***Pratham***

***Sr Data Engineer***

***+1 945 218 2002****prathampatil0895@gmail.com*

***PROFESIONAL SUMMARY***

*Results-oriented Data Analyst with over* ***9+ years of experience in AWS data engineering,*** *analytics, and enterprise data migration using AWS services. Understanding of SnowFlake cloud technology. Skilled in designing scalable data models, automating ETL pipelines, and developing insightful BI dashboards.* *Extensive experience in Text Analytics, generating data visualizations using R, Python and creating dashboards using tools like Tableau, PowerBI. Experienced in BI reporting tools like Tableau, Power BI and Cognos. Understanding of structured data sets, data pipelines, ETL tools, data reduction, transformation and aggregation technique, Knowledge of tools such as DBT, DataStage Expertise in AWS tools such as Redshift, S3, Glue, and QuickSight, with a strong foundation in SQL and Python for data validation and transformation. Proven ability to manage end-to-end data workflows, enforce data governance, and ensure secure handling of sensitive data. Adept at collaborating with cross-functional teams and aligning data solutions with business objectives in Agile environments.*

***Cloud Platforms:*** *AWS (Redshift, S3, Glue, Athena, Lambda, QuickSight)*

***Languages & Scripting:*** *SQL, Python, Shell scripting*

***ETL Tools:*** *AWS Glue, DataStage, Informatica, Apache Beam*

***BI & Reporting:*** *QuickSight, Tableau, Power BI, Excel Macros, Looker*

***Database Systems:*** *IBM DB2, Redshift, OLAP/OLTP Systems*

***Data Handling:*** *Data Modeling, Data Wrangling, Data Migration, Reconciliation, Data Validation, PII Masking*

***Project & Process:*** *Agile, SDLC, IAM, RBAC, Stakeholder Engagement*

***EDUCATION***

***Master of Science,*** *Computer and Information Sys****.*** *May 2024*

*New England College, NH*

***Bachelor of Engineering in Computer science*** *May 2015*

*The Oxford College of Engineering, VTU, India.*

***TECHNICAL SKILLS***

***PROFESSIONAL EXPERIENCE***

***Sr. Data Engineer/ Analyst, Cardinal Health, Dublin, OH September 2024 – Present***

* *Involved in Data Modeling and implemented scalable data models using AWS Redshift to support business intelligence, analytics and reporting needs across diverse domains.*
* *Worked on creating ER diagrams for the understanding of the structure, relationships and constraints of data and to validate the queries post migration.*
* *Created detailed data mapping specifications and transformation rules for migrating data from IBM DB2 to Amazon Web Services (AWS).*
* *Conducted data profiling, cleansing, and validation to ensure data integrity during the migration process.*
* *Created Excell macros for creating repetitive reports during and after the migration process to track the changes during the process.*
* *Created and maintained interactive dashboards and KPI reports post migration.*
* *Created view for reconciliation and ensuring that transformations didn’t alter data and to verify data integrity.*
* *Collaborated with data governance team to ensure the quality of the data and data validation.*
* *Utilized AWS cloud storage for staging data files and AWS Redshift for analytics.*
* *Creates QlikSense data extractor (ETL) that coverts relational and NoSQL databases into QVDs*
* *Worked with data governance team on identity and access management (IAM) to classify data (PII) and define user- based roles.*
