

Hacemos un proyecto Big Data

Estudiantes María Correas Crespo, Judith Urbina Córdoba

Profesor Marta Martín Moreno

Asignatura Análisis de Datos en entornos Big Data

1. Actividad: Definición Objetivos y KPIs

Descripción

En esta actividad, se propone que desarrolléis un pequeño proyecto de Big data utilizando las herramientas de esta plataforma. El proyecto se dividirá en dos entregas que cubrirán la definición de objetivos y KPIs y la implementación de la propuesta mediante herramientas de AWS. Finalmente, presentaréis vuestra propuesta de proyecto a vuestros compañeros del aula. La temática es libre; podéis elegir una área de interés o aplicar el aprendizaje a un problema real próximo a vosotros.

En esta primera fase del proyecto, tenéis que definir el alcance del proyecto, establecer objetivos claros y medibles, e identificar los KPI (Key Performance Indicators) que usaréis para evaluar el éxito del proyecto.

Objetivos

Se pide:

1. Selección de la temática o caso de estudio
 - ☐ área de interés o un caso de estudio que queráis abordar con el proyecto de Big data
 - ☐ temática tiene que ser relevante y aplicable a un contexto real.
2. Descripción del problema
 - ☐ Aseguraos de describir por qué este problema es importante
 - ☐ qué impacto puede tener su resolución
3. Establecimiento de objetivos. Los objetivos tienen que ser realistas y alineados con la resolución del problema identificado.
 - ☐ Definís objetivos específicos y medibles que queréis lograr con el proyecto.
 - ☐ Explicad cómo el análisis de datos contribuirá a conseguir estos objetivos.
4. Definición de KPI.
 - ☐ Definís tres KPI (Key Performance Indicators) que sean relevantes para medir el éxito del proyecto
 - ☐ Explicad por qué estos KPI son importantes
 - ☐ cómo se relacionan con los objetivos establecidos, y de qué manera se medirán.
5. Identificación de fuentes de datos.
 - ☐ Identifica y describe las fuentes de datos que has utilizado para abordar el problema.
 - ☐ Indicáis la naturaleza de los datos,
 - ☐ su procedencia
 - ☐ cómo se prevé que sean utilizadas en el análisis.
6. Asignación de tareas dentro del equipo.
 - ☐ Repartís las responsabilidades y tareas entre los miembros del equipo.
 - ☐ Aseguraos que cada miembro tiene un papel claro y definido,
 - ☐ tareas asignadas son coherentes con sus habilidades y roles dentro del proyecto.

Propuestas de datasets

Datos de buena calidad para minimizar la corrección. Datos que tengan información relevante, precisa y actualizada. Algunos enlaces que podrían ser útiles

1. Ofrecido por Amazon y con ejemplos Open data on AWS
2. estados unidos catálogo de datagov
3. europa data europa
4. Plataformas de código abierto creando cuenta: pewresearch
5. Contiene un listado mayor de datasets Open Access Directory
6. Otro listado mayor de datasets Universidad de Missouri
7. Si se sabe buscar datasets amplios Dat search
8. Muy bien documentada la manera de citar y otros detalles nature
9. Contiene lista de centros de datasets y tips para elegir sqream

No se puede abrir al entrar dentro del dataset o hay pegas

1. Migraciones de animales! movebank
2. varios re3data

2. Actividad: Implementación AWS

Descripción

En el proyecto Big Data tenemos que tener

1. Pruebas de proceso de datos.
 - ☐ Realizáis un conjunto de pruebas para asegurarnos que los datos se están procesando correctamente.
 - ☐ Describís los pasos seguidos para verificar la precisión y la consistencia del proceso de datos.
2. Validación de los KPIs.
 - ☐ Validáis que los KPIs definidos están correctamente medidos y que reflejan fielmente la realidad del análisis.
 - ☐ Explicáis los métodos utilizados para asegurar la precisión y la relevancia de los KPIs.
3. Incorporación de scripts o código relevante.
 - ☐ Incluis los scripts o código más relevantes que han sido utilizados durante la implementación del proyecto.
 - ☐ Aseguraos que están suficientemente descritos para facilitar la comprensión de las operaciones realizadas.

Objetivos

Por servicio de AWS usado

1. Por qué hemos usado este servicio? Cómo se relaciona con nuestros objetivos de proyecto?
2. Cómo lo describimos?
3. Cómo lo hemos usado? Con qué tipo de datos hemos trabajado en nuestro proyecto?

En detalle describimos

1. con un diagrama de Amazon diagrams
2. cómo realizaremos el procesamiento de los datos a AWS?
 - ☐ explicación de los flujos de trabajo
 - ☐ explicación de las transformaciones de datos
 - ☐ gestión de la integración entre los diferentes servicios de AWS

Usamos

Recomendaciones

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula. Maximizar el uso de los servicios PaaS:

1. servicios S3, Lambda, Kinesis, EMR, EC2, Glue, Athena o Redshift

y minimizar el uso de servicios IaaS