|  |  |  |
| --- | --- | --- |
| **Nombre**: José Guillermo Robles Sánchez | | **Matrícula**: 2905526 |
| **Nombre del curso:**  Infraestructura de big data | **Nombre del profesor**:  Miguel de Jesús Martinez Felipe | |
| **Módulo**:  2 | **Actividad**:  Evidencia 1 | |
| **Fecha**: 2023 | | |
| **Bibliografía**: | | |

TranStats

TranStats es un base de datos que ayuda a proporcionar información sobre el transporte aéreo, marítimo, ferroviario y terrestre de todo Estados Unidos y esta misma es administrada por la oficina de estadística de transporte.

Su base de datos se constituye por una serie de tablas que contienen información sobre diferentes aspectos para el transporte intermodal dentro de ellas están las siguientes:

* Air Carrier Summary Data: esta tabla proporciona información sobre aereolineas estadounidenses sobre sus finanzas flotas y operaciones
* Freight Traffic by Commmodity: esta table proporciona sobre el volume de carga transportada por modo de transporte
* Passenger Travel: esta tabla proporciona información sobre los viajes de pasajeros por modo de transporte

Instrucciones para TranStats:

* Visitar su pagina web
* Dar click al enlace de registrarse en la parte superior de la pantalla
* Registrarse dentro de su sitio web
* Complementar y llenar el formulario
* Una vez registrado y siguiendo todos los pasos
* Seleccionamos la tabla que queramos descargar y le damos click en descargar

Nos dará todos los datos comprimidos por medio de un .zip y tendremos que descomprimirlos con cualquier programa

Cada tabla que descarguemos incluirá los siguientes datos:

* Nombre de la tabla
* Descripción de la tabla
* Fecha de recopilación de datos
* Fecha de finalización
* Frecuencia de actualización de datos

Las bases de datos y su análisis es importante para que se pueda comprender el contenido y significado de cada uno de los datos.

El análisis de la descripción de las tablas da información sobre el contenido de cada tabla y sus columnas.

El análisis de las características de las tablas proporciona información sobre su estructura y diseño para ver como es que se relacionan entre si y como pueden ser utilizados para realizar varios análisis mas avanzados

En TranStats su análisis de base de datos es importante para que se comprenda todo el tipo de transporte en USA, sus tablas dan información detallada sobre todo su tipo de transportes incluidos carga, pasajeros y transportes de cargas

PowerShell es una línea de comandos y lenguaje de scripting que esta basado en tareas y diseñado para administrar sistemas que se basa en .NET common language runtime dentro de sus características son:

* Su automatización proporciona variedad de herramientas y funciones que automatizan tareas administrativas
* Información: proporciona variedad de herramientas para que se recopile información sobre su sistema
* Análisis: ayuda a proporcionar variedad de herramientas para realizar analisi de datos

Apache BeeLine  
sus finalidad de estos programas es proporcionar interfaces de líneas de comandos para interactuar con hive, beeline se basa en el protocolo JDBC permitiendo ejecutar consultas hive desde los comandos

¿Cómo usarlo?

* Instalar apache BeeLine
* Iniciar e cliente de línea de comandos de beeline
* Conectarse a hive
* Ejecutar consultas hive

HiveQL es un lenguaje de consulta SQL que se usa para interactuar con Hive, es un sistema de almacenamiento de datos distribuido que permite almacenar y analizar grandes cantidades de datos estructurados; Es similar a SQL en varios aspectos, se usa para interactuar con datos almacenados en HDFS que es un sistema de archivos

Permite realizar amplia gama de operaciones como:

* Creación de tablas
* Manipular datos
* Analizar datos