* *Converted Talend Joblets to support the snowflake functionality.*
* *Used AWS Glue during ETL process for batch data processing and informatica for simplified workflows.*
* *Worked on extracting structured data from DB2 database in to text(.txt) files for ingestion and processing in AWS services.*
* *Designed and deployed AWS Lambda functions to automate data validation tasks and streamline post-migration reconciliation processes.*
* *Created several types of data visualizations using Python and Tableau. Extracted Mega Data from AWS using SQL Queries to create reports.*
* *Develop reports and dashboards using Qlik Sense. Scrum Master of Agile scrum to track and organize team efforts (using Rally).*
* *Implemented the machine learning algorithms using python to predict the quantity a user might want to order for a specific item so we can automatically suggest using kinesis firehose and S3 data lake.*
* *Implemented AWS S3 lifecycle policies to optimize storage costs and ensure proper archival of historical datasets.*
* *Leveraged AWS CloudWatch for monitoring ETL jobs and setting up alerts for data pipeline failures and performance issues.*
* *Used Spark SQL for Scala & amp, Python interface that automatically converts RDD case classes to schema RDD.*
* *Loaded the tables from the azure data lake to azure blob storage for pushing them to snowflake*
* *Created reusable data modules and data sets in Cognos to empower self-service analytics for end-users.*
* *Used AWS Athena to query and analyze large-scale datasets directly from S3, reducing the need for additional infrastructure.*
* *Connected to AWS Redshift through Tableau to extract live data for real time analysis.*
* *Worked on writing, testing, debugging SQL code for transformations using data build tool (DBT).*
* *Created data sharing between two snowflake accounts.*
* *Automated nightly build to run quality control using Python with BOTO3 library to make sure pipeline does not fail which reduces the effort by 70%.*
* *Lead a team of Data Scientists and Engineers to develop and productionize various versions of Machine Learning (ML) and Natural Language processing models in novel microservices architectures of batch scoring and real-time API serving of predictions.*
* *Configured AWS Glue Data Catalog to manage metadata and improve discoverability and governance of datasets across analytics teams.*
* *Designed, developed, and managed Power BI, Tableau, QlikView, Qlik Sense Apps including Dashboard, Reports, Storytelling*
* *Created and maintained SQL Server scheduled jobs, executing stored procedures for the purpose of extracting data from Oracle into SQL Server. Extensively used Tableau for customer marketing data visualization*
* *Microservices based architecture for Machine Learning (ML) based applications. XGboost implementation.*
* *Built extract / load / transform (ETL) processes in the Snowflake Data Factory using DBT to manage and store data from internal and external sources.*
* *Develop Informatica mappings to be implemented based on client requirements and for the analytics team.*
* *Experience in ETL techniques and Analysis and Reporting including hands on experience with the Reporting tools such as Cognos.*
* *Create end-to-end solution for ETL transformation jobs that involve writing Informatica workflows and mappings.*
* *Collaborated in setting up cross-account data sharing using AWS Redshift Data Sharing for secure and real-time access to analytics across business units.*

