

Guía del Proyecto Final de Curso

Curso: Fundamentos de Arquitectura de Datos / Universidad Rafael Landívar

Modalidad: Trabajo en grupo (máximo 6 integrantes)

Fecha de entrega: Última sesión de curso

Título del Proyecto: Arquitectura de Datos para la Movilidad Inteligente de ACME EV

1. Objetivo General

Diseñar e implementar una Arquitectura de Datos moderna, escalable y eficiente que permita a la empresa ACME EV, dedicada a la venta de vehículos eléctricos, consolidar su estrategia de diferenciación basada en el valor de los datos, tanto para el cliente como para la empresa.

2. Caso de Negocio: ACME EV

ACME EV es una empresa de vehículos eléctricos con 10 sucursales en distintos países. Cada vehículo en circulación (actualmente 10,000 y creciendo un 20% anual) transmite datos periódicos:

- Cada 30 segundos: ubicación GPS (vigencia: 30 días).
- Cada minuto: fecha y hora, estado de batería (%), encendido/apagado, código de problema detectado (000-999,000=OK) y kilometraje (vigencia: 365 días).

Se sabe que en los próximos modelos de vehículos la información de estado que se envíe tendrá atributos adicionales, pero aun no conocemos su estructura.

Cada sucursal necesita acceder a estos datos para planificar mantenimientos, análisis de fallas y ofrecer valor agregado al cliente. ACME EV desea convertirse en un referente de analítica avanzada de movilidad eléctrica.

Cada propietario puede descargar la información GPS (VIN, fecha/hora, latitud, longitud) de los vehículos de su propiedad. Para lo cual tiene algún nivel de seguridad para control de acceso.

ACME EV debe garantizar que la información recopilada este disponible en los plazos indicados (30 días de GPS, 365 días de Estados). Necesitan asegurar la disponibilidad, la integridad y la confidencialidad.

3. Actividades del Proyecto

El grupo debe asumir dos roles principales:

A. Como Arquitectos de Datos:

- Diseño conceptual y lógico de la arquitectura de tal forma que sea un documento para entendimiento desde el punto de vista de negocio.
- Integración de fuentes de datos y flujo entre componentes.

- Uso de arquitecturas Lambda (batch y streaming) o Kappa (todo es streaming), justificando su elección.
- Plan de escalabilidad para soportar 5 años de crecimiento.
- Controles de seguridad y gobernanza de datos.
- Justificación de alineación con la estrategia empresarial.

B. Como Ingenieros de Datos:

- Selección de tecnologías (bases de datos, herramientas ETL/ELT, almacenamiento distribuido, etc.).
- Simulación de implementación mediante datos sintéticos.
- Carga de datos transaccionales desde editores simulados.
- Almacenamiento en SQL y NoSQL.
- Procesamiento de flujos con Kafka, Spark o similar.
- Serialización en JSON, CSV, Parquet.
- Comparación de rendimiento entre procesos batch y en tiempo real.

4. Entregables

Informe Técnico del Proyecto (máx. 10 páginas):

1. Diseño de Arquitectura (documento para discusión a nivel gerencial, evitando los tecnicismos)
2. Diseño de ingeniería (modelos que expliquen el “que” y no el “como” implementar la solución).

Prototipo:

1. Implementación de ingesta de datos GPS y Estados (datos generados por el vehículo).
2. Implementación de recuperación de datos para el cliente (datos GPS).
3. Implementación de recuperación de datos para la sede (datos de Estados).

5. Criterios de Evaluación

Criterio	Ponderación
Diseño de la Arquitectura Conceptual	20%
Justificación y alineación estratégica	10%
Implementación y simulación técnica	10%
Comparación y análisis de eficiencia	20%
Gobernanza y seguridad de datos	10%
Calidad del informe final	10%
Presentación y defensa del proyecto	20%

6. Tecnologías sugeridas (no limitativas)

- Bases de datos: PostgreSQL, MongoDB, Cassandra.
- Procesamiento: Apache Kafka, Spark, KNIME.
- Lenguajes: Python, SQL, Bash.
- Serialización: CSV, JSON, Parquet.
- Infraestructura: Puede ser local y/o nube.

