



**UNIVERSIDAD CATÓLICA DE HONDURAS**

**“NUESTRA SEÑORA REINA DE LA PAZ”**

**Bases de Datos Organizacionales.**

**Integrantes:**

Jessie Martel Gaboriell - 1104200300015

Gustavo Alfredo Cano - 0101200003655

Carlos Alejandro Navarro - 0101200003739

Hugo Daniel Sanabria - 0101200000147

David Fernando Morales - 1503199902656

**Grupo #2**

**2025**

## Tabla de Contenido

Introducción	3
Objetivos	4
Objetivo General	4
Objetivos Específicos	4
Análisis Técnico y Diseño del Sistema	5
Descripción del problema	5
Volumen de datos estimado	6
Justificación Técnica	7
Fuente de datos a utilizar.	8
Análisis Técnico y Selección de Tecnología	8
Modelo de Base de Datos: NoSQL (Column Family)	9
Formalismo técnico de Base de Datos Distribuidas	9
Relación con CAP Theorem	10
Análisis de Diseño de Base de Datos	11
Estructura de Datos	11
Diagrama ER Modelo de Trazabilidad de Productos	12
Implementación de Queries en Casandra	13
Diseño del Sistema	27
Trello	27
Github	27
Google Colab: Python + Pandas	28
Pinggy	28
Contenedores y Orquestación	28
Scripts de Inicialización	29
Arquitectura General	29
Ejemplos de Modelo de Datos en Cassandra	29
Desarrollo del Sistema	31
Ejecución del Proyecto	31
Google Colab	31
Estructura	31
Resultados	32
Conclusiones	39
Referencias	40

## **Introducción**

El presente documento corresponde al informe técnico del Proyecto Capstone del curso Bases de Datos Organizacionales, cuyo tema está enfocado en la trazabilidad de productos en la cadena de suministro, con el tema brindado podemos comprender que necesitamos encontrar o demostrar el recorrido de un producto a lo largo de su cadena de suministro, es decir, desde su origen hasta la entrega a su consumidor final. El punto principal de nuestro proyecto es poder implementar un sistema backend capaz de gestionar esta trazabilidad de productos en una cadena de suministro utilizando tecnologías modernas de bases de datos distribuidas.

La solución desarrollada por el equipo Bit Bang Crew consiste en utilizar un Notebook construido en Google Colab, el cual utiliza Apache Cassandra como base de datos NoSQL distribuida en la cual se pueden llegar a tener una gran cantidad de datos, los cuales son utilizados para el almacenamiento de eventos de trazabilidad.

Este sistema se ejecuta mediante Docker y Docker Compose, lo cual facilita el despliegue, la portabilidad y la reproducibilidad del entorno.

Dentro de este informe se encontrarán componentes los cuales consideramos esenciales tales como el análisis técnico, diseño, arquitectura, resultados y conclusiones, sirviendo como guía, plantilla y referencia técnica del proyecto.

## **Objetivos**

### **Objetivo General**

Desarrollar un sistema backend capaz de registrar, consultar y gestionar información relacionada a la trazabilidad de productos dentro de una cadena de suministro utilizando tecnología NoSQL distribuida

### **Objetivos Específicos**

Implementar un sistema de trazabilidad basado en Apache Cassandra como base de datos NoSQL distribuida, utilizando Docker y scripts de inicialización para garantizar un entorno estable, escalable y replicable.

Desarrollar un entorno de análisis en Google Colab capaz de conectarse a Cassandra, procesar datos externos provenientes de la API de Open Food Facts e integrarlos al sistema para realizar consultas, reportes y visualizaciones operativas.

Diseñar y documentar un modelo de datos funcional que permita registrar productos, proveedores, lotes y eventos logísticos, complementándolo con dashboards y herramientas colaborativas como GitHub y Trello para la gestión del proyecto.

## **Análisis Técnico y Diseño del Sistema**

### **Descripción del problema**

Hoy en día en las gestiones de cadenas de suministros el seguimiento de los productos juega un papel fundamental en el logro de una eficiencia y confiabilidad operativa en cada etapa del proceso de logística, sin embargo, muchas compañías se encuentran restringidas o no cuentan con sistemas capaces que les permitan registrar, almacenar y consultar con efectividad las rutas de sus productos desde sus respectivas fuentes de origen hasta su destino final.

En la actualidad algunos sistemas limitan en cierto punto la escalabilidad, disponibilidad u operación en grandes volúmenes de información que pueden ser generados en múltiples actividades, esto llega a un punto de limitar la obtención de datos en tiempo real, lo cual puede presentar una dificultad ante alguna incidencia o un control de auditorías, aplicándolo en este contexto surgió la necesidad de buscar un modelo que facilite la gestión de información en grandes cantidades con el cual se nos permita tener un registro confiable de cada evento que se realice en la cadena de suministro.

Con esto nos centramos en las bases de datos NoSQL debido a que estas nos ofrecen una arquitectura escalable, el cual también es tolerante a fallos, lo cual resulta de una forma casi perfecta para poder implementar en un sistema de seguimiento de productos eficiente, el problema principal que este proyecto trata de abordar es encontrar una solución basada en un sistema de bases de datos NoSQL el cual nos permita demostrar la trazabilidad de los productos optimizando las consultas y el seguimiento de toda información que sea solicitada.

## **Volumen de datos estimado**

Se estima que la base de datos gestionará un volumen de datos de más de 5 mil registros distribuidos entre las diferentes entidades. El proyecto está orientado al análisis de información y trazabilidad de los productos, lo que involucra un volumen considerable de información. Dado que el sistema es de trazabilidad y análisis de información, este volumen de datos tiende a ir incrementando progresivamente a medida se van incorporando nuevos productos.

Tipo de operaciones frecuentes.

- Consultas agregadas, por rango de fechas y códigos de barra.
- Representación analítica para medir los indicadores de trazabilidad y flujos de información adicional.
- Representación gráfica por medio de dashboards, para seguimiento operativo.

En consecuencia, el tipo de carga de trabajo predominante será de lectura intensiva y orientada al análisis de datos (OLAP), aunque también incluirá operaciones periódicas de inserción y actualización derivadas de la captura continua de eventos de trazabilidad.

