ROHIT SHARMA

Gurugram, Haryana • +91-8076242806 • rohit.singh999063@gmail.com

SUMMARY

- Accumulated 17.2 years of comprehensive experience as Technical Delivery Manager within esteemed IT establishments, including Accenture, Wipro, and Pitney Bowes.
- Specialized in crafting innovative solutions encompassing cloud data lakes, data marts, data vaults, data migration, data transformation, data preparation and modern data warehouses design.
- Proficient in project management and designing solutions, consistently showcasing a successful track record in formulating and executing cloud computing data strategies while ensuring robust data management and security across hybrid and multi-cloud environments.

SKILLS

Skills: Project Management, Team management, Stakeholder management, DLT, Delta Live Tables, Pyspark, SparkSQL, Big Data, Azure Data Factory, Azure Data Bricks, Synapse, Glue, Hive, HDFS, YARN, S3, EC2, AWS Lambda, Amazon Web Services, Statistical Modeling, Hive, Hadoop, ETL, Java, API Gateway, SNS, SQS, Lambda, and Step Functions, Modeling and Design, Data Analytics, Big Data Processing, PowerBI, DAX Scripting, DataModelling

CORE COMPETENCIES

Project Management Skills: Demonstrated expertise in Technical Project Management, employing SAFE Agile/Scrum Methodology. Proficient in steering project teams towards delivering high-quality results within defined timeframes and budgetary constraints. Managed end-to-end project phases, encompassing planning, design, scope definition, estimation, resource management, process oversight, and quality compliance.

Client Engagement and Sales Collaboration: I have achieved exceptional client engagement milestones, catering to Fortune 100 clients globally in diverse domains such as Telecom, and Banking, spanning regions including the US, UK, and Middle East. Through cultivating robust relationships and closely collaborating with Sales Teams, I've successfully secured multi-million-dollar deals. Additionally, I actively participate in RFPs across various project phases and contribute to various Proof of Concepts (PoCs) and consulting solutions, demonstrating a proactive and engaged approach throughout project lifecycles.

Architecture Expertise: Proficient in designing large-scale data warehouse solutions and middleware data preparation for applications or AI consumption. Expertise in managing OLTP data model changes and ensuring effective adaptation to evolving requirements. Well-versed in Decentralized Data Mesh frameworks. Extensive experience in executing projects spanning on-premise, multicloud, and on-premise to cloud data migration, covering both Greenfield and Brownfield scenarios. Specialized in solutioning for real-time data streaming and implementing effective data storage mechanisms.

Technology Proficiency: Proficiently skilled in utilizing cutting-edge technologies such as Data Bricks, Azure Data Factory & Synapse, and Data Vault 2.0 Modelling. Adept in designing non-relational Databases like Cassandra and employing ETL and ELT tools. Possesses brief knowledge of Data visualization tools.

Cloud Solutions: Proficient in delivering cloud consulting services on Microsoft Azure, specializing in designing architectures for data movement, implementing multi-region business continuity solutions, and facilitating on-premise to cloud data migration.

Data Management: Expertise in Data Warehousing, Relational Databases, OLTP, OLAP, Business Intelligence (BI), and Big Data Solutions. Proficiency with Synapse, Databricks, MPP databases.

Data Governance & Security: Implementing data governance frameworks like GDPR, ensuring data security, and complying with industry standards. Strong background in data security measures, including encryption and access controls.

Data Modeling: Proficient in designing ER, dimensional, data vault data models for different complex analytical needs.

ETL/ELT & Integration: Extensive experience with data pipelines, data integration, and data quality management with Data Factory, Databricks, Synapse. Designing complex data workflows for efficient data extraction, transformation, and loading.
Metadata & Catalog Management: Proficient in data catalog tools like Databricks Unity Catalog, Expert in metadata management to maintain data lineage and quality.

