

Mark Wan (万晶)

156-0198-1011 | wanjingjjj@gmail.com | github.com/wanjingjjj

EDUCATION

Shanghai Jiaotong University

Bachelor of Biology

Shanghai

Sep. 2003 – Jun. 2007

EXPERIENCE

Adevinta (ebay classifieds group)

Sr. Staff Engineer(L26)

Shanghai

Nov. 2020 – Present

- Design and maintain the data platform based on Airflow and Spark running on *EKS*.
- Develop and maintain in house Spark version with *AWS* glue, delta.io, multi-arch, and cloud native support.
- Lead and develop in house Notebook solution using *React* and *go* based on Jupyter and VSCode.
- Review and optimize the *AWS* cloud cost to meet the budget.

CraiditX

Data Engineering Manager

Shanghai

Sep. 2018 – Oct. 2020

- Lead the team on developing the data product for bank customers.
- On site deploy data product on customer's on-premise infrastructure.
- Coordination on people resource and communication between multiple projects.

Sr. Data Engineer

Jun. 2015 – Aug. 2018

- Created the data warehouse platform from scratch by using airflow and CDH on *AWS*.
- Created an unified data log service to collect data from site application.
- Created BI dashboard by using Superset. BU also have the flexibility to explore and create new report by self-service.
- Transform data and generate features for machine learning data model applications.

ebay

Sr. Software Engineer(L24)

Shanghai

Mar. 2011 – May. 2015

- Developed and maintained billing analytic data pipelines using *UC4(Appworx)* and *SAS*.
- Maintain and batch support on bi-monthly ebay billing run.
- Developed self-service tools to help BU to retrieve data from billing system.
- Automated workflow for billing operations.

Alibaba Group

ETL Engineer

Hangzhou

Sep. 2010 – Mar. 2011

- Developed and maintained data warehouse platform for *Ali-Finance*.
- Created data pipelines to generate features for online credit loan AI model.

ebay (dx solution contractor)

DW Developer

Shanghai

Jul. 2007 – Sep. 2010

- Build ETL process on *ebay* Teradata platform.
- SQL tuning and batch support on data warehouse data pipelines.

PROJECTS

Notebook Scheduling | *AWS(EKS), Kubernetes, Argo Workflow, React, Go*

Jul. 2023 – Jun. 2024

- Use *Argo Workflow* to run jupyter notebook files recurrently.
- Use *kubernetes operator* to store and manage the schedules.
- Web interface provided to let the user manage all the schedules.

DWH Migration | *AWS(EKS, S3, Glue), Airflow, Spark, Kubernetes, Python*

Jun. 2021 – Jun. 2023

- Use *EKS* as the foundation of the infrastructure, *Airflow* as the workflow orchestration engine, *Spark* as the main process engine. All services are cloud native.
- Created base docker image containing all of the tool chains and ETL handlers, which provided as the default run time of the all ETL jobs.
- Created *Spark* image with patches to connect to *AWS Glue* as hive metastore.
- Migrate more than 100 existing data pipelines from *ebay* in-house platform to this new platform.
- Design and implement the CI/CD service for the new platform.

AI Data Service | *Python, Go, CDH(Hive), Java*

Jun. 2016 – Aug. 2018

- Use *Python / Go* to build backend data service to serve the online ML model.
- Create data logging service to capture the intermedia and final result from ML model, which will be stored on Hadoop.
- Process the data from log, along with other sources offline on Hadoop using hive.

TECHNICAL SKILLS

Languages: Python, SQL (SparkSQL / Hive), Go, Java, JavaScript, HTML/CSS

Frameworks: Spark, Hadoop, Flask, pandas, Matplotlib, Airflow, Argoflow, React, Beego, Springboot

Developer Tools: Git, Kubernetes (EKS), Docker, AWS, Terraform, Datadog, Grafana