## Justificación Técnica

Se utiliza base de datos NoSQL por las siguientes razones:

**1. Volumen masivo de datos.** Cada producto y lote genera una cantidad enorme de datos ya sea por evento o estado, esto generará miles o millones de registros que a larga las bases de datos SQL no están diseñadas para ello, en cambio NoSQL sí.

**2. Costos.** En términos generales una base de datos NoSQL va a escalar horizontalmente con réplicas en situaciones de alta demanda de datos, en cambio SQL escala verticalmente lo cual impacta en nuestro bolsillo. Adicionalmente nuestra elección, Cassandra tiene escalabilidad lineal y por ende barata.

**3. Flexibilidad.** Podemos iniciar con una base de datos con pocos modelos y atributos, pero conforme el proyecto avanza surgen nuevas funcionalidades que impactan en la modificación de modelos, esto NoSQL lo trabaja a la perfección dando la flexibilidad suficiente.

**4. Disponibilidad.** Un sistema logístico como este requiere estar disponible el 99.9% del tiempo, por ello una base que pueda escalar horizontalmente es necesario, esto lo cumplirá Cassandra administrando varios nodos del aplicativo en distintos puntos geográficos.

## **Fuente de datos a utilizar.**

La API permite realizar consultas por código de barras o criterios específicos, las respuestas se obtienen mediante un formato JSON, el cual puede ser adaptado para ser integrado con la base de datos y así alimentar los análisis de trazabilidad que se tienen estipulados realizar. Esta API es de la propiedad de Open Food Facts, la cual es de uso abierto para diferentes propósitos, entre ellos tenemos la búsqueda de mejores selecciones en cuanto a comidas. Aparte de ofrecer su API, Open Food Facts ofrece opciones de descarga masiva de su dataset en varios formatos lo cual facilita la creación de dataset locales para realizar los análisis preferidos por los interesados. Enlace del api utilizada: <https://openfoodfacts.github.io/openfoodfacts-server/api/ref-v2/#get-api/v2/product/-barcode>

## **Análisis Técnico y Selección de Tecnología**

La trazabilidad en una cadena de suministro permite seguir el recorrido, origen y eventos asociados a un producto desde su proveedor hasta su entrega final. Esto requiere almacenar eventos en tiempo real y consultarlos de forma eficiente.

Apache Cassandra, como sistema NoSQL distribuido, ofrece:

- Alta disponibilidad (AP según CAP)
- Escalabilidad horizontal
- Escrituras rápidas
- Replicación automática entre nodos

Google Colab permite conexión con casi cualquier base de datos moderna en pocos pasos, además de tener un repertorio de librerías que facilitan su uso y una comunidad activa que nutre y comparte muchos casos de usos posibles en diferentes tipos de proyecto.

Docker facilita la creación de ambientes replicables y la base de datos dentro de contenedores aislados los cuales pueden ser implementados por cualquier miembro del equipo.

### **Modelo de Base de Datos: NoSQL (Column Family)**

Se seleccionó un modelo NoSQL distribuido, adecuado para almacenar eventos independientes, registrar grandes volúmenes de información y mantener disponibilidad constante.

### **Formalismo técnico de Base de Datos Distribuidas**

El sistema se implementa en un modelo NoSQL específicamente con Cassandra. Este modelo es adecuado para datos semiestructurados y orientado a operaciones de lectura/escritura de alta velocidad. Esta estructura permite distribuir automáticamente los datos entre nodos, manteniendo baja latencia incluso ante incrementos espontáneos de información a la base de datos.

La naturaleza del dataset el cual es relacionado a eventos independientes, registros de actividad y trazabilidad de entidades esto se ajusta a un almacenamiento distribuido donde cada evento puede almacenarse sin requerir transacciones complejas ni restricciones relacionales estrictas.

Cassandra fue seleccionado como sistema gestor de bases de datos distribuidas debido a sus características:

1. Debido a su descentralización de nodos, el cual permite operar incluso cuando alguno o varios nodos fallen.
2. Replicación automática de nodos.
3. Estructuras automatizadas, donde la escritura es altamente eficiente y por consiguientemente consistentemente eventual, además que permite registrar eventos de trazabilidad.
4. Escalabilidad horizontal, permitiendo añadir nodos con un costo menor que una escalabilidad vertical.

En síntesis, Cassandra fue seleccionado debido a su alta disponibilidad y tolerancia a fallos, en conjunto a una consistencia eventual la cual es permitida en esta naturaleza de proyecto.

### **Relación con CAP Theorem**

- A (Availability): el sistema nunca deja de aceptar escrituras.
- P (Partition Tolerance): tolera fallas entre nodos.
- C (Consistency): se maneja consistencia eventual, suficiente para trazabilidad

