2025.12.12

Portfolio

1992.09.09
Hyo youl Park

Index

1 About me

Hello,
My name is Hyoyoul Park, and I worked as a Data Architect at Krafton. For over 7 years, I've been building analytics infrastructure for major gaming titles. I primarily handled the operation and maintenance of Hadoop Ecosystem-based infra structure, and more recently, I've focused on modernization efforts using Databricks and Delta Lake. I contributed to the mo dernization of data pipelines and the construction of scalable analytics systems for AAA games like PUBG, Hi-Fi Rush, and S ubnautica 2.

A particularly important aspect of my work has been cross-studio collaboration. Working regularly with international teams such as Montreal Studio, Tango Studios, and Unknown Worlds, I needed not only technical expertise but also strong communication skills in both Korean and English. Recently, I've worked on Hi-Fi Rush's revenue forecasting system and PUB G's in-game currency verification analysis, processing data for over 2.8 million users across multiple platforms.
The core of my work is making data accessible and actionable. Beyond simply processing data efficiently, I aimed to build systems that provide clear insights to diverse stakeholders, from game developers to executives. My goal was not just to co nnect and create dashboards, but to design analytics solutions that understand business context and drive actual decision-making.

Although I've been working in the gaming industry since 2018, I maintain a strong curiosity about various technical topics both within and outside my work. In my professional domain, working with Delta Live Tables and Databricks has been a distinctly different experience from operating traditional Airflow and Hadoop Ecosystem. Outside of work, I explore diverse areas ranging from 3D printing and mechanical keyboards to modern web development, compiler theory, and more recently, various experiments with AI. I find peace of mind through mechanical keyboard maintenance and repair, and managing my collection of succulents like Pachypodium. Of course, I still enjoy gaming regularly!

When I'm not working on data pipelines, you'll find me tinkering with Next.js using the Claude + Cursor combination, study ing AI/ML topics, or translating technical documentation between Korean and English for global teams.

GitHub Pages

✓https://hyoyoul-blog.vercel.app

✓www.linkedin.com/in/hypark5540

1

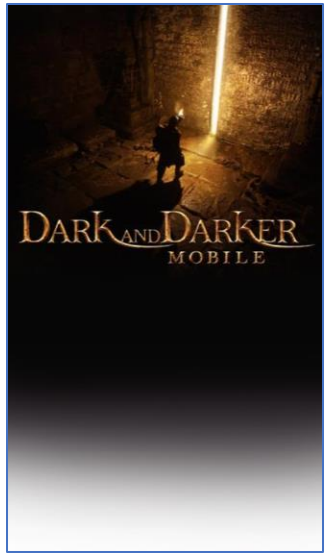
Work experience: 7y, 5months.

KRAFTON GAME UNION

✓ 2022.12 ~ 2026.01: **Full-time**,

-Data Architect, Analytics

Team member



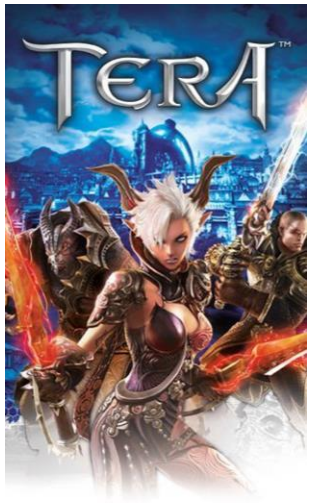
Bluehole

✓ 2018.06 ~ 2018.09: **Intern**

✓ 2018.09 ~ 2022.12: **Full-time**

- Data Architect, Data Engineering

Team Member



✓www.linkedin.com/in/hypark5540

GitHub Pages

✓https://hyoyoul-blog.vercel.app

Work Experience

Data Architect

- ✓ **Data Structure Reorganization and ETL Optimization**
  - ✓ Designed and optimized data architecture to improve ETL efficiency and analytical query performance across multiple game titles
  - ✓ Managed data structures utilizing Databricks features including Medallion Architecture, Delta Live Tables, and Liquid Clustering
  - ✓ Reorganized and redesigned data structure for enterprise-wide game data processing and management at Krafton, expanding beyond the initial 2-3 titles
  - ✓ Ensured consistency through integrated architecture analysis and code reviews with the Governance team for new title ETL structure design
  - ✓ Integrated Package Development and Team Collaboration for Multiple Game Titles
  - ✓ Developed internal data management tools using C# and LINQPad, improving work automation and efficiency
  - ✓ Implemented MSSQL server management, data extraction, and ETL processes
  - ✓ Provided integrated reporting metrics for collaborating overseas studios
- ✓ **Log Design and Verification**
  - ✓ Designed and verified logs based on device environments (PC, Mobile) and server infrastructure (Dedicated Server, Central Server)
  - ✓ Designed and delivered log schemas according to each game's characteristics and development studio requirements
- ✓ **Data Quality and Integrity Verification**
  - ✓ Migrated Hive and Presto-based queries and code resources to Databricks, verified compatibility, and modified code
  - ✓ Completed quality verification, final review, and data delivery for migrated Databricks ETL pipelines
  - ✓ Improved and migrated processes to support and verify PUBG primary currency accounting operations

