

万晶

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

教育背景

上海交通大学

生物技术理学学士

上海

2003 年 9 月 - 2007 年 6 月

工作经历

Adevinta (ebay classifieds group)

资深软件工程师 (L26)

上海

2020 年 11 月 - 至今

- 设计开发基于 Airflow 和 Spark 的数据平台，使用云原生技术运行于 EKS。
- 开发维护内部的 Spark 版本，基于开源社区项目，添加对 AWS glue, delta.io, Arm 架构和云原生的支持。
- 基于开源 Jupyter 和 VSCode 开发内部的 Notebook 工具，用于数据分析与机器学习开发。前端基于 React，后端基于 go。
- 定期复查与优化 AWS 资源使用成本，保证资源能够充分合理地使用。

氮信科技

数据开发部经理

上海

2018 年 9 月 - 2020 年 10 月

- 带领团队开发数据产品，主要服务于银行客户。
- 客户自建数据中心架构上的数据产品部署与维护。
- 沟通与协调多个实施项目的人员与技术方案。

高级数据工程师

2015 年 6 月 - 2018 年 8 月

- 在 AWS 云上从头搭建基于 Airflow 和 CDH 的数据仓库平台。
- 开发维护统一日志采集服务。
- 创建基于 Superset 的 BI 报表服务。销售与分析人员可自助式浏览与创建数据报表。
- 为机器学习模型处理数据与开发特征指标。

ebay

高级软件工程师 (L24)

上海

2011 年 3 月 - 2015 年 5 月

- 使用 UC4(Appworx) 和 SAS 开发维护账务数据分析批处理作业。
- 开发内部自助工具以帮助销售与客服更方便地获取账务数据。
- 自动化部分账户流程。

阿里巴巴集团

ETL 工程师

杭州

2010 年 9 月 - 2011 年 3 月

- 开发维护阿里金融数据仓库平台。
- 为在线小微贷款机器学习模型创建数据流程，开发数据指标特征。

ebay (诚归计算机)

数据仓库工程师

上海

2007 年 7 月 - 2010 年 9 月

- 在 ebay Teradata 平台上开发 ETL 流程。
- 数据批处理作业的 SQL 调优与问题解决。

项目

Notebook 作业调度 | AWS(EKS), Kubernetes, Argo Workflow, React, Go

2023 年 7 月 - 2024 年 6 月

- 使用 Argo Workflow 定时调度 jupyter notebook 文件。
- 使用 kubernetes operator 存储与管理调度定义。
- 基于 React 的前端网页界面，可供用户管理所有调度任务。

数据仓库上云 | AWS(EKS, S3, Glue), Airflow, Spark, Kubernetes, Python

2021 年 6 月 - 2023 年 6 月

- 使用 EKS 作为基础架构，Airflow 作为流程编排引擎，Spark 作为数据处理引擎。所有的服务都为云原生。
- 创建包含所有工具链与 ETL 框架的 Docker 镜像，提供所有 ETL 作业的运行时。
- 创建包含补丁可以 AWS Glue 作为 Hive 元数据的 Spark 镜像。
- 从原 ebay 数据平台迁移超过 100 个数据流程到新平台。
- 为新平台设计与创建持续集成与持续发布流程。

AI 数据服务 | Python, Go, CDH(Hive), Java

2016 年 6 月 - 2018 年 8 月

- 使用 Python 和 Go 创建后端数据服务，供在线机器学习模型调用。
- 统一数据采集服务，将机器学习模型的中间过程与最终结果保存到 Hadoop。
- 使用 Hive 将采集的日志数据与其他数据做处理，供分析模型迭代使用。

技术栈

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

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

开发工具: Git, Kubernetes (EKS), Docker, AWS, Terraform, Datadog, Grafana