## Análisis de Diseño de Base de Datos

### Estructura de Datos

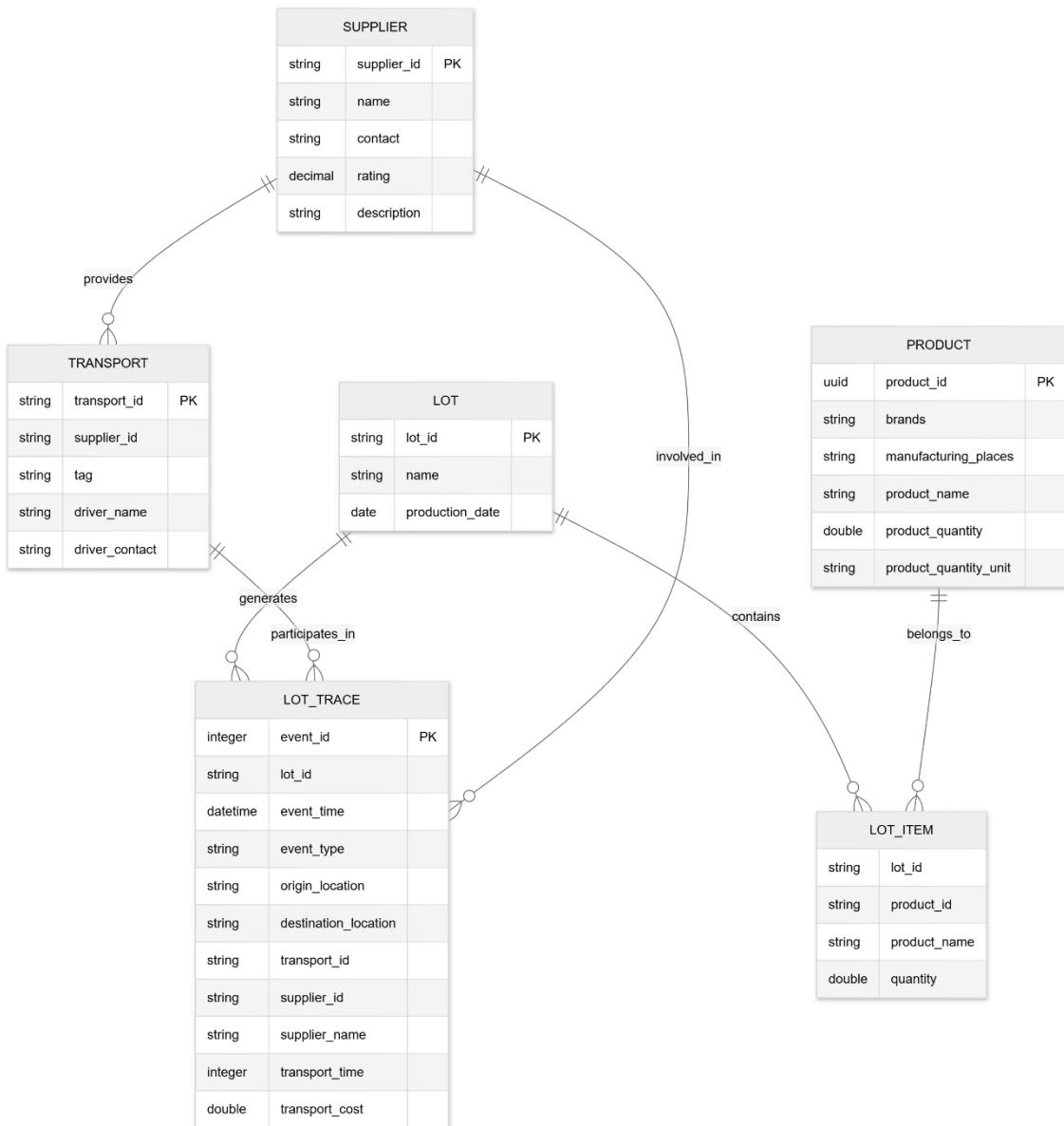
El sistema se compone de cinco tablas principales, las cuales almacenan información relacionada con proveedores, transportes, lotes de producción, productos asociados y eventos de trazabilidad.

- SUPPLIER: Contiene los datos de los proveedores (supplier\_id, name, contact, rating, description).
- TRANSPORT: Relaciona los transportes con cada proveedor (transport\_id, supplier\_id, tag, driver\_name, driver\_contact).
- LOT: Representa los lotes de producción (lot\_id, name, production\_date).
- LOT\_ITEM: Define la relación entre lotes y productos (lot\_id, product\_id, product\_name, quantity).
- LOT\_TRACE: Registra los eventos logísticos (event\_id, lot\_id, event\_time, event\_type, locations, costos, etc.).

## Diagrama ER Modelo de Trazabilidad de Productos

**Figura 1**

Diagrama ER del modelo de trazabilidad de productos



**Nota.** Elaborado por el grupo en <https://www.mermaidchart.com/d/7f113b26-68ce-4800-85c2-33758ad9a500>

## Implementación de Queries en Casandra

### Creación Tabla Products

```
session.execute("""  
    CREATE TABLE IF NOT EXISTS products (  
        product_id uuid PRIMARY KEY,  
        brands text,  
        manufacturing_places text,  
        product_name text,  
        product_quantity double,  
        product_quantity_unit text);  
    """")
```

**Tabla 1**

*Product Table*

product_id	brands	manufacturing_places	product_name	product_quantity	product_quantity_unit
7a734d25-acaa8-4567-b728-e910fbab7bd7	maroc.		Sidi Ali	330	ml
f91e73e8-b7c3-485ea4f8-ea646e2fd12c	perly	Maroc	perly	100	g

ef1d70fb-ff10-42aa-ae6c-1c041210af5e	sidi ali		Sidi Ali	2000	ml
75e51f38-2045-40bf-914c-6f37196806f9	Les Eaux Minérales d'oulmès,Sidi Ali		Eau minérale naturelle	1500	ml
26d18ae5-0604-4a32-bc42-bfdd721da8bf	AQUAFINA	Casablanca Morocco	اكوافينا	330	ml
17c3706a-ba3b-4fdc-b435-efc76dcacf4	Jaouda		Lait	450	ml
76f53234-8057-43aa-a0f2-7ec884b29635	Cristaline	Saint-Martin de Gurson, France,24610	Eau De Source	1500	ml
fd382374-4d4c-480a-	MILKY FOOD	Maroc	Fromage blanc nature	1000	g

9e19- ec44f2f1c bed	PROFESSIONAL				
ca7c557e- 84e5- 45d0- 9239- a693af47d 96d	Jaouda		uht jaouda 1L	1000	ml
bd385c1f- de2d- 4794- 84ab- 876cfec53 28c	La vache qui rit				
3b1806bf- 3393- 47e7- 8b47- 7de390c7c 0f2	Takis		Takis volcano	55	g
8df66d4f- 1ada- 4492- 915d- 607696b6 3e85	Gut & Günstig		Tomaten	500	g
074d0d79- 3f5e- 449e- b923-	Dobele		Cornetti rigatti		

5994b60b cbd2					
a6219914- 7a64- 4733- 8293- 6b73c46d 31bd	Schogetten		Crunchy Peanut Butter Chocolate	100	g
22bd6c44- 7e9b- 4b87- 9282- 6b0fb0b43 93b	Auchan		Muffins con pepitas de chocolate		

**Nota.** Se pone un mínimo de datos correspondientes a un archivo Excel con los datos completos

### Creación Tabla Suppliers

```
session.execute("""
CREATE TABLE IF NOT EXISTS suppliers (
    supplier_id text PRIMARY KEY,
    name text,
    contact text,
    rating double,
    description text
)
""")
```

