Kevin Williams Permana
Data Engineer

Jakarta, Indonesia | Phone: +49 176 58691506 | Email: kevinity310@outlook.com

SUMMARY

I am a Data Engineer with nearly 3 years of experience, specializing in on-premises big data ecosystems built on Hadoop and Kubernetes. I have supported clients in migrating to big data platforms, integrating systems, developing APIs, and ensuring production stability. Proficient in Hadoop, Python, Spark, Hive, Trino, Airflow, and Kafka, I also have experience building real-time data pipelines from Oracle databases to big data platforms for analytics. I'm passionate about solving complex problems with innovative data solutions.

EXPERIENCE

PT Bodha Dharmaja Aryadhana – Jakarta, Indonesia
Data Engineer (June 2020 – present)
- Big Data Platform Development & Maintenance
    - Researched and developed tools with team for building a big data platform using Kubernetes (on-prem) with Hadoop, Hive, Spark, Jupyter, Airflow, Kafka, and other technologies.
    - Performed testing, monitoring, and maintenance configuration to ensure optimal performance and platform stability.
    - Led the research and development of integrating Hadoop into a data lakehouse architecture by incorporating MinIO (on-prem object storage), Apache Iceberg, Apache Hudi, and Trino. This solution was developed to meet client needs and overcome Hadoop's limitations, adding transactional features and improving the platform's scalability.
    - Developed custom APIs to support client needs and integrate backend services with the frontend.
- Support Client (PT Gudang Garam Tbk.)
    - Data Migration & Archiving
        - Developed and migrated ETL/ELT processes from Oracle to Hadoop using Alteryx, reducing the Oracle database workload.
        - Developed and migrated ETL/ELT processes to our big data platform, helping the BI team reduce the workload on the Qlik Sense server.
        - Assisted the BI team in integrating the big data platform with Alteryx.
    - Change Data Capture (CDC) & Replication
        - Collaborated with Oracle DBAs to set up Oracle XStreams and implement CDC using Debezium for real-time table replication.
        - Monitored, troubleshot, and tuned the CDC connector to ensure stability.
    - Real-time Data Streaming
        - Built real-time data streaming pipelines using Spark Streaming, integrating with Apache Hudi and processing data from Apache Kafka.
- Support Client (PT Surya Madistrindo)
    - Support the client's data engineering team in migrating their data to our big data platform and integrating their existing service Airflow.
    - Contributed to developing a machine learning model with the data science team for the Traveling Salesman Problem (TSP), optimizing routes to reduce operational time and fuel costs.

EDUCATION

BUNG KARNO UNIVERSITY (2020 - 2024) | Indonesia
Bachelor's Degree in Information System - GPA 3.6

SKILLS

- Hadoop
- Python
- Spark
- Trino
- Kafka
- Apache Iceberg

- Apache Hudi
- Airflow
- CDC With Debezium
- Docker
- Kubernetes

ADDITIONAL ACHIEVEMENTS

- Successfully contributed to building a big data platform called "Sapujagad," overseeing testing, maintenance, and tool selection until deployment in production, and ensuring the platform is ready for client use.
- Successfully integrated the Hadoop ecosystem with modern data lakehouse tools (MinIO and Trino), combined with table formats such as Apache Hudi and Apache Iceberg, to enable many features in the data warehouse and provide flexibility for processed data.
- Implemented CDC in Oracle using XStreams (after replacing LogMiner due to high delays) and enabled real-time table replication with Debezium.
- Achieved success in building and running real-time analytics streaming pipelines using Spark Streaming, integrated with Apache Hudi.
- Complete the migration of ETL/ELT data processing for the client to reduce the high workload on their service database (Oracle) and visualization tools (Qlik Sense, Tableau).

CERTIFICATIONS

- Bootcamp at Dumbways.ID
  - Fullstack Javascript Developer | 21 March 2022 - 21 May 2022
- Bootcampt Digital Skola
  - Data Engineer | 21 March 2022 – 21 Mei 2022
- Apache Airflow: The Hands-On Guide | Udemy – Marc Lamberti

STRENGTHS

- Analitical Problem-Solver
  - Utilize data analysis tools to overcomes production issues, increasing efficiency
- Collaborate Team Player
  - Successfully collaborate with cross-functional teams to deliver projects on-time and optimal workloads
- Adaptable Learner
  - Happy learner and applied new data technologies to solve many problem on data engineer.