

EDUCATION

SIT-King Mongkut's University of
Technology Thonburi (IT)

Master degree

Faculty of Information technology

October 2021 – April 2023

RMIT University / Kasetsart University

Bachelor degree (Double Degree)

Faculty of Aerospace engineer September
2014- April 2018

INDUSTRIAL LICENSE CERTIFIED

-Azure Databricks Platform Architect

-Databricks Lakehouse Platform Essential

-Databricks Certified Data Engineer
Associate

-AWS Partner: Accreditation (Technical)

-Kafka Confluent fundamental Credential

-Aws Partner Accreditation Technical

-Azure Data Engineer Associated Certified
(4 times re-certificate since 2020)

-Azure Fundamental AZ-900

-Azure Security, Compliance SC-900

-Azure Data fundamental DP-900

-Tableau Desktop Specialist

-Talend Cloud Fabric certified

-IBM Big data fundamental

-Salesforce-Datorama Admins Certified

SKILLS

Programing language

Python Java Spark SQL Html CSS

Version Control – CI CD

Github Action / Git branch strategy

Container concept

Docker and orchestrating concept

Data visualization with BI tool (tableau
power-bi/data studio)/Streamlit

Big data understand of concept ETL and
data warehouse and OLAP operation,
parallel distributed computing, Hadoop
ecosystem/Spark

Networking concept ip4, DNS, DHCP,
routings model, VPC, NSG, subnet mask

CONTACT

PHONE:

+66-95739-0898

GitHub : [www.github.com/Panuvat-Dan](https://github.com/Panuvat-Dan)

EMAIL: dpanuvatz2m1994@gmail.com

Panuvat Danvorapong

<https://panuvatdport.onrender.com/>

TECHIN AL CAL CONSULTANT

• Technical Consultant POD Lead on migration sub contract for Databrick Asia pacific (6 months contract Oct 2024 - Mar 2025)

- Consult and Modernize the ETL Framework (ingestion – Curated – Outbound – OLAP – Reconciliation)
- Migrate HD-Insight workload (Hadoop) to Azure databrick workload (Configuration – data pipeline job – metadata sync – control table)
- Migrate Traditional Hive metastore to Unity catalog to enhance - modernize governance data platform
- Code conversion from Hql (Hive) to SparkSQL (Databrick) for both DDL – DML statement
- Populate data partition by partition (business date) to delta once data migration framework run with databrick workflow
- Reconciliation data from Hive meta-store and delta table in unity catalog
- Develop ETL modules (config manager – log handler – ingestion – transformation -inbound - outbound) using pyspark and unit test on spark submit
- Communicate with customer as on their requirements and test result
- Research spark structure streaming checkpoint to handle incremental source change instead of traditional watermarking CDC (last update from log)
- 50 % Manage and Mentor 4 juniors as incident requests

• Senior ETL Consultant for Generali Insurance Group (July 2023-Present)

- Develop and maintain existing ETL pipeline using Talend Studio
- Collaborating and gathering requirement from finance and actuarial and Business analyze team.
- Scheduling ETL job using TAC orchestration tool
- Data preparation for monthly report for IFRS team
- Research data ingestion processes leveraging delta Lakehouse technology, ensuring efficient and scalable solutions.
- Researched Incrementally load on Autoloader, Checkpoint, and Watermarking techniques to capture source changes during batch/micro-batch processing.
- Implemented Slowly Changing Dimension (SCD2) to store historical data for master target tables
- Explored and applied Databricks optimization techniques such as data skipping (optimize), z-ordering, partition, vacuum for improved read -write performance on delta tables.
- Designed and automated pipelines to manage job dependencies, streamlining workflow configurations.
- Conducted unit testing for Python code on Databricks notebook.
- Scheduled and managed data pipelines efficiently using Databricks workflow.
- POC Configured YAML files on Github Action workflow to create build a pipeline artifact after committing from developers
- Researching on how to use terraform to deploy databrick's workflows, job cluster
- Work and collaborate with data governance team on technical issues e.g. data catalog tool (Talend data catalogue), business glossary, and data dictionary, data mapping and data lineage from system