**Tabla 2***Suppliers Table*

supplier_id	name	contact	rating	description
SUP001	Sources Alma Distribution	contact@sourcesalma-distribution.com	4.7	Distribuidor europeo de aguas minerales como Cristaline y Oulmès.
SUP002	Ferrero France Commerciale	ventas@ferrero-france-commerciale.com	4.6	Distribuidor de confitería y cremas de cacao (Nutella, Kinder, Tic Tac).
SUP003	La Boulangère & Co Distribution	comercial@laboulangere-co.com	4.5	Panadería industrial francesa: panes, brioches y bollería.
SUP004	Barilla & Wasa Logistics	logistics@barilla-wasa.com	4.4	Distribuidor de pasta italiana, pesto y productos Wasa crujientes.
SUP005	Nestlé Waters Supply Europe	orders@nestle-waters-supply.com	4.6	Distribuidor de aguas S.Pellegrino, Perrier y Nestlé Pure Life.
SUP006	Nestlé France Food & Drinks	foodservice@nestle-france.com	4.5	Alimentos y bebidas Nestlé como Nesquik, Maggi, Ricoré y Nestlé Dessert.
SUP007	Danone Maghreb & Dairy Services	clientes@danone-maghreb.com	4.3	Distribuidor regional de lácteos y aguas

				Danone (Ain Saïss, Jebli, Centrale).
SUP008	Ethiquable Fair Trade Foods	contact@ethiquable-distribution.com	4.4	Distribuye productos de comercio justo: mantequilla de maní, cacao, cereales.
SUP009	Menguy's Nut & Spread Distribution	sales@menguys-foods.com	4.2	Cacahuete tostados, cremas de maní y snacks saludables.
SUP010	Léa Nature / Jardin Bio Service	service-client@leanature-bio.com	4.3	Distribuye alimentos orgánicos de la marca Jardin Bio y Léa Nature.
SUP011	Saint Hubert Plant-Based & Bio	info@sainthubert-bio.com	4.1	Distribuye margarinas vegetales y productos sin lactosa Saint Hubert.
SUP012	Atlas Maghreb Beverages & Snacks	ventas@atlas-maghreb.com	4.2	Distribuidor inventado: bebidas, lácteos y snacks de marcas marroquíes como Sidi Ali, Star, Jaouda, Chergui y Bahia.
SUP013	Global Snacks & Chips Import	contact@globalsnacks-import.com	4	Importador de snacks internacionales (Pringles, Tyrrell's, Kettle, Takis).
SUP014	Premium Chocolates &	orders@premium-choco-biscuits.eu	4.3	Distribuye chocolates y galletas finas

	Biscuits Europe			(Lindt, Milka, LU, BN, Poulain).
SUP015	Breakfast & Cereals International	support@breakfast-cereals-int.com	4.2	Cereales y desayunos (Kellogg's, Weetabix, Quaker, Jordans, Bjorg).

**Nota.** Se pone un mínimo de datos correspondientes a un archivo Excel con los datos completos

### Creación Tabla Transport

```
session.execute("""
CREATE TABLE IF NOT EXISTS transport (
    transport_id text PRIMARY KEY,
    supplier_id text,
    tag text,
    driver_name text,
    driver_contact text
)
""")
```

### Tabla 3

*Transport Table*

transport_id	supplier_id	tag	driver_name	driver_contact
TRANS001	SUP001	FR-4581-AA	Louis Moreau	+33 6 82 14 7721
TRANS002	SUP001	FR-8732-BX	Camille Roche	+33 6 77 12 3409
TRANS003	SUP002	FR-6742-ZB	Antoine Lefèvre	+33 6 90 33 1487
TRANS004	SUP002	FR-3921-LM	Lucien Dupont	+33 6 31 77 4559

TRANS005	SUP003	FR-9930-GR	Claire Petit	+33 6 33 20 1140
TRANS006	SUP004	IT-2145-KM	Marco Rossi	+39 342 556 1203
TRANS007	SUP004	IT-7780-NP	Giovanni De Luca	+39 349 883 2210
TRANS008	SUP005	FR-5591-CN	Claire Dufour	+33 6 71 95 4478
TRANS009	SUP005	FR-8102-KL	Pierre Durand	+33 6 52 13 4987
TRANS010	SUP006	FR-2023-BV	Jean-Luc Bernard	+33 6 54 11 2304
TRANS011	SUP006	FR-6011-MN	Alice Fournier	+33 6 98 32 4459
TRANS012	SUP007	MA-1290-XA	Abdel Karim Benali	+212 6 55 88 4411
TRANS013	SUP007	MA-1132-LS	Hicham Aït Omar	+212 6 78 43 2199
TRANS014	SUP008	FR-6675-PT	Sophie Laurent	+33 6 70 77 9821
TRANS015	SUP009	FR-8102-KL	Pierre Durand	+33 6 52 13 4987

**Nota.** Se pone un mínimo de datos correspondientes a un archivo Excel con los datos completos

### Creación Tabla Lot

```
session.execute("""
CREATE TABLE IF NOT EXISTS lot (
    lot_id text PRIMARY KEY,
    name text,
    production_date date);
""")
```

### Tabla 4

*LOT Table*

lot_id	name	production_date
LOT001	Cristaline Water Lot A25	2025-01-18

LOT002	Nutella Cream Lot 12B	2025-02-04
LOT003	La Boulangère Brioche Lot F03	2025-02-22
LOT004	Barilla Pesto Genovese Lot 07G	2025-03-01
LOT005	Nestlé Dessert Chocolate Lot C15	2025-03-12
LOT006	Danone Yogurt Natural Lot 05M	2025-03-27
LOT007	Ethiquable Peanut Butter Lot 04E	2025-04-02
LOT008	Menguy's Salted Peanuts Lot 10F	2025-04-14
LOT009	Jardin Bio Organic Spread Lot 03B	2025-04-30
LOT010	Saint Hubert Margarine Lot V21	2025-05-05
LOT011	Atlas Maghreb Mineral Water Lot 22M	2025-05-18
LOT012	Pringles Sour Cream Lot SC2025	2025-05-27
LOT013	Lindt Milk Chocolate Lot 25L	2025-06-04
LOT014	Kellogg's Corn Flakes Lot CF09	2025-06-20
LOT015	PepsiCo Hawai Tropical Lot H25	2025-07-02