Data Engineering

- ✓ **Hadoop Ecosystem Architecture & ETL Process Implementation and Maintenance Management**
  - ✓ Supported large-scale big data platforms across Azure, AWS, Apache Ambari Hadoop, Hive, Spark, and Presto environments
  - ✓ Contributed to DW system development including S3, MySQL, and Databricks DLT
  - ✓ Developed ETL workflow orchestration using Apache Airflow, Hadoop Workflow, Oozie, and Docker
  - ✓ Ensured data integrity through log QA and verification with development studios
  - ✓ Created interactive dashboards using Tableau, providing real-time insights to business teams
  - ✓ Supported rapid post-launch decision-making through data analysis and query optimization using Python and PySpark
  - ✓ Rapid Infrastructure Replacement Due to Open Source Policy Changes
    - ✓ Cloudera CDP -> Apache Ambari
- ✓ **External Sharing Operations Tool Development and Maintenance**
  - ✓ Developed and maintained Django-based LogQA tool, enabling rapid data delivery to external partners
- ✓ **ETL Process Optimization**
  - ✓ Consolidated and optimized multiple game ETL processes into a single ETL pipeline
  - ✓ Developed and maintained proprietary package (Calisto.py) for management
    - ✓ Redesigned and modified long-standing codebases with newly unified structure
- ✓ **Airflow Batch Automation Implementation**
  - ✓ Completed Airflow installation and configuration, containerization, and operational deployment
- ✓ **Log Collection Module Development**
  - ✓ Created AWS Kinesis SDK-based module tailored for Korea self-publishing
  - ✓ Completed pipeline (Kinesis-Firehose DataStream) creation, log storage verification, and metric delivery
  - ✓ Managed AWS Lambda-based AppAnnie (now data.ai) mobile revenue log collection
- ✓ **In-house Analytics Sharing Platform "Together" Development and Deployment**
  - ✓ Developed integrated structured/unstructured analytics environment based on PUBG Analytics Team
  - ✓ Built using PHP-based CMS XpressEngine
  - ✓ Created dedicated spaces for analytics requests and project-specific external analytics areas

Data Analyst

- ✓ **Predictive Analysis**
  - ✓ Responded to historical revenue forecast analysis requests for Hi-Fi Rush due to limited information availability
  - ✓ Applied case studies from other games to relevant AI models, trained models, and delivered predictions
- ✓ **Responding to Requests from Studios & Stakeholders**
  - ✓ Conducted early churn analysis, diagnostics, and FQ (First Quarter) analysis for LIVE games
- ✓ **Integrated KPI Metrics Delivery**
- ✓ **Responded to metric-related inquiries from development studios**

