

潘星星



[Online Version: <https://paxinla.github.io/my-online-resume/en/>]

☎ 137-9448-1350 | ✉ aksura@email.cn | 🌐 <https://paxinla.github.io>

Position : Data Engineer | Salary : Negotiable | City : Shenzhen or Remote

I have many years of professional experience in data development, familiar with the entire data processing process, including data cleaning, data storage, and solving various data problems (data backlog, data delay, data transmission efficiency, etc.). My main work areas include the design and implementation of enterprise data warehouses, ETL pipelines of massive data, design and construction of data processing platforms, design, implementation and maintenance of OLTP database architectures, SQL performance tuning, etc. I efficiently collaborate and coordinate in projects. I have designed and implemented complete data warehouse projects for a china large commercial bank, as well as implemented Big Data solutions for Internet companies.

The industries I have worked in include traditional IT, Internet, and AI companies. I prefer working with teams that are rigorous and practical. I hope to join a company that has long-term business development plans and is committed to creating continuous value for society.

TECHNOLOGY STACK

🦅 Practiced

- ➔ Python/Scala : ★★★★★
- ➔ OLAP / ETL : ★★★★★
- ➔ OLTP : ★★★★★☆

🦾 Competent

- ➔ Prefer Programming Language: SQL, Bash, Clojure
- ➔ Familiar: PostgreSQL, Spark/Hadoop, RabbitMQ, MongoDB, Kettle

> CERTIFICATE

- 🏆 PMP | Project Management Professional
- 🏆 OCP 11g | Oracle Certified Professional 11g
- 🏆 Software Design Engineer | Qualification Certificate of Computer and Software Technology Proficiency

> EDUCATION

- 🎓 2008~2012 Chongqing University of Posts and Telecommunications Computer Science and Technology - Bachelor
- 🎓 2008~2012 Chongqing University of Posts and Telecommunications Communication Engineering - Bachelor minor

> WORK EXPERIENCE

1. Starboard Networks Pte. Ltd. -- Data Engineer , 3 year(s), Since 2021
2. Badou Intelligent Technology (Guangzhou) Co., Ltd. -- Data Engineer , 1 year(s), Since 2019
3. Shenzhen Aipin Information Technology Co., Ltd. -- Data Engineer , 3 year(s), Since 2015
4. Shenzhen Forms Syntron Information Co., Ltd. -- Dev DBA , 3 year(s), Since 2012

PROJECT

【Oct 2021 - Jan 2025】 A Cloud-Native Data Lakehouse Project

- Data Lakehouse | Spark/Databricks

This project involved the establishment and continuous enhancement of a large-scale lakehouse system, providing robust data support for the company's research teams and various Web3 products.

My responsibilities in this project included:

1. Designing, planning, and implementing modern lakehouse and service facilities tailored to the company's business needs.
2. Analyzing and constructing efficient data models and ETL processes, continuously building and improving high-performance data services.
3. Participating in data research and analysis to meet product data requirements and support business decision-making.
4. Summarizing and maintaining technical documentation to promote the development of high-quality products.

The enterprise-level lakehouse system I was responsible for was built on a cloud-native solution using Databricks on AWS. It completely moved away from outdated Hadoop ecosystem tools, offering superior data management and processing capabilities. This solution significantly reduced the company's physical infrastructure and operational costs without increasing vendor lock-in risks. Using this lakehouse, I integrated and processed raw data from blockchain transactions (averaging around one million rows daily) and smart contracts, efficiently delivering high-quality data to the company's various Web3 data products and data science teams.

【Aug 2016 - Jan 2019】 Manage Production OLTP databases

- PostgreSQL | Alibaba Cloud

This project serves as the online database for the company's core product, "X志愿" and is deployed on Alibaba Cloud. It reliably supports the data management needs of the product.

