Qiu Honghao

Male | Age：34 years old | 15802026152 | honghaobeyond@163.com
10 year work experience | Position：Senior Data Engineer

Professional Summary

Graduated from a top-tier university with strong learning capabilities. Over 10 years of software development experi ence, with extensive backgrounds in both the internet and financial industries. Proficient in Java programming and e xperienced in big data and distributed system development. Familiar with data warehouse theory and skilled in Had oop, Hbase, Kafka, Spark, and other big data technologies. Developed data platforms, customer platforms, and user profiles with over ten million user data. Have 4 years of experience as a technical team leader, leading a team of 7-8 members, demonstrating team management and leadership abilities.

Employment History

Beijing DeepZero Technology Co., Ltd. Big Data Technical Manager 2022.08-2024.03

1. Responsible for implementing and developing big data platform products (CDP, MA, AI) for companies like Yuexiu Group, Clubmed, and I.T.
2. Constructed data warehouses based on customer business needs, conducted data governance tasks on cloud platforms such as Alibaba Cloud and Tencent Cloud.
3. Managed the development team, providing technical support, training, and enhancing team members' technical skills and collaboration.
4. Oversaw development task management, setting design, development, and documentation standards, ensuring project delivery with high quality.

Guangzhou DataStory Information Technology Co., Ltd. Development Team Leader 2019.08-2022.05

1. Led the product development team to build a one-stop KOL solution from scratch.
2. Responsible for the business requirements analysis, architecture design, system development, and iterations for media KOL selection products.
3. Formulated project plans, ensuring coding quality, progress, and product maintenance, coordinating resources for timely and high-quality deliverables.
4. Participated in the development, delivery, and maintenance of customized projects for major clients, including A.S. Watson's Data Center and Tencent TEG Data Aggregation Platform.

Chinasoft International（Outsourced to Hang Seng Bank） Senior Software Engineer 2018.03-2019.05

1. Key contributor to the successful launch of the "Hong Kong HKJC Credit Card Upgrade" project.
2. Implemented ARCAD, guiding the team on its effective use and deployment.
3. Participated in the agile development process, contributing to incremental releases through multiple Sprints, and ensuring project efficiency and timely delivery to Hang Seng Bank.

SoftStone Information（Outsourced to Amway China） AS400 Software Engineer 2016.10-2018.03

1. Took charge of the development and maintenance of Amway China's daily necessities inventory system and logistics system.
2. Performed root cause analysis and timely repairs for issues arising in the production environment.

HSBC Software Development Co., Ltd. AS400 Software Engineer 2014.09-2016.09

1. Worked for Asia Pacific ATMP projects, focused on project delivery on ATM & Cards System for HSBC Indonesia.
2. Collaborated with colleagues from Asia-Pacific countries to provide 7x24 fire-fighting support for ATMs .
3. Proficient in RPG/CL programming and passed SCS level AS400 assessment.

Project List

Yuexiu Group Data Center & Customer Center Project Technical Manager 2022.08-2024.03

In this project, we set up a data center, CDP, and MA tailored to the Yuexiu Group's private environment, leveraging Yuexiu Group's data. We integrated diverse customer data from Yuexiu's various sub-sectors, encompassing over 61 million customer records, 200 million user behaviors, and 21 million order details. Using the ONEID engine, we created 19 million ONEIDs and over 700 tags, enabling features like customer profiling, segmentation, and 360-degree analysis. The system also facilitates targeted marketing via SMS, mini-program messages, AI calls, and more.