**Nota.** Se pone un mínimo de datos correspondientes a un archivo Excel con los datos completos

### Creación Tabla Lot\_Item

```
session.execute("""
CREATE TABLE IF NOT EXISTS testks.lot_item (
    lot_id text,
    product_id uuid,
    product_name text,
    quantity double,
    PRIMARY KEY (lot_id, product_id)
);
```

""")

**Tabla 5**

*LOT\_ITEM Table*

lot_id	product_id	product_name	quantity
LOT001	76f53234-8057-43aa-a0f2-7ec884b29635	Eau De Source	8200
LOT001	e1fd70fb-ff10-42aa-ae6c-1c041210af5e	Sidi Ali	5600
LOT002	3517bdd2-a6ed-4cbb-8514-36fedecc649d	Nutella	4800
LOT002	43f528ba-71f4-4808-9c50-884fce2dc325	Kinder Bueno	6900
LOT003	4f2cf8e4-3165-4da8-b1e3-11e84d9d5b72	Pain de Mie	3700
LOT004	12e30cda-8b94-49c9-8b2b-4252fc376d7a	PESTO alla GENOVESE	5900
LOT004	e54a8a9c-05d9-44a8-9f4b-473cf5e7cda9	Barilla Penne Rigate	7200
LOT005	bdd9dbb7-dadc-4df6-9a5e-9c0c1e9e1a23	Nestlé Dessert Chocolat Noir	5100
LOT006	4661d56d-c4cf-4bf3-9d05-aafa593c0608	Yaourt Nature	8600
LOT006	b94392ff-ec3a-4514-8baf-b5f249b0cf19	Jebli Lait UHT	5400
LOT007	cfdfb87a-df31-480e-bbd0-ef5dd3fa7f22	Beurre de Cacahuète	3400

LOT008	3db6c018-3535-4571-a14c-f9447efacdd2	Menguy's Cacahuètes Grillées	7000
LOT009	a19b5633-22e3-41df-b7db-64d967f80ab2	Tartine Chocolat Bio	4100
LOT010	43e3f786-bdfa-44dd-a612-b707f84d2b9a	Tartine Doux	2500
LOT011	0c8126c8-9e74-4a42-bf38-9db0f4c9ad1d	Bahia Eau Minérale	7700

**Nota.** Se pone un mínimo de datos correspondientes a un archivo Excel con los datos completos

### Creación Tabla Lot\_Trace

```
session.execute(""""
CREATE TABLE IF NOT EXISTS lot_trace (
    lot_id text,
    event_id int,
    event_time timestamp,
    event_type text,
    origin_location text,
    destination_location text,
    supplier_id text,
    supplier_name text,
    transport_id text,
    transport_time int,
    transport_cost double,
    PRIMARY KEY (lot_id, event_id)
)
""")
```

**Tabla 6***LOT\_TRACE Table*

event_id	lot_id	event_time	event_type	origin_location	destination_location	supplier_id	supplier_name	transport_id	transport_time	transport_cost
1	LO T01 0	2020-05-23T01:43:00	Reprogramado	Agadir, Marruecos	Toulouse, Francia	SUP 008	Ethiqueable Fair Trade Foods	TRANS027	33	647.89
2	LO T00 2	2020-03-24T16:15:00	Reprogramado	Casablanca, Marruecos	Tánger, Marruecos	SUP 019	Unilever Maghreb & Africa Foods	TRANS028	14	504.7
3	LO T01 8	2020-08-13T21:23:00	Entrega final	Marsella, Francia	Lyon, Francia	SUP 030	Sultan North Africa Foods	TRANS022	65	1903.4
4	LO T00 8	2020-06-18T21:34:00	Entrega final	Paris, Francia	Hamburgo, Alemania	SUP 009	Menguy's Nut & Spread Distribution	TRANS026	32	1109.02
5	LO T00 3	2020-03-12T22:00:00	Salida de planta	Tánger,	Barcelona, España	SUP 009	Menguy's Nut &	TRANS023	51	1286.42

		2:16: 00		Marrue cos			Spread Distrib ution			
6	LO T00 6	2020- 02- 03T2 1:09: 00	Demor a por inspec ción	Marrak ech, Marrue cos	Lyon, Francia	SUP 007	Danon e Maghr eb & Dairy Servic es	TRA NS01 2	14	172.21
7	LO T00 5	2020- 11- 07T0 3:34: 00	Salida de planta	Agadir ,	Oporto, Portugal	SUP 030	Sultan North Africa Foods	TRA NS02 1	32	749.46
8	LO T01 6	2020- 03- 06T2 0:32: 00	Demor a por inspec ción	Marrak ech, Marrue cos	Paris, Francia	SUP 029	Excelo Biscuit Manuf acturin g	TRA NS04 7	58	501.45
9	LO T01 4	2020- 05- 26T2 1:26: 00	Entreg a final	Rabat, Marrue cos	Nice, Francia	SUP 021	Primev ère Dairy Altern atives	TRA NS05 0	6	191.12
10	LO T00 3	2020- 08- 24T1 4:08: 00	En tránsit o	Hambu rgo, Alema nia	Madrid, España	SUP 007	Danon e Maghr eb & Dairy	TRA NS03 0	3	115

							Servic es			
11	LO T01 5	2020- 08- 19T0 7:53: 00	Salida de planta	La Rochel le, Francia	Marseille , Francia	SUP 008	Ethiqu able Fair Trade Foods	TRA NS00 5	36	642.03
12	LO T00 2	2020- 06- 03T1 2:20: 00	Entreg a final	Marrak ech, Marrue cos	Marseille , Francia	SUP 006	Nestlé France Food & Drinks	TRA NS02 8	24	934.26
13	LO T01 0	2020- 08- 12T0 5:17: 00	Salida de planta	Brusel as, Bélgic a	Nice, Francia	SUP 014	Premiu m Chocol ates & Biscuit s Europe	TRA NS05 0	31	1198.0 4
14	LO T00 9	2020- 07- 12T0 8:46: 00	Repro grama do	Hambu rgo, Alema nia	Casablan ca, Marrueco	SUP 013	Global Snacks & Chips Import	TRA NS03 5	65	2432.2 7
15	LO T01 7	2020- 01- 27T0 1:14: 00	En tránsit o	Marsei lle, Francia	Madrid, España	SUP 008	Ethiqu able Fair Trade Foods	TRA NS02 4	20	728.99