My responsibilities in this project included:

1. Responsible for the design, implementation, daily operation, configuration, monitoring, and troubleshooting of the database system architecture across all environments, including development, testing, and production. In charge of monitoring backup status, developing recovery strategy plans, and creating and executing data backup and recovery plans.
2. In charge of database performance monitoring and tuning, promptly identifying issues and optimizing the way the backend accesses the database, as well as addressing poorly performing SQL queries.

3. Responsible for the logical structure design of databases in projects. This involves analyzing business requirements, designing, reviewing, and creating database objects, as well as maintaining them.
4. Generate and update core datasets required by products in bulk based on business needs, ensuring their quality is maintained.
5. Quickly locate and resolve database-related faults and performance issues.

"X志愿" is one of the company's core data products. It relies on historical enrollment data from colleges and universities nationwide, combined with AI algorithms, to predict the current year's admission outcomes for target school majors based on the personal circumstances of college entrance exam candidates. It provides intelligent and personalized advice for filling out volunteer applications. The company is one of only three volunteer application service providers in the Baidu App nationwide.

The company uses Alibaba Cloud as its cloud service provider, and all production environments are deployed on Alibaba Cloud. I am fully responsible for the design and implementation of the database architecture for this product.

After evaluating the performance and stability of Alibaba Cloud RDS, the company considered various factors such as price, data access characteristics of the "X志愿" product, and the overall product's requirements for database performance and stability. As a result, the decision was made to build a production database cluster based on PostgreSQL on ECS. I designed and implemented the database cluster architecture solution, which has been operating stably without any failures throughout the year. Compared to the previous generation of MongoDB-based database solutions, the new cluster offers significant improvements in read/write performance, development flexibility, feature richness, efficiency of batch data updates, and service stability. I have also automated the backup strategy implementation for this cluster, ensuring the security of product data.

The access to the database by the business system and the growth of basic business data experience a significant surge during the period around the college entrance exams every year, while the workload remains relatively stable during other periods. Investing resources to handle such peak access at all times would undoubtedly be a huge waste. Therefore, I have designed a solution that allows the cluster to smoothly scale up and down before and after business peak periods. During peak hours, the cluster stably supports approximately 8k TPS, while outside of peak hours, it efficiently supports the business in a cost-effective manner.

【Jun 2013 - Aug 2015】 A Business Data Loading Platform for a commercial bank

- Orchestration | Massive Data ETL

This project is a platform for implementing ETL processes and scheduling for various characteristic application systems of the Shenzhen Branch of a large commercial bank. It customizes, processes and delivers data from heterogeneous data sources to the online library of the business system according to the needs of various downstream business systems.

My responsibilities in this project included:

1. Responsible for maintaining and developing the orchestrating and core job procedures of the platform.
2. Design and maintain the ETL processes for data files transferred from the head office and data in the specialized system database on the platform. Handle any platform failures promptly.
3. Customize ETL processes according to new requirements raised by users of different specialized application systems, monitor and ensure the successful completion of daily batch processing.
4. Assist in retrieving historical data for specialized application system users as required.

The project involves multiple heterogeneous data sources belonging to other organizations' systems, and the destination is the application-level data marts and OLTP databases of various departmental specialized business systems. There are certain requirements for the execution efficiency of batch jobs. After me taking over the project, I completed the transformation of most of the existing historical job programs, improving their fault tolerance and the platform's ability to automatically handle errors.

For example, the platform has a source table with approximately 40 million account records, which serves as the data source for many downstream target systems. On quarterly interest settlement days, the significant increase in data volume often led to excessively long runtimes and frequent failures of the loading job (in merge mode), requiring manual on-site monitoring and intervention. After me taking over the project, I optimized the ETL process for this table by automatically splitting and slicing the source data files, analyzing incremental changes, and merging them into the existing snapshot table. This improvement enhanced the overall data loading efficiency and automation level, ensuring that the job runs stably and normally both on interest settlement days and regular days.

Acknowledgement

Thank you very much for taking the time to read my resume. I look forward to the opportunity to work with you.