**INFORME DE RESULTADOS**

**Unidad 3 - Tarea 6**

**Peso: 10%**

**Equipo/Grupo : 3**

**Estudiantes :**

**- *Luis Felipe Gil Gómez***

* ***Gerson Gustavo Fernández Badillo***
* ***Davidson Harley Rave Buitrango***
* ***David Pabón García***

**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”.**

Se requiere la convergencia de talento humano, planificación del trabajo, contratación de diversos servicios, instalación y configuración de aplicaciones y servidores locales y en la Nube, desarrollo de software, soporte técnico, entre otros.

Por lo tanto, es necesario la conformación de un gobierno de datos y la estructuración de un Proyecto de Big Data; así como una lista de verificación del Proyecto Big Data que incluya todos los ítems necesarios y los tiempos estimados para lograr el objetivo.

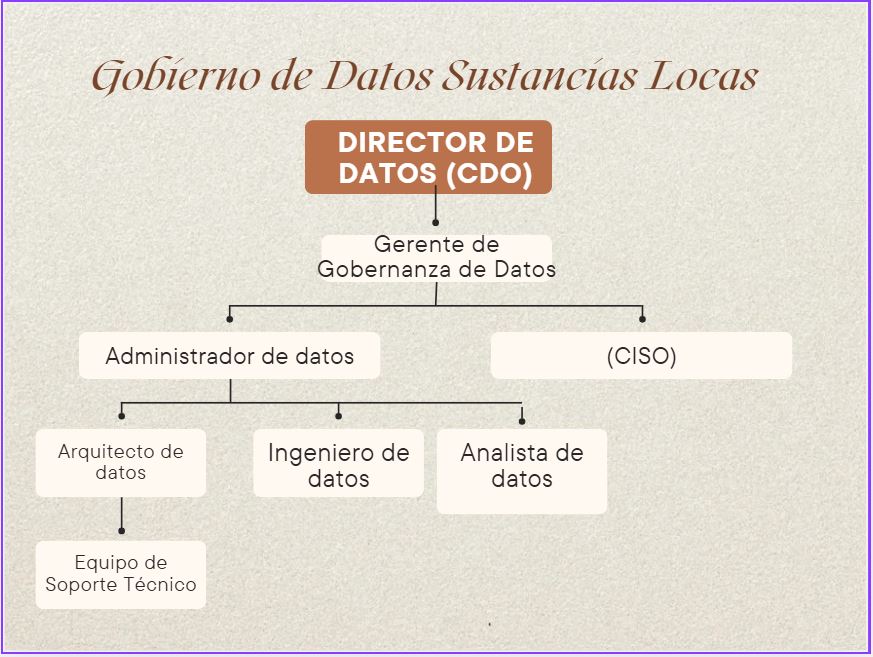
**1.-Diseño de un Gobierno de Datos**

*Después de analizar el enunciado y evaluar los requerimientos, proponga la estructura y roles que debe tener el gobierno de datos que se encargara del Proyecto Big Data*

**1.1- Roles del gobierno de datos**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Departamento o persona** | **Descripción** | **Rol, responsabilidades y/o funciones** |
| **1** | Director de Datos (CDO) | Miembro de la alta dirección o gerencia, responsable de la estrategia de datos general. | Liderar la estrategia de datos, coordinar roles, asegurar cumplimiento de políticas y alineación con objetivos corporativos. |
| **2** | Gerente de Gobernanza de Datos | Coordinador general de gobierno de datos. | Establecer políticas, procesos y estándares; coordinar personal. |
| **3** | Administrador de datos | Especialista en calidad y gestión de datos por área. | Supervisar la calidad de datos, garantizar metadatos correctos y coordinar con usuarios para mantener integridad de datos. |
| **4** | Arquitecto de datos | Diseñador de la infraestructura de datos. | Diseñar modelos de datos, arquitectura de almacenamiento, flujos de datos e integración con sistemas de terceros. |
| 5 | Ingeniero de datos | Especialista en extracción, transformación y almacenamiento de datos. | Desarrollar sistemas de datos, asegurar escalabilidad y disponibilidad del sistema de almacenamiento. |
| 6 | Analista de datos | Analista y modelador de datos. | Analizar las estadísticas para la detección y predicción. |
| 7 | Director de Seguridad de la Información (CISO) | Responsable de la seguridad y privacidad de los datos. | Definir políticas de acceso, cifrado, control de privacidad, cumplimiento legal y protección de datos sensibles. |
| 8 | Equipo de Soporte Técnico | Personal de infraestructura y soporte TI. | Instalación, configuración y mantenimiento de servidores, hardware y software. |