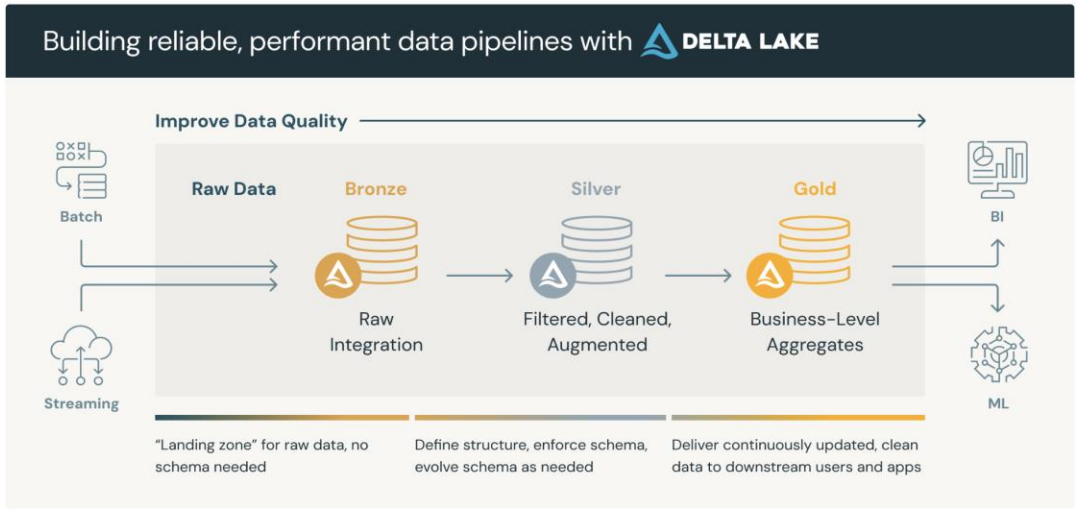
What I've done

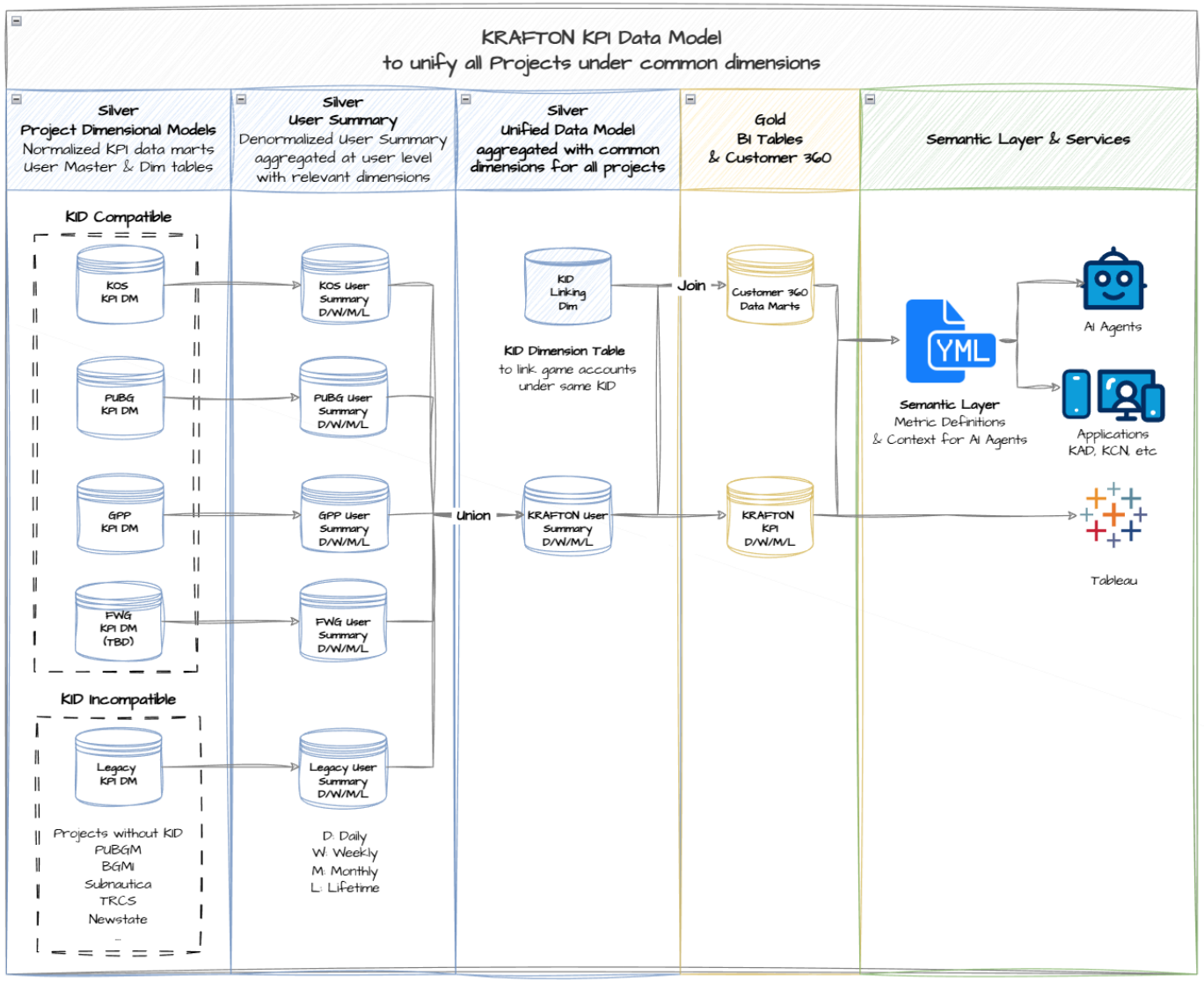
3

What I've done – **Data Structure Reorganization and Optimization in Accordance with Data Governance**

- ✓ **Project Description**
  - ✓ Applied policies based on Medallion Architecture structure proposed by the Data Governance team, streamlined integrated KPI management
- ✓ **Responsibilities**
  - ✓ Understood and planned Medallion Architecture, Naming Conventions, and table structures through collaboration with the Data Governance team, with documentation
  - ✓ Applied Liquid Clustering, Z-order, and Materialized Views appropriately for DLT situations within each table structure
    - Considered cardinality levels
  - ✓ Reorganized existing queries and data structures
  - ✓ Verified Krafton SDK-based telemetry logs
  - ✓ Managed integrated Git and conducted code reviews
- ✓ **Results**
  - ✓ Consolidated 3-5 pipelines per game project into 1 pipeline, reducing management costs
  - ✓ Improved and enhanced query speed through Liquid Clustering and Z-order application (reduced 1-hour tasks to 20 minutes)
- ✓ **Technologies Used**
  - ✓ **S3 / Databricks / DLT / Liquid Clustering / Materilized View / Z-order / Spark**

What I've done – **Data Structure Reorganization and Optimization in Accordance with Data Governance**





3

What I've done – **Integrated KPI Metrics Delivery**

- ✓ **Project Description**
  - ✓ Modified and unified metrics according to the direction of the Tableau TF team and team
- ✓ **Responsibilities**
  - ✓ Unified and integrated KPI dashboards that were previously managed in separate formats for each game project
  - ✓ Confirmed and improved structure through continuous collaboration with Data Governance for Databricks DLT connection to Tableau and connection architecture
  - ✓ Secured ease of connectivity through previously implemented Medallion Architecture-based reorganization
- ✓ **Results**
  - ✓ Consolidated multiple in-house metrics that were separately managed for each game project into 1 integrated metric
    - Excluding PUBG and some LIVE games
- ✓ **Technologies Used**
  - ✓ **S3 / DLT / Tableau / Databricks / Python / Spark**

3

What I've done – **Integrated KPI Metrics Delivery**

- ✓ **Project Description**
  - ✓ Operation and support for the following game services:
    - Subnautica2, The Bird That Drinks Tears (DEV)
    - Subnautica1, Hi-fi Rush (LIVE)
    - PUBG PC, PUBGM-KRJP (LIVE)
    - DNDM CBT (DEV)
    - TERA Console (LIVE)
    - TERA PC (Publisher Transition, LIVE)
    - ELYON PC (LIVE, CBT)
- ✓ **Responsibilities**
  - ✓ Data verification and deferred revenue related carry in-out calculation to support accounting operations related to PUBG primary currency system
  - ✓ Studio support for games in development such as Subnautica 2, The Bird That Drinks Tears, and DNDM (log design and verification)
    - Log design and verification based on device environments (PC, Mobile) and server infrastructure (Dedicated Server, Central Server)
  - ✓ Service support for LIVE-stage games such as TERA and ELYON
  - ✓ User matching and verification for business team events, PII identification through security compliance
  - ✓ Responded to various requests within constraints arising from publisher transition process
    - e.g., Estimated revenue measurement for Hi-Fi Rush due to Microsoft's conservative data-sharing approach
  - ✓ Performed various other tasks

