

PROYECTO BIG DATA

1.- BDA	1
2.- SBD	4

1.- BDA

Enunciado:

El CIPFP de Mislata ha contactado con nosotros para elaborar un cuadro de mando integral que recoja la información histórica de una serie de indicadores que ellos consideran interesantes. Están abiertos a todo tipo de sugerencias y propuestas de innovación.

El proyecto se desarrollará en varias fases:

1. Crear un cuadro de mando con un documento que recopila los datos históricos de los indicadores de calidad introducidos a mano. Este cuadro tendrá las gráficas de evolución de los distintos indicadores.
2. Algunos de estos indicadores se obtienen de datos extraídos de la aplicación de calificaciones de la Conselleria. Se quiere añadir más páginas al cuadro de mando con gráficas de evolución de las calificaciones por evaluación y ciclo formativo.
3. Investigar el origen del resto de indicadores y automatizar el proceso de obtención de los datos

De momento solo se pretende realizar las fases 1 y 2:

- Fase 1: Cuadro de mando del histórico de indicadores
- Fase 2: Cuadro de mando con la información del documento adjunto, pero con total libertad para mejorar el formato del informe y adaptarlo a Power BI

Evidentemente, existe un trabajo previo de estudio y análisis de la información recibida, los orígenes de datos.

Orígenes de datos:

Los orígenes de datos de los que disponemos:

- Indicadores de Calidad
 - CSV con los datos de los Indicadores
 - CSV con las líneas del PAA
 - CSV con los objetivos del PAA
 - CSV con los procesos de Calidad
- Calificaciones
 - CSV con los datos de los Alumnos
 - CSV con los datos de las Calificaciones
 - CSV con los datos de los Cursos
 - CSV con los datos de los Módulos
 - CSV con los datos de los Grupos
 - CSV con los datos de las Horas de cada Modulo
- Ejemplo de informe

KPI's:

- Alumnos con todos los módulos aprobados (Todos los que estén matriculados)
- Alumnos Evaluables (Aquellos que tengan nota > 0)
- % Alumnos con todos los módulos aprobados (Respecto de los evaluables)
- % Alumnos que promocionan a 2ª, aquellos con menos de 240 h suspendidas
- Umbral de % Alumnos con todo aprobado por curso = 30%
-

Trabajo:

- Fase 1 (Indicadores de Calidad):
 1. Realiza un análisis de los datos
 2. Realiza una ETL para preparar el modelo de datos.
 - Incluye: Validaciones de ficheros, Jobs y Transformaciones
 - El modelo de datos tiene que estar en BD
 3. Realiza el Cuadro de Mando en Power BI.
- Fase 2 (Calificaciones):
 1. Realizar un análisis de los datos
 2. Identificar relaciones entre los datos
 3. Entregar un borrador del diseño de la ETL y del Cuadro de Mando para su validación por el profesor
 4. Realiza una ETL para preparar el modelo de datos
 - ☐ Incluye: Validaciones de ficheros, Jobs y Transformaciones
 - ☐ El modelo de datos tiene que estar en BD
 5. Crear un Cuadro de Mando con Power BI
 - ☐ Gráficos
 - ☐ Filtros
 - ☐ Comparativas
 - ☐

Entrega:

- Documento análisis y diseño propuesto (Tiene que validarlo el profesorado antes de continuar con el proyecto)
- Ficheros Pentaho de las ETL
- Fichero Power BI con los Cuadros de Mando

2.- SBD

Duplicar el proyecto con servicios de AWS

Trabajo:

- Elegir una opción de almacenamiento
- Almacenar los datos
- Con Spark:
 - Realizar la limpieza de datos
 - Preparación de los datos con posibles nuevos campos
- Utilizar algún Job o Workflow para realizar todo el trabajo
- Realizar las consultas del informe modelo con algún servicio de AWS
- Crear un informe con QuickSight (Opcional)

Servicios AWS:

- S3
- Glue
- Lambda
- Athena
- QuickSight
-