

INFORME DE RESULTADOS

Unidad 3 - Tarea 6

Equipo/Grupo : 2

Estudiantes :

- Harlan Santiago Enciso Riaño
- Miguel Angeles Rojas Pabon
- Maria Camila Rodriguez Ortiz

Objetivo General

Diseñar un Gobierno de Datos y la estructura de un Proyecto de Big Data de envergadura para enfrentar la problemática en relación a la implementación de una Arquitectura de Plataforma de control y monitoreo de concentración de gases tóxicos en las fábricas de la empresa “Sustancias Locas”.

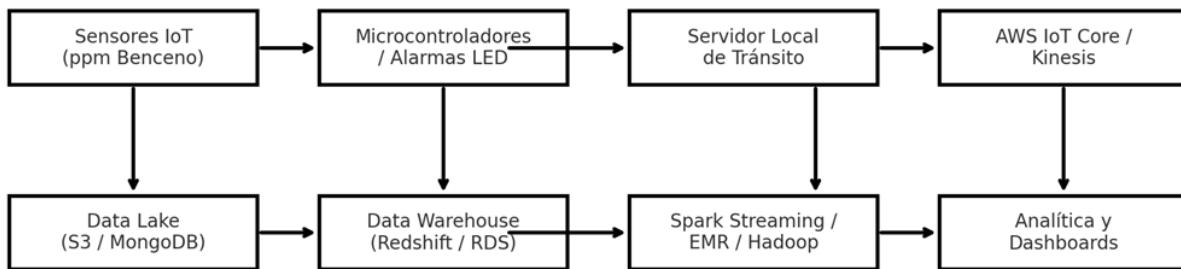
1.-Diseño de un Gobierno de Datos

1.1- Roles del gobierno de datos

	Departamento o persona	Descripción	Rol, responsabilidades y/o funciones
1	Chief Data Officer (CDO)	Líder estratégico de toda la gestión del dato en la empresa.	<ul style="list-style-type: none">- Define la política de datos.- Alinea el proyecto con la gerencia.- Toma decisiones críticas sobre datos y seguridad.- Supervisa cumplimiento normativo ambiental.
2	Comité de Gobierno de Datos	Grupo directivo que define reglas, políticas y lineamientos para todo el ciclo de vida del dato.	<ul style="list-style-type: none">- Define estándares de calidad.- Establece protocolos ante alarmas.- Valida arquitectura Big Data.- Controla la trazabilidad del dato.
3	Data Steward (por línea de producción)	Responsable del dato operativo generado por sensores y microcontroladores.	<ul style="list-style-type: none">- Verifica integridad y consistencia de los datos.- Administra catálogos (sensores, filtros, turnos).- Autoriza cambios en reglas de negocio.
4	Arquitecto de Datos	Diseña la arquitectura Big Data y define tecnologías a usar.	<ul style="list-style-type: none">- Define integración IoT → Servidor → Nube.- Diseña Data Lake y Data Warehouse.- Asegura interoperabilidad local/nube.- Valida flujos de ingesta.
5	Ingeniero de Seguridad de Datos	Encargado de la protección, cifrado y control de accesos.	<ul style="list-style-type: none">- Implementa cifrado en tránsito y reposo.- Administra permisos IAM.- Realiza auditorías y monitoreo.- Define políticas de retención.
6	Data Engineers (3 personas)	Construyen pipelines de datos y sistemas de streaming.	<ul style="list-style-type: none">- Desarrollan ingestión real-time (Kinesis / Kafka).- Implementan Spark Streaming.- Limpian y transforman datos.- Mantienen disponibilidad 24/7.
7	Ingeniero IoT /	Técnico encargado de	<ul style="list-style-type: none">- Instala sensores nuevos.

	Especialista en Sensores	sensores, microcontroladores y redes industriales.	- Calibra dispositivos. - Mantiene firmware actualizado. - Garantiza correcta transmisión local.
8	Administrador de Bases de Datos (DBA)	Responsable del Data Warehouse, bases NoSQL y clusters Hadoop.	- Optimiza consultas para el dashboard. - Mantiene índices y particiones. - Administra backups. - Garantiza integridad del almacenamiento.
9	Analista de Datos / Científico de Datos	Encargado del análisis predictivo e informes.	- Modelo predictivo de vida útil de filtros. - Análisis de picos y anomalías. - Reportes ambientales y KPIs. - Soporte al dashboard.
10	Equipo de Soporte Técnico 24/7	Grupo que atiende incidentes, fallos o alarmas críticas.	- Monitorea infraestructura. - Responde ante fallos de sensores o red. - Escala incidentes mayores. - Garantiza continuidad del servicio.

1.2.- Estructura gráfica del gobierno de datos



2.- Diseño del Proyecto Big Data

2.1- Elementos del Proyecto Big Data

	Elemento	Descripción	Observaciones
1	Sensores IoT de gases tóxicos	Sensores IoT de gases tóxicos	Reemplazo por sensores de última generación.
2	Microcontroladores y módulos de comunicación	Microcontroladores y módulos de comunicación	Envío de datos a servidor local (MQTT / WiFi).
3	Servidor local de tránsito	Servidor local de tránsito	Recibe datos de 4 líneas; valida y enruta.
4	Plataforma de Nube (AWS IoT Core / Kinesis)	Plataforma de Nube (AWS IoT Core / Kinesis)	Procesa millones de eventos en tiempo real.
5	Data Lake (S3 / MongoDB)	Data Lake (S3 / MongoDB)	Almacena datos crudos sin procesar.

