

Anilkumar Mithailal Patel

105 Richards Avenue 2406, Norwalk, CT, USA | anilpatel3699@gmail.com | +1 475-251-5831 | [in](#) [LinkedIn](#) | [GitHub](#)

SKILLS & OTHER

Programming skills, version Control & OS: Python, SQL, Rest API, Flask, Shell Scripting, Git, Linux.

Cloud: AWS, GCP, Azure & Databricks.

Big Data: Spark, Kafka, HDFS, MapReduce, Yarn, Hive, Zookeeper, Hadoop distribution, Apache Beam, Presto.

Data Warehousing: Data Lake, Data Marts & Data Modeling, Snowflake, Big Query, Redshift.

Workflow Orchestration: Apache Airflow, DBT Cloud, Streamsets.

SQL & No SQL Databases: MySQL, PostgreSQL, Oracle, Mongo DB, Elasticsearch, HBase.

DevOps Tools (CI/CD): GitHub Actions, Ansible, Jenkins, Docker, Kubernetes.

Visualization: Tableau, Power BI.

PROFESSIONAL EXPERIENCE

Data Engineer Intern

Shamrock Foods Company, Phoenix, United States

05/2023 to 08/2023

- Created and implemented robust ETL pipelines utilizing Azure Data Factory, transforming the data extraction process for Shamrock stores across different states; **reduced data processing time by 90%** and **increased data accuracy by 25%**.
- Transformed legacy ETL systems into Databricks Delta Tables, resulting in an impressive **80% decrease in processing time** and significant efficiency gains.
- Consolidated data from Microsoft REST API and Azure Active Directory API by incorporating Databricks, Spark SQL, and Delta; **reduced data errors by 20%**, guaranteeing reliable and precise data for augmented analysis.
- Spearheaded the implementation of **DBT Cloud and Snowflake**, resulting in a **75% reduction in ETL code** development, streamlining data transformation processes, and improving overall efficiency by creating robust Data Modeling and Data Mart (Star Schema).

Senior Data Engineer

Priceline.com INC, Mumbai, India

05/2020 to 08/2022

- Analyzed and assessed data engineering team's codebase for compliance with coding standards, performance optimization on SQL Queries, and issue identification; provided actionable feedback resulting in a **30% decrease in system downtime** and Elevated overall code quality.
- Streamlined the establishment of a state-of-the-art GCP Data Lake-house, transforming data management practices and elevating data-driven decision-making capabilities.
- Strategically optimized data workflows, enhanced data integrity, and fortified data security measures to ensure seamless **ingress/egress patterns, data governance, cataloging, database design, Dimensional Modeling, Data mart and retention policies**.
- Headed Geospatial team, migrating data to Cloud SQL Postgres & Big Query for a **40x performance boost**, ensuring compatibility with Cloud Composer Airflow on ArcGIS Pipeline.
- Architected migration frameworks (on-premises to GCS, Big Query, Cloud SQL & Bigtable), and Built **Compaction** on GCS storage, resulting in a **\$10,000 cost reduction**. Migrated On-prem Java App/Airflow Dag's to the Cloud, devised a booking propensity model, and Co-ordinated to DevOps CI/CD pipelines.
- Led the implementation of **migrating analytical data offloading to Starburst Presto using Kubernetes, enabling \$13000 cost** savings. Managed with Hive external tables in Presto, integrated with GCS for storage.
- Directed and executed the enhancement of customer review streaming processes utilizing GCP Dataflow, Big Query, and Kafka integration; delivered a **35% boost in data accuracy**, empowering adaptive customer feedback analysis.

Data Engineer

AWS via (Mactores Cognition Private Limited), Mumbai, India

12/2018 to 05/2020

- **At Seagate Technology (17 Months):** Initiated a data migration strategy, using Airflow as the scheduler, and employed Python/Pyspark for backend processing. This led to a significant cut in analytics **processing time by 3 days**. Leveraged AWS S3 Data Lake, Presto as the Query Engine for long-running queries, and Spark SQL for short-running queries during the successful migration of databases and factory files from on-premises infrastructure.
- **Godrej Consumer Products (8 Months):** Engineered an end-to-end ETL solution, seamlessly transitioning On-Prem data to AWS infrastructure (S3, Aurora, RDS & Athena) and **Established Data Mart for Tableau Business Reports via JDBC connector**. This comprehensive solution not only purify data accessibility but also resulted in a remarkable 50% reduction in data ingestion time, enhancing overall efficiency.
- Planned and Instituted strategic migration plans in Agile Methodology for **Blackhawk Network Holdings, Veritas, and Katerra** ensuring a smooth transition of critical data and systems, resulting in elevated operational resilience and a **25% reduction in post-migration** issues.

