultados de las consultas de las funciones operativas se han eliminado algunas características con el objetivo de mostrar solo la información más relevante. Se han obtenido los siguientes resultados para la query de la función operativa sobre las sancione generadas para este DNI = 63050842E:

dni_deudor	tipo	fecha_grabacion	cantidad	estado	matricula
63050842E	discrepancia carne	2013-05-09 18:53:14	1000	stand by	2955AUA
63050842E	discrepancia carne	2013-05-10 18:53:14	1000	stand by	2955AUA
63050842E	discrepancia carne	2013-05-29 10:35:28	1000	stand by	2955AUA
63050842E	impago	2008-03-12 00:46:59	140	stand by	2955AUA
63050842E	impago	2008-03-12 02:04:01	190	stand by	2955AUA
63050842E	impago	2008-09-09 14:26:53	370	stand by	2955AUA

Como se puede observar, se muestra la información más relevante relacionada con las sanciones del DNI: 63050842E

Se han obtenido los siguientes resultados para la query de la función operativa sobre los expedientes activos sin plazo de pago cerrado:

dni_deudor	tipo	fecha_grabacion	estado	matricula
63050842E	discrepancia carne	2008-03-12 02:04:01	stand by	2955AUA
63050842E	discrepancia carne	2008-08-05 09:10:31	stand by	2955AUA
63050842E	discrepancia carne	2008-09-09 14:26:53	stand by	2955AUA
63050842E	discrepancia carne	2008-09-14 20:23:03	stand by	2955AUA
63050842E	discrepancia carne	2009-02-13 14:08:57	stand by	2955AUA
63050842E	discrepancia carne	2009-02-27 20:08:44	stand by	2955AUA

Como se puede observar, todas las filas tienen el estado “stand by”, que cumple con lo pedido en el enunciado.

### 2.2. Querys del análisis estadístico 1

Se han obtenido los siguientes resultados para las queries del análisis estadístico 1:

- **Multas por marca y modelo**

marca	modelo	total_multas
Zitron	Tangerine	1930
Bemev	Berlin	2528
Rinaul	Fuente	3370
Zitron	Orange	1178
Escola	Tercius	841

Como se puede observar, se agrupan el número de sanciones por marca y modelo de vehículo de una manera correcta.

#### ■ Multas por color

color	total_multas
azul	1876
dorado metalizado	1352
ocre metalizado	162
gris	3989
marron metalizado	464

Como se puede observar, se agrupan el número de sanciones por color de vehículo de una manera correcta.

#### ■ Multas de velocidad por marca y modelo

marca	modelo	total_multas
Zitron	Tangerine	1775
Bemev	Berlin	2304
Rinaul	Fuente	3076
Zitron	Orange	1071
Escola	Tercius	753

Como se puede observar, se agrupan el número de sanciones de velocidad por marca y modelo de vehículo de una manera correcta. Además, el número de multas es menor al de la tabla de multas por marca y modelo, pudiendo asumir que la selección por el tipo “velocidad” ha sido efectiva.

## 2.3. Querys del análisis estadístico 2

Se han obtenido los siguientes resultados para las querys del análisis estadístico 2:

#### ■ Exceso de velocidad medio

carretera	media_velocidad_registrada	media_velocidad_radar
A2	116	100
A3	116	100
M30	66	50
M45	96	80
M40	96	80
A6	116	100
A4	116	100
A1	117	100
M50	117	100
A5	116	100

Estos resultados permiten obtener en la capa de aplicación un porcentaje dividiendo el resultado de “media velocidad registrada” entre “media velocidad radar”.

- **Tramo y sentido mas conflictivo**

carretera	kilometro	sentido	infracciones_tramo
M45	28	ascending	202
M40	43	descending	421
A1	218	ascending	184
A5	53	ascending	263
M30	11	ascending	702

La query muestra correctamente el número de infracciones por tramo y dirección de cada carretera, pudiendo escoger el mayor en la capa de aplicación.

## 2.4. Querys del análisis estadístico 3

Se han obtenido los siguientes resultados para las querys del análisis estadístico 3:

- **Conductores más infractores**

dni_deudor	num_multas
63050842E	525
55074832K	435
78135349W	472
12519376Q	431
28232810B	837

La query funciona correctamente, agrupando el número de sanciones por el dni del deudor.

- **Probabilidad de infracción cuando el conductor es distinto al propietario**

conductor_igual_propietario	count
False	29656
True	19069

La query funciona correctamente, devolviendo dos filas con las que calcular un porcentaje en la capa de aplicación.

### **3. Conclusiones**

Las tablas creadas y las consultas diseñadas cumplen con los casos de uso planteados inicialmente, devolviendo los valores esperados y mostrando que el modelo de datos es funcional y alineado con los objetivos establecidos.