3 What I've done – **Integrated KPI Metrics Delivery**

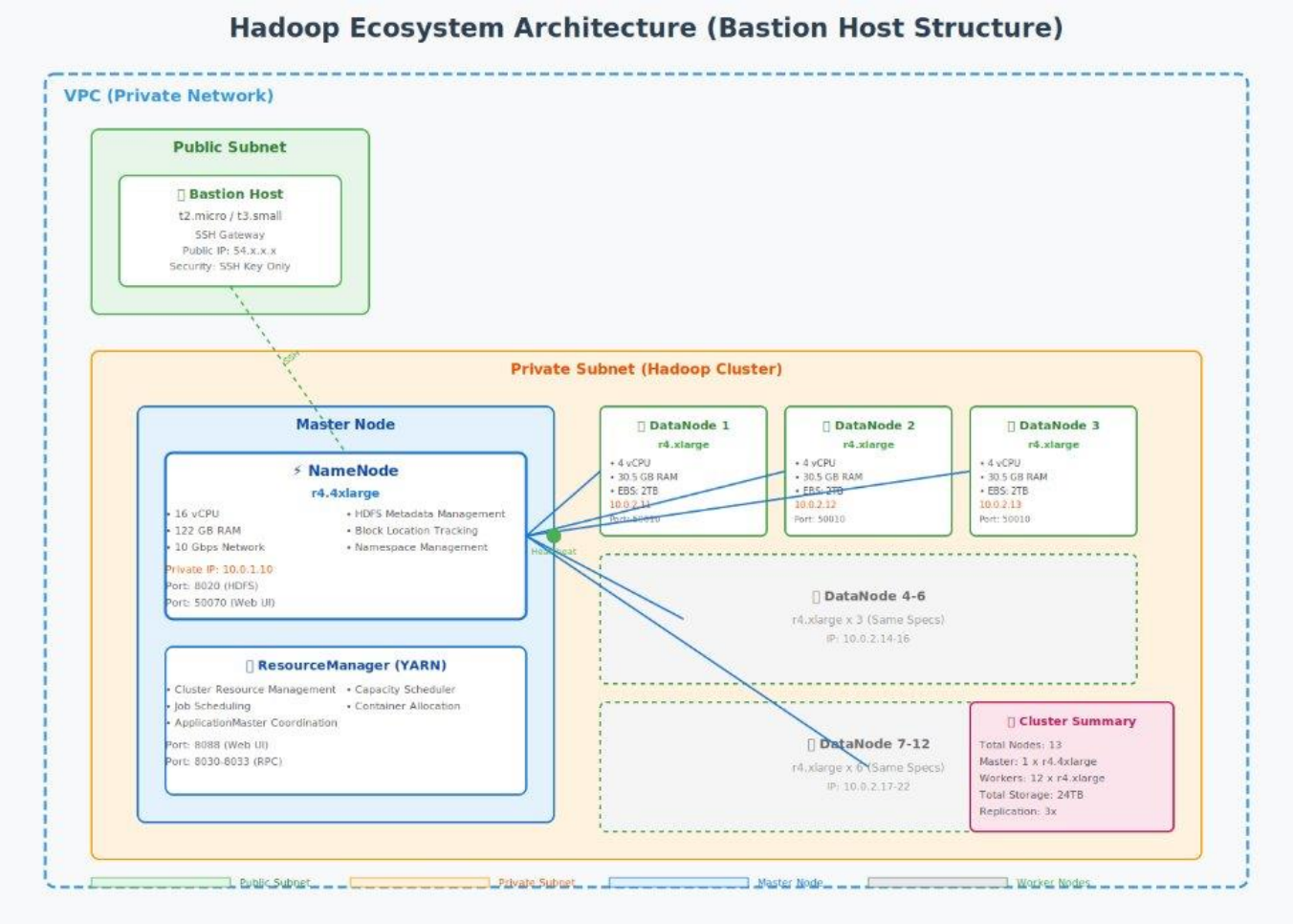
- ✓ **Results**
  - ✓ Gained experience in game launch support and operations
- ✓ **Technologies Used**
  - ✓ **S3 / Hive / Presto / Databricks / Opensearch / Unreal Engine Telemetry / Firebase / Kinesis / Python / Spark / Perforce(P4V) / Git**

3

What I've done – Infrastructure Migration

- ✓ **Project Description**
  - ✓ Migration of AWS ETL service (200 million KRW annually) operated by subsidiary Bluehole Studio to Krafton's integrated Databricks service within 6 months
  - ✓ Annual cost reduction
- ✓ **Responsibilities**
  - ✓ Migrated resources and data managed within AWS infrastructure to Databricks
  - ✓ Established policies after identifying legal data management and permission issues between subsidiary and parent company
  - ✓ Established and shared migration plan after reviewing Databricks DLT and Unity Catalog content → Execution
  - ✓ Completely reorganized and modified existing Hive and Presto-based syntax queries and code to align with Unity Catalog and DLT format
  - ✓ Modified batch operations and environment to minimize log data transfer costs
  - ✓ Version control through integrated GitLab
- ✓ **Results**
  - ✓ After completion of transition in December 2023, reduced annual AWS costs from 200 million KRW to 40 million KRW annually, approximately 80% cost savings (maintaining only S3 transfer and storage costs)
    - Note: This represents savings in additional redundant costs beyond Krafton Databricks' enterprise credit costs
- ✓ Technologies Used
  - ✓ **Databricks / DLT / Unity Catalog / Python / S3 / AWS / Spark**

