Jian Yang

Senior Data Engineer · Shanghai

📞 (+86) 18516191607 | ✉ ernestyj@outlook.com | 🔗 Ernestyj

Summary

Current senior data engineer at Paypal with around 3 years work experience. Developed C++ based large distributed NoSQL database; experienced with C++ developing. Designed and developed financial transaction risk control system, including decision rules analysis, automatic risk control platform, real time transaction data monitoring and alerting; experienced with Python full stack development, data analysis, model building and visualization.

Work Experience

PayPal · Decision Analytics and Management Department
Aug. 2017 - Now

Decision Rule Analytics - Senior Data Engineer
- Cooperated with global analysts and analyzed global transactions data (SQL, Pandas).
- Built Decision Rule Analytics System (Flask, React), including rule parsing (XML), near real time decision data monitoring, rule performance metrics evaluate, decision factors reference analysis (Neo4j). Saved about 80% decision analytics cost.
- Optimized read/write performance on risk big data structural storage (MySQL), supported full data update.

Automatic Risk Control Platform
- Developed unified data preprocessing platform to support Spark ETL from diverse data source (HDFS, Teradata), and feature statistics and feature encoding.
- Built decision tree model by XGBoost, and set up performance metrics for generating accurate risk solutions.
- Built standard SQL to risk solution (IBM decision rules) translate engine.
- Developed loss alerting and risk solutions monitoring web platform by React/AntD.
- The project effectively cut down 2-3 weeks for generating risk solutions, saved about 70% human resources and saved loss about $80k per month. Won Paypal Spot Award (5%).

Realtime Transaction Monitoring and Alerting
- Developed alert action consumption module (Apscheduler) for Risk Decision Engine, and post alert concentration analysis (Pandas, Spark), visualization (Matplotlib) and mail alerting (Jinja2).
- Implemented task chains and asynchronized prioritized task scheduling (Celery, RabbitMQ), improved the alerting service performance and reliability.

Baidu · Department of Infrastructure (INF)
Mar. 2017 - Jun. 2017

Mola Large Distributed NoSQL Database - Research and Develop Engineer
- Research on Mola distributed database architecture, master election on metadata server (Lease), data fragment auto-recover on node server.
- Fixed and enhanced the asynchronized write in data fragment auto-recover process by applying lock mechanism.

Baidu · Commercial Ecosystem Data Product Department
Aug. 2016 - Feb. 2017

Advertise Big Data - Intern Data Engineer
- Monitoring advertise big data from Fengchao and Holmes data source. Assisted in building funnel model for analyzing crowd transferring, flowing, and attribution analysis, to evaluate the effects of advertising and marketing campaigns.
- Assisted in upstream data ETL (Hive), implementing Spark UDF and optimizing ETL performance.

Extracurriculum Experience

Alexa Smart Home Group - Group Leader
Aug. 2018 - Oct. 2018
- Attended Paypal Innovation Lab Smart Home Group, lead to develop and integrate Alexa with Philips smart lamps. Organized innovation day activity and made a speech on smart home technology.

Quantitative Investment Analysis - Algorithm Research
Feb. 2016 - Oct. 2016
- Researched and proposed Maximal Information Coefficient based stock feature selection method, achieving 4.5% higher predicting precision compared with Principle Component Analysis and Singular Value Decomposition method.
- Designed an ensemble classification model named SRA-Voting, gaining 3.125% higher predicting precision compared with Support Vector Machine and Fandom Forest.

Education

SJTU (Shanghai Jiao Tong University)
Sep. 2014 - Mar. 2017

M.Eng. in Software Engineering

UESTC (University of Electronic Science and Technology of China) *Sep. 2010 - Jun. 2014*
B.Eng. in Software Engineering

Skills

Programming	Python, C++, Java, SQL, Shell
Framework	Pandas, Flask, Celery, React, Spark
Tool	Teradata, MySQL, MongoDB, Emacs, Git, Jenkins, Docker, Excel, Tableau
English	CET6, GRE(314), can be used as work language