PROFESSIONAL EXPERIENCE

Accenture Technology, India, IN Jan 2023 – Present
Azure Bigdata Architect

- Managing Azure Bigdata Cloud project setups, resource provisioning, team management, profit and loss management with respect to 2 end to end projects.
- Spearheading cross-functional teams of Bigdata engineers, DevOps professionals, developers, and business analysts, providing strategic leadership and support across the project lifecycle
- Implemented both batch and streaming pipeline simultaneously using Maintenance Views and Streaming in DELTA live tables and maintained the Data Quality with CONSTRAINTS and Monitor the pipelines for discarded tables by doing query one event logs captured as DELTA TABLE only.
- Designed and Implemented CDC in Delta tables to make sure that DELTA table is up to date with source table from KAFKA.
- Reduced redundant activities of 3 departments by building Data Warehouse, Data Modelling and Data Ingestion activities.
- Used DAX scripting in PowerBI to perform complex calculations, create new columns and measures, and filter and summarize data.
- Hands-On expert level knowledge in Draw.io for Dimensional Modelling with SQL Plugin for ER diagrams
- Implemented security model for governing data objects in Unity Catalog, defined and implemented data access rules and managed the data ownership.
- Implemented dynamic views to protect PII data in columns and rows with in a table.
- Designed and implemented data pipelines with architecture patterns such as ETL, ELT, CDC, Lambda, Serverless, Fan-Out/In and Microservices.
- Hands-on experience in designing and implementing e-commerce platforms with a SQL database like MySQL for transactional integrity and later integrate a NoSQL database Cassandra for user activity logs to manage scalability.
- Employed sophisticated data skipping techniques leveraging statistical insights within the Delta LakeHouse Architecture, seamlessly integrated with the Delta Engine.
- Devised an innovative solution for the efficient management of numerous small files through the strategic application of BinPacking and Compaction techniques, complemented by optimization and Z-ordering methodologies. This solution optimizes storage utilization and enhances data retrieval efficiency, showcasing a commitment to streamlined data management.
- Implemented a comprehensive Medallion architecture, harnessing Delta Lake's robust capabilities. This meticulously designed framework optimizes data for advanced applications in data science and machine learning at the silver layer, while employing a refined aggregation approach at the gold layer for Business Intelligence (BI) and reporting
- Streamlined ETL processes involve extracting files from the raw folder to populate the bronze table, creating a dynamic versioned view. Data modifications are then merged into the silver table, differentiating updates and inserts, followed by aggregations. The gold table receives complete overwrite updates, establishing it as the repository for BI and reporting, while the silver table serves as a dedicated resource for machine learning endeavors.

Cluster Cost Optimization

Identify and optimize the cost of hundreds of Azure Databricks Clusters.
First, we understood how to calculate the real cost of the cluster that we want to optimize, we discussed with the cluster/application owner, understand their needs (e.g., do they need to run 24/7, do they need lots of processing, or can they wait a few minutes), and optimize the Databricks cluster. Then a few days later, we examined the data and savings.

Identify the Total Monthly Cost of Each Cluster
We have used Microsoft Generate Cost Details Report API to get all the data for the Azure subscription where our Databricks clusters are running. With this we identified different components and their cost, including IP Address, storage, virtual network, virtual machines, Databricks DBU, etc. We have generated the raw and formatted data and saved the daily and total daily costs in local storage or Azure Blob Storage so it can be used to generate custom Power BI reports. In order to calculate the cost of the Databricks cluster, we have summed up cost of different components within a specific time frame. Because all the data was mixed up, we had used Tags to filter and sum up the right resources. Databricks adds tags to each Azure Resource so we can look at the tags and figure out which Databricks Cluster, Databricks SQL Warehouse, or Databricks Computer Job uses a specific resource. We have used JSON response from Databricks Clusters List API to extract important data such as Cluster ID and Cluster Name, Cluster Owner, Cluster Specifications (Cores and Memory), Spark Version, Configuration (min and max number of nodes). We have optimized cluster settings with number of hours for which cluster should run instead of 24/7, avoided photon acceleration, optimized VM size with spot instances and reduced the number of nodes. We have saved cost up to 92%.

