

Mohan Shashanka Chamарthy

+1 331-980-7684 | mschamarthy@gmail.com | linkedin.com/in/mschamarthy | Santa Clara, California

SUMMARY

Professional Data Engineer with 8+ years experience in Data modeling, Product Analytics, development of ETL pipelines, distributed systems using Python, SQL, Java, Hadoop, Hive, Spark, Presto, Kafka, Spring boot and AWS. Expert in producing datasets aiding product analytics and growth analytics. Masters' in Computer Science with competence in machine learning & big data application development. Specialized in data analytics, machine learning & cloud computing. Self-motivated professional with experience in leading multiple software development projects spanning multiple cross-functional development teams.

SKILLS

Big Data Technologies	: Presto, HiveQL, Hadoop, Apache Spark, Kafka, Oozie, SOLR, Apache Crunch, Map Reduce, Vertica DB, HDFS, HBase, Hue.
Programming & Web	: Python, Java, Scala, C, Bash, HTML/CSS, JavaScript, Ajax, JQuery, MongoDB, AngularJS, ReactJS, NodeJS (MEAN stack), PHP, WordPress.
Frameworks	: Hibernate, Spring (core, MVC, Boot, JPA), Jersey, JavaFX, Guice, Django, Ruby on Rail.
Machine Learning Tech.	: NLTK, scikit-learn, NumPy, Stanford POS Tagger & NER, Text Mining, word2vec.
Others	: Maven, Docker, AWS (EC2, EMR, S3, SQS, SNS, Lambda, RDS, CloudFormation, CloudWatch, ALB), MATLAB, GitHub, JIRA, SQL, Jenkins, Splunk, Tableau.

EXPERIENCE

META PLATFORMS INC.

Senior Data Engineer

5th Oct. 2020 - Present

- Re-designed the data warehouse data model for a user-facing product that boasts over 2 billion monthly active users resulting in creation of a robust, privacy compliant and highly secured data product (Datasets, dashboards and experimentation setup) enabling privacy safe environment for data-informed decision-making, enhancing our product insight capabilities and product's performance. This enabled effortless data discovery, exploration and self served analytics for cross functional teams.
- Spearheaded implementation of an agile, python-based product analytics suite that enabled rapid development of data assets to support product insights and experimentation. This pioneering effort led to a remarkable 75% reduction in development time, streamlining the process to just 1 week per feature and empowering our teams to drive innovation at an accelerated pace.
- Designed and implemented ETL pipelines in Spark, Presto and python to produce datasets for product analytics and growth analytics.
- Implemented data pipelines and built dashboards to visualize key product growth metrics like DAU, MAU, conversion metrics and Growth accounting metrics.
- Proposed, designed and built datasets for Retention analysis, channel analysis and funnel analysis of Cross Family Growth products.
- Implemented product experimentation metrics and build framework that enable A/B testing of UI changes and Funnel improvements.
- Design and revamped existing user behavior datasets to enable privacy compliant growth analytics without loss of existing analytical capabilities and historical data.
- Partnered with cross functional teams to build appropriate metrics and surface product improvements that helped grow product users by 220 Millions (i.e. 24%) in 1.5 Years.
- Optimized Spark ETL pipelines that processed 20+ Billion records that led to a runtime reduction from 8 hours to 40 mins (improvement of 91%)
- Mentored 6 Data engineers new to Meta, Managed an intern and onboarded 4 Senior engineers joining the team.

CERNER CORPORATION

Team Lead | Senior Software Engineer

11th Nov. 2018 – 2nd Oct. 2020

Associate Senior Software Engineer

8th Jan. 2018 – 11th Nov. 2018

- Designed and implemented ETL pipelines to produce data extracts from data lakes and load into data warehouses.

- Implemented high-quality ETL pipelines in Java, Guice, Hadoop and Apache Crunch that are scalable to process 100+ TB of datasets.
- Configured and optimized Map-Reduce jobs to improve execution time and resource utilization of data pipelines by more than 50%.
- Containerized REST services written in Java, MySQL & Spring Boot using Docker.
- Configured and deployed various parts of the software system into the cloud using various AWS services like EMR, S3, Route 53, ALB, IAM, RDS, CloudFormation, SNS.
- Architected and implemented a highly scalable multi-tenant distributed SaaS system using Kafka, Java, Spring boot.
- Proposed, designed and led implementation of a new framework to improve maintainability and reusability using multiple design and architectural patterns reducing overall development time of pipelines by 3+ months.
- Produced designs for various enhancement features using object-oriented design principles & software architectures.
- Proactively automated many manual steps to reduce human errors and reduce support time using Spring Boot & Java.
- Organized knowledge transfer sessions for the organization to discuss design patterns and new big data technologies.
- Lead, designed & implemented multiple (10+) high risk, high impact projects successfully with on-time delivery.
- Managed team of 4 software engineers with responsibility for development and maintenance of software solutions.
- Onboarded 3 software engineers by training them in software design, programming, testing & various technologies.
- Mentored 5 teams in design and implementation of POC software projects to build new SaaS platforms.
- Received 'Rising Star' award for innovation in workflow automation leading to resilience of the product.
- Received 'Night on the Town' award for quick resolution of complex support issues which had large financial impact.

SYMMETRY BIOSCIENCES INC.