**Nota.** Se pone un mínimo de datos correspondientes a un archivo Excel con los datos completos

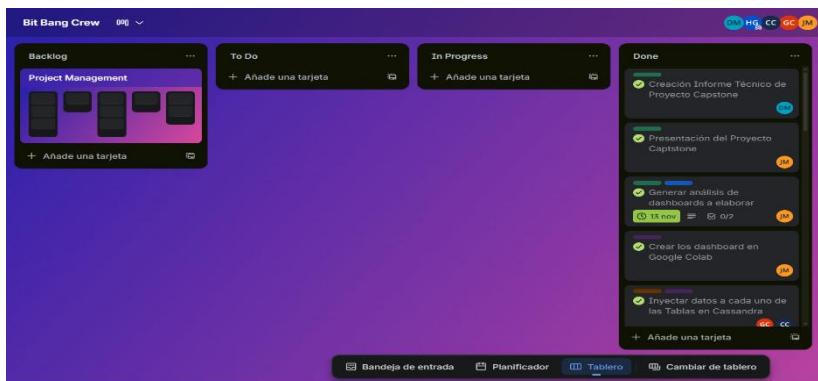
## Diseño del Sistema

### Trello

Tablero: [Bit Bang Crew](#)

### Figura 2

Captura del tablero de Trello

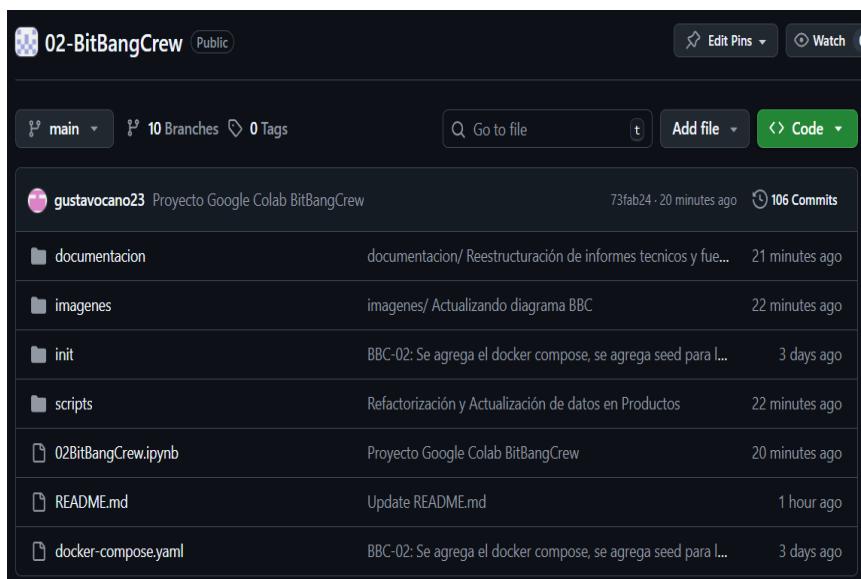


### Github

Repositorio: [02-BitBangCrew](#)

### Figura 3

Captura del repositorio de GitHub



## Estructura de Directorios

- documentación/
- imágenes/
- init/
- scripts/
- Readme.md
- docker-compose.yaml

## Google Colab: Python + Pandas

Framework utilizado para:

- Visualización de reportes.
- Conexión con Cassandra mediante drivers oficiales
- Validación básica de datos
- Escritura y lectura de datos en Cassandra.

## Pinggy

Pinggy se utilizó como herramienta de tunneling para exponer hacia Internet el servicio ejecutado desde Google Colab, permitiendo acceder a Cassandra en el notebook sin necesidad de desplegar infraestructura externa. Gracias a esto, el equipo logró conectarse y ejecutar scripts desde Google Colab para la generación de reportes, inyección de datos, creación de tablas, entre otros.

## Contenedores y Orquestación

El proyecto utiliza Docker Compose para:

- Levantar Cassandra
- Ejecutar scripts init/init.cql automáticamente

Esto garantiza un entorno uniforme para todos los integrantes.

## **Scripts de Inicialización**

El archivo init/init.cql contiene todas las instrucciones necesarias para preparar la base de datos Cassandra utilizada en el proyecto. Este archivo asegura que el sistema tenga el keyspace inicial configurado correctamente antes de ejecutar el Notebook en Google Colab.

## **Arquitectura General**

- Google Colab (Notebook) (Python + Pandas)
- Base de datos Cassandra
- Pinggy
- Contenedores Docker para ambos servicios

## **Ejemplos de Modelo de Datos en Cassandra**

*CREATE KEYSPACE IF NOT EXISTS supply\_chain WITH replication = {*

```
'class': 'SimpleStrategy',
'replication_factor': 1
};
```

*CREATE TABLE IF NOT EXISTS product (*

```
product_id TEXT PRIMARY KEY,
brands TEXT,
manufacturing_places TEXT,
```

```
producto_name TEXT,  
producto_quantity DOUBLE,  
producto_quantity_unit TEXT  
);
```

```
CREATE TABLE IF NOT EXISTS Lot (  
    lot_id TEXT PRIMARY KEY,  
    name TEXT,  
    production_date DATE  
);
```

```
CREATE TABLE IF NOT EXISTS Lot_item (  
    lot_id TEXT,  
    product_id TEXT,  
    quantity DOUBLE,  
    Product_name TEXT,  
    PRIMARY KEY (lot_id, product_id)  
);
```

```
CREATE TABLE IF NOT EXISTS Lot_trace (  
    event_id INT PRIMARY KEY,  
    lot_id TEXT,  
    event_time TIMESTAMP,  
    event_type TEXT,  
    origin_location TEXT,  
    destination_location TEXT,  
    supplier_name TEXT,  
    supplier_id TEXT,  
    transport_id TEXT,  
    transport_time INT,  
    transport_cost DOUBLE  
);
```

## **Desarrollo del Sistema**

### **Ejecución del Proyecto**

*docker-compose up --build*

Servicios que levanta:

- Cassandra
- Servicio de inicialización (init.cql)

### **Pinggy**

*ssh -p 443 -R0:localhost:9042 tcp@free.pinggy.io*

Permite publicar a la nube la base de datos cassandra por el puerto 9042

### **Google Colab**

Enlace del Proyecto:

[https://colab.research.google.com/drive/1cuB\\_CBYAJ\\_Rm4qsjaOHUkhsSLbmUnDh6?usp=sharing](https://colab.research.google.com/drive/1cuB_CBYAJ_Rm4qsjaOHUkhsSLbmUnDh6?usp=sharing)

### **Estructura**

- Instalación de librería
- Importación de librerías
- Conexión a Cassandra
- Función de inyección de Cassandra
- Creación de tablas Cassandra
- Inyección de información en Cassandra
- Análisis y reportes de información

## **Resultados**

- Notebook funcional ejecutándose en Google Colab
- Conexión estable con Cassandra
- Tablas creadas correctamente
- Registro y consulta de eventos de trazabilidad
- Documentación técnica del sistema
- Uso de GitHub y Trello como herramientas colaborativas

### **Análisis de los Gráficos a Implementar en Proyecto Capstone.**

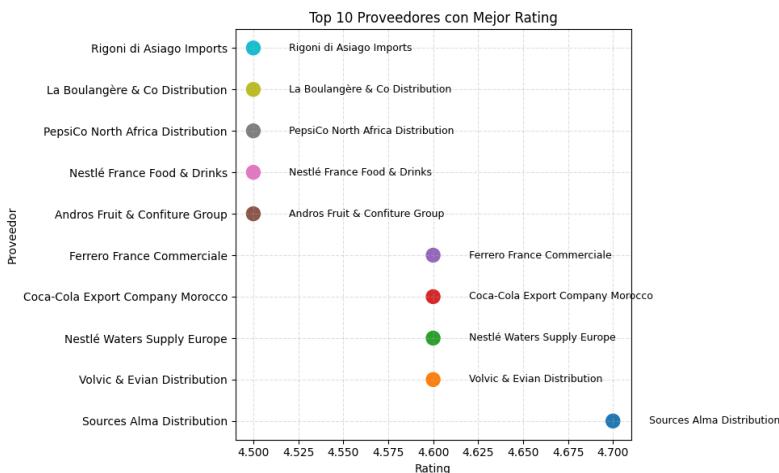
El proyecto capstone “trazabilidad de productos” tiene como requerimiento la visualización del flujo logístico, así como la calidad de los productos y el comportamiento operativo de los lotes a lo largo de su recorrido.

Se hizo una selección de gráficos los cuales son los siguientes:

1. Proveedores con mejor rating.
2. Volumen total por lote.
3. Distribución de eventos por tipos.
4. Proveedores más eficientes.
5. Tiempo vs Costo.
6. Lotes y su producto.

**Figura 4**

*Gráfico: proveedores con mejor rating*

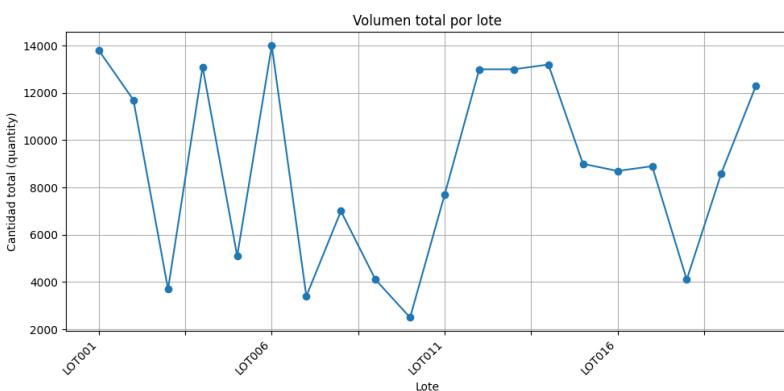


Análisis: El proveedor "Sources Alma Distribution" es quien destaca claramente con un rating alrededor de 4.70, el cual es el que se posiciona como el mayor entre los demás proveedores.

Esto nos es útil al momento de elegir los proveedores de nuestros productos, ya que estos pueden destacar por sus tiempos o calidad. Los ratings se encuentran entre 4.50 - 4.70, esto nos da a conocer que los proveedores de los productos son de alta satisfacción.

**Figura 5**

*Gráfico: Volumen total por lote*



Análisis: El gráfico muestra el volumen total procesado por cada lote, utilizando un line plot donde cada punto representa un lote y su cantidad total. El primer patrón como es visible muestra una gran variabilidad, con picos altos que van desde 13,000 - 14,000, de igual forma muestra valles bajos como 2,000 - 3,000.

Podemos destacar de esta gráfica los picos de:

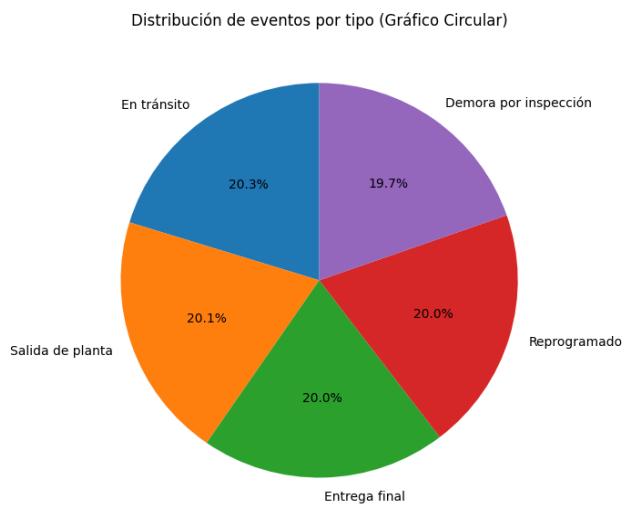
- LOT007/ LOT006(cerca de 14 000)
- LOT011(dos puntos alrededor de 13 000)
- LOT014(12 500 Aprox.)

Los valores más bajos aparecen en:

- LOT006 (punto cercano a 3 000)
- LOT011 (aprox. 4 000)
- Otros puntos intermedios alrededor de 5 000–6 000

## Figura 6

*Gráfico: Distribución de eventos por tipos*



Análisis: El gráfico circular muestra la proporción de eventos operativos registrados para los distintos tipos de evento dentro de la trazabilidad logística. Cada segmento representa la frecuencia relativa de cada tipo.

