

TRABAJO OBLIGATORIO: JULIO 2019

INTRODUCCIÓN Y REGLAS GENERALES

El proyecto obligatorio tiene como objetivo que el estudiante integre los conocimientos tanto teóricos como prácticos adquiridos en la materia “Herramientas de Big Data”.

PROPUESTA

PARTE 1: MODELO

Se debe seleccionar un conjunto de datos que cumpla con un volumen de más de un millón de registros, en el que se desee realizar consultas en menos de cinco segundos y con una alta variabilidad en las columnas, con este conjunto de datos se debe modelar una arquitectura de Big Data, que contemple una capa de obtención de datos (Collection tier), diferenciando se obtendrán los datos en batch, microbatch o en tiempo real; con una capa de manejo de datos (Message queuing tier), en la que se realizan procesos de extracción, ingesta, transformación y carga de los datos. Después de que se obtienen los datos, será necesario procesar los mismos a través de una capa de análisis (Analysis tier), esta capa deberá contemplar el almacenamiento para la generación de sumalizaciones y agregaciones sobre los datos, además de su indexación y la generación de estructuras de datos persistentes (Long-term storage tier) que permitan la ejecución de algoritmos para la identificación de patrones. Finalmente, es necesario contemplar una capa de acceso a los datos (Data access) y los reportes multidimensionales. Usted deberá seleccionar un aspecto específico que quiere analizar de su conjunto de datos, por ejemplo: sentimientos, campaña política, rastreo de comentarios, rankings, marcas, productos, tiendas, ubicación geográfica, rutas, utilización de palabras específicas, seguimiento de cuentas específicas, etc. A continuación, realizará las siguientes actividades:

- 1) Describa el proyecto en términos de los siguientes aspectos:
 - a) Planteamiento del Problema
 - b) Objetivo General
 - c) Objetivos Específicos
 - d) Arquitectura de la solución (describiendo cada una de las capas y su funcionamiento)
- 2) Defina las aplicaciones que utilizaría en cada una de las capas:
 - a) Descripción de cada aplicación y su utilización
 - b) Indique el por qué selecciono esta aplicación y no otra

PARTE 2: IMPLEMENTACIÓN

Se ha dispuesto de una máquina virtual con todos los productos previamente instalados y configurados, explique cada una de las aplicaciones y la configuración necesaria en cada una de ellas a fin de lograr implementar el modelo anterior. Deberá realizar un informe con ambas partes, de acuerdo con las normas establecidas en la Facultad de Ingeniería y publicarlo en el sistema automatizado de obligatorios antes de la fecha límite de entrega.