MANAS CHAKKA

LinkedIn: manas-chakka | Github: mchakka
Cumming, GA 30040 | 678-538-8330 | mchakka@gmail.com | US Citizen

EDUCATION

GEORGIA INSTITUTE OF TECHNOLOGY - MAJOR GPA: 3.78

B.S. COMPUTER SCIENCE
Expected Graduation: **Dec 2020**
Concentrations: **Intelligence** & **Information Internetworks**

SKILLS

PROGRAMMING
JAVA | PYTHON | SQL | SCALA | C
OBJECTIVE C | ASSEMBLY | BASH

DATA
SPARK | HBASE | HADOOP | R | KERAS
PANDAS | SCI-KIT LEARN

FRAMEWORKS
FLASK | DJANGO | REACTJS

SERVICES / TOOLS
AWS | JENKINS | DOCKER | AZURE |
GIT | RESTFUL SERVICES

RELEVANT COURSEWORK
OPERATING SYSTEMS AND NETWORKS
INFORMATION SECURITY
ARTIFICAL INTELLIGENCE
INTERNET SYSTEMS AND SERVICES

CERTIFICATIONS
AWS SOLUTIONS ARCHITECT
MACHINE LEARNING SPECIALIZATION

LEADERSHIP

ALPHA KAPPA PSI PROFESSIONAL CO-ED BUSINESS FRATERNITY

VP OF INTERNAL AFFAIRS
APR 2019 – APR 2020
Served on 8-member executive board to organize weekly chapter and campus initiatives across a 70+ brotherhood

PHILANTROPY DIRECTOR
DEC 2018 – APR 2019
Planned over 10 philanthropy events involving professional and business skills during the semester in local communities

WORK EXPERIENCE

CLOUDERA | SOFTWARE ENGINEERING INTERN   MAY 2020 – AUG 2020
HBASE TEAM **|** JAVA | PYTHON | SCALA | SPARK | JENKINS   SANTA CLARA, CA (REMOTE)

- Evaluated read & write efficiency of PySpark SQL on HBase Tables through 10+ regression testing scripts to provide customers an alternative means of access to HBase through Python.
- Built feature parity matrix to summarize current limitations between HBase Java Client API and PySpark while integrating support for 6 new HBase DML bulk operations from PySpark.
- Designed solutions for customers to use CDSW (Cloudera Data Science Workbench) and PySpark on data stored in Cloudera Operational Database to train ML/DS Models in real-time.

KPMG US LLP | BIG DATA SOFTWARE ENGINEERING INTERN   JUN 2019 – AUG 2019
LIGHTHOUSE DIVISION **|** PYTHON | SPARK | PANDAS | SQL | KERAS | BASH   ATLANTA, GA

- Automated parallel-processing pipelines to extract data from 1.5 million scanned jail documents to SQL database. Built CNN image classification model to sort scanned documents into 4 categories using Keras and Google Tesseract OCR.
- Created Regular Expression rules per category of documents to fix OCR errors along with rule-checking validation metrics. Constructed a scalable file list tracking database to eliminate possible data-loss throughout the entire data pipeline.

KALMED HEALTHCARE | SOFTWARE ENGINEERING INTERN   AUG 2017 – AUG 2018
JAVA | PYTHON | AWS SAGEMAKER | REACTJS   ATLANTA, GA

- Utilized hospital sample blood data to build regression models and visualizations for computing the probability of a patient having a blood-related disease with a 97% precision rate.
- Designed infrastructure of a cross-platform server storage system to update and display health information of a patient from the health kiosk to the patient's personal account.

PROJECTS

SLAPCAST | CREATE-X CAPSTONE   SEPT 2019 – NOV 2019
PYTHON | FFMPEG | JAVA | ANDROID STUDIO

- Processed video clips and extracted QR codes with object detection from frames and video metadata to create a local video file hierarchy structure.
- Constructed mobile application that relayed QR Code information to and from a web application using REST APIs for scalability and portability.

STAMPEDE | HOME DEPOT 2018 CODEATHLON TOP 6   OCT 2018
SWIFT | DJANGO | ARKIT

- Developed mobile application that organized when and where campus organizations had public tabling events. Integrated Augmented Reality for users to dynamically see where and how far an event is located with their camera.
- Implemented user-friendly UI/UX design and designed functional architecture for flexibility when adding more features.

RECEIPT SCRAPER   SEPT 2018
PYTHON | SELENIUM | JAVASCRIPT | AWS DYNAMODB | AWS ELASTIC BEANSTALK

- Automated receipt collection from Amazon.com to simplify invoice creation, used for expense reports and tax returns.
- Utilized AWS to build the application and exported collected data to a NoSQL database for horizontal scalability.
- Implemented configuration feature for users to specific the frequency of autonomous data collection.