What I've done – Infrastructure Migration



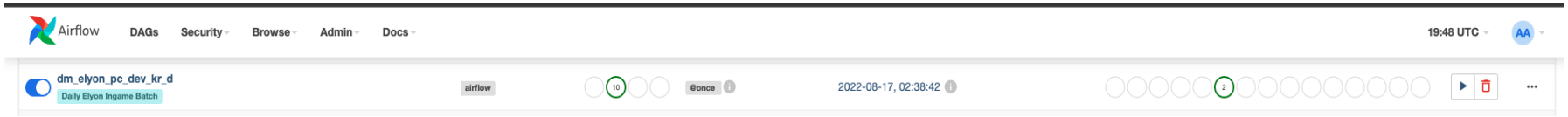
3

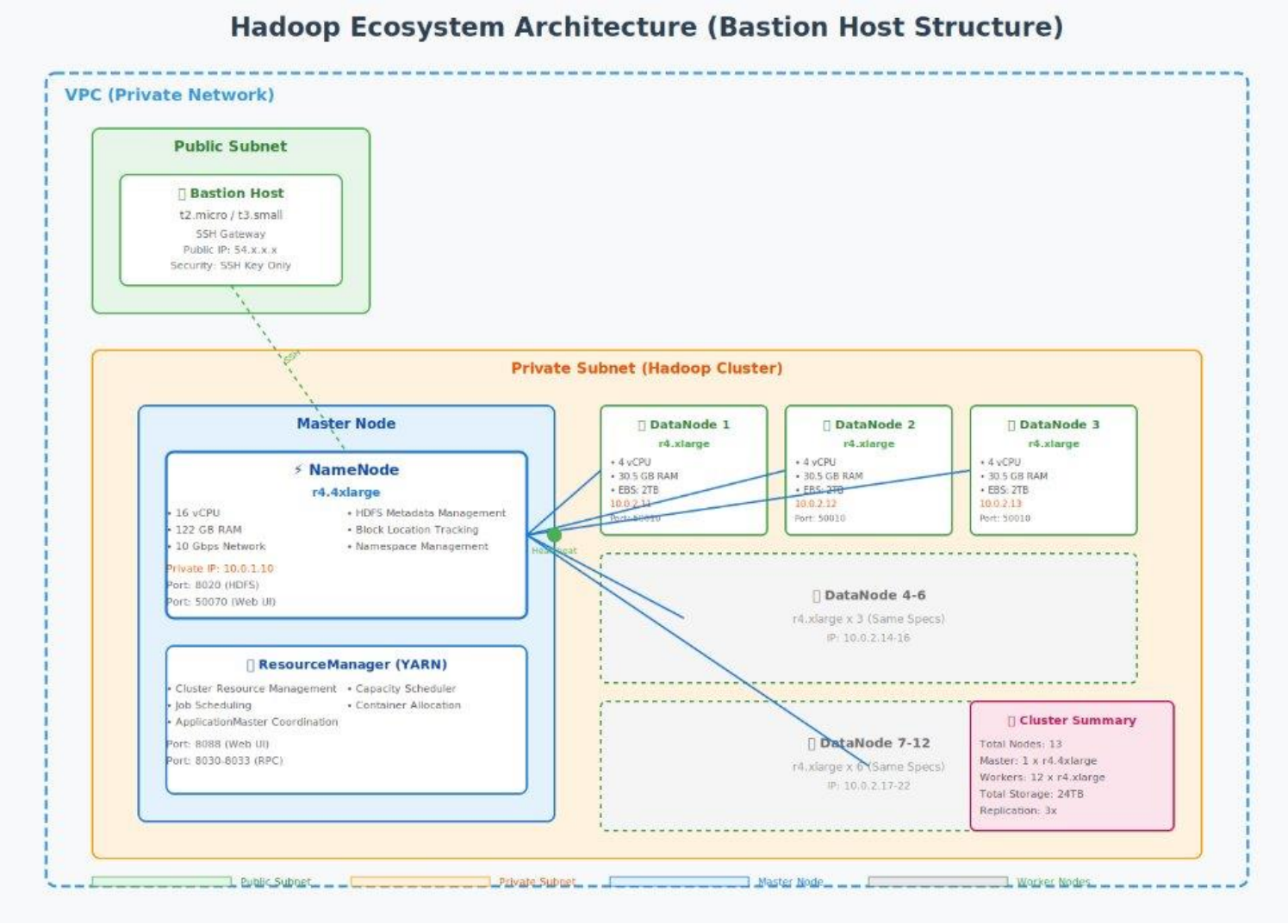
What I've done – Analytics Lab ETL Process Development and Maintenance

- ✓ **Project Description**
  - ✓ Maintained and contributed to Hadoop Ecosystem Architecture with 13-node structure using Bastion Host architecture
  - ✓ Initially applied Cloudera's CDP service, but switched to Apache Ambari due to open-source policy changes
  - ✓ Created internal ETL processes including writing DAGs per project (ELYON/TERA) using Airflow and Git management
  - ✓ Provided metrics for country-project data when batches were executed per country-project according to time zones
  - ✓ Developed and implemented self-verification module "Verify" for anomaly/missing value detection, implemented alarm system (SMTP, Slack)
  - ✓ Analyzed AWS RI purchase costs and monitored infrastructure expenses
- ✓ **Responsibilities**
  - ✓ AWS infrastructure maintenance and management
  - ✓ Development and maintenance of internal integrated project package (Calisto.py)
  - ✓ Consolidated and reorganized independent ETL processes per project (Elyon/TERA) into a single structure
  - ✓ Established and verified batch-specific query rules after modifications
  - ✓ Created self-verification package (Verify.py) and applied notification functionality across all channels
  - ✓ Applied feature to aggregate metadata from development network → update external network
- ✓ **Results**
  - ✓ Unified and integrated dispersed and difficult-to-manage ETL processes per project, reducing management costs
- ✓ **Technologies Used**
  - ✓ **Hadoop / AWS / Airflow / Presto / Python / S3 / Hive / RDBMS / Git / Tableau BI**