***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** | Gobierno de Datos | Conjunto de políticas, roles, estándares y métricas que aseguran la calidad, disponibilidad y seguridad de los datos. | Debe incluir roles como DataSteward, Data Owener y definir políticas de calidad |
| **2** | Arquitectura del Sistema | Infraestructura tecnológica que integra sensores, microcontroladores, servidores locales y servicios en la nube | Incluir arquitectura hibrida (on-premise + cloud) |
| **3** | Adquisición de Datos | Captura de datos desde sensores instalados en los filtros de las líneas de producción. | Sensores más actuales requeridos. |
| **4** | Ingesta de Datos | Procesos para recolectar y almacenar los datos en servidores temporales | Herramientas como APACHE NiFi o AWS loT Core. |
| 5 | Almacenamiento de Datos | Bases de datos distribuidas para almacenar datos estructurados y no estructurados en la nube y localmente | AWS S3, Amazon Redshift, MongoBD, Entre otros. |
| 6 | Procesamiento de Datos | Limpieza, transformación y modelado de datos para análisis. | Utilizar Spak, AWS Glue u otras herramientas de ETL. |
| 7 | Analítica de Datos | Aplicación de modelos para detectar anomalías y predecir el desgaste de los filtros | Algoritmos de machine learning. |
| 8 | Visualización de Datos | Desarrollo de dashboards en tiempo real para visualizar niveles de ppm, alertas de estado general | Usar herramientas como Power BI, Tableau, o dashboards personalizados. |
| 9 | Alarmas y Protocolos de Seguridad | Generacion de alertas visuales, sonoras y digitales según umbral de ppm. Activacion de protocolos | Definir niveles críticos |
| 10 | Gestion de Turnos y Usuarios | Registro de los operarios y supervisores por turno. Permitir trazabilidad de acciones por usuario. | Integracion con sistemas de autenticación. |
| 11 | Desarrollo de Software | Aplicaciones web o móviles para gestión del sistema, visualización de datos y alertas | Incluir pruebas y control de calidad del software |
| 12 | Soporte Tecnico y Mantenimiento | Gestion de incidencias, Monitoreo continuo del sistema, mantenimiento de hardware y software. | Establecer SLA para respuesta rápida. |
| 13 | Seguridad de la informacion | Proteccion de datos sensibles, cifrado de datos en transito y en reposo. | Aplicar políticas de ciberseguridad de AWS y en los servidores locales. |
| 14 | Formacion del Talento Humano | Capacitacion al personal técnico y operativo sobre el uso del sistema, interpresetacion de alertas y respuesta a emergencias | Incluir manuales y entrenamiento por turnos. |
| 15 | Evaluacion y Mejora Continua | Revision periódica del sistema y los resultados para ajustar el modelo de predicción y mantenimiento preventivo. | Plan de mejora cada 3 o 6 meses. |

