Wenqi Jiang

Data Engineer / Backend Engineer

About Me

Hello, my name is Wenqi Jiang.

I'm now studying a master's degree in **Data Science** at the Polytechnic University of Madrid (UPM) while working as an **ontology engineer** at Ontology Engineering Group (OEG).

Before that, I worked in mainland China for over three years as a **big data engineer** and **backend engineer**, so I have experience in handling Tbs data, and I am most skilled in: `Spark` / `Flink` / `Elasticsearch` / `Spring` and `Java` / `Scala` / `Python`.

Besides daily work, I am constantly looking for new challenges, I learned to `React` / `Flutter`. Recently, I have been working on **Web 3.0** in my spare time.

Category	Techniques
Big Data	Kafka, Hadoop, CDH, Sqoop, HBase, Hive, Impala, Spark, Flink, Elasticsearch
Backend	Maven, Gradle, Spring Boot, Dubbo, MyBatis, Hibernate, FreeMarker, Mysql, Redis, Nginx, Tomcat
Data Science	NLP, Ontology, Statistical Analysis, Machine Learning, Deep Learning, Time Series Mining
Frontend & Mobile	React.js, VUE.js, Scrapy, Flutter
Blockchain	Dapp, Solidity, Web3.js, Hardhat, Alchemy, Infura
DevOps	Docker, Github Action
Programming Language	Java, Scala, JavaScript, Python, Solidity, C++, R

Contact

- Mobile: KzM0IDY5MS03NTItMTU3
- Email: jiangwenqi1995@gmail.com
- Online Resume: jiangwenqi.info
- GitHub: github.com/jiangwenqi
- Blog: dev.to/jiangwenqi

Experience and Projects

Ontology Engineer (Scholarship)

OEG-UPM | *Dec. 2021 - Present*

LOOM-LD

Feb. 2022 - Present

A student **scholarship** for the development of linking algorithms and graph linkers.

- Analyzed the **ontologies** in different data sources;
- Specification and development of **linking discovery** algorithms, including text similarity and geographic relationship;
- Development of a `Sparql-based` knowledge graph linker using customized `Apache Jena` functions;
- Evaluation of the algorithms with **OAEI-2021** datasets;
- Complement of the master's thesis.

Keywords: `Ontology Graph Linker RDF Text Similarity DE-9IM WKT SPARQL Apache Jena OAEI`

Big Data Engineer

TOGIC | *May 2019 - Aug. 2021 · 2 yrs 4 mos*

Offline Data Warehouse

Dec. 2019 - Jun. 2020

- The data warehouse contains over **200 TB** of data and more than 150 tables, including user activity logs and system logs from the Togic App and Webox System, which generate over **300 GB** of data every day.
- The data warehouse structure is **STAR** modelling and is organized into four layers: ODS, DWD, DWS, and ADS, which saved over **80%** storage space and eased the small files' problem on `HDFS`.

Keywords: `CDH Kafka HDFS HIVE Spark`

Data Statistics System

May 2019 - Mar. 2021

- The statistics system was **restructured** based on the new Offline Data Warehouse, and almost all big data components **upgraded**, including `HDFS`, `Spark`, `Kafka` etc.;
- The code was rewritten by using **DataFrames**, and was more readable and maintainable, which **reduced the process time** to 1/4 of the previous one.
- Implemented more than **30 new statistical items** and generated reports;
- Fixed the bugs that the former system caused.

Keywords: `Sqoop Hive Spark SQL Scala Elasticsearch MySQL MongoDB Python`

Log Collection Service

Sept. 2019 - Sept. 2019

- **Restructured** the log collection service to improve performance and stability;

- Used `Nginx` and `Netty` to receive and forward logs, and handle error data;

- Produced all logs into the message queue of `Kafka`, which has five brokers, three data backups, and seven days of log redundancy to **improve the robustness** of this service.

Keywords: `Nginx Netty Java Kafka HDFS`

Backend Engineer

WEBSTUDIO | *Jul. 2018 - Apr. 2019 · 10 mos*

- Implemented backend APIs of big data visualization dashboards and general reports for the `BI` applications;
- Developed more than 30 **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.;

- Built and maintained the big data platform;
- Fixed bugs and developed **new features** for the previous company's projects.

Keywords: `Dubbo` `Spring Boot` `MyBatis` `Hibernate` `Java` `Nginx` `MySQL` `Redis` `Docker` `FastDFS` `CDH` `Hadoop` `HBase` `Impala` `Spark ML`

Software Engineer (Internship)

KEG-THU | *Feb. 2018 - Jul. 2018 · 6 mos*

- 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.

Keywords: `NLP` `CRF` `GROBID` `Neo4j` `MongoDB` `Java FX`

Open Source

Dapp-Learning

Dapp Learning DAO | *Apr. 2022 - Present*

- Contributions of various tutorial tasks;
- Translations of the basic tasks;
- Shares of the latest blockchain news and cutting-edge technics in community.

Keywords: `Blockchain` `web3.js` `Solidity` `Dapp` `Ethereum` `Hardhat` `Alchemy` `Infura` `DAO`.

Education

MS Data Science

Polytechnic University of Madrid | *2021 - Present*

- Big Data / Cloud Computing and Big Data Ecosystems Design;
- Data Processes / Information Retrieval, Extraction and Integration / Time Series Data Mining;
- Statistical Data Analysis / Machine Learning;
- Open Data and Knowledge Graphs / Intelligent Systems / Deep Learning;
- Introduction to Research Methodology / Ethical, Legal and Social Aspects etc.

BE Computer Science & Technology

Beijing Information Science & Technology University | *2014 - 2018*

- Principles of Computer Compositions / Computer Architectures;
- Compilation Principles / Practice of Embedded Application System;
- Operating System / Principles and Applications of Database;
- Advanced Math / Linear Algebra / Discrete Mathematics / Probability and Statistics;
- Programming Languages: C, C++, Java;
- Data Structure, Algorithm, etc.