7. Recomendaciones Finales

- Documentar cada decisión técnica con argumentos sólidos.
- Asegurar la trazabilidad del flujo de datos de punta a punta.
- Utilizar ejemplos reales o sintéticos acordes al volumen proyectado.
- Considerar eficiencia en costo, mantenimiento y escalabilidad.

8. Ejemplos de los datos que se generan

Información GPS generada y enviada por cada vehículo (VIN) cada 30 segundos a la sede central de ACME EV.

Explicación de la estructura:

VIN = Identificación única del vehículo (*Vehicle Identification Number*)

Fecha / Hora: Fecha y hora de generación de la ubicación del vehículo

Latitud / Longitud: Coordenadas GPS del vehículo a la “Fecha” y “Hora” indicada.

VIN	Fecha	Hora	Latitud	Longitud
VIN00009	2025-05-25	18:15:32.0000000	19.365659	-99.075285
VIN00005	2025-05-25	18:15:32.0000000	19.329478	-99.102546
VIN00007	2025-05-25	18:15:32.0000000	19.375532	-99.04688
VIN00031	2025-05-25	18:15:32.0000000	19.417649	-99.049235
VIN00004	2025-05-25	18:15:32.0000000	19.481425	-99.118248
VIN00033	2025-05-25	18:15:32.0000000	19.400359	-99.117152
VIN00008	2025-05-25	18:15:32.0000000	19.355704	-99.182905
VIN00065	2025-05-25	18:15:33.0000000	19.351488	-99.145514
VIN00034	2025-05-25	18:15:32.0000000	19.483686	-99.194185
VIN00028	2025-05-25	18:15:32.0000000	19.40999	-99.099288
VIN00018	2025-05-25	18:15:32.0000000	19.329816	-99.089542
VIN00029	2025-05-25	18:15:32.0000000	19.395762	-99.173816
VIN00058	2025-05-25	18:15:33.0000000	19.327956	-99.138348
VIN00077	2025-05-25	18:15:33.0000000	19.378069	-99.11854
VIN00079	2025-05-25	18:15:33.0000000	19.467511	-99.005345
VIN00020	2025-05-25	18:15:32.0000000	19.318262	-99.04607
VIN00017	2025-05-25	18:15:32.0000000	19.308524	-99.182807

Información de Estados generada y enviada por cada vehículo (IdDispositivo) cada minuto a la sede central de ACME EV.

Explicación de la estructura:

idDispositivo: VIN (*Vehicle Identification Number*)

FechaHora: Fecha y Hora en que se recibió se generó la información de Estado

EstadoCarga: Porcentaje de carga que tiene la batería del vehículo

OnOff: El vehículo esta Encendido=On=1, Apagado=Off=0

Estado: Codigo de desperfecto reportado. Valores entre 0 y 999; 0 indica "Sin problema"

Kilometros: Cantidad de Kilometros recorridos por el vehículo (odómetro)

IdDispositivo	FechaHora	EstadoCarga	OnOff	Estado	Kilometros
VIN00005	2025-05-26 20:30:51.000	58	0	0	470
VIN00004	2025-05-26 20:30:51.000	80	1	15	623
VIN00002	2025-05-26 20:30:51.000	31	0	0	477
VIN00003	2025-05-26 20:30:51.000	13	0	0	503
VIN00001	2025-05-26 20:30:51.000	18	1	0	825
VIN00005	2025-05-26 20:31:50.000	13	0	0	470
VIN00004	2025-05-26 20:31:50.000	28	0	0	623
VIN00005	2025-05-26 20:32:49.000	57	0	0	470
VIN00004	2025-05-26 20:32:49.000	86	0	217	623
VIN00001	2025-05-26 20:31:50.000	69	0	0	825
VIN00005	2025-05-26 20:33:48.000	1	1	0	470
VIN00004	2025-05-26 20:33:48.000	77	1	0	623
VIN00001	2025-05-26 20:32:49.000	54	1	0	827
VIN00005	2025-05-26 20:34:47.000	56	0	0	470
VIN00001	2025-05-26 20:33:48.000	8	0	0	827
VIN00004	2025-05-26 20:34:47.000	50	1	0	624
VIN00005	2025-05-26 20:35:46.000	37	0	0	470