6	Data Warehouse (Redshift / PostgreSQL)	Data Warehouse (Redshift / PostgreSQL)	Datos transformados para analítica histórica.
7	Sistema de procesamiento	Sistema de procesamiento en streaming (Spark / EMR)	Detección de picos, anomalías, alertas inmediatas.
8	Dashboards (Power BI / Grafana / QuickSight)	Dashboards (Power BI / Grafana / QuickSight)	Visualización de ppm, alarmas y tendencias.
9	Sistema de alarmas automáticas	Sistema de alarmas automáticas	Notificaciones en tiempo real a supervisores.
10	Modelo predictivo (vida útil del filtro)	Modelo predictivo (vida útil del filtro)	Reduce costos y evita reemplazos anticipados.
11	Protocolos de seguridad	Protocolos de seguridad de datos	Cifrado, IAM, auditoría del flujo de datos.
12	Equipo técnico y roles de gobierno de datos	Equipo técnico y roles del gobierno de datos	Operación y soporte 24/7.
13	Documentación, capacitación y puesta en marcha	Documentación, capacitación y puesta en marcha	Última etapa antes del despliegue final.

2.2- Lista de verificación del Proyecto Big Data

	Elemento	Estado	Tiempo
1	Conformación del Gobierno de Datos	Completado	Semana 1
2	Definición de políticas de datos y seguridad	Completado	Semana 1
3	Instalación y calibración de sensores	En proceso	Semana 2
4	Configuración de microcontroladores	En proceso	Semana 2
5	Instalación del servidor local de tránsito	Completado	Semana 2
6	Configuración de AWS IoT Core / Kinesis	Pendiente	Semana 3
7	Configuración del Data Lake	Pendiente	Semana 3
8	Configuración del Data Warehouse	Pendiente	Semana 3
9	Construcción de pipelines de ingesta	Pendiente	Semana 4
10	Implementación de	Pendiente	Semana 4

	Spark Streaming		
11	Pruebas de detección de alarmas	Pendiente	Semana 4
12	Modelos analíticos (predictivo)	Pendiente	Semana 5
13	Construcción del dashboard	Pendiente	Semana 6
14	Pruebas con supervisores de planta	Pendiente	Semana 6
15	Seguridad: accesos, cifrado, auditoría	Pendiente	Semana 7
16	Documentación y capacitación	Pendiente	Semana 7
17	Despliegue final 24/7	Pendiente	Semana 8
18	Cierre y entrega	Cierre y entrega	Fin del proyecto

3.- Conceptos y elementos de Big Data

#	Componente	Descripción (breve cita bajo Norma APA)	Referencia (APA)
1	Arquitectura de Sistema de Información	Una arquitectura de SI define la estructura tecnológica y los componentes que permiten procesar, almacenar y distribuir información en una organización (Laudon & Laudon, 2020).	Laudon, K., & Laudon, J. (2020). <i>Management Information Systems</i> . Pearson.
2	Hadoop	Hadoop es un framework de procesamiento distribuido para grandes volúmenes de datos, basado en el paradigma MapReduce (White, 2015).	White, T. (2015). <i>Hadoop: The Definitive Guide</i> . O'Reilly.
3	Spark	Apache Spark es un motor de procesamiento en memoria para análisis rápidos y procesamiento distribuido (Zaharia et al., 2016).	Zaharia, M., et al. (2016). <i>Apache Spark: The Definitive Guide</i> . O'Reilly.
4	PostgreSQL	PostgreSQL es un sistema de gestión de bases de datos relacional avanzado y open source (Momjian, 2001).	Momjian, B. (2001). <i>PostgreSQL: Introduction and Concepts</i> . Addison-Wesley.
5	Amazon Kinesis FireHose	Servicio de AWS que permite capturar, transformar y enviar flujos de datos en tiempo real a destinos de almacenamiento (Amazon Web Services, 2023).	Amazon Web Services. (2023). <i>Amazon Kinesis Data Firehose Documentation</i> .
6	Amazon QuickSight	QuickSight es un servicio de analítica y visualización de datos en la nube de AWS (AWS, 2023).	Amazon Web Services. (2023). <i>Amazon QuickSight Documentation</i> .
7	Amazon Glue	AWS Glue es un servicio de integración y ETL administrado que automatiza la preparación de datos (AWS, 2023).	Amazon Web Services. (2023). <i>AWS Glue Documentation</i> .
8	HTTP API	Una API HTTP permite la comunicación entre sistemas a través del protocolo HTTP para el intercambio de datos (Fielding, 2000).	Fielding, R. (2000). <i>Architectural Styles and the Design of Network-based Software Architectures</i> (Tesis doctoral). UC Irvine.
9	Boto3	Boto3 es la librería oficial de AWS para Python que permite interactuar con los servicios de la nube (AWS, 2023).	Amazon Web Services. (2023). <i>Boto3 Documentation</i> .
10	Amazon RedShift	Redshift es un data warehouse en la nube optimizado para consultas analíticas masivas (Amazon Web Services, 2023).	Amazon Web Services. (2023). <i>Amazon Redshift Documentation</i> .
11	Amazon Athena	Athena es un servicio de consulta interactiva que permite analizar datos en S3 usando SQL (AWS, 2023).	Amazon Web Services. (2023). <i>Amazon Athena Documentation</i> .
12	Amazon CloudWatch	CloudWatch es un servicio de monitoreo y observabilidad para métricas, logs y aplicaciones en AWS (AWS, 2023).	Amazon Web Services. (2023). <i>Amazon CloudWatch Documentation</i> .
13	Boto3	Librería de AWS para programación y automatización de servicios en Python (AWS, 2023).	Amazon Web Services. (2023). <i>Boto3 Documentation</i> .
14	Formato de datos JSON	JSON es un formato basado en texto para intercambio ligero de datos estructurados (Bray, 2017).	Bray, T. (2017). <i>The JavaScript Object Notation (JSON) Data Interchange Format</i> . IETF RFC 8259.
15	Formato de datos CSV	CSV es un formato simple de datos tabulares separados por comas, ampliamente usado en análisis (RFC 4180, 2005).	Shafranovich, Y. (2005). <i>Common Format and MIME Type for CSV Files</i> . IETF RFC 4180.
16	Formato de datos Parquet	Parquet es un formato de almacenamiento columnar optimizado para análisis de Big Data (Vohra, 2016).	Vohra, D. (2016). <i>Practical Hadoop Ecosystem</i> . Apress.
17	Microcontrolador Arduino	Arduino es una plataforma de hardware libre usada para control y adquisición de datos (Banzi & Shiloh, 2014).	Banzi, M., & Shiloh, M. (2014). <i>Getting Started with Arduino</i> . Maker Media.
18	Microcontrolador ESP8266	El ESP8266 es un microcontrolador WiFi usado en proyectos IoT para transmisión de datos (Espressif Systems, 2020).	Espressif Systems. (2020). <i>ESP8266EX Datasheet</i> .
19	Sensor MQ-135	Sensor MQ-135 detecta concentración de gases tóxicos como benceno, amoníaco y óxidos (Winsen, 2019).	Winsen Electronics. (2019). <i>MQ-135 Gas Sensor Datasheet</i> .
20	Benceno	El benceno es un compuesto orgánico volátil altamente tóxico clasificado como carcinógeno (ATSDR, 2007).	ATSDR. (2007). <i>Toxicological Profile for Benzene</i> . U.S. Department of Health.

