

## PROFILE SUMMARY

Accomplished AWS Data Engineer with over 7 years of extensive experience in conceptualizing, developing, and managing cutting-edge data solutions on the AWS platform. Adept at constructing scalable data pipelines, refining data storage infrastructure, and upholding impeccable data quality standards. Demonstrated success in steering multifaceted data projects to fruition within stipulated timeframes and budgets.

## TECHNICAL SKILLS

- Programming Languages: Python, SQL
- Cloud Platforms: Amazon Web Services (AWS)
- Data Warehousing: Snowflake, S3
- ETL Tools: Glue, Lambda, DBT, Apache Airflow, Redshift, Snowflake
- Database Technologies: Amazon Redshift, PostgreSQL, MySQL
- Big Data Technologies: Apache Spark
- Data Visualization: Tableau, Power BI
- Version Control: Git, Azure Dev-Ops(ADO)
- Operating Systems: Linux, Windows

## PROFESSIONAL SUMMARY

- |   |                             |
|---|-----------------------------|
| <b>Ernst &amp; Young LLP India Ltd., Pune</b> <ul style="list-style-type: none"><li>• Data Engineer</li></ul>           | <b>(Sept' 22 – Present)</b> |
| <b>Infosys India Ltd., Bengaluru</b> <ul style="list-style-type: none"><li>• Data Engineer</li></ul>                    | <b>(Dec '19 – Aug' 22)</b>  |
| <b>ITC InfoTech India Ltd., Bengaluru</b> <ul style="list-style-type: none"><li>• Data Analyst, Data Engineer</li></ul> | <b>(Aug '16 – Dec '19)</b>  |

## PROFESSIONAL EXPERIENCE

- |   |                             |
|---|-----------------------------|
| <b>Ernst &amp; Young LLP India Ltd.</b><br><b>Client: Novo Nordisk</b> <ul style="list-style-type: none"><li>• Environment : AWS, Glue, Snowflake, DBT, Lambda, Power BI, Python, Linux, S3, Redshift, PostgreSQL, Spark</li><li>• Duration : 10 Month</li><li>• Role : Data Engineer</li><li>• Business Domain : Pharmaceuticals</li></ul> | <b>(Sept '22 – Present)</b> |
|---|-----------------------------|

### Key Responsibilities:

- Collaborated across departments to deliver data-driven insights and solutions in the pharmaceutical domain.
- Provided pivotal technical guidance to pharmaceutical clients.
- Designed and executed end-to-end data pipelines on AWS, employing Glue, Lambda, Step Functions, and S3 for efficient data ingestion, transformation, and loading into Snowflake data warehouse.
- Orchestrated intricate data workflows using AWS Step Functions, enhancing operational efficiency.
- Ensured data quality via comprehensive checks and monitoring utilizing AWS services and custom scripts.
- Leveraged DBT for agile data modeling and transformation, facilitating rapid deployment of pipelines.
- Spearheaded the establishment and maintenance of a centralized data catalog and governance framework using Databub.
- Microsoft Power Flow to streamline Continuous delivery of data from different channels like Outlook, Teams.
- Implemented robust failure and error notification mechanisms using AWS CloudWatch Alarms and SNS.
- Collaborated with Azure DevOps team to automate deployment and version control of data engineering artifacts.
- Offered expert guidance on data engineering, management, and quality assurance best practices within the pharmaceutical sector.

- |   |
|---|
| <b>Client: AstraZeneca</b> <ul style="list-style-type: none"><li>• Environment : AWS, Glue, Snowflake, DBT, Lambda, Power BI, Python, Linux, S3, Redshift, PostgreSQL, Spark</li><li>• Duration : 8 Month</li></ul> |
|---|

- Role : Data Engineer
- Business Domain : Pharmaceuticals

#### Key Responsibilities:

- Led the design and implementation of data pipelines for processing and analyzing clinical trial data, ensuring compliance with regulatory requirements such as HIPAA and FDA regulations.
- Implemented real-time data streaming solutions using Apache Kafka for monitoring and analyzing drug manufacturing processes.
- Designed and optimized data models in Snowflake and DBT for efficient storage and retrieval of pharmaceutical data.
- Collaborated with data scientists and bioinformaticians to integrate genomic and biomarker data into pharmaceutical research and development workflows.
- Conducted data governance and quality assurance activities to ensure the accuracy, completeness, and consistency of pharmaceutical data.
- Collaborated with clients to understand their data requirements and design scalable data architectures on AWS.
- Implemented end-to-end data pipelines using AWS services such as S3, Glue, and Redshift, ensuring optimal performance and reliability.
- Developed and maintained ETL processes using DBT, transforming raw data into usable formats for analysis.
- Conducted performance tuning and optimization of data storage and retrieval processes, resulting in significant cost savings for clients.
- Provided technical leadership and mentorship to junior team members, fostering a culture of continuous learning and growth.

