

# Santosh Kumar Gouda

## Data Analytics Technical Lead

Bangalore • +91 7619264934 • [s.k.gouda21@gmail.com](mailto:s.k.gouda21@gmail.com) • [linkedin.com/in/sgouda0412](https://www.linkedin.com/in/sgouda0412) • [github.com/sgouda0412](https://github.com/sgouda0412)

More than 8+ years of experience working as a Data Engineer. I have built scalable data pipelines that process millions of records, delivering actionable insights, features, and scores at scale.

## WORK EXPERIENCE

---

- **Sears Holdings India** – Data Analytics Technical Lead (Nov 2022 – Present)
- **Enkash** – Data Engineer (May 2022 – Aug 2022, Contract-to-Hire)
- **Tech Mahindra** – Data Engineer (May 2021 – May 2022)
- **Alten India** – Data Engineer (Sep 2020 – Apr 2021, Contract-to-Hire)
- **Bristlecone** – Data Analytics Engineer (Nov 2019 – Sep 2020, Contract-to-Hire)
- **Accenture Solutions** – Data Analyst (Jan 2016 – Sep 2018)

## PROFESSIONAL SUMMARY

---

- Expertise in **Retail (FMCG, Home Appliances & Tools), Aerospace, Mining, and Fintech** sectors.
- Utilized **AWS and Azure** for large-scale data processing.
- Developed efficient **Apache Airflow** workflows with complex DAGs.
- Skilled in writing and optimizing **complex SQL queries**.
- Built and maintained **data pipelines and models**, driving seamless data integration.
- Led on premise to **cloud migrations**, improving scalability.
- Monitored and **resolved pipeline issues**, ensuring reliable performance.
- Proficient in Python (**NumPy, Pandas, PySpark**) and handling big data **formats (Parquet, CSV, JSON)**.
- Used **Apache Spark** for fast data processing.
- Strong **communicator and team player**, experienced in **AGILE** environments.

## SKILLS

---

- **Cloud Platforms:** AWS (S3, IAM, RDS, Glue, Redshift, Lambda, Athena, CloudWatch, EMR, QuickSight), Azure (Blob Storage, ADF, Databricks, DevOps) – 3.5 year's experience.
- **Data Solutions:** Data Pipelines, Data Warehousing, Data Lakes, ETL.
- **Programming/Query Languages:** Python (Proficient), SQL (Intermediate), PySpark (Intermediate), Scala (Basic), Airflow, Data Structures & Algorithms – 4 year's experience.
- **Version Control/Tools:** Git, JIRA – 4 year's experience.
- **Soft Skills:** Communication, Business Acumen, Critical Thinking, Lifelong Learner.

## EDUCATION

---

- Completed a B Tech in **Computer Science & Engineering** with a CGPA of 7.74 from the **National Institute of Technology, Nagaland**

## PROJECTS

---

**Sears Holdings India, Pune, Maharastra- Data Analytics Technical Lead**

**Nov 2022 – till date**

**Project:** Transformco (Data Migration, ETL, Automation & Reporting)

**Client:** Transformco

**Team Size:** 8 members

**Domain:** Retail

### **Project Overview:**

Transformco leverages analytics to optimize operations, enhance customer experience, and drive growth through data-driven insights.

### **Role and Responsibilities:**

- **Designed and implemented ETL pipelines** to migrate data from on-premises **OLTP systems (MySQL, PostgreSQL, and MongoDB)** and **SFTP/S3 storage** to a **Snowflake data warehouse**, enabling comprehensive analytics.
- **Led a data loading project**, optimizing the **ETL process** to load high-volume data from various sources into **Snowflake**, improving data ingestion speed by **30%** and ensuring near real-time data availability for analytics..
- **Automated** the data migration process by creating efficient **Airflow DAGs**, enhancing operational efficiency.
- **Crafted complex SQL queries** to address ad-hoc requests from business stakeholders, consistently delivering accurate and timely **data-driven insights**.
- **Developed daily sales and profit reports** for **Home Appliances** and **Tools repairs**, utilizing **Tableau** to visualize business performance metrics

**Tech Stack:** Python, SQL, Snowflake, PySpark, ETL, Data Migration, Airflow, Tableau.

**Enkash, Bangalore, Karnataka– Data Engineer**

**May 2022 – Aug 2022**

**Project:** Enkash (Data Migration & ETL)

**Client:** Enkash

**Team Size:** 6 members

**Domain:** Fintech

### **Project Overview:**

Enkash generates substantial volumes of Customer, Card, Non-Card, and Statement data, necessitating the migration of this data from an OLTP system to an OLAP system for unified analytics.

### **Role and Responsibilities:**

- **Developed ETL pipelines** to **migrate data** from an on-premises **MySQL OLTP system** to a **Redshift data warehouse**, enhancing **analytics capabilities**.
- **Engineered** custom AWS Glue scripts for **data extraction, transformation, and loading**, leveraging **PySpark, SQL, and Python** for optimized performance.
- **Constructed an analytics dashboard** in **AWS QuickSight**, enabling business users to visualize monthly, weekly, and quarterly **sales and marketing performance metrics**.

