**DOCUMENTACIÓN SPRINT 1**

**Almacenamiento de Datos en Amazon S3**

**Decisión sobre la Estrategia de Almacenamiento**

Como parte de la planificación del proyecto, se ha decidido utilizar un único bucket de almacenamiento en Amazon S3 para centralizar todos los datos. Esta elección se basa en la optimización del rendimiento y la eficiencia en el acceso a la información. Unificar el almacenamiento en un solo bucket permite reducir la latencia y mejorar los tiempos de consulta, evitando la sobrecarga de gestionar múltiples buckets. Además, desde una perspectiva de costos y rendimiento, acceder a un único clúster con los datos organizados es más eficiente que distribuir la información en múltiples clústeres separados.

**Estructura de Almacenamiento en S3**

Siguiendo los principios de organización modular y escalabilidad, se ha creado una carpeta denominada "data" dentro del bucket principal. En esta carpeta, los datos se almacenan de acuerdo con las siguientes convenciones:

* Clasificación por criptomoneda: Cada criptomoneda cuenta con su propio subdirectorio.
* Histórico de datos: Se almacenan los registros correspondientes a los últimos años para garantizar un acceso rápido a la información más relevante.

Esta organización permite mejorar el rendimiento y la facilidad de acceso, alineándose con las necesidades del equipo de desarrollo y análisis.

**Formato de Almacenamiento de Datos**

Los archivos se almacenarán en formato CSV, ya que este formato facilita su manipulación y análisis mediante Python. La elección de CSV responde a su compatibilidad con herramientas de análisis de datos y su facilidad de uso dentro del entorno de desarrollo del equipo.

**Relación con el Framework Scrum**

Durante el desarrollo de este proyecto, se han aplicado los principios de Scrum, organizando el trabajo en sprints y entregables incrementales. La definición de la estrategia de almacenamiento formó parte de una de las primeras iteraciones, en la que se priorizó la estructuración de los datos y la optimización del acceso.

A lo largo de las siguientes iteraciones, se continuará refinando el almacenamiento y optimizando la estructura de los datos en función de las necesidades del equipo de desarrollo y análisis.