***Sr. Data Engineer, Connecticut Department of Labor, Hartford, CT January 2023 – October 2024***

* *Contributed to the migration of legacy IBM Cognos systems to modern AWS Platform (AWS) services, enabling scalable, cloud-native analytics and reporting capabilities.*
* *Built automated data workflows using DataStage and Python to simplify data movement and cut down on manual tasks, making the migration process faster and more efficient.*
* *Tuned SQL queries during migration to make data transfer smoother and faster, improving overall system performance.*
* *Conducted rigorous data validation, reconciliation, and post-migration audits to ensure integrity and compliance.*
* *Created and published the Data Migration Project reports on daily and weekly progress.*
* *Created and delivered both regular and custom reports using AWS QuickSight to support decision-making for both executive and operational teams.*
* *Utilized SQL and Python to automate data validations and reconciliation during migration.*
* *Tested the migrated reports in AWS QuickSight from start to finish, making sure they met business needs, were accurate, and aligned with the KPIs—ensuring they were consistent with the reports from the old Cognos system.*
* *Streamlined data ingestion workflows to ensure Cognos reports are refreshed with low-latency, real-time data.*
* *Built automated data pipelines with Cloud AWS Glue and Apache Beam to simplify the migration process, reducing manual work and making batch data transfers from IBM DB2 to AWS faster and more efficient.*
* *Developed data warehouse model in snowflake for over 100 datasets using whereScape.*
* *Creating Reports in Looker based on Snowflake Connections*
* *Involved with development of Ansible playbooks with Python and SSH as wrapper for management of AWS node configurations and testing playbooks on AWS instances.*
* *Worked closely with system administrators to optimize Linux servers for temporary data storage, improving both performance and security during the data staging process in the migration.*
* *Build Dashboards using QlickView, QlikSense and integrate them into the portal with (Mashup and Dev-Hub)*
* *Analyzed huge volumes of data Devised simple and complex HIVE, SQL scripts to validate Dataflow in various applications. Performed Cognos report validation. Made use of MHUB for validating Data Profiling & Data Lineage.*
* *Built automated reporting solutions with AWS QuickSight, giving business users easy access to key performance metrics on-demand, so they no longer had to rely on manual reports and could make quicker, more informed decisions.*
* *Configured EMR cluster for data ingestion and used DBT (data build tool) to transform teh data in Redshift.*
* *Validating the data from SQL Server to Snowflake to make sure it has Apple to Apple match.*
* *Used data masking techniques to anonymize personally identifiable information (PII) during migration, making sure sensitive data was protected before being moved to the cloud and staying compliant with privacy regulations.*
* *Develop stored procedures/views in Snowflake and use in Talend for loading Dimensions and Facts.*
* *Developed a new data scheme for the data consumption store for the Machine Learning (ML) and AI models to quicken the processing time using SQL, Hadoop, and Cloud services.*
* *Involved in creating Unix shell scripts for Informatica work flow execution.*
* *Designs, develops and manages QlikSense and Tableau Apps including Dashboard, Reports, Storytelling*
* *Installed and managed Qlik Sense Enterprises Server*
* *Develop Informatica SCD Type-2 Mappings based on the requirements.*
* *Facilitated training sessions to demo DBT tool for various teams and sent weekly communications on different topics related to Data Engineering.*
* *Data visualization: Pentaho, Tableau, D3. Have knowledge of Numerical optimization, Anomaly Detection and estimation, A/B testing, Statistics, and Maple. Have big data analysis technique using Big data related techniques i.e.,Hadoop, Map Reduce, NoSQL, Pig/Hive, Spark/Shark, Mllib and Scala, NumPy, SciPy, Pandas, scikit-learn.*
* *Developed Python AWS serverless lambda with concurrent and multi-threading to make the process faster and asynchronously executing the callable.*
* *Configured IAM policies and RBAC to restrict access to sensitive data, ensuring that only authorized roles could access or unmask PII during the migration process.*

