

Assignment 1 - Big Data Cloud

Project Title: **Exploring the Non-Fungible Token Revolution: An Analysis of NFT Transactions from 2021.**

Group A:

Krystian Fikert, student ID 22268291

Marcio Vieira, student ID 22266459

Vinit Saini, student ID 21267544

In this project, we analyse a dataset of NFTs (Non-Fungible Tokens) transferred and sold from 2021 (Apr-Sept, 8GB). NFT stands for Non-Fungible Token. It is a form of digital asset that is unique, divisible (fractional NFTs), and stored on a blockchain. These tokens are used to represent ownership of a digital item, such as a digital art piece, a virtual collectable, or a digital game item. They are becoming increasingly popular for creators and collectors to own and trade unique digital assets. They have a wide range of use cases, including:

1. Digital Art & Collectibles: NFTs can represent digital assets such as art, music, videos, and other virtual items.
2. Digital Gaming: NFTs can represent tangible in-game assets such as characters, weapons, and skins.
3. Digital Ownership: NFTs can also represent ownership of physical assets such as web3 domains, real estate, cars, and other assets.
4. Digital Identity: NFTs can represent digital identities, allowing individuals to control and secure their data.
5. Digital Currency: NFTs can also be used to represent digital currencies, allowing for the tokenisation of cryptocurrencies.

For the project, we will focus on the data related to the following categories: Digital Art, Collectibles, Digital Gaming, and Digital Ownership (with a focus on Web3 domains).

NFTs can exist on several networks and platforms, including Ethereum, Solana, Polygon, Algorand, Cardano, Binance, and more. Each network has unique features and advantages, so it is essential to understand their differences when deciding where to store and trade NFTs. Ethereum is the most widely used platform for NFTs due to its support for smart contracts, scalability, and low fees.

The NFT data is not readily available without charge for research purposes. Due to DCU library constraints, we had to base our data analysis on data sourced free of charge from <https://www.kaggle.com/datasets/simiotic/ethereum-nfts>, which is a repository of NFT transactions on the Ethereum blockchain.

To support the consistency of the data, we have used an API connection through Moralis to gather additional data linked with the original data content (<https://moralis.io/>).

For NFT data analysis, we employ a collection of Google tools such as Google Dataproc, HDFS, Hive, and Spark. Google Dataproc is a managed service that executes Apache Hadoop and Apache Spark jobs. HDFS is a distributed file system that stores and manages large datasets, and Hive is a data warehousing solution that allows SQL-like access to big data. Spark is an agile in-memory data processing engine that can manage the processing and analysis of large-scale data.

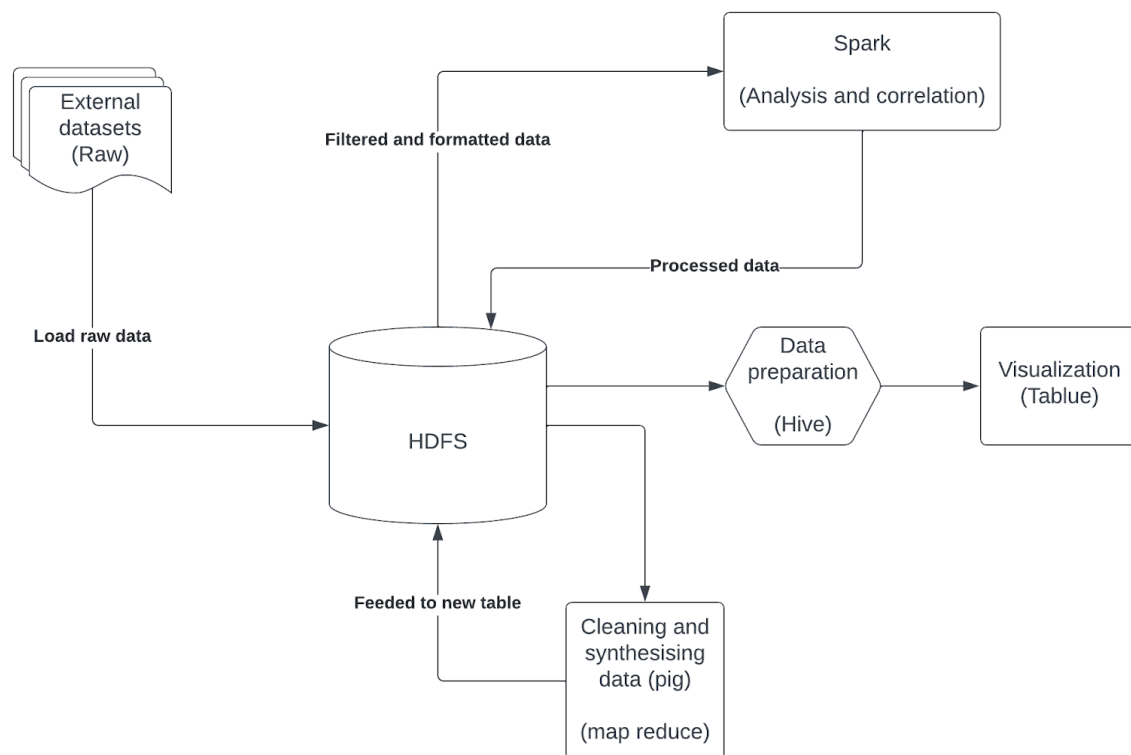


Diagram 1. Usage of the tools.

Our team of three members has various backgrounds, from a software engineer, data engineer, to data architect and analyst. We have divided initial work into several tasks allocated on a project basis.

- Finding relevant NFT data that could be downloaded and processed
- Cleaning the data and adjustments to project needs
- Conducting exploratory data analysis
- Gathering the complementary data from additional services (Moralis API)
- Setting up and configuring the data processing infrastructure
- Creating visual representations of the data allows stakeholders to understand the analysis results and trends over time easily.

To get the most out of our analysis, our focus was on the following data points:

- The volume of NFT transactions over time
- Average transaction value
- Most popular NFT collections (e.g. digital art, collectables)
- Most active NFT buyers and sellers
- The correlation between the popularity of a particular NFT category/collection and the average transaction value.

By focusing on these data points, we can gain valuable insights into the growth and evolution of the NFT market and identify trends and patterns in NFT buying and selling behaviour. Additionally, we can use this information to inform investment decisions and help predict future trends in the NFT market.

By working together and leveraging the strengths of each team member, we can ensure that this analysis project is completed efficiently and effectively. The analysis results will provide valuable insights into the growth and evolution of the NFT market, as well as inform investment decisions and help predict future trends in the NFT market.