1. **Organizar el Trabajo**

* **Identificar las Tareas:**
  + Investigación de conceptos clave: Big Data, Data Mining, Machine Learning.
  + Análisis de repositorios de datos: Data Lake, Data Warehouse, Data Mart.
  + Evaluación de herramientas para Big Data: sistemas de gestión de flujo de datos, motores de consulta (como Apache Hadoop, Apache Spark), y plataformas de gestión de datos (como Cloudera, AWS, Google BigQuery).
  + Revisión de requerimientos de hardware: servidores, almacenamiento, redes, GPUs, CPUs.
  + Presupuestación: estimar costos de software, hardware y recursos humanos.
  + Definición de cronograma: tiempo estimado para la implementación del sistema.
* **Agrupar las Tareas por Tipo:**
  + **Investigación Técnica:** Enfocada en entender Big Data y las herramientas tecnológicas.
  + **Análisis de Hardware:** Evaluar qué infraestructura es necesaria para soportar el sistema.
  + **Planificación del Proyecto:** Establecer un cronograma y estimar costos.
* **Distribuir las Tareas:**
  + Asignar tareas específicas a los miembros del grupo según sus habilidades e intereses.
* **Definir Fechas para Reuniones:**
  + Establecer reuniones regulares para revisar avances y coordinar ajustes necesarios.
* **Definir Fecha Límite de Entrega:**
  + Determinar la fecha final de entrega y asegurarse de que todas las tareas se completen a tiempo.

1. **Temas que Deben Conocer y Entender**

**Repositorios de Datos:**

* **Data Lake:** Almacena grandes cantidades de datos sin procesar y en su formato original. Permite flexibilidad y es ideal para análisis complejos y no estructurados.
* **Data Warehouse:** Repositorio estructurado que almacena datos procesados y organizados para la toma de decisiones y reportes empresariales.
* **Data Mart:** Subconjunto de un Data Warehouse, enfocado en una línea específica del negocio o departamento.

**Herramientas para Big Data:**

* **Gestión de Flujo de Datos:** Herramientas como Apache NiFi y Kafka ayudan a gestionar el flujo de datos en tiempo real.
* **Motores de Consulta:** Apache Hive y Presto permiten consultas rápidas sobre grandes conjuntos de datos.
* **Plataformas de Gestión de Datos:** Herramientas como Hadoop, Spark, y plataformas en la nube (AWS, Google Cloud, Azure) facilitan el procesamiento y análisis de grandes volúmenes de datos.

**Hardware:**

* **Servidores:** Deben tener alta capacidad de procesamiento y almacenamiento, con CPUs multicore y posibilidad de escalar.
* **Almacenamiento:** Unidades SSD para rapidez y alta capacidad para grandes volúmenes de datos.
* **Redes:** Conectividad de alta velocidad, con switches y routers capaces de manejar grandes volúmenes de tráfico de datos.
* **GPUs:** Para tareas intensivas de Machine Learning y procesamiento paralelo.

1. **Presupuestación y Planificación del Proyecto**

* **Definir Costos:** Evaluar costos de hardware, licencias de software, y recursos humanos.
* **Tiempo de Desarrollo:** Definir un cronograma detallado, incluyendo fases de instalación, configuración, pruebas y despliegue final.
* **Propuesta al Cliente:** Crear un documento que detalle los servicios ofrecidos, tiempo estimado de implementación, y costos asociados.

1. **Revisión y Presentación Final**

* Consolidar toda la información en un documento claro y profesional.
* Preparar una presentación para explicar al cliente los beneficios del sistema propuesto y responder posibles dudas.