**2.2- Lista de verificación del Proyecto Big Data**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Elemento** | **Estado** | **Tiempo** |
| **1** | Conformación del equipo técnico | Pendiente | 3 dias |
| **2** | Definición del Gobierno de Datos | En proceso | 5 dias |
| **3** | Instalación de sensores nuevos | Pendiente | 7 dias |
| 4 | Adquisición de hardware | Pendiente | 5 dias |
| 5 | Configuración de microcontroladores | Pendiente | 3 dias |
| 6 | Configuración de red y servidores | Pendiente | 4 dias |
| 7 | Contratación de servicios en la nube | En proceso | 2 dias |
| 8 | Desarrollo del sistema de ingesta | No iniciado | 6 dias |
| 9 | Diseño de base de datos | En proceso | 4 dias |
| 10 | Desarrollo de procesos ETL | No Iniciado | 5 dias |
| 11 | Implementacion de analitica | No iniciado | 7 dias |
| 12 | Diseño del dashboard | No iniciado | 5 dias |
| 13 | Desarrollo de alarmas y procotolos | No iniciado | 6 dias |
| 14 | Pruebas del sistema | No iniciado | 4 dias |
| 15 | Capacitacion al personal | No iniciado | 4 dias |
| 16 | Documentacion del sistema | No iniciado | 3 dias |
| 17 | Mantenimiento y soporte | Planificado | Permanente |
| 18 | Evaluacion final y entrega | No iniciado | 2 dias |

**3.- Conceptos y elementos de Big Data**

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Componente** | **Descripción (breve cita bajo Norma APA)** | **Referencia (APA)** |
| **1** | Arquitectura de Sistema de Información | Conjunto estructurado de componentes hardware y software que soporta el procesamiento, almacenamiento y análisis de datos. | Laudon, K. C., & Laudon, J. P. (2020). Management Information Systems (16th ed.). Pearson. |
| **2** | Hadoop | Framework de código abierto para almacenamiento distribuido y procesamiento masivo de datos usando el paradigma MapReduce. | White, T. (2015). Hadoop: The Definitive Guide (4th ed.). O'Reilly Media. |
| **3** | Spark | Motor de procesamiento distribuido en memoria, útil para procesamiento de datos en tiempo real y por lotes. | Zaharia, M., et al. (2016). Apache Spark: A Unified Engine for Big Data Processing. ACM. |
| **4** | PostgreSQL | Sistema de gestión de bases de datos relacional (SGBD) avanzado, usado para el almacenamiento local y en la nube. | PostgreSQL Global Development Group. PostgreSQL Documentation. https://www.postgresql.org |
| **5** | Amazon Kinesis FireHose | Servicio de AWS que permite la ingesta y almacenamiento de datos en tiempo real hacia S3, Redshift u otros destinos. | Amazon Web Services. *Amazon Kinesis Documentation*. <https://docs.aws.amazon.com> |
| **6** | Amazon QuickSight | Herramienta de Business Intelligence de AWS para visualizar datos y crear dashboards interactivos. | Amazon Web Services. Amazon QuickSight User Guide. https://docs.aws.amazon.com |
| **7** | Amazon Glue | Servicio de ETL (Extracción, Transformación y Carga) de AWS que automatiza el movimiento de datos desde S3 hacia Redshift o RDS. | Amazon Web services. |
| **8** | HTTP API | Interfaz de comunicación que permite el envío y recepción de datos entre microcontroladores y servidores mediante protocolo HTTP. | Fielding, R. T. (2000). Architectural Styles and the Design of Network-based Software Architectures. |
| **9** | Boto3 | SDK de Amazon para Python que permite interactuar con servicios como S3, Kinesis, Glue, entre otros. | Amazon Web Services. Boto3 Documentation. |
| **10** | Amazon RedShift | Almacén de datos (data warehouse) escalable de AWS diseñado para realizar análisis complejos de grandes volúmenes de datos. | Amazon Web Services. Amazon Redshift Guide. |
| **11** | Amazon Athena | Servicio de AWS para ejecutar consultas SQL directamente sobre datos almacenados en S3 sin necesidad de cargarlos a una base de datos. | Amazon Web Services. Amazon |
| **12** | Amazon CloudWatch | Servicio de monitoreo de infraestructura, aplicaciones y logs dentro de la nube de AWS. | Amazon Web Services. Amazon |
| **13** | Boto3 | SDK de Amazon para Python que permite interactuar con servicios como S3, Kinesis, Glue, entre otros. | Amazon Web Services. Boto3 Documentation. |
| **14** | Formato de datos JSON | Formato de texto ligero y legible utilizado para el intercambio de datos entre dispositivos (por ejemplo, sensores y servidores). | ECMA International. ECMA-404 The JSON Data Interchange Standard. |
| **15** | Formato de datos CSV | Formato de texto plano que almacena datos tabulados separados por comas, útil para hojas de cálculo y exportación/importación. | Shafranovich, Y. (2005). Common Format and MIME Type for Comma-Separated Values (CSV) Files. RFC 4180. |
| **16** | Formato de datos Parquet | Formato de almacenamiento columnar optimizado para grandes volúmenes de datos, ideal para compresión y consultas en Big Data. | Apache Software Foundation. |
| **17** | Microcontrolador Arduino | Placa de desarrollo con microcontrolador, utilizada en IoT para control y lectura de sensores. | Banzi, M., & Shiloh, M. Getting Started with Arduino (3rd ed.). Maker Media. |
| **18** | Microcontrolador ESP8266 | Microcontrolador con conectividad WiFi, usado para capturar y enviar lecturas de sensores a través de APIs HTTP. | Espressif Systems. ESP8266 Technical Reference. |
| **19** | Sensor MQ-135 | Sensor de gas que detecta concentraciones de benceno y otros gases nocivos en el aire. | Winsen Electronics. MQ-135 Gas Sensor Datasheet. |
| **20** | Benceno | Compuesto químico volátil, cancerígeno, monitoreado en la fábrica por representar un riesgo ambiental y ocupacional. | OSHA. Benzene - Occupationa |