**Tech Stack:** Python, SQL, AWS (S3, IAM, Athena, RDS, Glue, Redshift, EMR, QuickSight), PySpark, ETL, Data Modeling, Data Warehousing, Data Lakes, Data Pipelines, Airflow.

**Tech Mahindra, Bangalore, KA – Data Engineer**

**May 2021 – May 2022**

**Project:** eBay (Data Warehouse Migration & Data Pipelines)  
**Duration:** 10 Months  
**Team Size:** 5 members  
**Domain:** Retail

**Project Overview:**  
At eBay, the organization generated substantial volumes of **Customer** and **Product Data**, necessitating the migration of this data from an **OLTP** system to a **Data Warehouse** for unified analytics.

**Role and Responsibilities:**

- **Engaged** with business users and stakeholders at various levels to understand requirements and translate them into effective **Data Solutions**.
- **Developed** a critical data pipeline that processed over **5 TB** of data by consolidating information from multiple disparate sources into a single destination, enabling swift data analysis for reliable business insights.
- **Automated** the manual data reconciliation process across different sources, resulting in a **90% increase** in data accuracy and efficiency.
- **Streamlined** the ETL process, reducing data wrangling time by up to **40%** through automation.
- **Enhanced** data retrieval efficiency by approximately **30%** through query optimization and indexing techniques.
- **Designed** a dashboard for the operations department to provide transparency into customer data, utilized by over **50** staff members and decision-makers to drive deliverables.

**Tech Stack:** Python, AWS (S3, Athena, IAM, Glue, Redshift, QuickSight), PySpark, SQL, ETL, Data Modeling, Data Warehousing, Data Lakes.

**Alten India, Bangalore, KA – Data Engineer** **Sept 2020 – Apr 2021**

**Project:** Anglo American (Data Migration & ETL)  
**Team Size:** 7 members  
**Domain:** Mining

**Project Overview:**  
At Anglo American, the organization generated substantial volumes of **Sensor Data** related to mining materials, necessitating the migration of this data from **MySQL** to a **Data Warehouse** for unified analytics.

**Role and Responsibilities:**

- **Migrated** multiple **SSIS** packages (stored procedures) from **SQL Server** to **Azure Databricks**, enhancing data processing capabilities.
- **Developed** pipelines in **Azure Data Factory (ADF)** using Linked Services, Datasets, and Pipelines to **Extract, Transform, and Load (ETL)** data from various sources, including **Azure SQL, Blob Storage, and Azure SQL Data Warehouse**.
- **Wrote** notebooks in **SQL, PySpark, and Python** to transform and optimize data workflows.
- **Implemented** build and release processes for deploying pipelines, triggers, linked services, and notebooks utilizing **Azure DevOps** and **PowerShell** for streamlined operations.

**Tech Stack:** Python, PySpark, SQL, Azure Blob Storage, ADF, Data Warehousing, ETL, Azure Databricks, Azure DevOps

**Bristlecone, Bangalore, KA – Data Analytics Engineer**

**Nov 2019 - Sept 2020**

**Project:** SAP Labs (Pipeline Investigation and ETL Monitoring)

**Team Size:** 7 members

**Project Overview:**

Conducted comprehensive pipeline investigations and monitoring to ensure the accuracy and availability of analytics data within the **Data Warehouse**.

**Role and Responsibilities:**

- **Implemented** a systematic approach to plan, create, and maintain **data architectures**, ensuring alignment with business requirements.
- **Developed** data pipelines and conducted investigations into pipeline spikes occurring in **Azure Data Lakes**, enhancing data reliability and performance.

**Tech Stack:** Python, PySpark, SQL, Azure Blob Storage, ADF, Data Warehousing, ETL, Azure Databricks.

**Accenture Solutions, Bangalore, KA – Data Analyst**

**Jan 2016 – sept 2018**

**Project:** Accenture Solutions (Data Extraction and Analysis)

**Team Size:** 7 members

**Domain:** Fintech

**Project Overview:**

Conducted data analysis using automated tools to identify and rectify errors, thereby maintaining high **data quality standards** for improved business decision-making. Developed a **Financial Market Dashboard** to provide a comprehensive overview of business performance.

**Role and Responsibilities:**

- **Employed** automated tools to extract data from primary and secondary sources, enhancing data acquisition efficiency.
- **Identified** and removed corrupted data while resolving coding errors and related issues to ensure data integrity.
- **Developed** and maintained databases and data systems, reorganizing data into a readable format for ease of use.
- **Performed** analyses to assess the quality and significance of data, ensuring accurate insights.
- **Filtered** data by reviewing reports and performance indicators to identify and rectify coding problems.
- **Utilized** statistical tools to identify, analyze, and interpret patterns and trends within complex datasets, facilitating diagnosis and prediction.

**Tech Stack:** Python, SQL, Excel, Power BI, Data Visualization.