1. Overall Technical Planning:
   Led the ETL and Java development teams, ensuring the smooth achievement of project objectives.
2. Yuexiu Data Center Development:
   a) Aligned customer needs with solution designs.
   b) Reviewed module designs, data access, and layered designs.
   c) Enhanced data quality, monitoring, and alert systems.
   d) Integrated data center with CDP, optimized data adaptation, ClickHouse writes, and Spark/Hive scheduling.
   e) Implemented a real-time data processing framework using Kafka and Flink for effective customer outreach.
3. CDP & MA Platform Development:
   a) Managed Yuexiu's cloud resources, deploying a microservices-based system on TKE (Tencent Kubernetes Engine), bolstering stability and availability.
   b) Facilitated CDP data integration, ONEID generation, and comprehensive user profiling.
   c) Provided technical expertise for the MA platform, enabling multi-channel customer engagement.
   d) Ensured security compliance through vulnerability assessments and penetration testing.
4. Customized Solutions & Design:
   a) Crafted the CDP's customer homepage, 360° customer views, and tag displays.
   b) Integrated MA with Yuexiu's WeChat Open Platform Third-Party services for targeted messaging.
   c) Developed new contact methods for the MA, including AI calls and mini-program integrations.
   d) Linked CDP with Yuexiu Real Estate's systems for segmented marketing and loyalty programs.
   e) Streamlined the MA approval process and integrated it with a collaborative platform for efficient teamwork.
   f) Introduced a CAS-based SSO portal for seamless authentication.
5. Issue Resolution:
   Troubleshooted and resolved problems related to CDP and MA operations during data processing.

I.T Mini Program Mall Intelligent Recommendation Platform Technical Manager 2023.11-2024.03

1. Assessed and deployed Tencent Cloud resources for the intelligent recommendation platform.
2. Lead data warehouse design, including splitting and assigning ETL tasks to big data engineers, and monitoring, using DolphinScheduler for workflow automation, enhancing data accuracy and processing speed.
3. Optimized yarn clusters and ETL processes to improve data processing efficiency.

DataStory KOL One-Stop Solution Development Team Leader 2020.12-2022.05

With social listening, DataStory KOL solution helps enterprise to reduce cost and increase efficiency for marketing. It covers the whole flow of KOL marketing that including Strategy stage、pre-Assessment、Post - Assessment and Asset precipitation. It consists of multiple apps, such as Strategy Insight, Content Exploration, KOL Ranking List, Post Monitor, Launch evaluation, Historical database and so on.
DataStory KOL database covers all popular social media platforms like Weibo/Wechat/Douyin/Red/Bilibili/ Kwai. The system monitors 1.5 million KOL accounts and stores 2 billion+ KOL posts for nearly 3 years, with daily increment of posts more than 2 million.

Technical selection
Java1.8、Spring Boot、Spring Cloud(Zuul + Nacos)、Spark1.6.3、Yarn、Hadoop2.6.0、HBase0.98、ZooKeeper3.4.5、MySQL5.6、Redis、Elasticsearch5.6.0、MyBatis-Plus、Spark Streaming

Key Responsibilities

1. Responsible for the design and implementation of the entire ETL (Extract, Transform, Load) process, from data acquisition, cleansing, processing to visualization.
2. Take the lead in the core coding work of the ETL process, including:
 a) Implementing the data acquisition layer, utilizing Spark Streaming technology to consume data in real-time from Kafka and storing it in Elasticsearch and HBase databases.
 b) Using Spark Core to tag Key Opinion Leaders (KOLs) in the data intermediate layer.
 c) Optimizing code architecture through design patterns like Singleton and Strategy, improving code reuse and stability.
 d) Encapsulating Elasticsearch query tool classes through custom annotations and reflection mechanisms, facilitating query operations and result acquisition by constructing entity classes.
 e) Developing a permission management system based on the Role-Based Access Control (RBAC) model.
 f) Utilizing Spring AOP to implement class and method-level monitoring aspects for logging, integrating with the company's ELK (Elasticsearch, Logstash, Kibana) log analysis system.
 g) Integrating with the company's proprietary visual DAG workflow task scheduling system.
3. Conduct comprehensive performance tuning of the application from the interface, batch processing, and cluster levels, with specific achievements including:
 a) Interface level: Improving the response time of the KOL real-time information query interface from 8 seconds to less than 1 second; rewriting the content exploration application interface to support billion-level full-text search, reducing data return time from over 1 minute to within 5 seconds.
 b) Batch processing level: Reducing daily batch processing time by 3 hours by addressing Spark data skewness issues and optimizing Spark parameters.
 c) Cluster level: Optimizing Elasticsearch cluster parameters, designing multiple indexes for the KOL content library, and implementing separate storage for hot and cold data.

4. Develop detailed project plans and take full responsibility for the coding quality, work progress, and coding standards of the development team, allocating human resources appropriately to ensure high-quality and timely product delivery.