R-system Consultant on IFRS17 project for KT-AXA group (May2023 – Oct2023)

- Gather (non) functional requirement from account and actuarial team
- Design change data capture technique and design data loading strategy (Full and Incremental load) on multiple data zoning when data is moving from source to destination
- Design and implement log and meta data driven pipeline using job log and job configuration
- Prepare the data from existing warehouse using aws-redshift as IFRS17 to server data reporting
- Develop and ETL batch processing (data ingestion – data transformation for daily and monthly job using ETL Talend studio tool)
- Maintain data integrity when data moving before generating IFRS17 report
- Optimize the data pipeline execute time by tuning SQL or implement a new solution columnar index – format – data virtualize technique

Data Engineer Consultant (July 2022 – June 2023 1 year contractor) at Adastra Consultant

- Develop a ETL pipeline by using azure synapse pipeline
- Unit test/UAT data pipeline on test dataset and reconciliation data between the hop
- Implement Data modelling from staging to curated layer using store-procedure and orchestrate by azure synapse pipeline.

- Optimize read performance by design distribution type on external tables (hash/round-robin table) on azure synapse cloud data warehouse
- Adjusting store-procedure to improve log framework to handle with both existing and new jobs
- Create Change data Capture logic with watermarking technique to capture when source has been changed (incremental load)
- Manually Deploy a synapse analytic by azure resource manager template to migrate all pipelines/datasets/linked service from dev to UAT
- Researching on Event -Driven Architecture kafka and spark structure streaming/ checkpoint to handle streaming workload

• **Data Engineer to Tesco Migration Project at Lotus Co,ltd (Sep 2021 – July 2022 as 6 months contractor)**

- Data ingestion on Delta architecture from data-sources (API – Table – SharePoint – Manual Files) to raw layer (Tier1) on Azure databricks platform
- Deduplicate and Cleansing data from Bronze layer to Silver layer
- Support Business unit e.g. Conceptual-Logical-Physical Database design and architect dataflow included full/incremental loading type – data type – schema table type.
- Develop and maintain ingestion - transformation Framework included job detail and job dependency configuration to parameterize the work-flow.
- Administrate task like set up and config ADF integration runtime/Linked service - Dataset
- Design data lake storage access's authentication approach (service principle – access key – sas key),
- Research on CI/CD solution on Azure devops tool to create a build – release pipeline

• **Project Engineer Toyota Motor Thailand (July 2019 – Oct 2021)**

- Initiated and planned strategic data projects (used car price prediction and inspection application based on historical data) aligning with business objectives.
- Conducted project status and progress, ensuing deadline and delivered all projects within established timelines and budgets.
- Collect and Prepare data from Relational database system to generate business intelligent report which is interactive and automatically refreshed every day (batch processing)
- Researching data technology and summarize report to the team.

Collaborating with Software engineer to catch program bug and find the way to error handling that bug.

Project Contribution and Education Publish Work

1. SiamETLFlow python package (<https://pypi.org/project/siamflowetl/1.0.0>)

- Owner and Contributor for python package to help ingestion and data quality
- Create Ingestion – Quality handler
- Create Dimensional Modeling Suggestion for dimensional – fact table from RDBMS system
- The Package version 1.0.0 only handle with pyspark and pandas dataframe
- Available via pypi by pip install siametlflow==1.0.0

2. Publish Master Independent study during master degree study

- Develop a machine learning model to predict on heart disease and deploy an interactive web app via streamlit
<https://panuvat-dan-mldeployment-6513070033-app-main-65130700332-xdteva.streamlit.app/>
- Develop ETL Modules (ingestion – cleansing – transformation – config manager – log handler modules) using tech stack postgres – docker – github action – airflow – spark – pandas
<https://deproject-cj2llr8mehkwxcaqjac74.streamlit.app/>
- Publish analytic sport complex web apps via KMUTT library using tech stacks flask – pandas – postgres – html – css