Wipro, Gurugram, Haryana, India March 2015 – Jan 2023

Technical Delivery Manager- Azure Bigdata

- The end-to-end architecture for this project involved reading raw data from the source via API method using Azure Data Factory and importing it into ADLS Raw containers in JSON format. Then, Databricks was used to perform data ingestion to ADLS Ingestion Layer and transformation to the ADLS Presentation layer. Later, data was analyzed and visualized through Databricks and Power BI. Azure Data Factory served as an orchestration tool to monitor and schedule the pipeline. On the Databricks website, three different layers processed through DataBricks notebooks were described as Bronze (where raw data was loaded), Silver (where data was filtered and cleaned), and Gold (where data was transformed through business logic).
- Gathered requirements and collected source data
- Designed solution architecture for our requirements
- Setup environments on Azure and initiated Databricks cluster
- Grant access for Databricks to Azure Data Lake
- Loaded source data into Azure Data Lake Storage
- Ingested and transformed data into Parquet formats on Databricks
- Created data pipelines in Azure Data Factory
- Scheduled and monitored data pipelines on Azure Data Factory
- Created database and tables from Parquet files in Databricks
- Analyzed data in Databricks and visualized data in Power BI
- Load the processed data into Hive External table and to create Partitioning and Bucketing techniques in Hive to improve the performance, involved in choosing different file formats like ORC, parquet format reducing 50-man hours per activity
- Design and Implemented Data Pipeline, reducing data processing time by 25%
- Implemented data quality checks, resulting in a 15% improvement in data accuracy.
- Collaborated with stakeholders and Data Scientist Team to Integrate their models into production workflows.
- Conducted performance tuning on SQL queries, optimizing data retrieval by 20%.
- Migrated On prem Data Distribution to Azure Cloud Distribution and Hive Data Warehouse using PySpark
- Orchestrated collaboration among stakeholders to delineate precise technical requirements, enabling the creation of innovative solution designs aligned with overarching business goals; achieved a 20% increase in revenue through improved conversion rates.
- Architected and maintained software components, harnessing the power of Python, PySpark, and Hadoop, to enable seamless processing and analysis of extensive datasets, generating a 60% improvement in efficiency and driving superior business performance.
- Harnessing SQL for comprehensive data querying, insightful analysis, and optimizing database performance to ensure efficient data retrieval.
- Architected and implemented highly scalable data solutions, integrating batch and real-time processing, resulting in a 40% increase in system performance and 50% reduction in data processing time.
- Created HLD and LLD for complete solution using different AWS and DevOps Tools and Services.
- Managed Team and Project with Program Manager from DevOps Perspective.

- Automated complete Infra and RPA UiPath Applications and services using Packer, GitLab CICD pipelines and terraform. These codes helped Wipro and Customer to save their more than 375+ man-hours of work every time in this project and for future deployments.
- Automated Custom AMI creation using Packer with builder, provisioner, winrm, different Json files, .ps1. sh files. These custom AMI's have 16 complex business applications including RPA applications from UiPath.
- Automated Patch Management and Back-Up services using AWS Patch Manager (Scan and Install for Kb articles) and Backup services with Terraform (IaC) within the region and across the region for EBS and RDS snapshots.
- Designed and tested restoration of RDS across the region using Automated backup and within the region using continuous point in time recovery.
- Developed and implemented Python Boto3 scripts to automate the copying of AMIs to the Disaster Recovery (DR) production region upon EC2 instance creation in the source Ireland region. Utilized Terraform scripts as automation tools for EC2 instance creation in the DR scenario, effectively managing instance operations, including stopping, volume detachment, snapshot selection, volume conversion, volume attachment, and instance startup, ensuring a reliable DR setup.
- Created and Implemented resources and practices like custom AMI's, EC2, S3, RDS, Back-up, Patch Management. SSM start stop Instances, DR, using tools like PACKER, TERRAFORM, GITLAB CICD.
- Architected and deployed a cutting-edge middleware solution leveraging AWS services API Gateway, SNS, SQS, Lambda, and Step Functions. By seamlessly integrating DataDog Alarms and Monitors with CloudByz Ticketing Tool, this solution eliminated manual ticket creation, resulting in a 60% reduction in incident response time and a 75% increase in team productivity.

TCS , India June 2006- Feb 2015

Solution Consultant (AWS, Data Enterprise)

- Perform Data Cleansing, Data Enrichment and Data Integration of Telecom BSS and NSS related data related to telecom Operators using SQL query and Partners Data Enterprise Platform.
- Leveraged technical expertise to customize core products for pre-sales demos, POCs, and pilots, addressing specific requirements of various use cases such as CEM, CCM, Data technologies, and
- Perform the Operation and Management activities on Linux Servers to supervise, manage and operate nodes which transmit and receive signals. Also write down and edit advanced Linux scripts to collect info from nodes.

EDUCATION

MDU BTech, ECE May 2002-May 2006