4.- Diseño de la estructura de la hoja de cálculo del microcontrolador “lecturas”

4.1. - Diseño de la estructura de la hoja de cálculo para almacenar lecturas del microcontrolador

Diseño de la estructura de la hoja de cálculo para almacenar lecturas del microcontrolador

#	Nombre Columna	Descripción	Observaciones
1	id_lectura	Número consecutivo que identifica cada lectura	Autoincremental
2	fecha	Fecha en la que se realiza la lectura	Formato AAAA-MM-DD
3	hora	Hora exacta de la lectura	hh:mm:ss
4	ppm	Valor de concentración del benceno en ppm	Lectura del sensor MQ-135
5	id_sensor	ID del sensor asociado	A1S01
6	id_microcontrolador	ID del microcontrolador que genera la lectura	A1M01
7	linea_produccion	Línea de producción donde está instalado	A1, A2, A3, A4
8	fabrica	Fábrica a la que pertenece	A, B o C
9	latitud	Geolocalización del sensor	Decimal
10	longitud	Geolocalización del sensor	Decimal
11	estado_ppm	Nivel de alarma según la clasificación	Normal, Riesgo, Crítico
12	filtro_cercano	Identificador del filtro asociado	F-A1-01

4.2. – Poblamiento de la hoja de cálculo del microcontrolador

Poblamiento de la hoja de cálculo del microcontrolador

A	B	C	D	E	F	G	H	I	J	K	L
id_lectura	fecha	hora	ppm	id_sensor	id_micro	linea	fabrica	latitud	longitud	estado_ppm	filtro_cercano
1	2025-11-22	8:00:01	0.8	A1S01	A1M01	A1	A	6.251	-75.563	Normal	F-A1-01
2	2025-11-22	8:00:11	1.2	A1S01	A1M01	A1	A	6.251	-75.563	Riesgo	F-A1-01
3	2025-11-22	8:00:21	0.6	A1S02	A1M02	A1	A	6.252	-75.564	Normal	F-A1-01
4	2025-11-22	8:00:31	3.5	A1S03	A1M03	A1	A	6.250	-75.566	Riesgo	F-A1-01
5	2025-11-22	8:00:41	10.1	A1S04	A1M04	A1	A	6.249	-75.561	Crítico	F-A1-01
6	2025-11-22	8:00:51	0.9	A1S05	A1M05	A1	A	6.248	-75.562	Normal	F-A1-02
7	2025-11-22	8:01:01	1.0	A1S06	A1M06	A1	A	6.247	-75.563	Normal	F-A1-02
8	2025-11-22	8:01:11	2.8	A1S07	A1M07	A1	A	6.246	-75.564	Riesgo	F-A1-02
9	2025-11-22	8:01:21	5.9	A1S08	A1M08	A1	A	6.245	-75.565	Riesgo	F-A1-02
10	2025-11-22	8:01:31	12.3	A1S09	A1M09	A1	A	6.244	-75.566	Crítico	F-A1-03
11	2025-11-22	8:01:41	0.7	A1S10	A1M10	A1	A	6.243	-75.567	Normal	F-A1-03
12	2025-11-22	8:01:51	1.4	A1S11	A1M11	A1	A	6.242	-75.568	Riesgo	F-A1-03
13	2025-11-22	8:02:01	0.3	A1S12	A1M12	A1	A	6.241	-75.569	Normal	F-A1-03
14	2025-11-22	8:02:11	7.5	A1S13	A1M13	A1	A	6.240	-75.570	Riesgo	F-A1-04
15	2025-11-22	8:02:21	9.9	A1S14	A1M14	A1	A	6.239	-75.571	Crítico	F-A1-04
16	2025-11-22	8:02:31	0.4	A1S15	A1M15	A1	A	6.238	-75.572	Normal	F-A1-04
17	2025-11-22	8:02:41	2.2	A1S16	A1M16	A1	A	6.237	-75.573	Riesgo	F-A1-04
18	2025-11-22	8:02:51	14.0	A1S17	A1M17	A1	A	6.236	-75.574	Crítico	F-A1-05
19	2025-11-22	8:03:01	0.5	A1S18	A1M18	A1	A	6.235	-75.575	Normal	F-A1-05
20	2025-11-22	8:03:11	3.0	A1S19	A1M19	A1	A	6.234	-75.576	Riesgo	F-A1-05

