

# ASHWIN CHAFALE

(213) 551-4836 | [chafale@usc.edu](mailto:chafale@usc.edu) | Los Angeles, CA – 90007  
<https://www.linkedin.com/in/ashwin-chafale> | <https://github.com/ashwinchafale>

## SUMMARY

Senior Software Engineer with **4+ years** of experience building data-intensive applications and tackling challenging architectural and scalability problems in the field of Health Science, Finance, and E-commerce. My interests lie in System Design, Machine Learning architecture, and Distributed Parallel Computing, and I am looking for an opportunity to work in core software development, ML/Data engineering, and related fields.

## EDUCATION

<b>University of Southern California, Los Angeles</b> Master of Science, Computer Science	Aug 2021 - May 2023
• <i>Coursework:</i> Algorithms, Machine Learning, Natural Language Processing, Databases, Web Technologies & Mobile Apps Development, Information Retrieval & Web Search Engines, Operating Systems	
<b>RCOEM, Nagpur - India</b> Bachelor of Engineering, Computer Science & Engineering • <b>Ranked 6<sup>th</sup></b> in the CS department out of over 1100 students • <i>Coursework:</i> Data Structures and Algorithms, Distributed Database Management, Data Warehousing, & Data Mining, Web Intelligence & Big Data	August 2013 – May 2017

## SKILLS

**Languages:** Python, Java, JavaScript, Scala, C/C++, SQL, Unix-shell scripting

**Front-end:** HTML, CSS, SwiftUI, React.js, Dart, Flutter,

**ML Libraries:** PyTorch, TensorFlow, Pyspark, Keras, NLTK, NumPy, Pandas, Scikit-learn, Matplotlib, Plotly

**Frameworks & Libraries:** Spark, Airflow, Flink, Kafka, Node.js, Spring Boot, RESTful web services, GraphQL, Flask

**Databases:** MySQL, Oracle, PostgreSQL, MongoDB, Cassandra, Amazon Redshift

**AWS Cloud Services:** EC2, RDS, IAM, S3, AWS Lambda, Athena, Glue, Step-Function, Sagemaker

**Tools:** Git, Docker, Kubernetes, AWS Cloud, Unity 3D

**Mobile Development:** iOS Development, Swift, Dart

## WORK EXPERIENCE

### Apple Inc.

*Data Engineering Intern – Research Health Software Team*

California, USA

May 2022 – August 2022

- Developed and maintained aggregated **ETL pipelines** using **Spark SQL** and **PySpark** on Hadoop file systems as part of Apple's Health Software team.
- Build an **end-to-end data application** that would allow QA/QE teams to validate the user data or detect anomalies in the user data streams uploaded by the research health app.
- Processed 10 TB of large data in a distributed parallel computing architecture and extracted meaningful features from it to promote data analytics
- Responsible for **scheduling, maintaining, and orchestration of data pipeline** using Apache Airflow
- Skills:** **Data Modeling, Data Analytics & Visualization**, Workflow orchestration, and ETL processes

### Amazon.com

*DevOps Contractor – AWS Codeguru Team*

Remote, India

April 2021 – June 2021

- Implemented **static code analysis** recommendation rules in **Java** for python files to improve code quality, application performance and detect code smells in repositories. The detection percentage of the python rules was 78 % in the amazon GitHub repository.
- Incorporated unit test in **JUnit** cases to validate python rules
- Led junior software engineering interns and collaborated with the Amazon Codeguru Seattle team for proper implementation and testing of python rules

### Persistent Systems

*Senior Software – Data Engineer*

Maharashtra, India

August 2020 – April 2021

- Implemented **PII Data Masking** and Obfuscation Pipeline using Python on client databases. Optimized data masking pipeline performance by 50% and obfuscated 450 GB of large data.
- Ingested data from various data clusters using a combination of **AWS Athena** and **AWS Glue** to create data views to be used in **BI** tools.
- Worked closely with ML solution architects for proper implementation of data-preprocessing pipeline and to derive insights from the data which could be further useful for the ML models.
- Responsible for the creation of various dashboards on **Tableau** for data analytics.

### Senior Software – ML Engineer

December 2019 – August 2020

- Responsible for deploying Facebook's Deep Learning **NMT Transformer** model on torch-serve (open-source tool for serving **PyTorch** models) and created generalized custom handlers for all transformer models enabling fast and easy deployment of the model
- Responsible for deployment and testing of various **Pytorch** models on container technologies like **Docker & Kubernetes**
- Acted as a liaison between teams in Amazon-Sagemaker and Facebook ML teams for the proper deployment of the transformer model.

### Software Engineer – Full Stack Developer

October 2017 – December 2019

- Implemented several time-critical features while migrating IBM users from mainframe-based to a **distributed computing architecture**
- Consistently deliver structured and maintainable **front-end** codes to build an intuitive user interface and reusable components on the **React.js**
- Implemented REST services by **Golang** with microservices architecture.
- Excellent knowledge of **NO-SQL** on Mongo and Cassandra DB and expertise in working with SpringIOC, SpringMVC, and SpringAOP
- Contributed to developing an event-driven and time-triggered report generation module using Java and Spring
- Leveraged Apache **Spark** for processing and aggregating high-volume historical data for generating reports
- Interacted with clients across various global locations like the USA, Canada, China, Australia
- Leveraged knowledge in Java, J2EE, Spring-MVC, Spring Boot, Golang, JavaScript, HTML, SQL, DOJO, Node.js, React.js

## **PROJECTS**

---

### **Digital Music Service Provider with Song Recommendation** | HTML/CSS, React.js, Python, MySQL, Swift

- Implemented the functionality to stream music, build personalized playlists and search by artist/song/album name
- Developed a song recommender system built using item-based collaborative filtering

### **Text Classification via Sentiment Analysis** | Python, PyTorch, NLTK, Gensim, Numpy, Pandas

- Performed text preprocessing on 4.2M Amazon Reviews and generated self-trained Word2Vec and TF-IDF features.
- Devised 20 different Perceptron, SVM, FFN, MLP, RNN & GRU models with accuracies > 0.82 on binary & multi-class classification.

## **PUBLICATIONS**

---

- “Are we spreading Reach or Hate? Towards Generating Contextually Relevant Sequence of Reactions for LinkedIn”. (Ashwin .C, et al.) EACL 2023 Submitted – ongoing evaluation.

## **ACHIEVEMENTS**

---

- Consecutively awarded **Top-Talent of the Year** award by Persistent Systems for the years 2020 and 2021
- Awarded Merit Rank Holder and **Merit Scholarship** by RCOEM, India – August 2014