

Nombre: Jesús Alejandro Valencia Valadez	Matrícula: 3013480
<ul style="list-style-type: none">• Nombre del curso: Infraestructura para Big Data	
Evidencia 1	

TranStats es una base de datos de transporte intermodal que proporciona información sobre el desempeño de vuelos, aerolíneas y aeropuertos en los Estados Unidos. Para descargar datos de TranStats, sigue estas instrucciones:

1. Acceso a TranStats:

- Visita el sitio web oficial de TranStats.
- Busca la sección que ofrece instrucciones para descargar datos.
- Asegúrate de leer las indicaciones detalladamente para realizar la descarga correctamente.

2. Perfil de la base de datos:

- Investiga la estructura de la base de datos.
- Examina los metadatos, como las tablas, columnas y tipos de datos.
- Comprende cómo están organizados los datos y qué información contienen.

3. Descarga de datos:

- Descarga al menos dos o tres meses de la información más reciente disponible.
- Asegúrate de obtener datos relacionados con el desempeño de vuelos, aerolíneas y aeropuertos en los Estados Unidos.

Importancia del análisis de la base de datos

El análisis de datos es crucial para cualquier organización. Algunos puntos destacados son:

- **Toma de decisiones:** El análisis de datos proporciona información relevante para tomar decisiones informadas.
- **Mejora de procesos:** Ayuda a optimizar operaciones y mejorar la eficiencia.
- **Gestión de clientes:** Permite comprender mejor a los clientes y ofrecerles un mejor servicio.
- **Automatización:** Facilita la automatización de tareas repetitivas.

Consola PowerShell

PowerShell es una solución de automatización de tareas multiplataforma que consta de tres componentes principales:

1. **Shell de línea de comandos:** Acepta y devuelve objetos .NET en lugar de solo texto. Incluye características como historial de comandos, autocompletado y canalización de comandos.
2. **Lenguaje de scripting:** Basado en .NET Common Language Runtime (CLR). Permite crear scripts para automatizar tareas de administración.
3. **Plataforma de automatización:** Habilita el uso de módulos de PowerShell para administrar diversas tecnologías, como Microsoft Azure, Windows, SQL, aplicaciones de terceros y más.

Apache Beeline con Apache Hive

Beeline es un cliente basado en JDBC para HiveServer2. Se utiliza para ejecutar consultas por línea de comandos contra Hive. La finalidad de usar Beeline con Apache Hive incluye:

- **Conexión remota:** Beeline permite conectarse a HiveServer desde cualquier lugar.
- **Independencia de Hive:** Beeline no requiere las dependencias completas de Hive.
- **Ejecución de consultas:** Puede ejecutar consultas y comandos directamente desde Beeline.

HiveQL y operaciones

HiveQL es el lenguaje de consulta utilizado en Hive. Permite crear y consultar tablas, optimizar consultas y más. Algunas operaciones comunes incluyen la creación de tablas, la extracción de datos y la optimización de consultas.

Espero que esta información te sea útil. Si tienes alguna otra pregunta, no dudes en preguntar. 😊