***Data Engineer, Linksfield’s Innovation Ltd, India September 2019 – July 2022***

* *Dug into sales and customer behavior data using AWS Redshift to spot trends in how businesses were buying electric appliances, helping the marketing team focus their efforts where it mattered most.*
* *Built interactive dashboards in AWS QuickSight that gave teams real-time insights into what was selling, how fast, and through which channels—making it easier to stay on top of performance.*
* *Grouped B2B customers into segments based on how often and what types of products they bought, helping sales and marketing teams tailor their messaging more effectively.*
* *Pulled data from CRM, AWS Pinpoint, and AWS tools to paint a full picture of the customer journey—from first website visit to final purchase—so teams could better understand what drives sales.*
* *Looked at past sales and inventory trends in AWS Redshift to spot patterns in product demand, helping the team plan better and avoid running out of stock or over-ordering.*
* *Tracked how well marketing campaigns were performing by creating clear, easy-to-read reports, helping the team make smarter decisions about where to spend their budget.*
* *Proficient in developing Data Visualization's using Tableau Software from diverse sources.*
* *Kept an eye on the competition by bringing in external pricing and market data, helping shape smarter pricing and positioning strategies for our products.*
* *Chunking the data to convert them from larger data sets to smaller chunks using python scripts which will be useful for faster data processing.*
* *Saved time by automating regular reports, using AWS Redshift and AWS QuickSight to make sure everyone had the latest data without having to manually update anything.*
* *Worked closely with product and sales teams to understand which appliances were most popular with different types of customers, helping plan promotions and bundling strategies more effectively.*
* *Responsible for writing Autosys JILs and scheduling the Informatica workflows on the Autosys server*
* *Analyzed customer churn patterns to figure out which clients were likely to stop buying and why, so the business could reach out in time and keep those relationships going.*
* *Looked at customer retention over time through cohort analysis, giving the team insight into how well we were keeping clients engaged and coming back.*
* *Automated Cognos report delivery and archival using Python and Cognos SDK/API.*
* *Lead and Owned all the steps of a Machine Learning (ML) service development Lifecycle from exploratory analysis to deployment of model in production.*
* *Real - time experience in using Visualization tools like Power BI, Tableau.*
* *Developing Spark applications using Scala and Spark-SQL for data extraction, transformation, and aggregation from multiple file formats. Using Kafka and integrating with the Spark Streaming. Developed data analysis tools using SQL and Python code.*
* *Conducted Data blending, Data preparation using Alteryx and SQL for Tableau consumption and publishing data sources to Tableau server.*
* *Performs advanced analytics by integrating QlikSense with R and Python*
* *Worked with informatica support for fixing Informatica Linux server issues.*
* *Migrated the entire Oracle databases into Qlik repository (QVD file) and automated the data flow*
* *Worked closely with the marketing team to see how campaigns were performing in different regions and with different types of customers, helping them fine-tune their messages and reach the right people.*
* *Update Strategy, Aggregator, Expression, Joiner Transformations and then loaded into data Warehouse using Informatica BDM 10.2.2.*
* *Created custom dashboards in AWS QuickSight that made it easy for the sales team to track which products were doing well, which ones weren’t, and how customer buying habits were shifting.*
* *Integrated Cognos BI with enterprise data warehouse (Snowflake/Redshift/BigQuery) to unify data reporting and governance.*
* *Evaluate Snowflake Design considerations for any change in the application*
* *Build the Logical and Physical data model for snowflake as per the changes required*
* *Looked at how customers interacted with the website alongside sales data to understand what they were interested in before buying—helping the team improve product pages and marketing content.*
* *Brought together data from tools like CRM, Google Ads, and product systems to give a full picture of each customer, making it easier to spot chances for upselling or offering related products.*

***Data Analyst, Inspire India Pvt Ltd, India August 2015 – September 2019***

* *Analyzed SIP investment trends and customer behavior to help teams better understand investor needs and support more informed financial planning.*
* *Created and maintained regular reports on SIP inflows, redemptions, and fund performance using Excel and SQL, helping stakeholders track key metrics with ease.*
* *Built dashboards in Power BI and Tableau to visualize portfolio trends and provide real-time insights for business and advisory teams.*
* *Worked closely with finance and product teams to clean and validate investment data, making sure reports were accurate and aligned with business goals.*
* *Conducted portfolio analysis for SIP clients, helping advisors recommend personalized investment adjustments based on goals and market shifts.*
* *Assisted in developing simple financial models to project returns for various SIP strategies, making it easier for clients to plan long-term.*
* *Collaborated with support teams to identify common SIP issues and provide data-backed suggestions for improving customer experience.*
* *Transformed raw investment data into meaningful summaries and visuals for business reviews and decision-making.*
* *Monitored performance KPIs like SIP cancellation rates, retention, and investor value, sharing insights to guide product and marketing strategies.*
* *Helped ensure compliance by supporting data documentation and cross-checking historical records for accuracy during audits.*
* *Authoring Python (PySpark) Scripts for custom UDF’s for Row/ Column manipulations, merges, aggregations, stacking, data labeling and for all Cleaning and conforming tasks. Migrate data from on-premises to AWS storage buckets.*
* *Supported financial analysis for new product launches, helping identify key customer segments and ensuring that the right investment products were marketed to the right audiences.*
* *Worked on Oracle Databases, RedShift and Snowflakes*
* *Define virtual warehouse sizing for Snowflake for different type of workloads.*
* *Collaborated with the marketing team to analyze customer acquisition data and assess the effectiveness of campaigns promoting SIPs, contributing to strategy refinement.*
* *Provided insights into SIP performance across different regions to assist in understanding regional demand and optimizing product offerings.*
* *Regularly worked with large datasets to extract actionable insights, improving reporting efficiency and supporting data-driven decision-making for senior leadership.*