---

**Infosys India Ltd.**

**(Dec '19 – Sep' 22)**

**Client: Levi Strauss & Co.**

- Environment : Power BI, Excel, Python, Linux, S3, AWS Glue, AWS Redshift, Lambda, PostgreSQL, Spark
- Duration : 26 Month
- Role : Data Engineer
- Business Domain : Fashion, Retail Footwear and Apparel

#### Key Responsibilities:

- Developed and maintained Digital Product Creation and Colorway Predictive Solution framework using AWS services like Glue, Redshift, and S3 for efficient data processing and predictive analytics.
- Collaborated with Levi's Denim team for demand forecasting and style planning, utilizing AWS Glue and Redshift for data extraction, refinement, and transformation.
- Implemented robust data pipelines on AWS Glue and S3 to automate large-scale dataset ingestion, ensuring data quality and integrity.
- Led the development of interface and API solutions for outbound data processing, integrating AWS Lambda functions for real-time processing.
- Engineered ETL processes for image processing through scripting automation, optimizing data workflows for efficiency and scalability.
- Designed and implemented end-to-end data pipelines using AWS Glue, S3, and Redshift, ensuring seamless data integration.
- Conducted batch processing and data transformation using AWS Glue and Spark for efficient data processing.
- Collaborated in developing business analytics dashboards on Power BI for comprehensive product attribute management.
- Utilized AWS data services for consumer demand analysis, enabling insights into style preferences based on cluster and region data.
- Supported retail merchandising initiatives with AWS data services, facilitating promotional effectiveness analysis and assortment planning.

---

**ITC InfoTech India Ltd.**

**(Aug '16 – Dec '19)**

**Client: Marks & Spencer**

- Environment : Python, Linux, Power BI, AWS S3, AWS Glue, AWS Redshift, SQL, AWS Lambda
- Duration : 18 Month
- Role : Data Engineer
- Business Domain : Retail, Fashion.

**Key Responsibilities:**

- Implemented predictive analytics and recommendation systems for colorways using AWS Glue, Redshift, and S3, facilitating data-driven product development decisions.
- Contributed to data mining efforts by extracting and transforming raw data using AWS Glue and Redshift for analysis and modeling.
- Developed predictive visualization dashboards for product components, leveraging AWS data services for scalable and interactive visualization.
- Played a pivotal role in central merchandising product management, utilizing AWS solutions to automate and streamline product lifecycle processes.
- Contributed to enhancing and upgrading existing systems as part of the SAAS+ team, leveraging AWS services to implement new functionality and enhance system performance.
- Participated in code retrofitment projects for system upgrades, ensuring seamless integration and compatibility with existing infrastructure using AWS data services.

---

**Client: Amazon Fashion and Tools**

- Environment : Power BI, Microsoft Excel, Python, Linux, AWS (S3, Athena, Lambda, SNS, SES, Dynamo DB)
- Duration : 18 Month
- Role : Data Analyst
- Business Domain : E-Commerce

**Key Responsibilities:**

- Worked as part of Demand Forecasting Team to Support Budgeting and planning for Listed Items by reliable forecasting.
- Quantities, Price and Size Forecasting as per customer based features, Location.
- Automated Build Deployment Process in Amazon Web Servers modules.
- Prepared Attribute Data Modeling with Client requirement and Layout Group Creation.
- Involved with Solution Design and preparing ADMs Document and Layout designing requirements.
- Responsible to implementation of Workflow and lifecycle specific to object in system.
- Coordinated with onsite team for System Configuration, validation and implementation.
- Involved in workshops with client to ensure understanding of Business requirement for system.
- Data Creation for Analytics Visualization and workshops.
- Worked on Mapping of Objects and layout configuration, Enumeration.

---

**INTERNSHIPS****Full Time Intern****(Mar '17 - Jul '20)****The School Of AI**

Bengaluru, Karnataka

- Developed web pages and conducted image tagging for traffic data.
- Engaged in computer vision model building and image classification

---

**CERTIFICATIONS**

- **Computer Vision** Certificate from The School of AI
- **AWS Solution** Architect – (Ongoing with LinkedIn Learning)
- **Machine Learning** Certification from Coursera
- **Python Developer** Certificate from The School of AI
- **Data Engineer** Certificate from LinkedIn Learning

---

**EDUCATION****C-DAC, Bangalore****(Jan '16 - Jul '16)****Development of Advanced Computing**

Bengaluru, Karnataka

- PG-Diploma with 6.4 CGPA

**B.E. Electronic and Communication****(Apr '11 - Jun '15)****Indore Institute of Science and Technology**

Indore, Madhya Pradesh

- Bachelors with 7.83 CGPA

**Declaration:** I hereby declare that all the above information furnished is true to best of my knowledge.**Date:****Place:** Pune**MAYANK SHARMA**