Big Data Engineer

Formcept Technologies & Solutions Private Limited, Bengaluru, India

07/2017 to 12/2018

- Initiated the development of real-time data pipelines utilizing Apache Kafka, with a primary focus on data ingestion and pre-processing, optimizing performance with AWS Cloud services.
- Formulated MapReduce programs to process unstructured data, populate staging tables, and store refined data in partitioned tables within the Enterprise Data Warehouse; **decreased data processing time by 40% and improved data accuracy by 25%**.
- Collaborated with front-end engineers, leveraging AWS Cloud services (Lambda, S3, Glue) for efficient data processing using Rest API and Swagger, **achieving a 20% improvement in overall ingestion speed**.
- Employed AWS Glue and PySpark to load data into S3 Buckets, implemented data filtering with Elasticsearch, and managed data storage in Hive external tables.
- Enriched API infrastructure with AWS API Gateway, Elastic Load Balancer, and CloudFront, **optimizing data accessibility by 80% in Data Lake**. Applied ELK and EFK stacks for log analysis, **resulting in a 30% increase in actionable insights from log data**.

Big Data Engineer

Infrasoft Private Limited, Mumbai, India

06/2016 to 07/2017

- Installed, configured, and utilized components of the Hadoop ecosystem for data processing and storage & Formulated MapReduce programs for parsing raw data and storing refined data in partitioned tables in the EDW.
- Revised a Batch data pipeline using Streamsets, Python and Apache Kafka, processing over 1 million events daily, **resulting in a 30% reduction in processing time**.
- Facilitated in Data Acquisition with the Data Engineer team to extract historical and real-time data by using Sqoop, Pig, Flume, Hive, MapReduce, and HDFS.
- Fabricated a solution reducing data ingestion latency by 40%, optimizing ETL processes for improved reporting accuracy and faster decision-making.
- Designed and Enacted a **star schema for a retail analytics** platform, **resulting in a 40% increase in query performance** and managing ETL pipelines for **5+ TB of daily data from sources**.

CERTIFICATIONS / PUBLICATIONS (BLOGS)	
<ul style="list-style-type: none">- GCP Professional Data Engineer.- AWS Certified Solutions Architect – Associate.- AWS Certified Developer – Associate.- GCP Professional Machine Learning Engineer.- AWS Academy Cloud Architecting.- AWS Academy Cloud Operations.- AWS Academy Cloud Security Foundations.	<ul style="list-style-type: none">- AWS - Optimizing Presto SQL on Amazon EMR to Deliver Faster Query Processing.- AWS - Lower TCO and Increase Query Performance by Running Hive on Spark in Amazon EMR.- AWS - Migrating from Oracle to Amazon Redshift.- Medium - Configuration Management and Continuous Deployment.
ACADEMIC PROJECTS	

Streamlined Food Crate Tracking with Raspberry Pi.	09/2023 to 11/2023
<ul style="list-style-type: none">- Orchestrated the successful integration of Raspberry Pi technology into food distribution crate management, resulting in a remarkable 95% improvement in lifecycle tracking efficiency and substantial reduction in financial losses.- Real-time Analytics Platform Development: Innovated a real-time analytics platform on AWS using Lambda, SQS, Glue, and Athena, leading to a 90% improvement in data processing and delivery for actionable insights on delivered orders.	
Rental Car Application.	02/2023 to 04/2023
<ul style="list-style-type: none">- Designed rental car system to optimize user experience, improve operations, and personalize recommendations. Constructed intuitive UI for easy booking, real-time availability, with automated serverless data lake on AWS S3, Glue, Lambda, API Gateway with streamline Analytical workflows.- Harnessed AWS services (RDS, DynamoDB, Dax, Route 53, and S3) to architect and deploy the application, ensuring high scalability and rock-solid reliability.	
Developed a Web Application For IT Asset Management.	09/2022 to 11/2022
<ul style="list-style-type: none">- Operationalized a ServiceNow-based ITAM system, reducing software licensing costs by 90% through accurate asset tracking with automated workflows for asset onboarding and disposal within the ITAM system, improving asset recovery by 15% and utilizing data analysis techniques within the ITAM system to identify underutilized assets, leading to a 10% reduction in hardware expenditure.	
NLP Text Summarization.	09/2022 to 11/2022
<ul style="list-style-type: none">- Engineered and launched a web application featuring an array of NLP text summarization models for efficient content analysis and deployment on GCP Cloud Run, resulting in a 25% improvement in processing speed and deployment efficiency.	

EDUCATION	
MASTER OF SCIENCE: INFORMATION TECHNOLOGY Arizona State University, Tempe, Arizona – IRA A. Fulton Schools of Engineering	08/2022 to 05/2024
BACHELOR OF ENGINEERING: COMPUTER ENGINEERING University Of Mumbai, India	07/2013 to 05/2016