Práctica Módulo Cloud Big Data

Una bodega ha comenzado un proyecto de despliegue de sensores por todos sus cultivos para monitorizar la humedad, temperatura, presión atmosférica y acidez de la tierra. En paralelo quieren comenzar a desplegar una solución basada en un Data Lake para centralizar todos los datos de la bodega para la explotación de la información de esta. Para ello en el Data Lake quieren centralizar la información de las siguientes fuentes:

* ERP (Enterprise Resource Planning) - Base de datos relacional
  + Pedidos
  + Facturas
  + Gastos
  + Clientes
* CRM (Customer Relationship Management) - Base de datos relacional
  + Clientes
  + Interacción con los clientes (llamadas y correos)
* Marketing - Plataforma online de Marketing
  + Campañas de marketing
  + Redes sociales (Seguidores)
* Sensores bodega
  + Temperaturas
  + Humedad
  + Presión atmosférica
  + Acidez tierra
* AEMET  API
  + Temperaturas
  + Lluvias

Con todos estos datos obtenidos y centralizados en un Data Lake se deberá diseñar un Data Warehouse para la explotación de las siguientes dimensiones:

* Pedidos
* Clientes
* Gastos
* Facturas
* Campañas
* Redes Sociales

Además se quiere aplicar un modelo matemático para la realización de un pronóstico de ventas basándose en un dataset que indique los siguientes campos:

* Cosecha
* Temperatura media/cosecha
* Acidez media/cosecha
* Presión media/cosecha
* Humedad media/cosecha
* Ventas

Para la explotación de los datos del DataSet se deberá implementar una API que podrá ser explotada de manera segura por una aplicación de desarrollo propio que está encargada a una empresa externa. Por otro lado será necesario el despliegue de una herramienta que permita la explotación de datos de manera dinámica y permita el diseño de dashboards personalizados que podrán ser consultados por usuarios clave.

* Diseñar una arquitectura que represente una solución del Data Lake basada en los servicios de AWS. En dicha arquitectura definir cada servicio empleado para qué va ser empleado y sus relaciones con el resto de los servicios.
* Además del diseño de la arquitectura será necesaria una explicación detallada de todas sus funciones, además de identificar qué servicios se han empleado para ingestión de datos, procesamiento, almacenamiento, análisis y visualización.

Nota: No se pretende que se realice el diseño de los datasets. Pero hay que tener en cuenta muchos de ellos necesitarán algún proceso de transformación antes de ser analizado.