

Saidarao Chirumamilla

San Jose, California | 470-519-8285 | [Email](#) | [LinkedIn](#) | [GitHub](#)

TECHNICAL SKILLS

Programming Languages: Python, JAVA, SQL, Scala.
Database Systems: MySQL, PostgreSQL, MongoDB, ClickHouse, Teradata.
Cloud Technologies: AWS (EMR, RedShift, S3, EC2, Glue, Kinesis, CloudWatch, Athena, Lambda).
Data Warehousing: Data modeling, ETL processes, data integration.
Visualization Tools: Tableau, Power BI, QlikView, D3.js, Python – Matplotlib, Seaborn, Plotly
Bigdata Technologies: Apache Spark, Hive, HBase, Hadoop, Kafka, Flink, Airflow, YARN.
Tools: Git, Docker, Kubernetes, Jupyter Notebook, SQL Workbench, Tableau.

PROFESSIONAL EXPERIENCE

Data Engineer at TikTok Mountain View, USA

Oct 2021 – Present

- Successfully provided end-to-end data services to support TikTok's live streaming platform, encompassing both batch and real-time data processing methodologies.
- Delivered actionable insights to optimize TikTok's recharge business operations, achieving a 26% reduction in channel fee costs and increased self-owned channel recharge volumes by utilizing advanced data gathering and analysis techniques.
- Strategically optimized data pipelines, meticulously aligning resource allocation with actual requirements, yielding a remarkable 60% reduction in operational costs by 200,000 dollars.
- Developed and maintained over 200+ data pipelines, leading to a 25% reduction in data latency, which contributed to a 15% reduction in operational costs by optimizing resource utilization and automating repetitive tasks.
- Ensured data accuracy through meticulous analytical techniques, collaborating effectively with cross-functional teams to deliver timely, high-quality solutions, leading to a 10% improvement in data reliability.
- Improved development and maintenance processes of data systems by creating comprehensive documentation, procedures, and user guides, resulting in a 20% increase in team adoption and a 15% boost in productivity.
- Led the integration of diverse data sources and regional datasets, architecting and deploying robust compliance protocols for secure data access during critical security periods, enhancing data security measures by 30%.
- Designed and executed multiple automation projects optimizing workflows, resulting in a 20% increase in team efficiency, a 25% reduction in costs, and a 15% time savings, while increasing the stability of data pipelines by 30%.

Data Lake Infrastructure Analyst Intern at Sun life Financial, Atlanta, USA

June 2021 - Aug 2021

- Migrated the data from multiple source systems, including both relational and non-relational databases, to HDFS as Hive tables, resulting in improved data processing efficiency and a 25% reduction in data processing time.
- Transformed the data using Spark and map Reduce loaded into data lake according to requirements and made available for different teams.
- Authored and published technical documentation for projects. Translated R scripts for metadata management into PostgreSQL and automated scheduling.

Graduate Research Assistant at Georgia State University, Atlanta, USA

Jan 2020 - April 2021

- Adult Census Income - Built a Machine Learning model to predict the annual income range of US population, by analyzing various parameters. Used Machine learning models like Logistic regression, Naive Bayes, K-nearest Neighbors.
 - Fully connected neural network - Built a 3 layered fully connected neural network to identify the handwritten numbers using basic python Programming and analyzed the performance of model with varied parameters for Iris and MNIST datasets with model accuracy of 89% for optimal combination of parameter.
 - Build a CNN model based on Cifar-10 dataset and improved the model efficiency by applying data augmentation, batch normalization, dropout and Ridge/L2 regularization techniques.
-

ACADEMIC PROJECTS

Recommendation System (Python, Big Data, HDFS)

- Designed a recommendation system using Spark's alternating least squares (ALS) method to learn latent factor representations for users and items in python and evaluated the system using the ranking metrics provided by spark.
 - Used an Alternative model formulation such as log compression and dropping low count values, on top of the baseline collaborative filter model to increase the accuracy of the recommendation system.
-

EDUCATION

The Georgia State University at Atlanta

Aug 2021

Master of Science in Computer Science

GPA: 3.66

RVR&JC College of engineering, Guntur, India

May 2019

Bachelor of Technology, Electronics and communication Engineering