

# NARESH UPADHYAY

## Data Engineer & Architect

[info.naresh74@gmail.com](mailto:info.naresh74@gmail.com) | 908-435-1312 | [LinkedIn](#) | Rancho Cucamonga, CA 91730



### Professional Summary

Seasoned Data Engineer with extensive experience across AWS, Azure, and GCP platforms, specializing in the development of scalable data architectures and ETL pipelines for healthcare and consumer applications. Proven track record in optimizing data workflows, enhancing analytical capabilities, and ensuring compliance in highly regulated environments. Adept at driving innovations and leveraging advanced analytics to deliver actionable insights, leading to significant business growth and operational efficiency.

### Professional Experiences

#### Cigna Healthcare, Franklin Lakes, NJ

August 2021- Till Date

##### Sr. Data Engineer

Cigna is a global health service company dedicated to improving the health, well-being, and peace of mind of those it serves. The project involved the development and maintenance of a unified Medicaid Claims & Enrollment system, utilizing cutting-edge AWS data solutions such as Redshift, S3, and Glue to optimize data flows and enhance reporting capabilities. This initiative significantly improved data accessibility and analytical processes, supporting Cigna's goal to provide seamless health services.

- Engineered and maintained robust data pipelines, focusing on the integration of AWS technologies including S3, Redshift, and Glue to enhance data storage, transformation, and retrieval processes.
- Led the migration of on-premises Data Lake solutions to AWS, utilizing S3 for scalable data storage and Glue for serverless data integration and ETL services, thereby optimizing data processing workflows.
- Implemented Apache Airflow for advanced batch scheduling, significantly improving the management and automation of data workflows on AWS.
- Developed and optimized data aggregation and transformation processes within a Medallion architecture using Databricks, facilitating efficient data operations across bronze, silver, and gold data layers.
- Authored SQL and Python scripts for data cleansing and validation in AWS Redshift, ensuring high data integrity and supporting advanced analytics initiatives.
- Participated in the design and execution of scalable data models on AWS, enhancing data accessibility and analytics capabilities across the organization.
- Streamlined and unified legacy data management processes by integrating AWS data services, improving data flow and system efficiency.
- Contributed to strategic data architecture discussions, emphasizing the use of AWS solutions in achieving optimal data storage, processing, and distribution.
- Produced comprehensive documentation on data architecture and ETL strategies, including detailed usage of AWS services, to guide future enhancements and maintenance.
- Engaged in continuous learning and application of AWS technologies to stay ahead of trends and improve project outcomes in data engineering.
- Supported the creation of predictive models and data analytics frameworks using AWS Redshift and Python, driving data-driven decision-making and strategic insights.

**Tool Stack:** Spark, Kafka, AWS S3, AWS Redshift, AWS Glue, Apache Airflow, Databricks, SQL, Python, Medallion Architecture, Stream Processing, Batch Processing, Lambda Architecture.

#### Amway, Ada, MI

Aug 2020 – July 2021

##### Data Engineer

Amway is an entrepreneur-led health and wellness company based in Ada, Michigan, dedicated to improving lives in over 100 markets globally. As a Senior Data Engineer, I designed & developed Google Cloud Platform (GCP)-based data architecture for clinical trial management, directly enhancing the company's ability to launch and analyze new health products efficiently and compliantly.

- Designed and implemented a data architecture for clinical trial management using Google Cloud Platform (GCP), enhancing compliance and analytical capabilities for health product trials.
- Developed automated ETL pipelines leveraging Google Cloud Dataflow and BigQuery, achieving a 30% reduction in processing time for critical analytics in cancer medicine sales.

