THANATHORN BOONYAKIJJINDA

Address: 77/193 CONDO REGENTHOME 22, Bangkok , Phrakhanong , Bangchak 10260
Phone : 0945603471
Email : Thanathorn.boon.cp@gmail.com

ABOUT ME

May I introduce myself? My name is Thanathorn Boonyakijjinda, but you can call me Chopper. I was born on 11 April 2000, making me 25 years old. I'm from Thailand and currently living in Bangkok. For my educational background, I graduated from Thammasat University with a bachelor's degree in Electrical Power Engineering with a GPA of 3.58. I have a passion for continuous learning and am always eager to enhance my skills. I am easygoing and cheerful, which helps me build strong relationships and collaborate effectively with others.

Website for additional information about my work. https://choppertnt.github.io/Thanathorn_Boonyakijjinda_webprofile/

EDUCATION

Thammasat University 2018–2023

- Power Electrical major , Engineering with GPAX at 3.59

EXPERIENCE

2023 Aug–2025 Jan

Alot tech • Data Engineer

ETL Process (Extract, Transform, Load)

- Extract: Automatically retrieve .xlsx files from Google Drive using Python daily.
- Transform Clean data by removing inconsistencies, null values, and unnamed columns using Pandas.
- Perform data transformations with SQL to calculate daily sales and group data by product type, while comparing it to the previous month.
- Store transformed data in a PostgreSQL database.
- Upload daily sales data to Google Sheets using the Google Sheets API.
- Analyze: Create visual dashboards in Power BI to display daily sales figures and compare cumulative sales with the previous month.
- Report: Export Power BI reports as PDFs for easy sharing. Automate email distribution of reports to executives daily using Python.

Monthly Report

- Create a monthly marketing report that includes the following sections: sales, engagement, and consumer behavior. The data will encompass YoY (Year-over-Year), MoM (Month-over-Month), YTD (Year-to-Date) , growth rate and an analysis of trends
- There are two formats: the first format is visualization, where in the past, reports were created using Looker Studio, but now we have switched to using Power BI for visualization.
- The second format is reporting, where we use Canva to create presentations for conferences.

Key Skill: Bs4, request, GCP API, JupyterLab, Colab, SQL, PowerBI

Bluebik Vulcan • Data Engineer 2025 Jan – Current

Cost Optimization

I have been commissioned to optimize costs in Google Cloud Storage (GCS), to reduce storage costs by analyzing data usage patterns and implementing storage policies that align with data access frequency

- Storage Usage Patterns: Reviewed object access logs to identify inactive (cold) data, monitored storage growth and usage trends using Cloud Monitoring and Logging.
- Data by Access Frequency: Categorized data into Hot (frequently accessed), Cold (rarely accessed), and Archived (long-term retention).
- Lifecycle Policies: Configured GCS lifecycle rules to automatically move or delete files based on access time.
- Redundant or Unused Files: Wrote Python scripts to detect and delete duplicate or obsolete files, coordinated with data owners to validate file necessity.

Vector Search

Built a scalable and intelligent search system that combines full-text search and vector search (Hybrid Search) using Google Cloud Spanner as the backend, with data pipeline orchestration via Dagster.

- Built a Dagster Data Pipeline: Developed a pipeline to extract data from CMS-stored tables in BigQuery, transformed and loaded the data into Google Cloud Spanner.
- Spanner Schema for Hybrid Search: Defined schema to store both text fields and vector embeddings (index, partitions), ensured data consistency and optimized schema for hybrid querying.
- Full-text and Vector Search: Enabled full-text indexing within Spanner, integrated vector similarity search using approximate nearest neighbor (ANN) techniques.
- Hybrid Search Logic in App Layer: Combined full-text and vector search results into a unified ranking system, built an API layer to expose the search functionality to client applications.

Key Skill: GCP, GCS, GKE, Dagster, Datahub, BigQuery, Spanner, VectorSearch, Full-TextSearch

PROJECT

ETL with Airflow + Pyspark + GCP + Bigquery

- Pick up a CSV file from a GCP bucket using PySpark, check for duplicates, and clean the data to make it ready for use. Upload the cleaned data to BigQuery, and use Power BI to connect the data to make a visualization dashboard. The entire process is managed using Airflow on Docker.

