***Proyecto - Ingeniería de datos***

**Necesidad:**

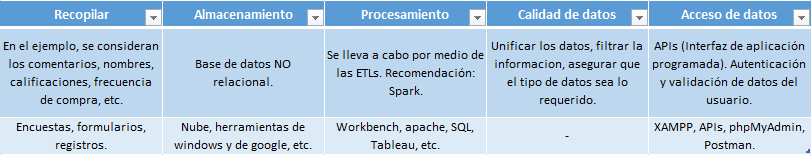
Establecer un sistema con la información relevante para mejorar la organización del stock de guitarras acústicas en la tienda.

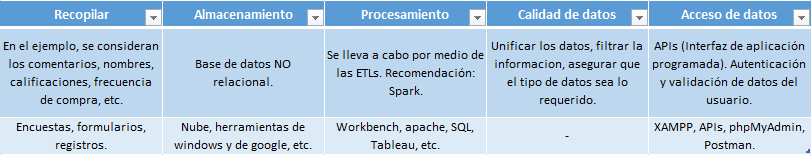
**Objetivo general:**

Crear una base de datos que permita mejorar la eficacia y rapidez en el acceso a la información correspondiente a las características de los instrumentos; como sus modelos, proveedores, ubicación (vitrina/bodega) y precios.

**Objetivos específicos:**

* Agrupar la información correspondiente a las guitarras acústicas en una única base de datos con el fin de obtener un acceso más eficiente y consolidado a las especificaciones de los productos, y de esta manera, facilitar el servicio de los asesores al momento de atender a los posibles clientes.
* Utilizar las herramientas (...) con el propósito de recopilar, almacenar, procesar, verificar la calidad y acceder a los datos de forma óptima.
* Establecer una base de datos completa y eficiente que cumpla con los requisitos establecidos por el cliente y almacene la información necesaria para el desarrollo y acceso al stock.
* Llevar a cabo pruebas indispensables para la verificación del funcionamiento de la base de datos, asegurando que el sistema unifique el formato, devuelva datos de calidad, evite duplicados y garantice que la información sea precisa de acuerdo con las necesidades del cliente.

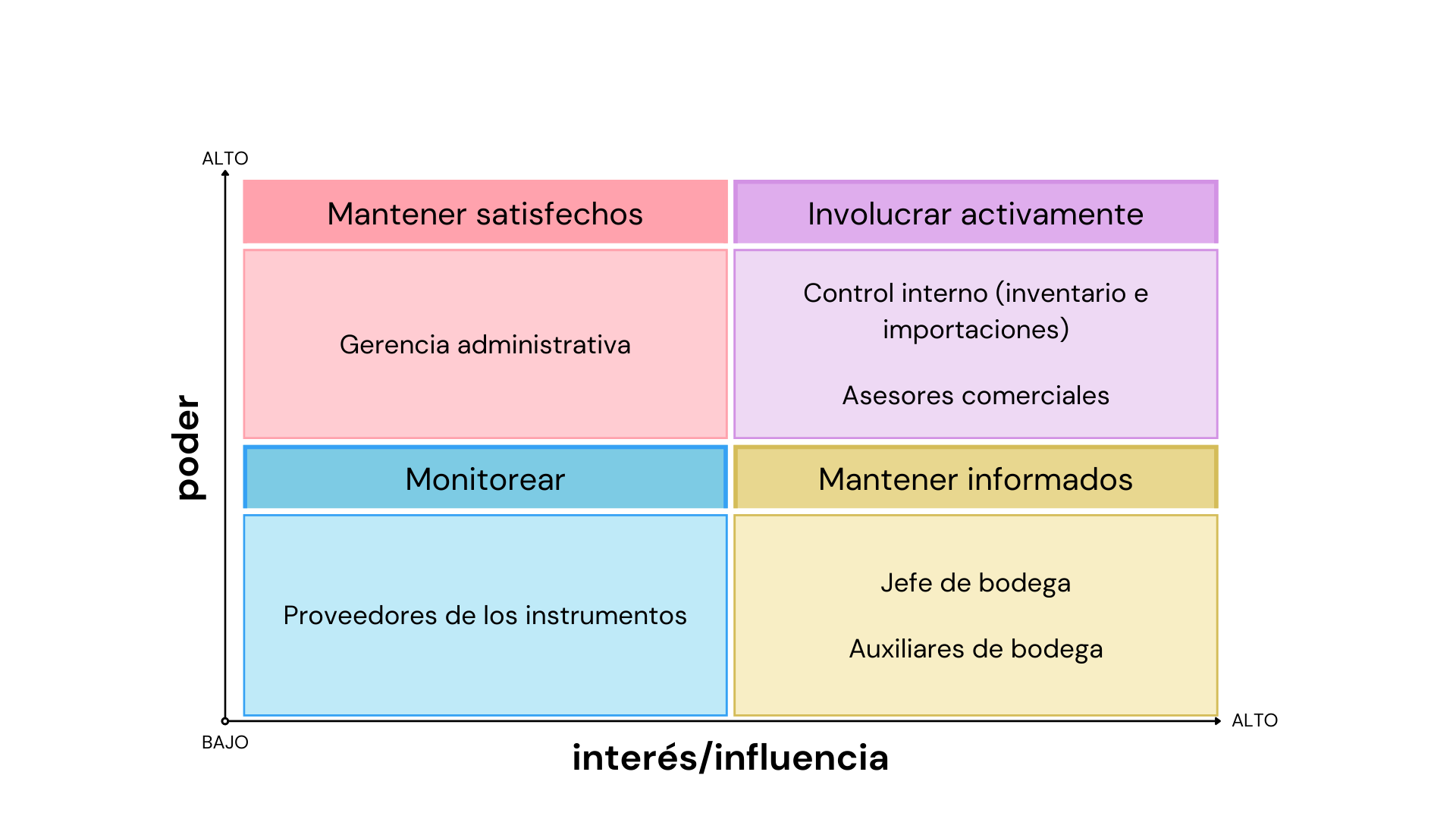




**ETL:**

Spark:

* Apache Spark es un motor rápido de uso general para computación de clústeres.
* Spark permite que los programadores escriban aplicaciones rápidamente en Java, Scala, Python, R y SQL para que puedan acceder a ellas los desarrolladores, los científicos de datos, etc.
* Con Spark SQL, los usuarios se pueden conectar a cualquier fuente de datos y presentarla en forma de tabla para que puedan utilizarla los clientes de SQL. Por si fuera poco, resulta muy sencillo desplegar algoritmos interactivos de aprendizaje automático en Spark. En cambio, con un motor de solo SQL, como Apache Impala, Apache Hive o Apache Drill, los usuarios solo pueden utilizar SQL o lenguajes semejantes para consultar los datos almacenados en distintas bases de datos. Por lo tanto, son frameworks más limitados que Spark.
* Ventajas: Velocidad (Ejecuta las cargas de trabajo 100 veces más rápido que con Hadoop), facilidad (cuenta con más de 80 operadores generales que facilitan el desarrollo de aplicaciones en paralelo), uso general (Spark permite usar una pila de bibliotecas que incluye SQL, DataFrame, MLlib para aprendizaje automático, GraphX y Spark Streaming), es de código abierto.



El método seleccionado es de tipo interactivo, con el fin de tener una comunicación directa con el cliente para entender claramente sus necesidades y requisitos. Esto se llevará a cabo mediante la técnica encuesta, considerando que es de vital importancia la participación de múltiples áreas pertenecientes al contexto laboral y esta técnica nos permite conocer diversos puntos de vista en torno al sistema de stock actual con el fin de identificar los puntos de mejora principales, almacenando sus respuestas y opiniones de forma clara y organizada. Algunos posibles ejemplos de preguntas que harían parte de la encuesta serían los siguientes:

1. ¿Actualmente qué cargo desempeña dentro de la empresa?

2. ¿Considera que el sistema actual de stock para las guitarras acústicas podría ser mejorado? *(totalmente en desacuerdo / en desacuerdo / neutral / de acuerdo / totalmente de acuerdo)*

3. ¿Cree que la información relacionada con la disponibilidad de instrumentos es de fácil acceso y entendimiento en la actualidad? *(totalmente en desacuerdo / en desacuerdo / neutral / de acuerdo / totalmente de acuerdo)*

4. Con el fin de optimizar el acceso a los datos de mayor importancia para facilitar su labor, ¿Qué información considera relevante incluir en el stock de guitarras acústicas de la tienda? *(Checklist: modelo, marca, proveedor, precio, cantidad, ubicación, otro - ¿cuál?)*

5. ¿Tiene algún comentario adicional con respecto a lo que desea ver evidenciado en la optimización del sistema actual?

Tarea: Investigar sobre las técnicas de estimación en SCRUM **Planning Poker y T-Shirt Sizes** y responder:

1. ¿Cómo se puede definir la estimación en SCRUM?
2. Para qué se utilizan las técnicas de estimación en SCRUM
3. En qué consiste cada técnica de estimación indicada

Sol: <https://www.institutoagile.com/post/estimaciones-con-story-points-en-scrum#:~:text=User%20Story%20es%20la%20t%C3%A9cnica,relativa%2C%20conocida%20como%20Story%20Points>.

1. La estimación es una herramienta útil para la planificación y para tener un buen desarrollo de un proyecto, pero se entiende que es una aproximación que puede ajustarse a medida del avance del proyecto y la nueva información disponible, así hacer correcciones tempranas.
2. Cuando implementamos Scrum podemos utilizar la técnica de estimación que mejor se adapte a nuestro proyecto, ya que existen varios tipos de técnicas de estimaciones que se pueden utilizar en un proyecto ágil, y el tipo adecuado dependerá de las necesidades y circunstancias del proyecto y del equipo. Cuando se implementa una en específico esta nos sirve para varios propósitos clave que ayudan en la planificación y gestión del trabajo en equipo.
3. **Planning Poker**: el proceso de planning poker se realiza al principio en el proceso de planificación del sprint, de modo que los Scrum masters y los gerentes de producto puedan tener una idea precisa de cuánto trabajo puede completarse en cada sprint.
4. **T-Shirt Sizes:** Las tallas de camiseta son una medida de la complejidad de una tarea o entregable en términos de tamaño. Se utilizan para establecer una línea base para la estimación y para comparar tareas entre sí. Esta técnica ayuda al equipo a estimar el trabajo sin entrar en detalles numéricos específicos, utilizando tamaños de camiseta como una metáfora para representar diferentes niveles de esfuerzo o complejidad.