

## Contents

1. Project Objectives
2. Project Phases
3. Phase I
4. Objectives
5. Methodology
6. Infrastructure
7. Indexing and Deliverables
8. Costs
9. Objectives
10. Methodology
11. Infrastructure
12. Indexing and Deliverables
13. Costs
14. Phase II
15. Objectives
16. Deliverables
17. Deliverable ExamplesHello community!
18. Objectives
19. Deliverables
20. Deliverable ExamplesHello community!

## Project Objectives

Helika, a gaming analytics and infrastructure company, has partnered with Plurality Labs to address the following questions:

- Who are the most valuable users and devs on Arbitrum?
- What is the Arbitrum user's path to purchase from web2 to web3?
- Are there alternative paths to drive engagement and retention?
- Where is the biggest drop-off in the FTUE funnel? (first time user experience)
- What other parts of the web3 ecosystem gamers are currently using or are likely to engage in (DeFi, etc.)
- What Arbitrum user archetypes are the most engaged? Least? Why? What segments drive the most monetization and TVL?
- What are the most common reasons for churn or for users to become disengaged?
- Where should Arbitrum invest to acquire the best users?
- Where should Arbitrum invest to incentive the most engaging and sustainable content?

In order to start the analysis of the following areas, it was necessary for Helika to index underlying Arbitrum data (Phase I) that will be used then for the analysis to address the questions above (Phase II).

## Project Phases

- Phase I: Data indexing (grant approved)
- Phase II: Analysis, insights and recommendations (to be applied and approved in the future)

## Phase I

### Phase I Objectives

Objective Summary: Indexing Arbitrum One and Nova so that relevant and timely insights can be generated to answer the questions in Phase II. All data must be cleansed, indexed, categorized and structured in easily accessible tables so it is ready for Phase II analysis.

### Phase I Methodology

Multiple steps must be taken in order for all of Arbitrum One and Nova data to be refined into a state that can be used by a sophisticated analytics team. To do this, Helika has cleansed, indexed, categorized and structured all data from Arbitrum One and Nova and it is ready for Phase II analysis. Custom KPIs has been developed based on the underlying data and example data visualizations have been provided at the end of this paper.

### Phase I Infrastructure

Helika setup the following infrastructure critical to indexing Arbitrum One and Nova and setting us up to deliver the insights as part of Phase II.

#### ETL:

The Extract, Transform, Load (ETL) process is commonly used in DS to gather data from various sources, transform it into a format suitable for analysis and load it into a database or data warehouse for further use. This process is crucial for preparing and consolidating data, making it accessible and usable for analytics and decision-making. To setup the ETL for this project Helika used Airflow and Kafka streaming.

#### Data Lake:

Helika created a centralized repository that allows for the storage of structured, semi-structured, and unstructured data at scale. In the context of data science, it provides a flexible environment to store large volumes of diverse data, which can then be used for various types of analytics including big data processing, machine learning, and real-time analytics. To setup the data lake for this project Helika used AWS S3.

#### Data Warehouse:

Helika created a data warehouse as the central repository to store, organize, and analyze the Arbitrum data. This helps to consolidate the data, optimize it for querying and analysis and support the Phase II objectives in a scalable way. For the warehouse Helika used AWS Redshift and Postgres Cluster.

### Phase I Indexing Process and Deliverable

#### Index

Labeled the data from different sources so it can easily be interpreted and categorized. Example labels include transfers, mints, burns, wallets, contracts, tokens, stakes and more.

#### Cleanse

Removed unnecessary formatting and information not relevant to the objectives above. Standardized the data from different sources so that it's interoperable.

#### Categorize

Grouped the data based on its type and relationship to other data points for intuitive access and analysis.

For example; Helika's wallet summary solution is our wallet indexing and categorization system, giving Helika partners, including Arbitrum comprehensive wallet analytics. It is including transaction counts, active wallets by day, wallet value (in crypto or fiat), assets held (NFTs, tokens), DEX interactions, how much volume these wallets traded on different marketplaces, blue-chip collections traded and other NFT segments, etc.

Other categories we have created include marketplaces, transactions, collections, tokens, NFTs and more.

#### Structure

The last step is Helika has structured the categories for easy joining and analysis. In addition to the raw tables, Helika will aggregate related data to create custom KPIs and analytical structures for easy querying. For example wallet-level roll-up, marketplace-level roll-up, contract-level retention roll-up.

Example tables created include: NFT transfers, NFTs sales, DEX swaps, DEX liquidity and more. This structure allows analysts to join the data quickly by collection, wallet, token, marketplace, DApps and more for fast analysis.

## Phase I Costs

[

Chart

1200×742 19.3 KB

](https://global.discourse-cdn.com/standard17/uploads/arbitrum1/original/2X/d/d6688aa0a7b392c5e1202a3096d69b15f4065350.png)

## Phase II

### Phase II Objectives

Using the indexed and structured data from Phase I Helika will deliver the insights to answer the following questions for Phase II. Phase II is awaiting approval.

These questions are work-in-progress and open to feedback/requests:

- Who are the most valuable users and devs on Arbitrum?
- What is the Arbitrum user's path to purchase from web2 to web3?
- Are there alternative paths to drive engagement and retention?
- Where is the biggest drop-off in the FTUE funnel? (first time user experience)
- What other parts of the web3 ecosystem gamers are currently using or are likely to engage in (DeFi, etc.)
- What Arbitrum user archetypes are the most engaged? Least? Why? What segments drive the most monetization and TVL?
- What are the most common reasons for churn or for users to become disengaged?
- Where should Arbitrum invest to acquire the best users?
- Where should Arbitrum invest to incentive the most engaging and sustainable content?

### Phase II Deliverables

Note: These are open to feedback and will expand. Phase II is not yet approved.

For Phase II Helika will customize the visuals below among others with Arbitrum data to generate the insights reports that will answer the questions. These visuals are a limited example of the full scope the analytics that will be provided.

### Phase II Deliverable Examples

These will be used to inform the insights report and convey the answers to the questions from the objectives. The visuals can't be included in this Forum Post because the image sizes are too big (we also tried shrinking them). See [here](#) starting on page 5.