Willi Oktavega

Email: willi.oktavega@gmail.com   **Website**: willioktavega.github.io   **Phone**: +62 81220562142

PROFESSIONAL EXPERIENCE

VIDA, Remote, Indonesia   January 2023 - Present

- Migrated batch jobs from cron in on-premise server to managed apache airflow service in AWS. Provides better monitoring and alerting for the jobs.

  Technologies/Tools: Apache Airflow, AWS MWAA, AWS CloudWatch, Slack API
  Programming: Python, SQL, Shell Script

- Improved engineering team and data scientists productivity by developing a job to decode base64 image data into an image file format and organize the image files for machine learning or engineering purposes.

  Technologies/Tools: Apache Spark, AWS S3, AWS EMR, Encord
  Programming: Scala, Python

- Developed a self service tool for data analysts/scientists to create AWS Athena table.

  Technologies/Tools: DBT, AWS Athena, AWS Glue
  Programming: Python, Shell Script

GoTo Financial, Jakarta, Indonesia   October 2019 - December 2022

- Revamped reporting service data pipeline for end users (merchants) by delivering 10x at least increased performance on average compared to the legacy reporting service. Bringing near-real time data processing combined with batch data processing as a backup. Provides a better experience to generate reports smoothly in the given period without experiencing time-out anymore.

  Technologies/Tools: Apache Airflow, Apache Spark, Apache Flink, Apache Kafka, AWS, Docker, ClickHouse, PostgreSQL, Apache Livy
  Programming: Python, Scala, Shell Script, SQL

- Migrated Apache Spark jobs run at AWS EMR to Kubernetes without downtime. The goal is to align with engineering plans to migrate from AWS to GCP.

  Technologies/Tools: Apache Spark, AWS, Kubernetes, Gitlab-CI
  Programming: Java, Shell Script

- Adopted a new data warehouse platform in the Commerce Enablement business unit to leverage data governance, data security, and data sharing between business units.

  Technologies/Tools: Debezium, Kubernetes, Apache Airflow, Raystack Optimus, Big Query, Gitlab-CI
  Programming: Python, Shell Script, SQL

Female Daily Network, Jakarta, Indonesia   March 2019 - September 2019

- Enable data pipelines and ready to use data for data analysts. Improve data analysts productivity as they didn't have to wait for the engineering to provide their requests.

  Technologies/Tools: Apache Airflow, AWS S3, Bitbucket Pipeline, Docker
  Programming: Python, Shell Script

- Proposed product recommendation for the end users and offers 70% in accuracy obtained by manual validation by the product team.

  Technologies/Tools: AWS S3, Docker, Bitbucket Pipelines, Apache Spark, Redis, Flask, Shell Script
  Programming: Python

Detik Network, Jakarta, Indonesia   January 2018 - February 2019

- Revamped most popular service to give better performance and data quality. The result is reduced potential production issues by 99% and easy to resolve the common issue in less than an hour.

  Technologies/Tools: Apache Spark, Apache Airflow, Apache Tomcat, Elasticsearch, Dataproc Google Cloud Platform, Gitlab CI, Docker, Google Cloud Pub/Sub
  Programming: Scala, Python, Java

- Proposed user profiling service to personalize user behavior when accessing our news platform and offers 80% in accuracy obtained by manual validation by the product team.

  Technologies/Tools: Apache Spark, Apache Airflow, Elasticsearch, Dataproc Google Cloud Platform, Gitlab CI, Docker, Google Cloud Pub/Sub
  Programming: Scala, Python

- Proposed tags suggestion using *Named Entity Recognition* for news writer and offers 70% in accuracy obtained by manual validation by the product team.

  Technologies/Tools: Django, GraphQL, SpaCy NLP, Apache Airflow, PostgreSQL, Gitlab CI
  Programming: Python, Shell Script

EDUCATION

Telkom University, Indonesia   Aug 2013 - Jul 2017

Bachelor of Informatics Engineering
Thesis: Aspect-Level Sentiment Classification on Product Reviews using Random Forest

SKILLS AND INTERESTS

Programming: Python, Java, Scala, Golang, SQL, Shell Script
Languages: English, Indonesia
Interests: Machine Learning, Natural Language Processing, Data Engineering, Software Engineering

VOLUNTEERING AND SERVICE

Data Science Indonesia, Indonesia   2020

In collaboration with **Commerce Enablement - GoTo Financial**, share about the experience as data engineer in the Commerce Enablement team.

Telkomsel, Indonesia   2019

Instructor for hands-on introduction to data engineering and machine learning.

Detikcom, Indonesia   2019

Instructor for hands-on batch processing with Apache Airflow and Apache Spark.