Tipos de evento incluidos:

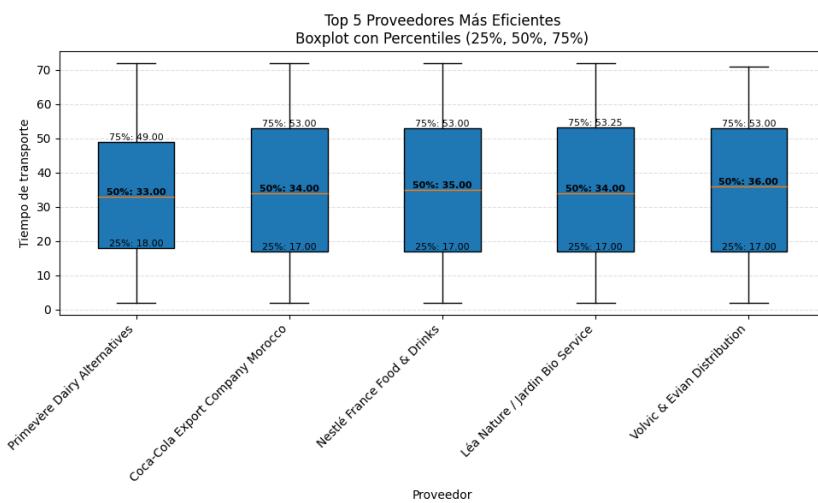
- Demora por inspección
- Reprogramado
- Entrega final
- Salida de planta
- En tránsito

Las proporciones están extremadamente parejas, oscilando entre 19.7% y 20.3%, lo cual indica que:

- No hay un tipo de evento que domine.
- Los procesos registran eventos de una manera balanceada.

### **Figura 7**

*Gráfico: Top 5 proveedores más eficientes*



Análisis: El gráfico muestra la distribución del tiempo de transporte por proveedor usando un boxplot, incluyendo percentiles clave:

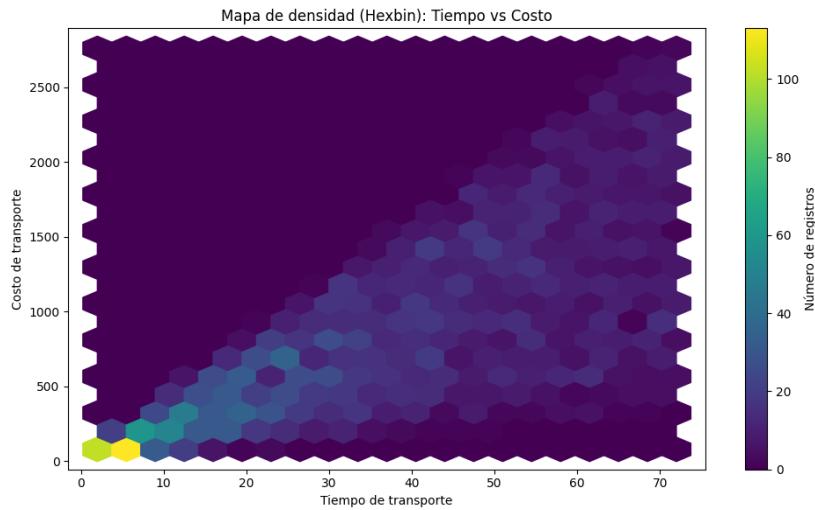
- 25% (Q1)
- 50% (Mediana)
- 75% (Q3)

Esto permite comparar variabilidad, consistencia y eficiencia real de cada proveedor. Más consistente:

- Coca-Cola Export Company Morocco
  - Mediana: 33.00
  - IQR:  $53 - 17 = 36$
- Proveedor con menor mediana absoluta:
  - Primavère Dairy Alternatives
  - Mediana: 33.00 (igual que Coca-Cola), pero Q1 más alto (18), lo que sugiere menos entregas ultra rápidas.
- Proveedor más rápido en promedio:
  - Volvic & Evian Distribution
  - Mediana más baja del grupo (36) → aunque ligeramente mayor, muestra uniformidad)
  - Q1 bajo (17) → tiene entregas rápidas frecuentes.

**Figura 8**

Gráfico: *Tiempo VS Costo*



Análisis: Este gráfico representa la relación entre tiempo de transporte (eje X) y costo de transporte (eje Y), utilizando un hexbin para mostrar dónde se acumulan más registros. Los colores indican densidad:

- Amarillo / Verde claro → muy alta concentración
- Verde / Azul → media
- Morado oscuro → baja o nula

El patrón general forma una diagonal ascendente, donde:

- Cuando el tiempo aumenta, el costo también tiende a aumentar.
- La nube de densidad se expande proporcionalmente.

La forma triangular sugiere que:

- Para tiempos bajos, hay una gran variabilidad en costo (0–300).
- Para tiempos altos (50–70h), los costos tienden a ser de 800+.

**Figura 9**

*Gráfico: Lotes y sus productos*



Análisis: El gráfico Sunburst representa una relación jerárquica entre:

- Nivel 1 (centro): Lotes
- Nivel 2 (anillos exteriores): Productos asociados a cada lote

Este tipo de visualización es ideal para mostrar cómo están distribuidos los productos dentro de cada lote, y si ciertos lotes contienen más o menos variedad que otros.

Cada producto tiene un color distinto, lo que facilita:

- Identificar rápidamente qué producto pertenece a qué lote
- Evitar confusiones entre categorías

## **Conclusiones**

El proyecto permitió construir un sistema práctico de trazabilidad utilizando Apache Cassandra, Google Colab y Docker, integrando datos reales y generando reportes útiles para analizar el comportamiento de los lotes y proveedores. La arquitectura distribuida seleccionada facilitó la gestión eficiente del volumen de información, mientras que el uso de herramientas colaborativas como GitHub y Trello permitió organizar y documentar el trabajo de forma efectiva. En conjunto, el proyecto demostró la viabilidad de implementar un sistema moderno, escalable y orientado al análisis dentro de una cadena de suministro.

## **Referencias**

Brewer, E. A. (2012). CAP Twelve Years Later: How the “Rules” Have Changed. *IEEE Computer*, 45(2), 23-29.

DataStax (o sitio oficial de Apache Cassandra). *Documentación de Arquitectura y Modelo de Datos*.

Rabl, T., et al. (2012). Solving Big Data Challenges for Enterprise Application Performance Management. *Proceedings of the VLDB Endowment*.