

Keshav Raj

keshav.rajsspn@gmail.com | +91-9576640200 | linkedin.com/in/k-e-s-h-a-v | github.com/k-e-s-h-a-v

SUMMARY

Ambitious Data Engineer with 3+ years of experience in building high-performance data platforms, specializing in decentralized architectures (DataMesh) with Data Governance and optimizing business intelligence systems. Successfully led teams, was quick to adapt to challenging environments and contributed to product acquisition

EDUCATION

Bhagalpur College of Engineering, Bhagalpur

Jul 2018 - Sept 2022

B.Tech in Electrical Engineering

CGPA: 8.2

Sainik School Punglwa Nagaland, Ministry of Defence

May 2010 - June 2017

Science, High School Education

Percentage: 81.2

EXPERIENCE

Infor Jul 2024 – Present

Hyderabad, India

• Software Engineer

Jul 2024 – Present

- Integrated **Apache Superset** using **Delta Lake** to the **DataMesh** platform for real-time **data visualization**.
- Built an **invoice processing system** with **Qwen** to extract data that feeds an **event-driven, configurable data entry pipeline**, reducing SLA from **48 hours to 1.5 minutes**.

albanero (acquired by Infor) Jun 2021 – Jul 2024

Hyderabad, India

• Software Engineer II (6 months)

Jan 2024 – Jul 2024

- Key contributor to a **state-of-the-art DataMesh platform**, later acquired by Infor.
- Slashed memory usage by 90%** in the **CSV-to-Excel pipeline**. eg: **52 GB to 4 GB** for \approx 3 million records.
- Enabled **multi-CSV merging** into a single Excel file, with each CSV as a **dedicated sheet**.

• Team Lead (8 months)

May 2023 – Jan 2024

- Led a team of **data analysts** to deliver **high-impact products**, including two key projects—one was live in just **6 weeks**.
- Built a **data mastering system** using the **Jaro-Winkler algorithm** for duplicate detection.
- Standardized entity names using **NER** with **spaCy**, and integrated **Loquite** for **address verification**.
- Conducted internal training sessions on our **DataMesh platform**, teaching teams how to use its **decentralized architecture** and **scalable workflows**.

• Software Engineer I (8 months)

Sep 2022 – May 2023

- Built a dynamic **SQL query generator** to search/filter **millions of records**.




• Software Engineer Intern (1 year 4 months)

Jun 2021 – Sep 2022


- Developed **Identity Access Management (IAM)** for DataMesh, including role-permission logic.
- Authored feature for real-time **status updates** for data transformation workflows using **Socket.io**.
- Implemented **multi-theme support** in a **SaaS platform**, standardized **SCSS practices**, and gained hands-on experience with **JavaScript**.

PROJECTS

VS Code Extensions

- Explorer Plus** : Supercharges the **VS Code** file explorer with **file/folder stats** and **sort**.
- Code Folder Summary** : Shows a **quick overview** of **file types** in a project.
- File Size and Count** : Displays **file/folder size and count** on hover.

Chrome Extension

- Read Canvas** : A **Chrome Extension** for reading the captcha from canvas. *See it in action*

TECHNICAL SKILLS

Languages: Python, SQL, JavaScript


Libraries/Frameworks: PySpark, FastAPI, Flask, Docker

BI/Analytics Tools: Apache Superset, AWS Athena, Delta Lake (Parquet), MongoDB

Tech: Git, Data Modeling, ETL Pipelines, Data Security, Data Federation

CERTIFICATES

An Introduction to Artificial Intelligence 

Programming, Data Structures and Algorithms using Python 

Data Analytics at E & ICT Academy 

IIT Delhi

IIT Madras

IIT Kanpur