Srujit Varasala

New York
862-237-5382 srujit.v@gmail.com linkedin.com/in/srujitvarasala/ github.com/srujit12091997

Professional Summary

Big Data Engineer with almost 3 years of experience in designing, developing, and optimizing large-scale data pipelines and ETL processes. Proficient in **Apache Spark**, **Hadoop**, and **cloud technologies**, with a focus on improving data processing performance and delivering high-quality insights. Experienced in handling terabytes of data and managing data integration across distributed systems. Adept at working with **Python**, **PySpark**, and **Scala** to drive efficiency and scalability in data operations.

Technical Skills

Big Data Technologies: Apache Spark, Hadoop, HDFS, Airflow, ETL pipelines, MongoDB, Neo4j
Programming Languages: Python, Spark (PySpark), Scala, Java
Databases: SQL, PostgreSQL, MongoDB, Neo4j
Cloud & DevOps: AWS, Docker, CI/CD pipelines, Git, GitHub
Data Processing Tools: Apache Kafka, Spark Streaming
Testing & Automation: Unit Testing (JUnit, PyTest)

Experience

Client: DLK Technologies Private Limited **Nov 2018 – Dec 2018**
Full Stack Web Development Intern *Chennai, Tamil Nadu*

- **Description:** Worked on end-to-end development of web applications, starting from frontend design with **HTML** to the deployment of the application. Contributed to various stages of the development lifecycle, gaining experience in both frontend and backend development.
- **Responsibilities:**
- Developed and maintained full-stack web applications using **Django**, improving overall performance and usability for client websites.
- Built and managed models, views, and templates for dynamic web pages, reducing database query times by 20%.
- Created and integrated REST APIs to streamline data flow, reducing latency and improving response times by 15%.
- Designed and implemented responsive UIs using **HTML**, **CSS**, **JavaScript**, and **Bootstrap**, leading to a 30% increase in mobile compatibility.
- Performed CRUD operations to maintain high data integrity, reducing data inconsistencies by 10%.
- Collaborated with team members using **Git** for version control, leading to the successful deployment of 5+ new features over 2 months.
- Debugged and optimized web applications, cutting page load time by 25% and resolving critical bugs in real-time.

Client: Techniki services Startup **Aug 2019 – Dec 2019**
Full Stack Developer Intern *Hyderabad, Telangana*

- **Description:** Worked on end-to-end development of web applications, starting from frontend design with **HTML** to the deployment of the application. Contributed to various stages of the development lifecycle, gaining experience in both frontend and backend development.
- **Responsibilities:**
- Developed and maintained web applications using **HTML**, **CSS**, **JavaScript**, and **Bootstrap**, enhancing cross-browser compatibility and responsiveness by 30%.
- Built and integrated **REST APIs**, improving communication between frontend and backend systems, reducing response time by 20%.
- Optimized backend logic and database management using **Java**, reducing server load by 15% and improving database efficiency.
- Led deployment processes using **Docker** and **CI/CD**, reducing deployment time by 50% and ensuring consistent production environments.
- Debugged, tested, and optimized web applications, reducing load times by 25% and minimizing downtime during releases.

- Collaborated with cross-functional teams, delivering 3 key projects on schedule over 4 months, ensuring smooth production deployment.

Client: Jio Platforms Limited **July 2020 – December 2022**
Big Data Engineer *Bengaluru, Karnataka*

- **Description:** Jio Platforms Limited is a leader in telecommunication services all over India and digital services providers, Providing businesses and consumers with Telecommunication services insights. I worked as a big data engineer on the JAMP (Jio Audience Management Platform) team. Managed data pipelines, and ETL processes, optimizing performance and ensuring seamless data integration across various platforms.
- **Responsibilities:**
- Designed and developed scalable data pipelines using **Apache Spark**, **Python**, and **Spark Scala** to process 10+ million records daily, reducing processing time by 40% and boosting data accuracy by 30% through automated data quality checks and data cleansing routines.
- Built and maintained batch processing systems for 1TB of data daily using **Apache Hadoop**, **HDFS**, and **Airflow**, ensuring seamless data integration.
- Developed and deployed the **Jio Maps** project, integrating geographic data from **Google Maps**, **HERE Maps**, and **OpenStreetMap**, increasing user navigation accuracy by 20%.
- Improved navigation accuracy by 20% through integrating 1B+ places of interest.
- Gathered geographical data via API keys and **Python** scripts, utilizing **Neo4j** for data analysis and visualization to uncover relationships, while using **MongoDB** for accurate and relatable mapping.
- Created efficient ETL pipelines using **Python** and **Airflow**, integrating data into **Oracle** databases, improving data reliability and flow.

Projects

Real-Time IoT Insights Hub | *Apache Kafka, Spark Streaming, HDFS, Python* **March 2023 – June 2023**

- Developed a real-time data streaming platform using **Apache Kafka** and **Spark Streaming**, processing over 2TB of incoming data daily from multiple IoT devices across the globe.
- Integrated with **HDFS** for scalable data storage and archiving, improving data accessibility and redundancy.
- Optimized the streaming architecture, reducing data processing latency by 30%, ensuring real-time decision-making for connected devices.
- Created automated **ETL pipelines** for cleaning, transforming, and loading data into a data warehouse, enabling efficient business analytics.

Places of Interest | *Apache Spark, Python, Neo4j, MongoDB, HDFS* **January 2023 – May 2023**

- Built a data pipeline for ingesting, processing, and analyzing geo-location data to provide real-time location-based insights. Used **Apache Spark** for large-scale data processing and **Neo4j** for graph-based data storage and analysis.
- Processed over 500GB of geo-location data using **Apache Spark**, enabling faster real-time analysis and insights generation.
- Stored data in **Neo4j** to build and analyze spatial relationships and networks, improving data retrieval times by 30%.
- Developed queries in **Cypher** to visualize connections between geographical entities, such as restaurants, hotels, and user activity points.
- Achieved a 20% increase in processing efficiency by optimizing data loading from **MongoDB** into **Spark**.

Web Data Scraping and Analysis | *Python, BeautifulSoup, Hadoop, MapReduce, HDFS, MongoDB* **September 2020**

- Developed a Python-based web scraper using **BeautifulSoup** to collect data from multiple websites, extracting information such as product prices, reviews, and ratings.
- Processed over 1TB of unstructured data using **Hadoop MapReduce**, improving data processing efficiency by 60%.
- Stored the cleaned data in **HDFS** and **MongoDB** for scalable and quick retrieval.
- Analyzed data using MapReduce and visualized key insights, such as price trends and product ratings, using **Matplotlib** and **Seaborn**.
- Achieved a 50% reduction in data processing time by leveraging distributed computing.

Education

New Jersey Institute of Technology **Newark, NJ**
Master of Science in Computer Science *Jan 2023 – Present* *CGPA: 3.0*

SRM Institute of Science and Technology **Chennai, Tamil Nadu**
Bachelor of Technology in Computer Science and Engineering *Aug 2016 – May 2020* *CGPA: 3.2*