

Siriwat Wisetpakdeewong

Data Engineer

 swisetpakdeewong@gmail.com  (+66) 97-231-5595  <https://github.com/MekWiset>

PROFILE SUMMARY

Data Engineer with 2.5 years of experience building and maintaining data pipelines across cloud platforms and large-scale migration projects, with a proven track record of delivering technical solutions and leading teams to drive efficiency across complex data workflows.

WORK EXPERIENCE

Inteltion

Nov 2024 - Present

Kasikorn Business Technology Group - Data Logic Migration

- Led the team's ingestion spec preparation, building a flagging script that automatically identified the 450 of 900 objects (50%) requiring no manual intervention, allowing the team to concentrate manual effort on the remaining objects.
- Validated and cleaned the remaining objects requiring manual intervention, resolving outdated entries and column inconsistencies, and set team standards for handling the technical specs before distributing to downstream teams.
- Conducted job-chain analysis across shell scripts and IBM DataStage jobs, mapping nested dependencies and establishing the correct execution order across 900 target objects prior to deployment.
- Converted data population logic from shell scripts and IBM DataStage into Databricks notebooks, handling deeply chained dependencies to ensure full functional equivalence with the legacy implementation.

Kasikorn Business Technology Group - Modern Data Platform Migration

- Led a sub-team of 6 data engineers within a 20+ person data migration project, delivering a full schema of about 1,000 objects (out of 5,000 objects project-wide) ahead of the project timeline.
- Responsible for tracking team delivery status, coordinating across sub-teams, and reporting progress to management.
- Developed an automated sample-data extraction script that eliminated manual intervention entirely and resolved mixed load frequency issues, saving ~2-3 hours/day across ~70% of the project.
- Built on-demand utility scripts and tools, enabling teammates to handle repetitive tasks independently rather than relying on manual effort.
- Established communication standards and execution-order protocols across all 5 sub-teams to keep work synchronized and reduce coordination overhead.
- Trained 5 new joiners on working standards and tooling, bringing them to independent productivity within 2 weeks.
- Extended support into the PROD env, resolving ~50 incidents from 5-10 recurring issues and supporting the reconciliation team through final sign-off.

Bluebik Digital (Internship)

Mar 2024 - Jun 2024

Big Data Migration Across Cloud Platforms

- Modeled a 19M-row raw dataset into normalized tables using SCD Type 2 and mapping tables to resolve data conflicts and preserve historical records, with big data cleaning and transformation to ensure consistency across all tables.
- Leveraged containerization for Airflow and Spark via Docker to build a reproducible dev environment for full pipeline testing at zero cloud cost.
- Built an automated daily batch pipeline with Airflow and PySpark, processing ~50,000 rows/day into a Google Cloud Storage landing zone with backfill recovery, then orchestrating into Google BigQuery for future business consumption.
- Refactored procedural scripts into dynamic, config-driven OOP classes, cutting pipeline runtime by 30% while improving maintainability with production-grade coding standards and best practices.
- Implemented a scheduled Azure Data Factory pipeline into an ADLS Gen 2 landing zone, with Databricks reconciliation that caught non-ASCII encoding mismatches in product-description fields after loading into the target Azure SQL Database.

National Electronics and Computer Technology Center - NECTEC (Internship)

Jul 2023 - Sep 2023

Data Pipeline Development

- Independently architected, built and delivered a greenfield ETL pipeline for medical product data, bringing a previously unhandled data flow into fully automated daily operation.
- Extracted and denormalized data from 4 relational tables (~450K rows) via pyscopg2 and Pandas, producing a clean, MongoDB-optimized document schema that matched the client's structure and format requirements.
- Built checkpoint tracking, selective row updates, and structured error handling into the pipeline, ensuring resilient and unattended daily execution in the production environment.
- Orchestrated the full pipeline with Apache Airflow, eliminating manual intervention as new data continuously lands in the source.
- Validated data integrity end-to-end through reconciliation logic across PostgreSQL and MongoDB, confirming consistency between source and target.

EDUCATION

2020 - 2024

- Civil Aviation Training Center (CATC)
 - Bachelor Degree in Aircraft Electronics Engineering

CERTIFICATES

- Databricks Certified Data Engineer Associate
- Microsoft Certified: Fabric Data Engineer Associate (DP-700)
- Microsoft Certified: Security Operations Analyst Associate (SC-200)