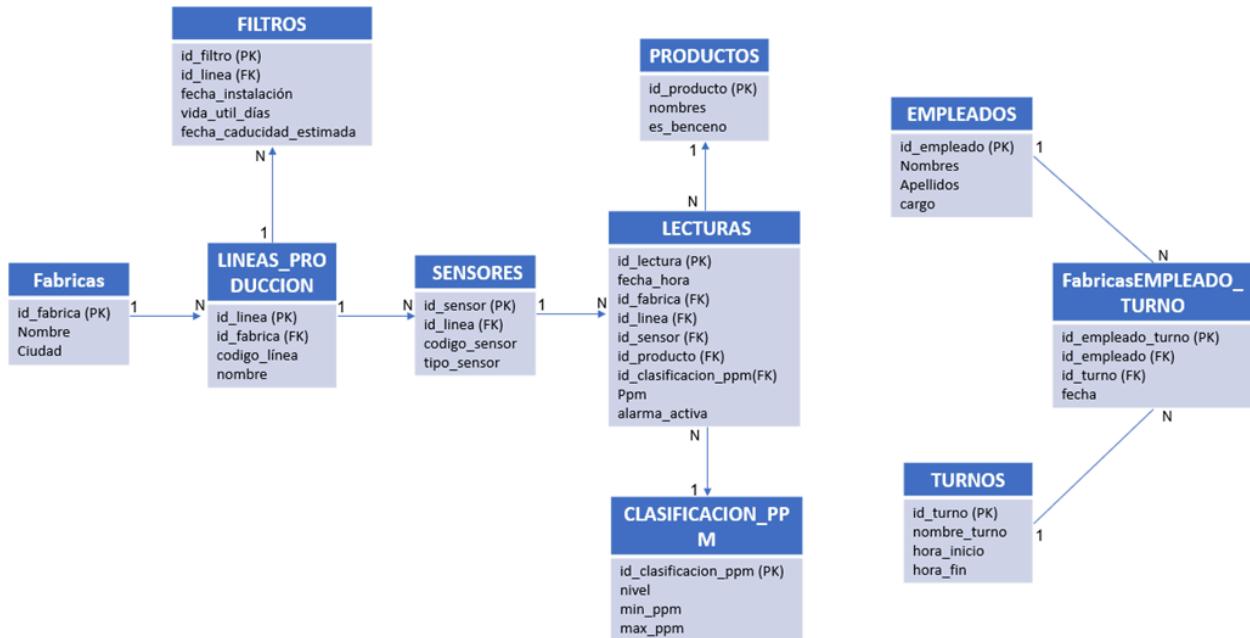
5.- Diseño de la estructura de la tabla de “lecturas”

 Diseño de la estructura de la tabla lecturas

#	Nombre Campo	Tipo de dato	Tamaño (Bytes)
1	id_lectura	BIGSERIAL	8
2	fecha	DATE	4
3	hora	TIME	4
4	ppm	REAL	4
5	id_sensor	VARCHAR(10)	10
6	id_microcontrolador	VARCHAR(10)	10
7	linea_produccion	VARCHAR(5)	5
8	fabrica	VARCHAR(2)	2
9	latitud	DOUBLE PRECISION	8
10	longitud	DOUBLE PRECISION	8
11	estado_ppm	VARCHAR(10)	10
12	filtro_cercano	VARCHAR(15)	15

6.- Diseño y creación de la Base de Datos “monitoreo-producción”

6.1.- Diagrama Entidad-Relación (entidades, relaciones, atributos y cardinalidades)



6.2- Diccionario de Datos (tipos de dato, tamaño, claves primarias y foráneas)

FÁBRICA

Campo	Tipo	Descripción
<code>id_fabrica</code>	SERIAL	Identificador único
<code>nombre</code>	VARCHAR(50)	Nombre (A, B, C)
<code>ubicacion</code>	VARCHAR (100)	Ciudad / dirección

LINEA_PRODUCCION

Campo	Tipo	Descripción
id_linea	SERIAL	Id único
id_fabrica	INT	FK → fabrica
codigo_linea	VARCHAR(10)	A1, A2, A3...