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

*Colocar en esta sección los datos que se generan en el sensor y se almacenan en la hoja de cálculo local “lecturas” en el microcontrolador. El mismo formato de hoja de cálculo se utiliza en el servidor local.*

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

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Nombre Columna** | **Descripción** | **Observaciones** |
| **1** | id\_lectura | Identificador único de la lectura | Valor incremental automático |
| **2** | fecha\_hora | Fecha y hora en que se toma la lectura | Formato: AAAA-MM-DD HH:MM:SS |
| **3** | ppm\_benceno | Nivel de concentración de benceno en partes por millón (ppm) | Dato numérico con hasta dos decimales |
| 4 | id\_sensor | Identificador del sensor que realiza la lectura | Relación con base de sensores |
| **5** | estado\_alarma | Estado de la alarma generada (normal, advertencia, crítica) | Se genera automáticamente por lógica de control |
| **6** | id\_linea\_produccion | Identificador de la línea de producción donde está ubicado el sensor | Puede ser 1, 2, 3 o 4 |
| **7** | usuario\_turno | Usuario o supervisor responsable en el turno actual | Asignado por autenticación del sistema |

**4.2. – Poblamiento de la hoja de cálculo del microcontrolador**

*Colocar en esta sección veinte (20) registros con datos ficticios. Nota: puede colocar el “pantallazo” de la hoja de cálculo o incrustar “copia” y “pega”. Las columnas son los datos*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **#**  (id\_lectura) | **Dato Columna 1**  (fecha\_hora) | **Dato Columna 2**  ppm\_benceno | **Dato Columna 3**  ((id\_sensor) | **Dato Columna 4**  (estado\_alarma) | **Dato Columna 5**  (id\_linea\_produccion) | **Dato Columna 6**  (usuario\_turno) |
| **1** | 2025-05-21 08:00:01 | 0.5 | S01 | normal | 1 | turno1 |
| **2** | 2025-05-21 08:05:15 | 1.1 | S02 | advertencia | 1 | turno1 |
| **3** | 2025-05-21 08:10:32 | 2.6 | S03 | crítica | 2 | turno1 |
| **4** | 2025-05-21 08:15:45 | 0.8 | S04 | normal | 3 | turno1 |
| **5** | 2025-05-21 08:21:10 | 1.7 | S01 | advertencia | 1 | turno1 |
| **6** | 2025-05-21 08:25:50 | 3.4 | S02 | crítica | 1 | turno1 |
| **7** | 2025-05-21 08:30:30 | 0.9 | S03 | normal | 2 | turno1 |
| **8** | 2025-05-21 08:35:00 | 1.3 | S04 | advertencia | 3 | turno1 |
| **9** | 2025-05-21 08:40:45 | 0.4 | S01 | normal | 1 | turno1 |
| **10** | 2025-05-21 08:45:30 | 2.1 | S02 | crítica | 1 | turno1 |
| **11** | 2025-05-21 08:50:00 | 0.6 | S03 | normal | 2 | turno1 |
| **12** | 2025-05-21 08:52:10 | 1.5 | S04 | advertencia | 3 | turno1 |
| **13** | 2025-05-21 08:55:25 | 3.0 | S01 | crítica | 1 | turno1 |
| **14** | 2025-05-21 08:58:42 | 1.0 | S02 | advertencia | 1 | turno1 |
