



HITO GRUPAL PROGRAMACION

Daniel Isaac Flores , Sabin Maniuc, Borja Pérez, Roberto Anero



19 DE MARZO DE 2023

CAMPUSFP
Humanes de Madrid

INDICE

FASE-1 (INVESTIGACION DE CONCEPTOS)

FASE-2(IMPLEMENTACION DE LA INVESTIGACION)

FASE-3(EVALUACION Y CONSIDERACIONES)

FASE-1

Esta fase es una explicación “teórica” de cuatro puntos fundamentales:

Para poder investigar los conceptos relacionados con paradigmas de programación y analítica se tendrá que primero averiguar los requisitos que se deberán cumplir para la investigación y desarrollo de los siguientes puntos:

1-Fuentes de datos: Cuando queremos hablar de Big Data la forma de encontrar información puede ser Estructurada, Semiestructurada y No estructurada en la cual se compone de diversas bases de datos como, por ejemplo;

Bases de datos NOSQL, Las propias redes sociales, Transacciones en línea etc.

Estos tipos de datos se pueden dividir en estructurados y No estructurados

En los cuales los datos estructurados son aquellos que se almacenan en un formato ya predefinido y de manera organizada mientras que en los datos no estructurados son aquellos que no se guardan en un formato predefinido ya que representa un gran problema para las empresas ya que no pueden gestionar volúmenes de datos.

2-Hadoop y Spark: Estos dos tipos de tecnología de código abierto muy utilizadas en el procesamiento de Big Data ya que estas dos tecnologías se complementan entre si para llevar una gestión y procesamiento de datos de manera más distribuida.

Hadoop es un marco de trabajo que permite el almacenamiento y el procesamiento de grandes cantidades de volúmenes de datos que se encuentran en clústeres de servidores.

Spark es un tipo de herramienta que nos ayuda a procesar datos de manera mas escalable en el cual proporciona un mejor rendimiento a hadoop también podría ayudar a otras aplicaciones para uso de aprendizaje y análisis en tiempo real.

3-Lenguajes de programación para gestión de datos: Las aplicaciones que mas destacan en este campo de gestión son Python Y Scala ya que sus usos se han convertido en los programas mas usados ya que pueden permitir al usuario poder gestionar Big Data a través de ellos.

Python es un lenguaje de programación de fácil aprendizaje ya que cuenta con una amplia gamma de bibliotecas en la cual se podría utilizar para seguir guías de utilización de este lenguaje.

Scala es un lenguaje programación de código abierto en el cual se ejecuta en la plataforma de Java Virtual Machine en el cual su diseño se basa en poder abordar problemas concurrentes y manejar estabilidad en los sistemas ya distribuidos.

4-Herramientas PowerBI Y Tableau:

Estas dos herramientas son principalmente para la visualización de datos en los cuales permite a los usuarios poder analizar y visualizar grandes volúmenes de datos de manera intuitiva y eficiente.

PowerBI se podría describir como una herramienta solucionadora en el tema de visualización de datos ya que este crea paneles de visualización en tiempo real.

Tableau Esta herramienta es muy parecida a la anterior solo que esta hace lo mismo, pero permite al usuario poder crear gráficos y paneles interactivos en los cuales cualquier usuario puede acceder en línea.

FASE-2

En esta segunda fase se realiza la implementación de la investigación. En concreto sería acceder a un volumen de datos y mostrarlo. Podríamos utilizar Scala o Python y mostrar el resultado en PowerBI o Tableau. La idea es que sea algo muy impactante por la calidad de contenido tratado, velocidad de acceso, volumen de datos....