What I've done – Analytics Lab ETL Process Development and Maintenance

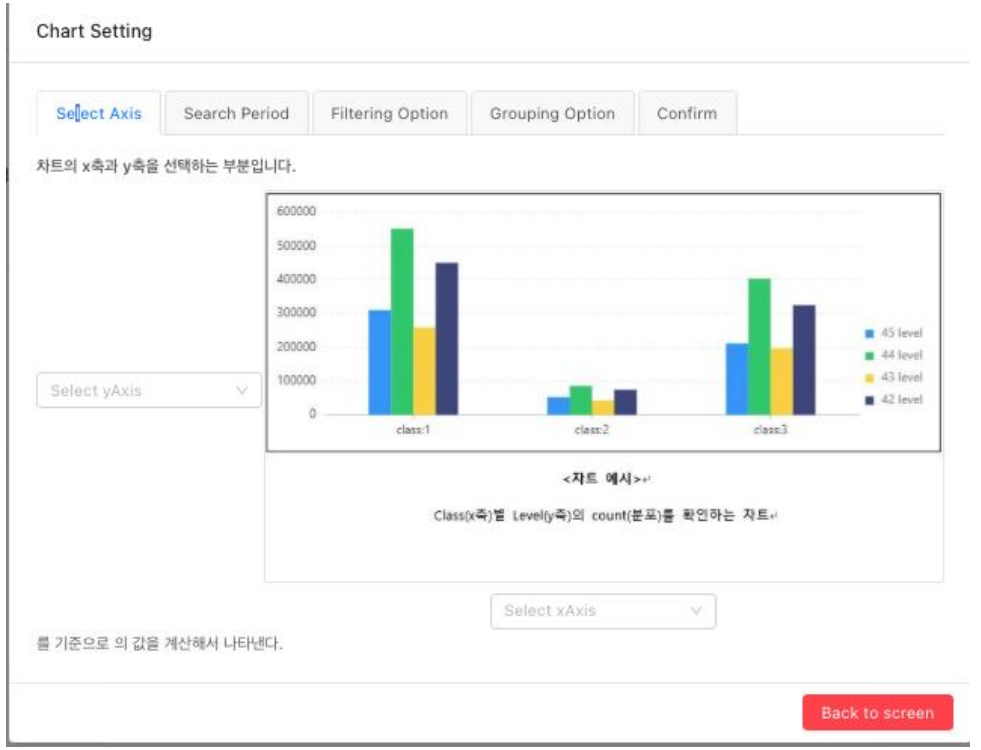
- airflowBatchSchedule.txt
- airflowUtil.py
- airflowUtilDev.py
- analysis 예시.ipynb
- Calisto.py
- config.json
- config.txt
- dateList
- Elyon.py
- monitor.json
- monitor.py
- newFile.csv
- newFile.csv.gz
- restore.ipynb
- return_day.csv
- runner.ipynb
- runner.py
- runner.sh
- runnerErrorCheck.py
- Tera.py

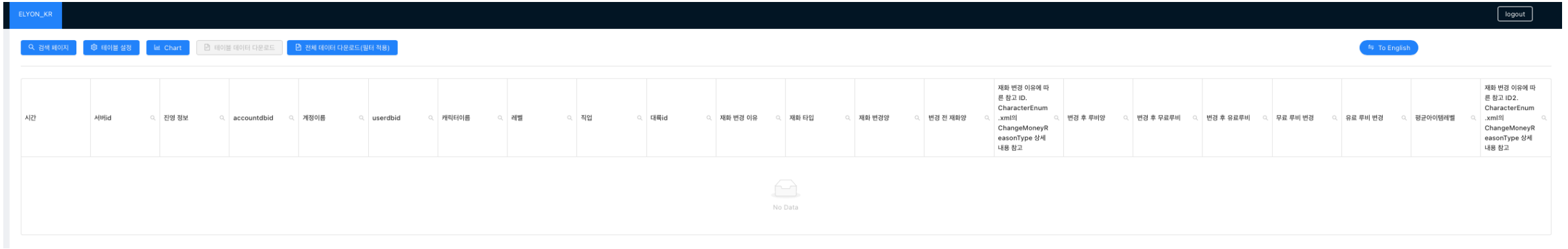




What I've done – Admin Tool : LogQA

- ✓ **Project Description**
  - ✓ Django-based external tool to help publishers view logs for their respective game projects
  - ✓ Publishers from each country can only view logs for their designated country
- ✓ **Responsibilities**
  - ✓ Server DB backup and version control
  - ✓ Update DB connection information for country-specific logs
  - ✓ Account management including user updates
- ✓ **Technologies Used**
  - ✓ **Python / django / Docker / React / npm / nginx / git**