SENSOR

Campo	Tipo	Descripción
id_sensor	SERIAL	Id
id_linea	INT	FK
codigo_sensor	VARCHAR(10)	A1S01, etc
modelo	VARCHAR(20)	MQ-135

CLASIFICACION_PPM

Campo	Tipo	Descripción
id_clasificacion	SERIAL	Id
rango_min	INT	Inicio del rango
rango_max	INT	Fin del rango
nivel	VARCHAR(20)	Bajo, Moderado...
accion	VARCHAR(100)	Qué hacer

LECTURAS

Campo	Tipo	Descripción
id_lectura	SERIAL	Id
id_sensor	INT	FK
ppm	INT	Lectura
clasificacion	INT	FK → clasificacion
fecha	DATE	Fecha
hora	TIME	Hora

CREAR BASE DATO

```
CREATE DATABASE "monitoreo-producción";
```

TABLAS

```
CREATE TABLE fabricas (
```

```
    id_fabrica SERIAL PRIMARY KEY,  
    codigo     VARCHAR(10) NOT NULL UNIQUE,  
    nombre     VARCHAR(100) NOT NULL,  
    direccion  VARCHAR(150),  
    ciudad     VARCHAR(80),  
    latitud    NUMERIC(9,6),  
    longitud   NUMERIC(9,6),  
    estado     VARCHAR(20) DEFAULT 'ACTIVA'
```

```
);
```

```
CREATE TABLE lineas_produccion (
```

```
    id_linea   SERIAL PRIMARY KEY,  
    id_fabrica INT NOT NULL REFERENCES fabricas(id_fabrica),  
    codigo_linea VARCHAR(10) NOT NULL,  
    descripcion VARCHAR(100),  
    estado     VARCHAR(20) DEFAULT 'ACTIVA'
```

```
);
```

```
CREATE TABLE sensores (
```

```
id_sensor      SERIAL PRIMARY KEY,  
  
id_linea       INT NOT NULL REFERENCES lineas_produccion(id_linea),  
  
codigo_sensor  VARCHAR(10) NOT NULL,  
  
tipo_sensor    VARCHAR(50) NOT NULL,  
  
precision_ppm  NUMERIC(6,2),  
  
unidad         VARCHAR(10) DEFAULT 'ppm',  
  
fecha_instalacion DATE,  
  
estado         VARCHAR(20) DEFAULT 'ACTIVO'  
);
```

```
CREATE TABLE tipos_filtro (  
  
id_tipo_filtro SERIAL PRIMARY KEY,  
  
nombre          VARCHAR(50) NOT NULL,  
  
descripcion     VARCHAR(150),  
  
vida_util_dias INT,  
  
costo_usd      NUMERIC(10,2)  
);
```

```
CREATE TABLE filtros (  
  
id_filtro       SERIAL PRIMARY KEY,  
  
id_linea        INT NOT NULL REFERENCES lineas_produccion(id_linea),  
  
id_tipo_filtro  INT NOT NULL REFERENCES tipos_filtro(id_tipo_filtro),  
  
codigo_filtro   VARCHAR(20) NOT NULL,  
  
fecha_instalacion DATE NOT NULL,  
  
fecha_cambio_prevista DATE,  
  
fecha_cambio_real DATE,  
  
estado          VARCHAR(20) DEFAULT 'EN USO'  
);
```

```
CREATE TABLE productos (
```

```
id_producto SERIAL PRIMARY KEY,  
nombre VARCHAR(80) NOT NULL,  
descripcion VARCHAR(150),  
es_benceno BOOLEAN DEFAULT FALSE,  
unidad_medida VARCHAR(10) DEFAULT 'kg'  
);
```

```
CREATE TABLE clasificacion_ppm (  
    id_clasificacion_ppm SERIAL PRIMARY KEY,  
    rango_min      NUMERIC(10,2) NOT NULL,  
    rango_max      NUMERIC(10,2) NOT NULL,  
    nivel          VARCHAR(30) NOT NULL,  
    descripcion    VARCHAR(200),  
    accion_recomendada VARCHAR(200),  
    color_alerta   VARCHAR(20)  
);
```

```
CREATE TABLE configuracion_fabrica (  
    id_config      SERIAL PRIMARY KEY,  
    id_fabrica     INT NOT NULL REFERENCES fabricas(id_fabrica),  
    frecuencia_lectura_seg INT NOT NULL,  
    ppm_max_permitido  NUMERIC(10,2),  
    dias_max_uso_filtro INT,  
    fecha_creacion DATE NOT NULL,  
    observaciones   VARCHAR(200)  
);
```

```
CREATE TABLE empleados (
```

```
id_empleado SERIAL PRIMARY KEY,  
documento VARCHAR(20) NOT NULL UNIQUE,  
nombres VARCHAR(80) NOT NULL,  
apellidos VARCHAR(80) NOT NULL,  
cargo VARCHAR(50),  
telefono VARCHAR(20),  
email VARCHAR(100),  
id_fabrica INT REFERENCES fabricas(id_fabrica)  
);
```

```
CREATE TABLE turnos (  
    id_turno SERIAL PRIMARY KEY,  
    nombre_turno VARCHAR(20) NOT NULL,  
    hora_inicio TIME      NOT NULL,  
    hora_fin   TIME      NOT NULL  
);
```

```
CREATE TABLE empleado_turno (  
    id_empleado_turno SERIAL PRIMARY KEY,  
    id_empleado   INT NOT NULL REFERENCES empleados(id_empleado),  
    id_turno       INT NOT NULL REFERENCES turnos(id_turno),  
    fecha          DATE NOT NULL,  
    es_supervisor  BOOLEAN DEFAULT FALSE  
);
```

```
CREATE TABLE lecturas (
```

```
id_lectura      BIGSERIAL PRIMARY KEY,  
  
fecha_hora      TIMESTAMP      NOT NULL,  
  
id_fabrica      INT      NOT NULL REFERENCES fabricas(id_fabrica),  
  
id_linea        INT      NOT NULL REFERENCES lineas_produccion(id_linea),  
  
id_sensor        INT      NOT NULL REFERENCES sensores(id_sensor),  
  
id_filtro        INT      REFERENCES filtros(id_filtro),  
  
id_producto      INT      NOT NULL REFERENCES productos(id_producto),  
  
ppm              NUMERIC(10,2) NOT NULL,  
  
temperatura      NUMERIC(5,2),  
  
humedad          NUMERIC(5,2),  
  
latitud          NUMERIC(9,6),  
  
longitud          NUMERIC(9,6),  
  
id_clasificacion_ppm INT      NOT NULL REFERENCES clasificacion_ppm(id_clasificacion_ppm),  
  
alarma_activa    BOOLEAN     DEFAULT FALSE,  
  
origen           VARCHAR(20)  DEFAULT 'local'  
);
```