Complete Project Code: Airflow Project on GitHub [https://github.com/Choppertnt/airflow_project]

ETL with Dagster + GoogleDrive + postgreSQL + DBT

- Pick up a CSV file from Google Drive with google API thought python . Upload data to PostgreSQL , and DBT for Tranfrom data and plot graph . The entire process is managed using Dagster

Complete Project Code: Dagster_Project on GitHub [https://github.com/Choppertnt/dagster_project]

Aws and Databricks

- Use AWS S3 as a Data Lake and Databricks clusters as a Data Lakehouse. Upload a CSV file to AWS S3 and mount it to DBFS. Use Databricks notebooks to generate code and process data with Delta Lake, following the bronze (raw data), silver (clean data), and gold (visualization/dashboard) stages. Finally, create workflows and set schedules to run every day.

Complete Project Code: Aws and Databricks on GitHub [https://github.com/Choppertnt/databricks_project]

Application Ordersales By Flask , Vercel , MongoDB , Pyspark and PostgreSQL

- Flask Web Application: A web form for order details including customer name, phone number, sales date, product, quantity, and price. The data will be semi-structured (JSON-like).
- Deployment to Vercel: You'll deploy your Flask app to Vercel for live access.
- MongoDB Storage: The form data will be uploaded and stored in MongoDB as semi-structured data.
- PySpark Data Transformation: Data from MongoDB will be loaded, transformed to structured data, and cleaned using PySpark.
- PostgreSQL for Storage: Transformed data will be loaded into PostgreSQL, where a trigger is set up to further clean the data.

Complete Project Code: MongoDBproject on GitHub [https://github.com/Choppertnt/databricks_project]

KEY SKILLS

- Python (Google API, bs4 selenium, tkinter, pyinstaller, matplotlib, seaborn, Flask)
- R program (dplyr, dummies, tidyverse, syuzhet)
- Database (PostgreSQL (Trigger, Function), BigQuery, Spanner, SQLite, MongoDB)
- Excel (VLookup, Pivot-table, IF, Query, Google Sheets)
- JavaScripts (APP SCRIPTS)
- Docker, Kubernetes
- GitHub, Vercel
- ChatGPT
- Google cloud platform (GKE, GCS, ETC.)

KEY SKILLS VISUALIZATION

- Looker Studio
- Power BI

KEY SKILLS TOOLS OF DATA

- Data Lake (AWS S3, GCS, MinIO)
- Data warehouse (Databricks, Snowflake, Bigquery, Spanner)
- Data pipeline (Fivetran)
- Data transform (DBT, PySpark)
- Data Orchestration (Airflow, Dagster)

CERTIFICATES/COURSE

2024 | DATACAMP
DATA ENGINEER
- Pass Exam Practice in Data engineer management , Exploratory Analysis

2024 | TRUE ACADEMY
DigiProof Assessment : Basic Data Analyst
- Pass Exam demonstrating their skills and knowledge in key areas for the of role junior data analyst.

2024 | DATACAMP
SQL ASSOCIATE
- Pass Exam SQL in Data management , Exploratory Analysis

2024 | DAGSTER
Dagster&Dbt
- Knowledge Asset , data orchestration and use DBT transform

2024 | DATA TH
ROAD TO Analytics Engineer
- Tools of AE such as databricks , AWS , Snowflake , Dagster , DBT

2024 | Google coursera
Google Advanced Data Analytics
- Learned about data science , python , data insight , statistics

2023 | Google coursera
Google Data Analytics
- Learned about Ask , Prepare , Process , analyze

2023 | Microsoft
Azure AI Fundamentals
- Pass Exam Knowledge of AI Fundamentals

2023 | Microsoft
Azure Data Fundamentals
- Pass Exam Knowledge of Data Fundamentals

2023 | Ultimate python
Web Scraping
- Knowledge how to Extracted data from website