| **15** | 2025-05-21 09:00:55 | 0.7 | S03 | normal | 2 | turno1 |
| **16** | 2025-05-21 09:03:00 | 1.8 | S04 | crítica | 3 | turno1 |
| **17** | 2025-05-21 09:06:10 | 0.3 | S01 | normal | 1 | turno1 |
| **18** | 2025-05-21 09:08:25 | 2.9 | S02 | crítica | 1 | turno1 |
| **19** | 2025-05-21 09:10:40 | 0.6 | S03 | normal | 2 | turno1 |
| **20** | 2025-05-21 09:13:00 | 1.2 | S04 | advertencia | 3 | turno1 |

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

*Colocar en esta sección la estructura de la tabla de “lecturas” dónde se almacenan los datos de las diferentes hojas de cálculo que provienen de los sensores. Esta tabla se encuentra dentro de la base de datos “monitoreo-produccion”*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Tabla** | **lecturas** | |
| **#** | **Nombre Campo** | **Tipo de dato** | **Tamaño (Bytes)** |
| **1** | id\_lectura | SERIAL (PK) | 4 |
| **2** | fecha\_hora | TIMESTAMP | 8 |
| **3** | ppm\_benceno | DECIMAL(4,2) | 4 |
| **4** | id\_sensor | VARCHAR(10) | 10 |
| **5** | estado\_alarma | VARCHAR(15) | 15 |
| **6** | id\_linea\_produccion | INTEGER | 4 |
| **7** | usuario\_turno | VARCHAR(20) | 20 |
| **Tamaño de un (1) registro en bytes** | | | 65 bytes |

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

*Colocar en esta sección el diagrama de Entidad-Relación sin atributos y el Diccionario de datos de la base de datos de gestión. Esta es la base de datos que contiene, además de la tabla lecturas,*

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

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

**6.3- Scripts de creación de las tablas de la Base de Datos “monitoreo-produccion”**

**7.- Poblamiento de la base de datos “monitoreo-produccion”**

*Colocar en esta sección un pantallazo de los resultados de DML (INSERTS) en pgAdmin4 realizados en todas las tablas de la base de datos.*

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

**7.2- Pantallazos de consultas SELECT de las tablas pobladas**

**8.- Conclusiones.**

*Elabore las conclusiones planteando la importancia y utilidad de esta tarea y su relación con el contenido de la asignatura. ¿Cuáles eran sus expectativas al inicio de la asignatura y en este momento final? ¿Cómo cree Ud. que el conocimiento aprendido en la asignatura afectará sus oportunidades laborales y desempeño profesional una vez obtenga su título en el Pascual Bravo?*

*Adicionalmente a estas conclusiones generales del grupo de trabajo, cada estudiante debe expresar sus propias conclusiones.*

**9.- Video de sustentación:**

*Elabore un video de sustentación con la participación de todos los integrantes (si es en equipo). Este vídeo debe informar sobre las actividades realizadas en general . Coloque el enlace en esta sección*