7.- Poblamiento de la base de datos “monitoreo-producción”

7.1.- Scripts de inserción de registros en todas las tablas de la base datos (INSERTS)

Fábricas

```
INSERT INTO fabricas (codigo, nombre, direccion, ciudad, latitud, longitud)
```

```
VALUES
```

```
('A', 'Fábrica A', 'Zona Industrial 1', 'Medellín', 6.2518, -75.5636),  
('B', 'Fábrica B', 'Parque Industrial Norte', 'Bello', 6.3373, -75.5581),  
('C', 'Fábrica C', 'Vía al mar km 5', 'Turbo', 8.0926, -76.7285);
```

Líneas de producción

```
INSERT INTO lineas_produccion (id_fabrica, codigo_linea, descripcion)
```

```
VALUES
```

```
(1, 'A1', 'Línea 1 Fábrica A'),  
(1, 'A2', 'Línea 2 Fábrica A'),  
(2, 'B1', 'Línea 1 Fábrica B'),  
(3, 'C1', 'Línea 1 Fábrica C');
```

Tipos de filtro

```
INSERT INTO tipos_filtro (nombre, descripcion, vida_util_dias, costo_usd)
```

```
VALUES
```

```
('Filtro Premium', 'Filtro de alta capacidad para benceno', 15, 300.00),  
('Filtro Estándar', 'Filtro estándar de línea', 7, 180.00);
```

Filtros

```
INSERT INTO filtros (id_linea, id_tipo_filtro, codigo_filtro, fecha_instalacion, fecha_cambio_prevista, estado)
```

```
VALUES
```

```
(1, 1, 'A1-FILT-001', '2024-05-01', '2024-05-16', 'EN USO'),  
(2, 2, 'A2-FILT-001', '2024-05-05', '2024-05-12', 'EN USO');
```

Sensores

```
INSERT INTO sensores (id_linea, codigo_sensor, tipo_sensor, precision_ppm, unidad, fecha_instalacion)  
VALUES  
(1, 'A1S01', 'MQ-135', 0.50, 'ppm', '2024-04-20'),  
(1, 'A1S02', 'MQ-135', 0.50, 'ppm', '2024-04-21'),  
(2, 'A2S01', 'MQ-135', 0.50, 'ppm', '2024-04-22');
```

Productos

```
INSERT INTO productos (nombre, descripcion, es_benceno, unidad_medida)  
VALUES  
('Producto X', 'Producto químico con emisión de benceno', TRUE, 'kg'),  
('Producto Y', 'Producto químico sin benceno', FALSE, 'kg');
```

Clasificación PPM

```
INSERT INTO clasificacion_ppm (rango_min, rango_max, nivel, descripcion, accion_recomendada, color_alerta)  
VALUES  
(0.00, 0.99, 'Bajo', 'Nivel seguro', 'Operación normal', 'Verde'),  
(1.00, 4.99, 'Moderado', 'Requiere vigilancia', 'Revisar filtros próximamente', 'Amarillo'),  
(5.00, 9.99, 'Alto', 'Peligro para la salud', 'Detener línea y revisar filtros', 'Naranja'),  
(10.00, 1000.00, 'Crítico', 'Riesgo extremo', 'Detener planta y activar protocolo de emergencia', 'Rojo');
```

Configuración fábrica

```
INSERT INTO configuracion_fabrica (id_fabrica, frecuencia_lectura_seg, ppm_max_permitido,  
dias_max_uso_filtro, fecha_creacion, observaciones)  
VALUES  
(1, 10, 5.00, 15, '2024-04-01', 'Configuración inicial Fábrica A'),  
(2, 10, 5.00, 15, '2024-04-01', 'Configuración inicial Fábrica B');
```

Empleados

```
INSERT INTO empleados (documento, nombres, apellidos, cargo, telefono, email, id_fabrica)
```

```
VALUES
```

```
('100000001', 'Carlos', 'García', 'Operario', '3000000001', 'carlos@fabricaA.com', 1),  
('100000002', 'Ana', 'López', 'Supervisor', '3000000002', 'ana@fabricaA.com', 1);
```

Turnos

```
INSERT INTO turnos (nombre_turno, hora_inicio, hora_fin)
```

```
VALUES
```

```
('Turno 1', '08:00:00', '16:00:00'),  
('Turno 2', '16:00:00', '00:00:00'),  
('Turno 3', '00:00:00', '08:00:00');
```

