Brainstorm

* Temas: Fish farm, videojuegos, libros de Sci-Fi
* La página de Kaggle tiene datasets de fish farming que podría usar:

<https://www.kaggle.com/datasets/ogbuokiriblessing/sensor-based-aquaponics-fish-pond-datasets> o con imágenes e información recopilada por IoTs en una granja en producción.

* Predictor del peso del pez.
* Esta página tiene datos de videojuegos: <https://www.3djuegos.com/juegos/>.
  + Predictor de sold-out de juegos o un sistema de recomendaciones
  + Puedo utilizar técnicas de scraping/crawling para extraer datos no estructurados.

Diseño del DAaaS (Data Analytics as a Service)

### Definición la estrategia del DAaaS

* Crear un servicio de dashboard que actualice semanalmente los datos de la granja y que entregue predicciones del peso promedio de los peces para los próximos meses.
* El servicio Dashboard proveerá visualizaciones y reportes que permitan al usuario realizar análisis de los datos semanales, establecer correlaciones entre variables, tendencias y obtener pronósticos de crecimiento a futuro.

### Arquitectura DAaaS

* Dashboard con Elastic-Kibana (o Plotly) en Google Compute Engine para las visualizaciones y reportes semanales.
* VM con Mongo DB para manejar la base de datos que contiene:
  + mediciones de sensores ambientales.
  + metadata (uri.address) de imágenes capturadas por las cámaras.
* Google Storage para almacenamiento de imágenes. (??)
* Cloud Function que realiza el procesamiento de imágenes y cálculo de predicciones.
* CSV en Google Storage con los resultados de analitica de datos.
* Hadoop/Cluster/Job con Dataproc que genere el CSV y lo guarde en un Bucket de Google Storage.
* Cloud Function/Scheduler para descarga automatica del dataset (metadata imágenes/sensores) al Google Storage.

### DAaaS Operating Model Design and Rollout (EN PROCESO)

* Al final de cada día a las 7:00 pm se ejecuta el Cloud Function (***Load\_Info***) para la descarga del dataset. Para ello se establece la conexión a los servidores externos IoTs y se importan las imágenes y mediciones recopiladas durante el día a un repositorio creado en Google Storage.
* Al terminar la función ***Load\_Info*** se ejecuta el job en el cluster/Hadoop para generar el CSV (base) y guardarlo en el Bucket asociado.
* Una vez terminado el job, se ejecuta el Cloud Function (***Predictions****)* de Python con librería *statsmodels o Prophet* para generar nuevos cálculos y predicciones basado en un modelo de pronósticos con series de tiempo.
* Al terminar la función ***Predictions***, con la opción de trigger se guardan los resultados de la analítica de datos.
* Google Cloud Platform (Compute Engine):
  + VM con Elastic-Kibana o Plotly para las visualizaciones y reportes semanales.
  + VM con Python y librería (statsmodels o Prophet) para hacer predicciones.
  + Cloud SQL: Instancia de Postgres para manejar la base de datos.
  + Proceso con código Python para procesar la imagen, calcular el tamaño y peso promedio del pez. Esto puede ser un Cloud Function ??
  + Cloud Function que cargue las predicciones en la BD.
  + CSV en Google Storage con resultado de analitica de datos.
  + .
* Dashboard de Kibana o Plotly para las visualizaciones y reportes semanales. El usuario final podrá analizar los datos y ver las proyecciones para los próximos meses

(EN PROCESO)

### 

### Desarrollo de la plataforma DAaaS. (ligera descripción del desarrollo)

*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.*

* Código para las funciones ***Load\_Farm\_Info*** y ***Transf\_Forecast***
* Ejemplo: crawling creado para extraer datos de la página web

Link a Diagrama:

* Google drive:

