



# Parth Soni

Data & AI Engineer | AWS Community Builder | Content Creator

**Ambitious and self-driven Technology Development Professional** with 8+ years of experience across multi-cloud environments, renowned for leading transformative initiatives in **Generative AI, Data Engineering, and Intelligent Automation**. Demonstrated expertise in building scalable solutions across **Data Warehousing, Data Analytics, AI/ML Integrations, and Modern Data Architectures**. Proficient in **SQL Development, Python, Java, .NET, and Core Data Structures & Algorithms**, with hands-on experience across a variety of databases and cloud-native ecosystems. A proactive innovator, lifelong learner, and collaborative communicator, dedicated to solving complex problems and driving AI-first digital transformation.



Parth.Soni78738@gmail.com



9913978738



Vadodara, India



linkedin.com/in/parthasoni



## SKILLS



## WORK EXPERIENCE

### Senior Data Engineer

Planet

06/2022 - Present

Achievements/Tasks/Projects/POCs

- Design, develop, and implement **real-time data processing solutions** by harnessing **Change Data Capture (CDC), Apache Kafka, Apache Flink, PyFlink, and Debezium** integrated on AWS EC2 instances, using Avro and Protobuf serialization formats, with Kraft for orchestration and Amazon DocumentDB for storing Payment Events.
- Developed scalable data processing pipelines using **Python, PySpark, and Pandas**, handling multi-terabyte datasets in distributed environments.
- Optimized PySpark transformations and actions to reduce job runtime by **30–50%**, improving overall pipeline throughput.
- Built modular, production-grade Python utility packages for data ingestion, validation, and schema enforcement.
- Designed and deployed end-to-end ETL/ELT workflows on **AWS EMR, Glue, Lambda, and S3**, enabling high-performance batch and real-time data processing.
- Implemented metadata-driven pipelines using **AWS Glue (PySpark/Scala)** with optimized job bookmarks, partitions, and dynamic frames.
- Built cost-efficient Spark clusters on **AWS EMR**, leveraging autoscaling, spot instances, and cluster re-use strategies.
- Designed NoSQL ingestion and querying patterns using **DynamoDB** with efficient partition keys and GSI-based lookups.
- Configured secure data access using **IAM roles/policies**, VPC endpoints, and fine-grained permissions for production workloads.
- Designed and deployed end-to-end ML pipelines using AWS SageMaker and Azure ML, including data ingestion, training, evaluation, and CI/CD integration.
- Designed and implemented **Databricks Medallion Architecture (Bronze, Silver, Gold layers)** to standardize data architecture.
- Integrated and applied **GenAI** models and techniques into real-time data processing workflows, enhancing automation and accuracy.
- Built and productionized **LLM-powered RAG workflows** with **LangChain, Pinecone/PGVector, and OpenAI embeddings**, improving retrieval accuracy and response quality. Optimized **data processing and cost-efficiency** across **AWS services (S3, Glue, Redshift, Lambda)** and **Databricks** clusters, while leveraging **IDMC** for workflow automation, governance, and compliance.

- Implemented model monitoring, drift detection, and governance using **MLflow** and Azure ML pipelines, ensuring compliance and reliability.
- Design and Build an Industry standard **ETL** solutions, for **Data Warehousing**. Involved in Debugging, Monitoring, developing and Maintenance of existing **BI** solutions (**SSIS, SSRS, SSAS, SQL Server**) SQL Server 2019. Integrating **Data Dog** APM Agents into On-Prem legacy solutions and configuring the **YAMLs**.
- Optimized large-scale **PySpark data pipelines** using **dbt** for feature engineering and training datasets, reducing job runtime by 30%.
- Building all POCs and Architecture using **Terraform IaaS, GitHub** for **CI/CD** pipelines for workflows. Automating deployment processes using **shell scripting**, proficiently **Dockerize** applications, and manage **containerized** environments. Craft custom **Docker** images, lead **Dockerized** projects, and demonstrate scalability benefits of **containerization**.
- Understand & Research planet products business to develop **ML algorithm** for the opportunity of increasing revenue using **Azure ML, AWS Sagemaker, scikit-learn**. Collaborate with Data Science team to develop **ML and Deep Learning Models** to achieve payment fraud detection algorithms and recommendation system. Work with architects on extending existing ML libraries and frameworks.



## WORK EXPERIENCE

### Senior Application Engineer

Thomson Reuters

04/2020 - 06/2022

ONESOURCE GTM - SaaS platform for International Global Trade Management.

Innovator of the year | SME of TechOps Team

- Designed, developed & optimized highly complex workflow interfaces, with flexible & scalable data and robust **ETL** architecture by utilizing **Spark & Hadoop** technologies to support ONESOURCE clients application and business users.
- Automated manual & monotonous tasks by making an centralized **data warehouse with star and snowflake schema** collaborating with product development team to save users efforts by 75%. Built architecture for high scalability streaming system.
- Research & Analysis on database of global trade networks to tune query performance by optimizing **Indexes of Transactional & Analytical** workloads.
- Creating/Modifying **Stored Procedures, Triggers, Routines, Views, Functions, SSIS** packages and complex SQL scripts to facilitate projects & business operations.
- Proposed **innovative ideas** on implementation of 6-data points requirement gathering over ONESOURCE GTM Application to facilitate support operations to reduce data load times up to 80%.
- Worked on migration of ONESOURCE GTM Products from on-prem to cloud databases **Azure Synapse Analytics (DW), Azure SQL DB & good experience working with Azure DF, Data lake storage, AWS S3, AWS Lambda & AWS Glue, AWS Kinesis, AWS EMR.**
- Tuned Spark configurations (executors, memory, shuffle partitions) to reduce compute cost by **25–40%**.
- Optimized Glue/EMR jobs by implementing partition pruning, broadcast joins, and storage-efficient file formats (Delta, Parquet, ORC).
- Collaborated with data architects, BI teams, and application developers to define data contracts, SLAs, and integration strategies.
- Supported analytics teams with curated, high-quality datasets enabling real-time insights and ML workloads.

### Software Engineer

Integration Point - Thomson Reuters

04/2018 - 03/2020

World's leading GTM IT industry, later acquired by Thomson Reuters

Achievements/Tasks

- Develop **C#** and **.Net** code to build next-gen GTM products managing all aspects of the **SDLC** process-requirement analysis, design, development, testing, and support/maintenance.
- Design and develop **Data Synchronization** from **Salesforce** to **GTM Platform** by writing **Data APIs in C# and .Net** to integrate with **Azure Data factory** for data orchestration using Azure SDKs and develop **APIs** to persist data in **NoSQL** database.
- Development of SQL Monitor tool, which dealt with creating new production alerts for client failures with **SQL scripts** collaborated with DBA team.
- Monitoring of web servers and databases using **DataDog** with **JSON** and **SQL**.
- Lead ConfigRequests team which dealt with implementation of workflow configuration changes related to client environments validations on **Azure DevOps**.
- Information Security Ambassador at Vadodara Site** : Development over security applications by different methodologies and contribution towards cloud technology collaborated with **DevOps**. Performed security checks on existing architecture in order to prevent breach of data.

# Software Developer Intern

## L&T Hydrocarbon Engineering Ltd

06/2017 - 07/2017

### Achievements/Tasks

- Full stack development over HR Junction and Resource booking project by developing modules with the help **ASP.Net, JavaScripts, T-SQL and Azure Cloud.**

## EDUCATION

### Bachelor of Engineering – Information Technology

Gujarat Technological University

2014 - 2018

CGPA - 8.73

### Higher Secondary Education - Science

Kendriya Vidyalaya / CBSE

2002 - 2014

10th CGPA - 8.2 & 12th - 74 %

## CERTIFICATES

### AWS Certified – Solutions Architect Professional

Cleared SAP-C02 Professional Exam.

### Microsoft Certified: Azure Data Engineer

Cleared DP-203, AI-900 and DP-900 Microsoft Exams.

### AWS Machine Learning Nanodegree

Completed AWS Machine Learning Foundations in AWS ML Scholarship.

### AWS Certified Machine Learning - Specialty

Cleared MLS-C01 Specialty Exam.

### AWS Certified Cloud Practitioner

Validates foundational, high-level understanding of AWS Cloud, services, and terminology.

### Scrum Fundamentals Certified

Validates the fundamental knowledge of Agile and Scrum Methodologies

## ACHIEVEMENTS

### AWS Community Builder

Selected as AWS Community Builder for contribution towards AWS and building community stronger.

### Secured 2nd Rank at “SARJAN” - State Level Project Exhibition

Built, designed, developed TnP Integument with Laravel and NoSQL with predictive ML algorithms

### Innovator of the Year - 2020

Proposed Ideas which made positive impact on Thomson Reuters business processes.

## PUBLIC SPEAKING

### AWS Community Builder – AWS Events

#### Role

- An active **AWS Community Builder**, recognized for delivering impactful public talks at AWS events, sharing insights on cloud-native AI, data platforms, and real-world GenAI applications.

### Planet | TechTalks - AWS Data Engineering Services

#### Planet Payment

#### Role

- Mentorship and knowledge sharing to built and grow AWS cloud community.

Reference : <http://tinyurl.com/ParthAWS>