Brainstorm

* Fórmula 1
* Posiciones ganadas en carrera: por piloto, por nacionalidad.
* Tweets (sentimiento): piloto bueno/malo
* Gráficas de los resultados (kibana – tableau?)

Diseño del DAaaS

### Definición la estrategia del DAaaS

*Definir el catálogo de servicios que proporcionará la plataforma DAaaS, que incluye incorporación de datos, limpieza de datos, transformación de datos, datapedias, bibliotecas de herramientas analíticas y otros.*

* ¿Qué piloto ha ganado más posiciones durante a carrera? ¿Importa la nacionalidad del piloto con el número de adelantamientos?
* ¿A los aficionados les gusta más los pilotos que adelantan más?
* Dashboard para posibles patrocinadores de pilotos/equipos en referencia al estudio de los adelantamientos y la repercusión que puedan tener los mismos.

### Arquitectura DAaaS

*Definir la selección de componentes, la definición de procesos de ingeniería y el diseño de interfaces de usuario. Diseño y ejecución de Proofs-of-Concept (PoC) para demostrar la viabilidad del enfoque DAaaS.*

* Dataset Formula 1 World Championship (1950 – 2021) <https://www.kaggle.com/rohanrao/formula-1-world-championship-1950-2020>

No utilizo un API para extraer datos, porque este dataset ya utiliza el API de donde quería extraer los datos: <http://ergast.com/mrd/>

* Análisis de 3 horas de los tweets para ver que tendencia hay dentro de los aficionados en referencia a los pilotos.
* Google Storage para almacenar los datos de los Dataset.
* Hadoop con Hive para hacer las queries de los Dataset que tenemos.
* Google Cloud Platform, para crear la clúster de Hadoop con hive del Dataset (es una base de datos relacional).
* Máquina virtual que transforme los resultados y los alce con dos tablas en un PosgreSQL en Google Cloud.
* Tableu para la representación de los datos.

### DAaaS Operating Model Design and Rollout

*Personalizar los modelos operativos DAaaS para cumplir con los procesos, la estructura organizacional, las reglas y el gobierno de los clientes individuales. Realizar seguimiento de consumo y mecanismos de informe.*

* Un operador descarga del Dataset y lo sube a Google Storage, para introducir los datos en Hive en Google Clud.
* El mismo operador Descarga/actualización del Dataset después de cada carrera de F1 y actualizar Hive cuando se suba el Dataset actualizado.
* El mismo operador pasados 4 días después de la carrera, realiza un análisis de 3 horas sobre los tweets que ha habido en referencia a la misma. Se esperan 4 días debido a que siempre hay algunos días de repercusión sobre las mismas.
* Se suben los datos a Google Storage, de ambos dataset y después se añaden/actualizan al clúster de Hadoop.
* Se lanza un Job de Hive, para que con las queries requerida, nos arroje dos tablas (csv). Una con los resultados de la query de los pilotos, posiciones ganadas y nacionalidad, y otra con los resultados de los tweets.
* Con una máquina virtual, cargamos los datasets en un PosgreSQL en Google Cloud.
* Usamos el conector de Tableau para importar los datos desde PosgreSQL.
* Una historia con dos Dashboards en Tableau lee la información del índice y despliega:
  + Un gráfico con los pilotos con más adelantamientos.
  + Un gráfico con adelantamientos y nacionalidad.
  + Una nube de palabras con los pilotos que se está hablando más, y poner en verde los que tengan mejor valoración y en rojo los que peor valoración tengan.
* Se envía al cliente un acceso a Tableau Viewer para que pueda visualizar los resultados de las historias.

### Desarrollo de la plataforma DAaaS.

*Construcción iterativa de todas las capacidades de la plataforma, incluido el diseño, desarrollo e integración,* ***pruebas****, carga de datos, metadatos y población de catálogos, y despliegue.*

Link a Diagrama:

<https://docs.google.com/drawings/d/12BEkq-qbXj9HKcyYk6j0hRfnXQ89J1FY17apjjOknJA/edit?usp=sharing>

Diagrama

Descripción generada automáticamente