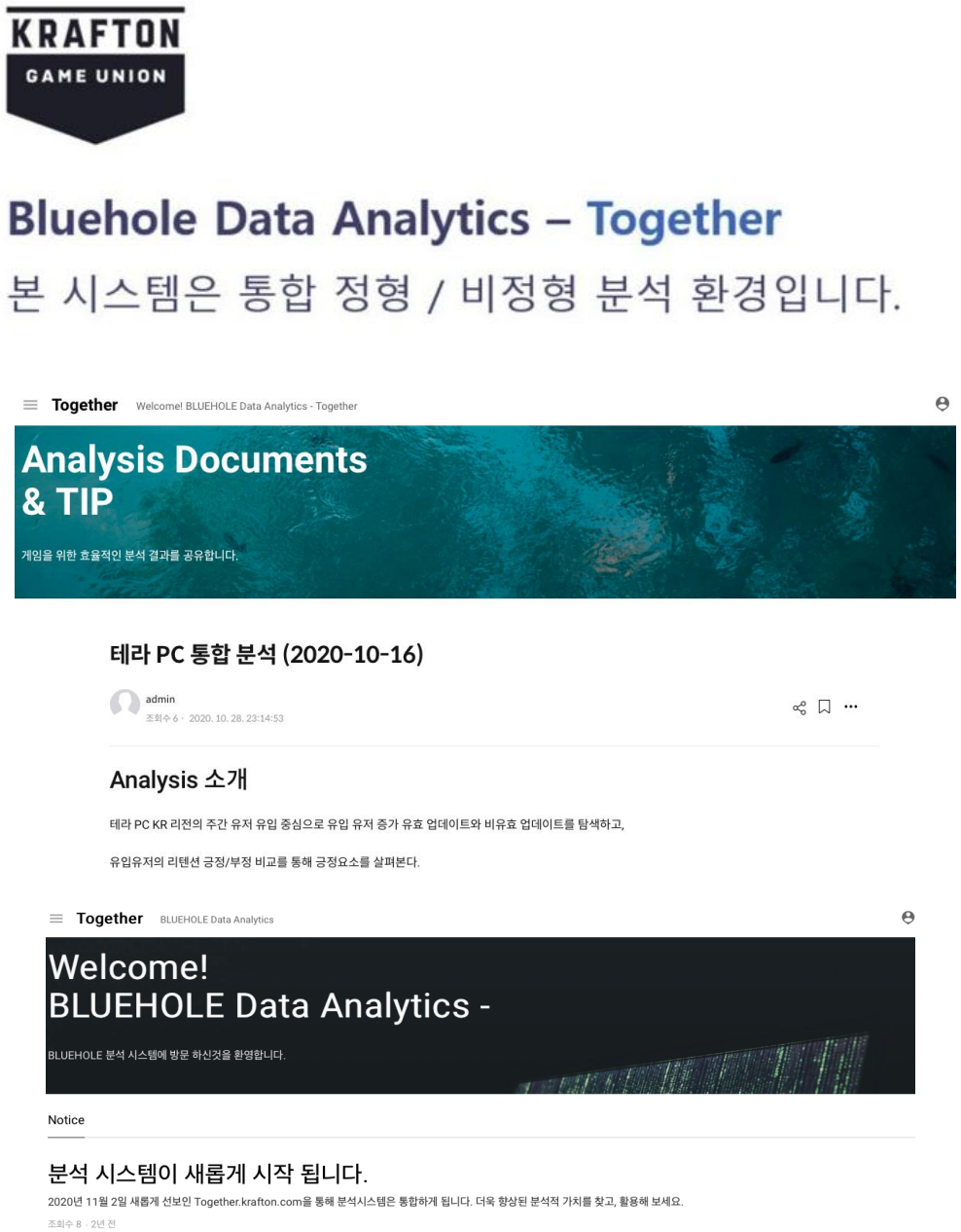




3

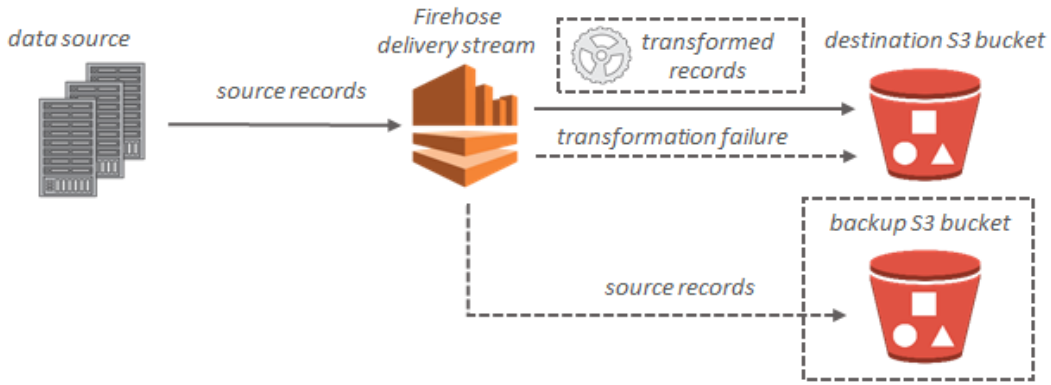
What I've done – Developed Web called 'Krafton Together'

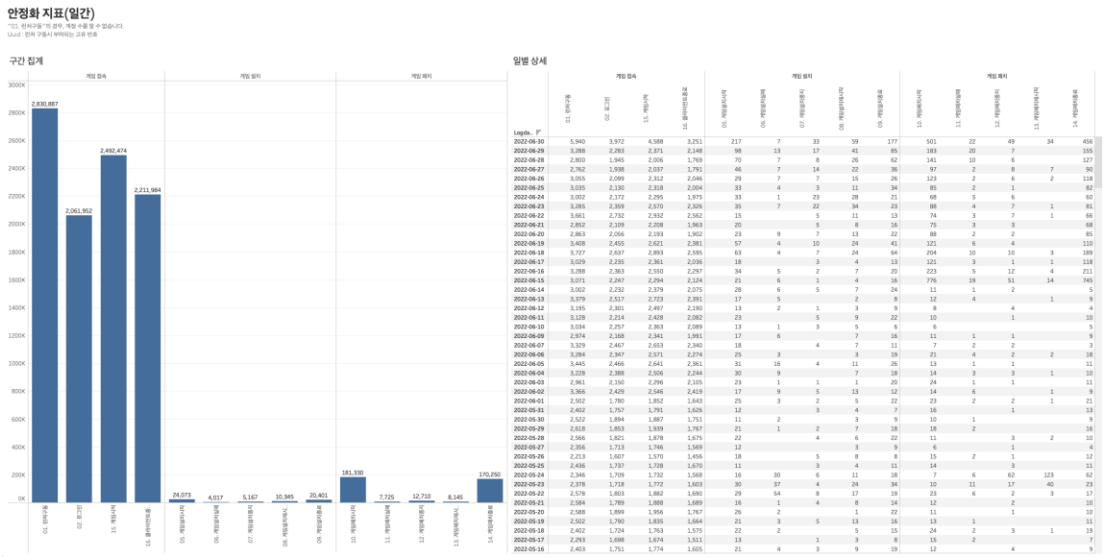
- ✓ **Project Description**
  - ✓ Developed integrated structured/unstructured analytics environment based on BHS
  - ✓ Built using PHP-based CMS XpressEngine
  - ✓ Created posting space for analytics-related requests & project-specific analytics spaces for external users
- ✓ **Responsibilities**
  - ✓ XpressEngine installation and development
  - ✓ DB backup automation
  - ✓ Account management including user updates
  - ✓ Plugin updates
- ✓ **Technologies Used**
  - ✓ **php / RDBMS / nginx / composer / jupyter / git**

