Wenqi Jiang
Data Engineer

(+34) 691-752-157
Madrid, Spain
jiangwenqi1995@gmail.com

jiangwenqi.info
github.com/jiangwenqi
linkedin.com/in/jiangwenqi

Having 3 years of experience as a software engineer in big data infrastructure and backend services, mainly focused on specializing in scalability, high availability, and low latency distributed systems design and implementation. Proficient in Java and Scala.

SKILLS

Techniques	Java, Scala, Hadoop, Kafka, Hive, HBase, CDH, Spark, Flink
	Python, Elasticsearch, MySQL, MongoDB, Docker
Languages	English (B2), Mandarin (Native)

WORKING EXPERIENCE

Ontology Engineer | OEG-UPM　　　　　　　　　　　　　　　　　　　　**Jan. 2022 - Jul. 2022**
LOOM-LD　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　*Madrid, Spain*

- Analyzed ontologies from different data sources, like DBpedia, Wikidata;
- Specialized 7 linking discovery algorithms, including string-based similarity and geometry relationships;
- Implemented a Sparql-based linking application by using Apache Jena, and evaluated it with the OAEI-2021 datasets;
- Written the master's thesis.

Big Data Engineer | TOGIC　　　　　　　　　　　　　　　　　　　　　　**May 2019 - Aug. 2021**
Offline Data Warehouse　　　　　　　　　　　　　　　　　　　　　　　　*Shenzhen, China*

- Maintained the Hadoop cluster containing over 200 TB of data, including user and system logs from 3 different sources;
- Upgraded big data components, like Hadoop, Spark, Kafka, etc., and monitored them on CDH;
- Designed and implemented a four-layer data warehouse based on the STAR model by using Hive, which had near 150 tables and received over 300 GB per day;
- Rewrote the ETL pipeline, which saved over 80% of storage and fixed the small files' problem on HDFS.

Data Statistics System

- Restructured the statistics system based on the new offline data warehouse;
- Rewrote the code by using Spark DSL, making it more readable and maintainable;
- Optimized the system's performance, reduced the process time by 25% compared to the previous version;
- Implemented more than 30 new statistical demands and fixed bugs.

Log Collection Service

- Designed and implemented the log collection pipeline using Nginx and Netty, improving performance, stability, and scalability;
- Processed all logs into the Kafka cluster, which has 5 brokers, 3 replications, and 7 days of redundancy, improving robustness.

Backend Engineer | WEBSTUDIO　　　　　　　　　　　　　　　　　　　**Jul. 2018 - Apr. 2019**
- Implemented backend APIs of big data visualization dashboards and general reports for the BI applications;　　*Beijing, China*
- Developed more than 25 data operators, such as summation, date conversion, string processing etc.;
- Implemented the one-click function of syncing all user information;
- Expanded 16 data sources, including MySQL, SQL Server, MongoDB etc.;
- Maintained the big data cluster;
- Fixed bugs and developed new features for the previous projects.

Software Engineering Intern | KEG-THU　　　　　　　　　　　　　　　**Feb. 2018 - Jul. 2018**
- Organized and annotated critical illness insurance documents;　　　　　　　　　*Beijing, China*
- Trained models to extract the logical structure of insurance documents by using GROBID;
- Created a desktop application using above models;
- Maintained datasets in NoSQL database, like MongoDB, Neo4j;
- Completed the thesis.

OPEN SOURCE CONTRIBUTIONS

Dapp Learning DAO　　　　　　　　　　　　　　　　　　　　　　　　　**Apr. 2022 - Present**
- Contributed and translated various tutorial tasks;
- Shared the latest blockchain news and cutting-edge technics in the community.

EDUCATION

MS. Data Science · *Polytechnic University of Madrid*	2021 - 2022
BE. Computer Science and Technology · *Beijing Information Science & Technology University*	2014 - 2018