



Tecnológico de Monterrey

Escuela de Ingeniería y Ciencias

Campus LATAM

Nombre del trabajo:

Fase 1. Historias de usuarios (SCRUM)

Curso:

Análisis de Grandes Volúmenes de Datos

Alumnos:

Juan Carlos Alvarado Carricarte, A01793486

Bryan Rodolfo Alvarado Cruz, A01793670

Eduardo Gabriel Arévalo Aguilar, A01793897

Profesor:

Alberto De Obeso Orendain

Fecha de entrega:

29 de enero

Resumen

El objetivo de este trabajo es describir las historias de usuario (HUs) que nos sirvan para generar un perfil socioeconómico de una región de Jalisco mediante la recolección y limpieza de datos, la creación de una interfaz de usuario y el desarrollo de funcionalidades de visualización y exportación de información. El equipo estará utilizando el sistema de ficheros distribuidos HDFS en las historias de usuario. En esta primera fase se genera el backlog con HUs que nos puedan servir para el desarrollo del proyecto, además se hace una proyección del tiempo estimado para desarrollar el proyecto.

Palabras clave: Perfil socioeconómico, Jalisco, recolección de datos, limpieza de datos, interfaz de usuario, visualización de información, exportación de información, HDFS, ficheros distribuidos, historia de usuario, HU.

Historias de usuarios (SCRUM)

Backlog de HUs:

1. Como usuario, quiero seleccionar las bases de datos más relevantes del proyecto para poder ser procesados y analizados.
2. Como usuario, deseo generar el diccionario de datos de las distintas fuentes o bases de datos que se usarán en el sistema, para poder identificar los campos y sus características.
3. Como usuario, deseo poder almacenar grandes volúmenes de datos en el sistema HDFS para poder procesarlos de forma eficiente y escalable.
4. Como usuario, deseo poder acceder a los datos almacenados en el sistema HDFS de forma rápida y segura, para poder realizar análisis en tiempo real.
5. Como usuario, deseo poder replicar los datos almacenados en el sistema HDFS en múltiples nodos, para garantizar la disponibilidad y tolerancia a fallos.
6. Como usuario, deseo cargar los datos almacenados en el sistema HDFS, utilizar herramientas Apache HIVE y Apache PIG para realizar transformaciones a los datos, para poder realizar análisis de gran escala.
7. Como usuario, quiero crear modelos con Apache Spark de regresión lineal y árbol de decisiones para poder analizar el crecimiento económico y la proyección de los sectores de Jalisco.
8. Como usuario deseo poder monitorear y administrar el estado y rendimiento del sistema HDFS, para garantizar su correcto funcionamiento y escalabilidad.
9. Como usuario, quiero poder seleccionar una región de Jalisco para poder visualizar su perfil socioeconómico.

10. Como usuario, quiero poder ver la información demográfica de la región seleccionada, como la población, edad, género, entre otros.
11. Como usuario, quiero poder ver la información económica de la región seleccionada, como el PIB, empleo, industrias importantes, entre otros.
12. Como usuario, quiero poder ver la información educativa de la región seleccionada, como el nivel de escolaridad de la población, instituciones educativas, entre otros.
13. Como usuario, quiero poder ver la información de vivienda de la región seleccionada, como el tipo de viviendas, acceso a servicios básicos, entre otros.
14. Como usuario, quiero poder ver la información de servicios de la región seleccionada, como hospitales, servicios de emergencia, entre otros.
15. Como usuario, quiero poder exportar la información visualizada en distintos formatos (PDF, Excel, etc.) para poder tenerla guardada para futuras consultas.
16. Como usuario, quiero poder comparar la información de diferentes regiones de Jalisco para poder tener una visión más amplia del estado.

Usuarios:

- Miembros del equipo 22 de la materia de Big Data encargados de procesar y analizar los datos.
- Miembros del equipo 22 de la materia de Big Data encargados de procesar y analizar los datos.
- Empresas que tengan el interés de invertir en la región.
- Investigadores y Universidades interesados en el análisis socioeconómico de Jalisco.

Tiempo estimado:

- Para implementar el sistema distribuido HDFS, se estima un tiempo total de 2 semanas, como entregables se espera la configuración y pruebas del sistema.
- Recolectar y limpiar los datos: 2 semanas.
- Crear interfaz de usuario: 1 semana.
- Desarrollar funcionalidades de visualización de información: 3 semanas.
- Implementar funcionalidad de exportación: 1 semana.
- Desarrollar funcionalidades de comparación: 2 semanas

Referencias

Atlassian. (s. f.). *User Stories | Examples and Template*.

<https://www.atlassian.com/agile/project-management/user-stories>

Geografía, E. D. N. I. Y. (s. f.-a). *Directorio Estadístico Nacional de Unidades*

Económicas. DENU. <https://www.inegi.org.mx/app/mapa/denue/default.aspx>

Regiones de Jalisco | Gobierno del Estado de Jalisco. (s. f.).

<https://www.jalisco.gob.mx/es/jalisco/regiones>