

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

# Wenqi Jiang

## Data Engineer / Java Developer

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

My name is Wenqi Jiang. I'm a software engineer with 3 years' experience in **big data infrastructure** and **backend services**, specializing in scalability, high availability, and low latency distributed systems design and implementation. I am proficient in Java and Scala and I have studied React and Flutter to find interesting challenges. Furthermore, I enjoy powerlifting and traveling in my spare time.

### 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, which generated 300 GB per day;
- Upgraded big data components, including 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 over 150 tables;
- Rewrote the ETL process, 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;
- 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.

Beijing, China

#### Software Engineering Intern | KEG-THU

Feb. 2018 - Jul. 2018

- Organized and annotated critical illness insurance documents;
- 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.

Beijing, China

### EDUCATION

MS. Data Science · Polytechnic University of Madrid

2021 - 2022

BE. Computer Science and Technology · Beijing Information Science & Technology University

2014 - 2018

### 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.

### SKILLS

Programming Languages    Java, Scala, JavaScript, C++, Python, Solidity

Frameworks    Hadoop, Spark, Flink, Kafka, Hive, Jena, Spring Boot, React, VUE, Web3.js

Languages    English (B2), Mandarin (Native)