Taobao Live Analytics Developer 2020.05-2020.08

Taobao Live Analytics aids brands in selecting most suitable live streamers for sales. Our data app analyzes live streams' effectiveness, focusing on traffic, content, and performance. It offers regular and real-time monitoring, visualizing key metrics and trends to enhance sales strategies.
Project Architecture:
Spring Boot, Kafka, Flink, Yarn, Hadoop2.6.0, MySQL5, Elasticsearch5.6.0, MyBatis-Plus

1. Responsible for monitoring the start of live streams by streamers, as well as submitting and scheduling live stream room web crawlers. Based on the MySQL table monitoring KOLs, regularly submit web crawlers for the Taobao Live app to scan for live streams. When a live stream is detected, submit crawlers for collecting live stream details and product information. Finally, write the streamer information, basic live stream details, live stream interaction details, and live stream product information into the Kafka message queue.
2. In charge of real-time message ingestion and data cleansing for the data interlayer. Use Flink to consume Kafka data in real-time, and collaborate with the algorithm team to identify product brands and categories through NLP based on product text information. Write live stream content data, interaction data, and live stream product data into three Elasticsearch databases. Maintain a unified approach for real-time and offline data for downstream business use.
3. Utilize Flink's window functions to capture live stream content data every minute (with crawlers writing an event every 5 seconds). Calculate trend indicators such as the number of people entering the live stream per minute, new followers, comments, etc., and store them in Elasticsearch.
4. Responsible for developing microservices interfaces for the real-time dashboard web end. Develop interface programs to aggregate and analyze the aforementioned data, format and return it to the front-end dashboard for display, including various trend charts and cumulative statistics.

Watsons Data Center Project Developer 2020.01-2020.05

The Watsons Data Center Project leveraged the power of a big data platform to establish a centralized data hub, featuring an enterprise-level data warehouse, SSO single sign-on integration, VMI dashboard and reporting implementation, APIs for VMI sales and inventory data, VMI permission integration, and real-time access to order and inventory data, which unified data sources and metrics.

1. Involved in building Watson's data warehouse using dimensional modeling, organizing data into five layers for efficient storage and retrieval.
2. Participated in the construction of the data collection platform, synchronizing data from multiple heterogeneous data sources to the ODS layer of the data warehouse from 23 systems and 180+ data tables, with a total of 5TB of original business data.
   a. User behavior data synchronization: Sonsors Kafka data source -> Flume -> HDFS -> Hive
   b. Offline full and incremental synchronization of business data: Sqoop script/Haina -> HDFS -> Hive
   c. Real-time synchronization of business data: Oracle -> OGG -> Kafka -> Spark Streaming or MySQL -> Binlog -> Canal -> Kafka -> Flume -> HDFS -> Hive
3. Responsible for the development of the member domain and member theme, including the construction of the member DIM table, the ADS layer's indicators such as the number of new users in the past N days, the total amount

of orders, the number of order users, and user retention rate, etc.

4. Participated in data governance to reduce redundant metric construction and improve data quality.

a. Model governance: Developed according to the standard ODS -> DWD -> DWS -> ADS data flow.

b. Specification governance: Used common warehouse word roots and followed naming conventions for databases, data tables, and cleaning specifications at each data layer.

c. Metadata governance: Utilized Atlas for metadata categorization, search, viewing data lineage dependencies, and other unified management tasks.

5. Use Presto to enable ad-hoc queries, and support users in generating corresponding statistical reports.

6. Schedule tasks on Haina to achieve visualization of workflows.

Education

Guangdong University of Foreign Studies Bachelor English(Information Management) 2010-2014

Skills

- Skilled in Java,Spring Boot, Spring MVC, MyBatis-Plus.
- Deep understanding of cloud-native tech, microservices, containers, CI/CD.
- Skilled in Kubernetes, deploying and managing complex apps.
- Experienced with Tencent Cloud and Alibaba Cloud.
- Proficient in Spark memory management, task execution, and RDD development.
- Knowledgeable in data warehouse modeling, Hive optimization.
- Familiar with Flume, Sqoop, Kafka, Zookeeper.
- Deep understanding of Redis, MySQL, Elasticsearch principles.
- Skilled in AS400 development, RPG, SQL, CL languages.
- Years of experience managing technical teams.