- Implemented scalable and robust data quality frameworks using Google Cloud services, significantly improving data handling capacities for complex health product analyses.
- Enhanced data extraction and integration processes with Google Cloud Pub/Sub and Data Fusion, ensuring high-quality data for accurate and reliable healthcare decision-making.
- Managed comprehensive GCP data solutions, maintaining a 95% success rate in project implementations, which supported real-time and predictive analytics crucial for healthcare strategies.
- Established stringent data governance protocols using GCP's security and compliance services, ensuring adherence to regulatory requirements for sensitive healthcare data.
- Optimized data storage solutions using Google Cloud Storage and BigQuery, resulting in substantial cost efficiencies and enhanced performance through strategic data management practices.
- Ensured continuous availability of analytics-ready data by integrating Google Cloud Composer for workflow management, supporting advanced data science applications in healthcare.
- Directed initiatives for data integrity and accuracy using Google Cloud's data quality tools, setting high standards for organizational data practices.
- Demonstrated flexibility and expertise in cloud-based data management by incorporating GCP into existing data infrastructure, facilitating scalable and secure data solutions tailored for healthcare needs.

**Tool Stack:** ETL, Spark, Kafka, Hadoop, Google Cloud Dataflow, Google BigQuery, Google Cloud Pub/Sub, Google Data Fusion, Google Cloud Storage, Google Cloud Composer, GCP security and compliance services.

**Nandi Technology Kathmandu, NP**  
**Software Engineer**

Jan 2020 - July 2020

- Spearheaded the development of Android applications, including Mahajodi, GaadiBazaar, and Education Tech, enhancing user engagement through intuitive interface design and robust functionality.
- Utilized Node.js to build and maintain scalable and efficient server-side applications, ensuring seamless data handling and application logic.
- Developed and maintained data pipelines using SSIS for efficient data staging and aggregation, contributing to the robust ETL processes essential for application performance.
- Engaged in comprehensive back-end development, creating core database schemas and data integration services for multiple applications, improving data consistency and access.
- Implemented automated scripts in Python for job scheduling, data cleansing, and sample data analysis, optimizing system operations and supporting continuous integration.
- Designed user-friendly interfaces for mobile applications, focusing on multi-sensory engagement and immersive user experiences.
- Contributed to the creation of detailed documentation for application operations, testing protocols, and data integration specifications, enhancing team understanding and project reproducibility.
- Managed the integration of diverse data sources, including non-relational data and flat files, into targeted databases through both SSIS tools and manual processes.
- Developed rigorous test cases and documentation to ensure high-quality software performance across various Android devices and versions.

**Tool Stack:** Node.js, Android, Python, SQL, SSIS, JavaScript, REST API, ETL

**Educa Technology, Kathmandu, NP | Application Developer Intern**  
**NCIT, Kathmandu, NP | ML Engineer Intern**

March 2019- Dec 2019  
July 2018- March 2019

Academic Qualifications & Certifications

**Bachelor's in Computer Engineering (2020)**  
**Pokhara University, NCIT, Kathmandu, NP**  
  
Microsoft Certified: Azure Data Engineer Associate

Microsoft Certified: AI Fundamentals  
Microsoft Certified: Data Fundamentals  
AWS Certified: Cloud Practitioner

## Technical Skills

<b>Data Engineering &amp; Architecture:</b>	Spark, Kafka, AWS S3, AWS Redshift, AWS Glue, Apache Airflow, Databricks, Medallion Architecture, Stream Processing, Batch Processing, Lambda Architecture, Google Cloud Dataflow, Google BigQuery, Google Cloud Pub/Sub, Google Data Fusion, Google Cloud Storage, Google Cloud Composer, SSIS.
<b>Cloud Solutions Implementation:</b>	AWS S3, AWS Redshift, AWS Glue, Apache Airflow, Google Cloud Platform (GCP), Google Cloud Dataflow, Google Cloud Pub/Sub, Google Data Fusion, Google Cloud Storage, Google Cloud Composer, GCP security and compliance services.
<b>Software Development:</b>	Node.js, Android, Python, SQL, JavaScript, REST API, Android Studio, Android SDK.
<b>Data Integration and ETL Processes:</b>	SSIS, Spark, Kafka, Hadoop, Google Cloud Dataflow, Google BigQuery, AWS Glue, Apache Airflow.
<b>Predictive Modeling and Analytics:</b>	Python, SQL, PyTorch, Pandas, NumPy, Matplotlib
<b>Database and Data Warehouse Management:</b>	SQL, AWS Redshift, Google BigQuery, Databricks, Apache Airflow.
<b>Project Documentation and Compliance:</b>	Confluence, JIRA, MS Office, Google Docs.

## References Available Upon Request