Asignación empleado-turno

```
INSERT INTO empleado_turno (id_empleado, id_turno, fecha, es_supervisor)
```

```
VALUES
```

```
(1, 1, '2024-05-20', FALSE),  
(2, 1, '2024-05-20', TRUE);
```

Lecturas

```
INSERT INTO lecturas (
```

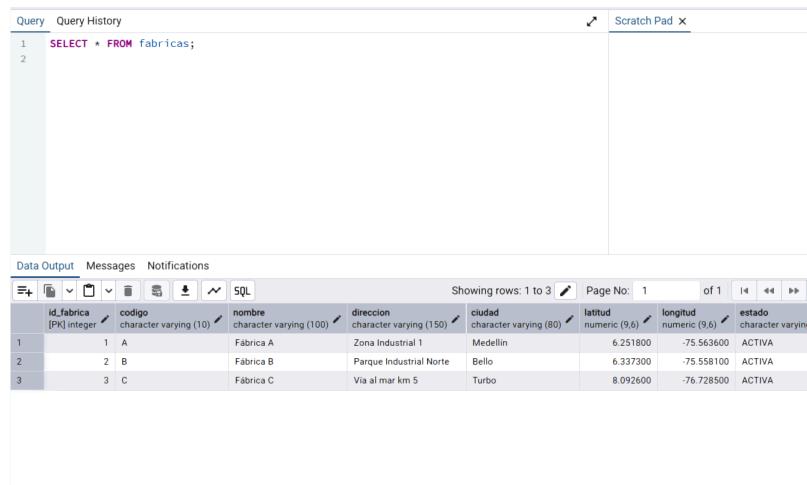
```
fecha_hora, id_fabrica, id_linea, id_sensor, id_filtro, id_producto,  
ppm, temperatura, humedad, latitud, longitud, id_clasificacion_ppm, alarma_activa, origen
```

```
) VALUES
```

```
('2024-05-20 08:00:10', 1, 1, 1, 1, 1, 0.80, 24.5, 55.0, 6.2518, -75.5636, 1, FALSE, 'local'),  
('2024-05-20 08:00:20', 1, 1, 1, 1, 1, 5.50, 24.7, 54.8, 6.2518, -75.5636, 3, TRUE, 'local'),  
('2024-05-20 08:00:30', 1, 1, 2, 1, 1, 10.20, 24.9, 54.5, 6.2518, -75.5636, 4, TRUE, 'local');
```

7.2- Pantallazos de consultas SELECT de las tablas pobladas

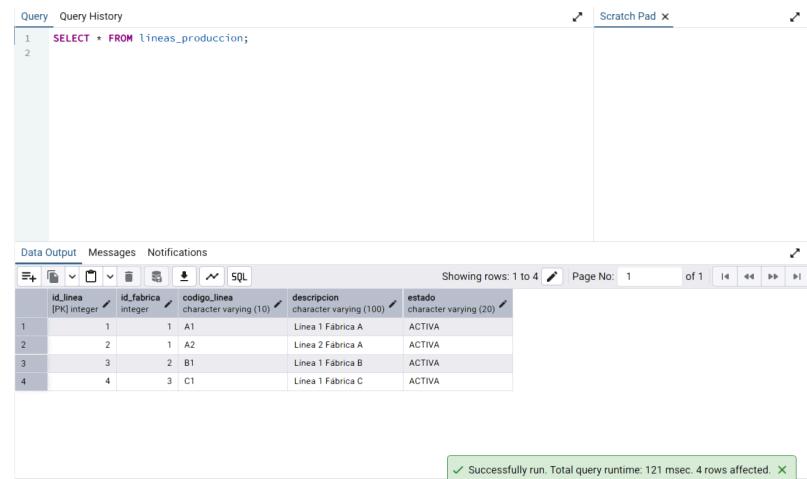
Primero verifico la tabla de fábricas:



```
Query Query History
1 SELECT * FROM fabricas;
2

Data Output Messages Notifications
SQL
Showing rows: 1 to 3 | Page No: 1 of 1 | 1 4 << >>
+----+----+----+----+----+----+----+----+
| id_fabrica | codigo | nombre | direccion | ciudad | latitud | longitud | estado |
+----+----+----+----+----+----+----+----+
| 1 | A | Fábrica A | Zona Industrial 1 | Medellin | 6.251800 | -75.563600 | ACTIVA |
| 2 | B | Fábrica B | Parque Industrial Norte | Bello | 6.337300 | -75.558100 | ACTIVA |
| 3 | C | Fábrica C | Vía al mar km 5 | Turbo | 8.092600 | -76.728500 | ACTIVA |
+----+----+----+----+----+----+----+----+
```

Luego revisó las líneas:

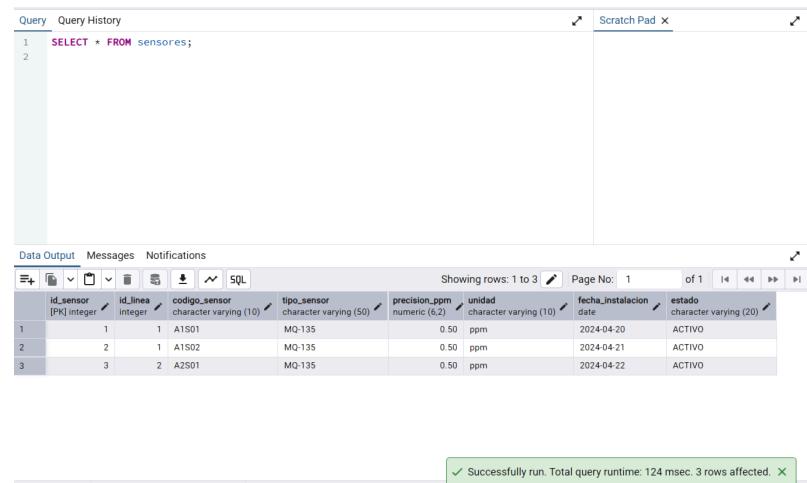


```
Query Query History
1 SELECT * FROM lineas_produccion;
2

Data Output Messages Notifications
SQL
Showing rows: 1 to 4 | Page No: 1 of 1 | 1 4 << >>
+----+----+----+----+----+
| id_linea | id_fabrica | codigo_linea | descripcion | estado |
+----+----+----+----+----+
| 1 | 1 | A1 | Línea 1 Fábrica A | ACTIVA |
| 2 | 1 | A2 | Línea 2 Fábrica A | ACTIVA |
| 3 | 2 | B1 | Línea 1 Fábrica B | ACTIVA |
| 4 | 3 | C1 | Línea 1 Fábrica C | ACTIVA |
+----+----+----+----+----+
```

✓ Successfully run. Total query runtime: 121 msec. 4 rows affected.

Después reviso los sensores:



```
Query Query History
1 SELECT * FROM sensores;
2

Data Output Messages Notifications
SQL
Showing rows: 1 to 3 | Page No: 1 of 1 | 1 4 << >>
+----+----+----+----+----+----+----+----+
| id_sensor | id_linea | codigo_sensor | tipo_sensor | precision_ppm | unidad | fecha_instalacion | estado |
+----+----+----+----+----+----+----+----+
| 1 | 1 | A1S01 | MQ-135 | 0.50 | ppm | 2024-04-20 | ACTIVO |
| 2 | 1 | A1S02 | MQ-135 | 0.50 | ppm | 2024-04-21 | ACTIVO |
| 3 | 2 | A2S01 | MQ-135 | 0.50 | ppm | 2024-04-22 | ACTIVO |
+----+----+----+----+----+----+----+----+
```

✓ Successfully run. Total query runtime: 124 msec. 3 rows affected.

Y finalmente reviso que las lecturas hayan quedado almacenadas correctamente:

The screenshot shows a MySQL Workbench interface. The 'Query' tab contains the SQL command: `SELECT * FROM lecturas LIMIT 20;`. The 'Data Output' tab displays the results of the query, which are three rows of data from the 'lecturas' table. The columns and their values are as follows:

	id_lectura [PK] bigint	fecha_hora timestamp without time zone	id_fabrica integer	id_linea integer	id_sensor integer	id_filtro integer	id_producto integer	ppm numeric (10,2)	temperatura numeric (5,2)	humedad numeric (5,2)	latitud numeric
1	1	2024-05-20 08:00:10	1	1	1	1	1	0.80	24.50	55.00	
2	2	2024-05-20 08:00:20		1	1	1	1	5.50	24.70	54.80	
3	3	2024-05-20 08:00:30		1	1	2	1	10.20	24.90	54.50	

Below the table, a message bar indicates: **Successfully run. Total query runtime: 117 msec. 3 rows affected.**

8.- Conclusiones.

Maria Camila Rodriguez:

La implementación del Gobierno de Datos y la arquitectura Big Data propuesta para la empresa “Sustancias Locas” me permitió organizar y entender mejor cómo una empresa puede usar los datos para mejorar sus procesos y tomar decisiones más acertadas. Al diseñar un sistema que recoge información de los sensores, la analiza y la muestra de manera clara, se logra un control más seguro de las emisiones y se facilita la labor del personal en planta.

Este proyecto no solo moderniza la infraestructura tecnológica, sino que también sienta las bases de una gestión responsable y eficiente del dato como activo estratégico para la organización.

Harlan Enciso:

Con base al trabajo me pasó es que mientras llenaba los 20 registros ficticios me di cuenta de lo repetitivas que son las lecturas cuando vienen cada 10 segundos y eso me ayudó a entender por qué el almacenamiento y la estructura importan tanto, si la tabla no está bien pensada, se llena rapidísimo o se vuelve difícil de manejar.

Lo aprendido en esta asignatura realmente es de gran valor ya que abre un camino para muchas oportunidades laborales en el mundo de los datos, en algún momento me interesé por este mundo ya que conseguí un trabajo donde trabaja manipulando datos a través de python y en ese momento me di cuenta que este campo tendría mucho futuro, pero realmente me apasiona otro campo, pero es algo que no descarto y los conocimientos obtenidos son una ventaja.

Miguel Agel Rojas:

Este proyecto me permitió entender de forma práctica cómo se integra una arquitectura Big Data completa, desde las fuentes de información hasta el procesamiento, almacenamiento y visualización de los datos. Trabajar con bases de datos, procesos ETL y diseño de estructuras me ayudó a valorar la importancia de planear bien cada componente para que el sistema funcione de manera continua y confiable. También reforzó la necesidad del trabajo en equipo y de una buena organización para poder complementar cada parte del proyecto. En general, fue una experiencia útil que fortaleció mis conocimientos y me dio una visión más clara de cómo se aplican estas tecnologías en problemas reales.

9.- Video de sustentación:

https://drive.google.com/file/d/1trQVjycVED6z5cQ9N3hKNFHKTayVEN_Y/view?usp=sharing