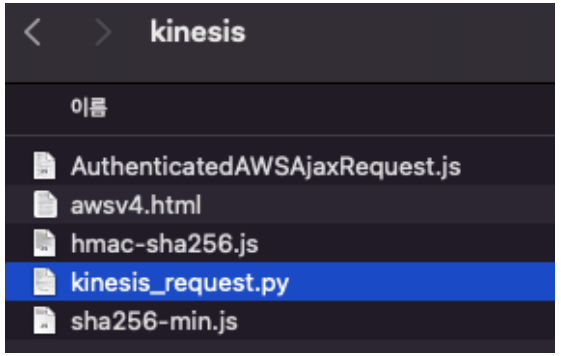


What I've done –TERA PC Self-Service Launcher Stability Metrics Development

- ✓ **Project Description**
  - ✓ After the start of TERA PC self-publishing, received request to create stability metrics for the self-developed launcher
  - ✓ Designed and delivered detailed log information including launcher success/termination
  - ✓ Created and delivered log generation and collection module (based on Kinesis SDK)
  - ✓ Provided metrics from collected log information
- ✓ **Responsibilities**
  - ✓ Stability log design
  - ✓ Log generation module creation and deployment
  - ✓ Log collection and storage
  - ✓ Metric creation
- ✓ **Technologies Used**
  - ✓ **AWS Kinesis / AWS Firehose / python / js / Tableau BI / git**



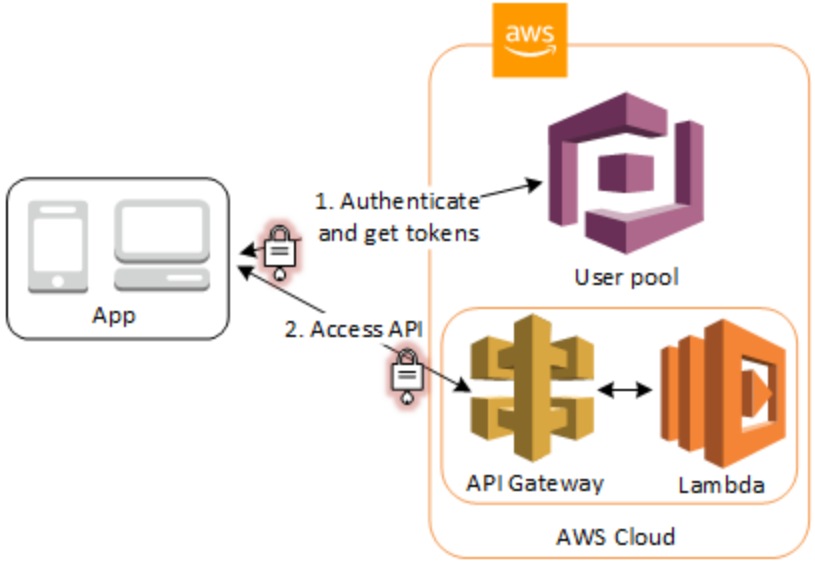


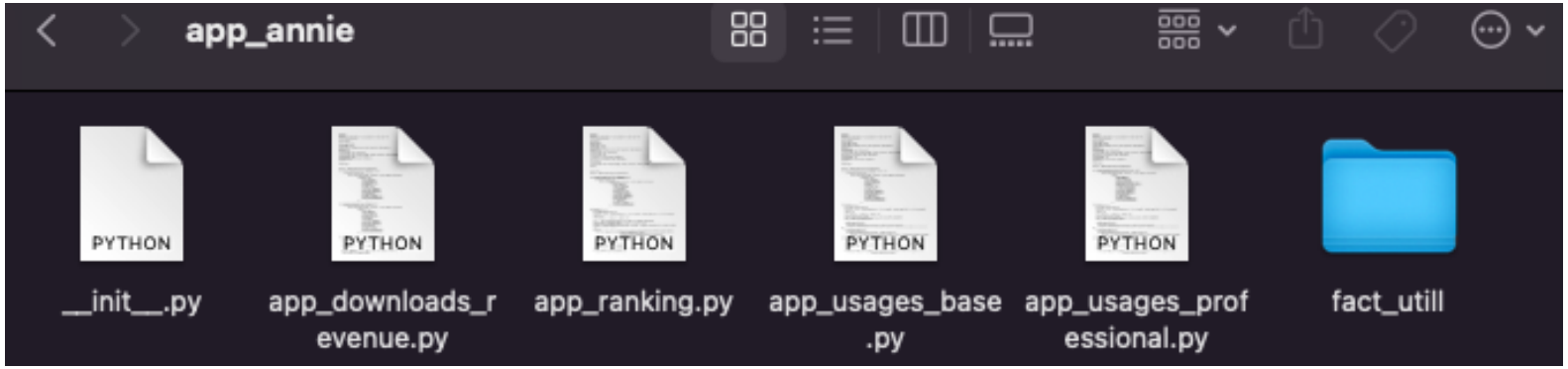


What I've done –Internal Mobile Game Revenue API Development

- ✓ **Project description**
  - ✓ Collection and storage of mobile game market revenue data from AppAnnie (now data.ai)
  - ✓ Developed and deployed custom API to enable internal Krafton access to stored data
  - ✓ Bulk data collection and processing for optimal execution cost and efficiency

  (refer to API documentation)
- ✓ **Responsibilities**
  - ✓ API development and deployment
  - ✓ Data collection and storage
- ✓ **Technologies used**
  - ✓ **AWS API Gateway / AWS Lambda / RDBMS / python / jupyter / git**





2025.12.12

Contact

✓www.linkedin.com/in/hypark5540

✓https://hyoyoul-blog.vercel.app

KRAFTON GAME UNION