Código que permita el acceso a fuente de datos utilizando Python.

```
main.py 1 x
main.py > ...
1 import pandas
2
3 resultado = pandas.read_csv('catalogo_cf.csv')
4
5 print(resultado)
6 # con este código se muestra el acceso a datos.
7
```

```
2          DMZ10S1100Z.Q          2214937          CF_2_10A.3 ... Cuentas Financieras/Metodologia SEC2010/Saldo/... Banco de España NaN
3          DMZ10S1200Z.Q          2215030          CF_2_10A.4 ... Cuentas Financieras/Metodologia SEC2010/Saldo/... Banco de España NaN
4          DMZ10S4000Z.Q          2216082          CF_2_10A.5 ... Cuentas Financieras/Metodologia SEC2010/Saldo/... Banco de España NaN
...          ...          ...          ...          ...          ...
4642         DMZ10S8000H.Q          2217814          CF_2_99.5.99 ... Cuentas Financieras/Metodologia SEC2010/Saldo/... Banco de España NaN
4643 DSPC102016CP70000 ES1_RS2.T          3777791          CF_2_99.6.2 ... Descripción Familia: Contabilidad nacional de... Instituto Nacional de Estadística NaN
4644 DSPC102016CP60000 ES2_RS1.T          3777777          CF_2_99.6.4 ... Descripción Familia: Contabilidad nacional de... Instituto Nacional de Estadística NaN
4645 DSPC102016CB11000 SS2.T          3777241          CF_2_99.6.7 ... Descripción Familia: Contabilidad nacional de... Instituto Nacional de Estadística NaN
4646 DSPC102016CD10000 ES2_RS1.T          3777344          CF_2_99.6.9 ... Descripción Familia: Contabilidad nacional de... Instituto Nacional de Estadística NaN

[4647 rows x 17 columns]
PS C:\Users\Sabin\Desktop\proyecto1>
```

En Python utilizando este sencillo código accedes a los archivos csv y los muestra por pantalla.

Hemos decidido escoger Python ya que tenemos algo de conocimientos previos del lenguaje junto con la librería pandas.

Pandas es bastante popular para el análisis de datos proporcionando una gran cantidad de herramientas y funciones en la cual manipula y analiza los datos estructurados CSV hojas de cálculo o bases de datos

Además, no es complicado de utilizar y se aprende bastante rápido por lo cual es una opción muy buena si necesitas analizar grandes cantidades de datos.

FASE-3

Para finalizar, realizamos una evaluación o consideraciones de cómo han evolucionado el acceso a datos en los últimos años. Desde acceso a ficheros, pasando por base de datos y consumiendo APIs.

EL acceso a datos ha experimentado una notable evolución en los últimos años por el crecimiento exponencial de la cantidad de datos disponibles y la necesidad de procesarlos de manera eficiente.

Mayor volumen de diversidad de datos, la diversidad de fuentes de datos ha crecido incluyendo datos estructurados y no estructurados datos en tiempo real y datos en streaming, datos de redes sociales etc. Esto ha requerido el

desarrollo de nuevas técnicas y herramientas para acceder, procesar y analizar datos de manera eficiente y efectiva.

La necesidad de acceder a grandes volúmenes de datos de manera rápida ha impulsado el desarrollo de tecnologías de procesamiento de datos a gran escala, como Apache Hadoop, Apache spark y otras herramientas de procesamiento paralelo y distribuido de datos en clústeres de pcs.

Con el aumento de la conciencia sobre la seguridad y privacidad de datos, ha habido un enfoque creciente en garantizar la protección de los datos en el acceso a los mismos, han desarrollado técnicas y herramientas para cifrar datos y cumplir con el GDPR.

El consumo de las APIs se ha vuelto una práctica común en la integración de datos y aplicaciones permitiendo acceder y consumir datos de manera estandarizada y programática a través de interfaces web lo que ha simplificado la integración de datos de diferentes fuentes y ha facilitado la creación de aplicaciones y servicios.

La automatización y el uso de técnicas de Machine Learnig (es una disciplina del campo de la inteligencia artificial que a través de algoritmos , dota a los ordenadores de la capacidad de identificar patrones de datos masivos y elaborar predicciones) en el acceso a datos a ganado terreno en los últimos años por lo tanto se han desarrollado algoritmos y técnicas de Machine Learning para el descubrimiento auto matico de esquemas de dato y la optimización de consultas y proceso de acceso a datos.

DOCUMENTACION APORTADA

Ahora que tenemos concluido nuestra investigación e ideas de proyecto tendremos que gestionarlo y juntarlo con los demás proyectos en mente y para eso tendremos una absoluta gestión sobre estos tipos de ideas y opiniones compartidas subidas a la plataforma de GitHub en el cual se nos hará más fácil poder sacar una conclusión final, todos los miembros tendrán acceso a este proyecto y lo podrán modificar a su gusto teniendo en cuenta la opinión del grupo para poder cambiarlo.