Sathwik Kondapuram

E-mail: sk67015@uga.edu   Ph: +1 (470) 437-7322

EDUCATION

University of Georgia, *Athens, GA*   Aug. 2019 – Dec. 2021
Master's in computer science   *GPA: 4*

Courses: Algorithms, Data Science, Advanced Data Analysis, Computer Vision, Database Management, Advanced Database Management, Operating Systems, Advanced System Security, Advance Cyber Security, Datawarehousing

SKILLS

Programming: Java, Python, JavaScript
Database Technologies: Oracle Database (SQL, PL/SQL), PostgreSQL, Mongo DB, Neo4j (Cypher).
Web Development: HTML, CSS, JavaScript
Big Data: Apache Spark, Pyspark, Spark SQL, Spark Streaming(Kafka) , Apache Hudi, Apache Hive
AWS: EMR, Glue, S3, EC2, VPC, IAM, Athena, SNS, SQS, Redshift, Lambda and step functions
Other: GIT, Intellij, Anaconda

WORK EXPERIENCE

Amazon

Data Engineer Intern   *June. 2021 – Aug. 2021*
- Designed and implemented an end-to-end solution for GDPR on S3 Datasets using Apache Hudi, AWS EMR, Glue catalog, Athena, Redshift and Dynamo DB, DMS, SNS, Apache Spark and Hive. This project enables us to execute data deletion requests in an optimal way.
- Tested and deployed end to end pipeline. Due to need for data freshness once in every 7 data commits are appended with inline compaction. Using this metrics on average 90% of computation and I/O costs are saved compared to old GDPR implementation mechanism.
- Designed and drafted standards to follow for implementing deletion on datasets stored in S3 efficiently and optimally.

Carl Vinson Institute of Government

Graduate Research Assistant   *Aug. 2020 – Current*
- Work on ad-hoc requests, address the problems, create reports and dashboards using Qlik(QlikSense, Nprinting).
- Address the questions with deep dive analysis on huge amounts of data using Python, SQL, Excel and Qlik.
- Performing ETL, Data Modeling, Data Analytics, Data Science using Qlik Sense, Python

Apps Associates

Associate Software Engineer   *Aug. 2017 – Jun. 2019*
- Implementing Oracle SAAS applications with modules spanning across Financials, Supply Chain and HCM. Implementing Oracle E-Business Suite 12.1+/12.2+ across several modules in Finance, Supply Chain, Procurement and Core HR.
- Developed Oracle Reports and Oracle Forms. Built complex Oracle SQL queries, PL/SQL programs and web pages using Oracle Apex. Created Views by building complex queries on multiple domains using Oracle SQL.

PROJECTS and RESEARCH: *Aug. 2019 – Apr. 2021*

- COVID contact tracing using Neo4j and Face recognition. This project helps in finding nth degree relations among people in O(n).
- Implemented A three step accident assistance: prevention, assistance, and response to avoid any delays in helping a victim using IOT, Android Studio, and Machine learning.
- Performed Data ingestion, Data Storage, Resource management, Performing Statistical analysis and processing on both Batch and streaming data, Building Schema, Generating and performing operations on Datasets, Data Frames and RDD.
- Hands-on experience on Multiple Databases like Cassandra, PostgreSQL, Neo4j, MongoDB. Experience of Dealing with cloud storage systems(Amazon S3) along with Amazon EC2.
- Ability to solve hard SQL queries, build Data Lakes, build schema, conceptual understanding of Normalization and other database concepts along with parallel processing and distributed systems.
- Performed ETL, Data preprocessing, Exploratory Data Analysis and then performed different Regression and Classification techniques to compare performance of different models on different types of data sets.
- Implemented Strawberry detection system for assisting drones to pluck or manage them. Designed an image segmentation system using Python OpenCV Package.
- Implemented URL Classification into Malicious and Benign using WordtoVec method. Worked with 3.6M instances and achieved 88% accuracy with Random Forests and 92% accuracy with Convolutional Neural Networks
- Designed a Database along with a website to handle online food ordering from local restaurants.
- Implemented Inter process communication with RPC, Multi-threaded Web Server, CPU Scheduler Simulator, Page Handling with different Page replacement algorithms in Linux.
- Conducted Research on Control Flow Integrity and potential drawback of different Mechanisms that implement CFI.