Client: Capital One

Software Engineer

19th Jan. 2017 – 5th Jan. 2018

- Proactively designed, architected & implemented a system for managing and running Spark jobs on transient EMR using Python, AWS Lambda, AWS SNS which is widely adopted by teams across the enterprise.
- Architected and implemented automated remote triggering of various stages in a distributed application system.
- Implemented distributed job monitoring and logging tools for the application using MySQL & AWS RDS.
- Developed Hydrograph (open source tool) based ETL workflows for data filtering, transformation & migrations.
- Implemented custom maven plugins for automated artifact packaging and deployment to nexus.
- Automated the Infrastructure setup of various AWS resources including EC2, EMR, Lambda, SNS, SQS, etc.
- Designed and implemented micro services using Kafka connectors to perform data collection into the application.
- Automated code deployment into various environments using GitHub, Jenkins, Artemis.
- Mentored an Intern with design and development of Workflow visualization web app.

CERNER HEALTHCARE SOLUTIONS PVT. LTD.

Software Engineer

2nd Jun. 2014 – 31st Jul. 2015

- Received an award for developing the in-house knowledge base and team competency on Apache Crunch.
- Designed and developed Hadoop analytical pipelines for Key Performance Indicators (KPIs) for Revenue Cycle.
- Implemented high-quality ETL pipelines in Java, Hadoop and Apache Crunch.
- Developed dashboards for visualizing KPIs using JavaScript, Ruby on Rails, D3.js.
- Proactively **improved runtime** from hours to minutes by business logic optimization for Hadoop cluster.
- **Trained 6 associates** on functional and technical aspects including Hadoop, Apache Crunch, GitHub, SOLR.
- Released artifacts and deployed REST services, web servers and ETL pipelines using Chef.

EDUCATION

ILLINOIS INSTITUTE OF TECHNOLOGY

Aug. 2015 – May 2017

Masters of Computer Science, Specializing in Data Analytics.

Graduation Date: 13th May 2017

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

Aug. 2009 – May 2014

Bachelor of Technology (Honors) in Electronics and Electrical Communication Engineering

*Master of Technology in RF and Microwave Engineering (Dual Degree Program). **Graduation Date: 26th July 2014***

STUDENT ORGANIZATIONS & EVENTS

TEDxIIT

Advisor, Web-Master Committee

Jan. 2017 – Present

Co-Chair, Web-Master Committee

Jan. 2016 – Jan.2017

- Responsible for mentoring, knowledge transfers and guiding the webmaster committee.
- Integrated real-time feeds from Facebook with the site using Facebook API, JavaScript and Ajax.
- Designed and implemented URL shortening system to create a dynamic subdomain forwarding system.
- Added information extraction techniques to Twitter, Instagram & Facebook feeds to generate dynamic word clouds.
- Maintenance and upgrading of the domain www.tedxiiit.com using JavaScript, PHP, SQL, HTML5/CSS, WordPress.
- Managed the access and permission for various users of the site.

PROJECTS

YELP REVIEW RATING PREDICTION USING CONVOLUTIONAL NEURAL

Jan. 2016 – May 2016

- Trained on 1 million reviews using a convolution neural network to achieve an accuracy of 61%.
- Generated word2vec model using the 1 million reviews. Vectorized the reviews using the trained word2vec model.
- Developed the convolutional neural network using Theanet library(with custom modifications for rectangular filters)
- Trained the CNN on a cluster of GPUs by setting up Theanet on EMR of GPU intensive nodes.

DEEP CONVOLUTIONAL NEURAL NETWORKS FOR SENTIMENT ANALYSIS OF SHORT TEXT

Jan. 2016 – May 2016

- Analyzed 2 million classified tweets using a convolution neural network to achieve an accuracy of 85%.
- Generated character level and word level embeddings using char2vec & word2vec.
- Derived fixed length Sentence vectors by combining character & word embeddings using unsupervised CNN modeling trying to derive the context from the otherwise context-less tweets.
- Derived sentiment values from the sentence vectors using 2 layer fully connected feed-forward neural network.

LIBRARY USER AND BOOK MANAGEMENT SYSTEM

Dec. 2015 – Dec. 2015

- Designed and implemented a configurable Java application using MVC in JavaFX and MySQL database schema.

AUTOMATIC TEXT SUMMARIZATION SYSTEM USING LEXICAL CHAINS

Nov. 2015 – Nov. 2015

- Performed Lexical chain analysis on a text article to extract the summary of the article.
- Used Stanford POS tagger for tagging words in a sentence and generating lexical chains for each noun to identify the most relevant noun representing the Article.
- Extracted top 5 sentences from the article that convey the most information about the article.

ACHIEVEMENTS

- Secured All-India 531 rank/300,000 students in Joint Entrance Examination for Indian Institute of Technology (2009)
- Achieved **3rd place** in the campus round of ITA Tech Challenge coding competition in September 2015.
- Won a **3rd place** in the robotics competition for developing a Smart Classroom system with Braille Printer.
- Active contributor, leader and a key member of the teams that developed **5 prototype robotic** systems over 5 years.
- Spearheaded a team of 15 as a **Captain** in the 2012 – 13 Hardware modeling competition.

Keywords: Data Engineer, Python, Java, Presto, Spark, Hadoop, Kafka, AWS, Data Mining, Machine Learning.