



Narendra .
Senior Data Engineer

Professional Summary

- Senior Data Engineer with 12+ years of global experience developing and maintaining scalable data pipelines and ETL processes for companies like IBM, Deloitte, and AWS. Extensive experience designing/ implementing cutting-edge enterprise application systems and data lake solutions/pipelines for obtaining data from rest APIs and database sources. Garnered Deloitte Applaud, R&R, and TMT Outstanding Awards, as well as IBM India Excellence & Eminence Award for data warehouse.
- Extensive Knowledge in designing of data pipelines using **AWS, IAM, Redshift, Athena, Cloud watch, Boto 3 SDK, S3, KMS, Lambda, Dynamo db, AWS Glue, Lambda, Sage maker, Ec2, VPN, subnet, cloud trail, SNS, QuickSight, Azure blob storage, Azure Data bricks.**
- Strong demonstration of creation of **streaming and batch** data pipelines using **Data bricks, HDFS, Hive, Spark, Scala, Kafka, Pyspark and Python** and managing **governance of objects in Unity catalog**, configuring and developing of **CI/CD pipelines** in Azure dev ops.
- Experienced working on **delta tables, auto loader, workflows, data bricks workspace, Azure git repo, data bricks access connector.**
- Expert in implementation of well versed pipeline to **apply data quality checks** to cleanse and produce finite data to down stream layers **Silver, Gold tables in medallion architecture.**
- Proficient in implementation of Business Intelligence solutions using Data Warehouse/Data Mart Design, OLAP, BI, Client/Server applications.
- Strong understanding of **Data warehouse principles and methodologies** - Ralph Kimball and Bill Inman and implemented start and snowflake dimension models.
- Highly skilled in designing of dimension and fact tables in star and snow flack dimension modeling.
- Demonstrated implementation of Design using of various ETL technologies like **IBM DataStage, Informatica and Tera data.**
- Solid experience of understanding complex logic and performance tuning of **SQL queries, PL/SQL procedures, Functions and triggers** to reduce query execution times.
- Extensively Designed and developed common **utility functions, re-usable ETL components** for data validation and data quality, audit processes, Job monitoring and error logging processes.
- Excellent knowledge orchestration of data pipelines with **Airflow, AWS Glue workflows, Auto sys and TWS.**
- **Highly hands on experience** integrating wide variety of data sources including **XML, JSON, PARQUET, AVRO, ORC, Kafka topics, Cloud watch logs and Dynamo Db** and ingest data into data lake.
- Experienced in **importing and exporting data** using Sqoop from HDFS to RDBMS and integrating **Hive with Spark environment using Spark SQL** and used Hadoop distributions like Cloudera and Horton works, worked on to transform data using **RDD, Data frames and Datasets.**
- Worked in fast faced projects in **Agile, waterfall** using **JIRA** and execution of sprint user stories. Have shown ability to continuously learning new technologies, look after to apply learned techniques in implementation of projects.
- Expertise in preparing of comprehensive **Design Documents / Test Case Specifications / Performance/Best practices** to follow to build Data Pipelines, other components.

Technical Skills

Cloud	<i>AWS Glue, S3, Redshift, DynamoDb, CloudWatch logs, S3 encryptions, IAM, KMS, Lambda, Ec2, EMR, RDS, Athena, CloudTrail, VPC Endpoint, ELB, Rout53, Subnet, Azure Data bricks, Azure blob storage, Azure Data Lake, Data factory.</i>
Big Data	<i>HDFS, Kaafka, Hive, Impala, Sqoop, Pig, HBase, Flume, Zookeeper, Oozie, Databricks</i>
Databases	<i>Oracle, MySQL, SQL Server, MongoDB, HBase, Teradata, DB2, Netezza.</i>
Programming Languages	<i>SQL, Shell, HQL, Scala, Python, C, C++, Bash, Pyspark</i>
Spark Components	<i>Pyspark, Spark Core, Spark SQL (RDD, Data Frames and Dataset), Scala, Glue Dynamic Data frames</i>
ETL tools	<i>IBM DataStage, Informatica, Teradata Utilities</i>
IDE	<i>Eclipse, NetBeans, IntelliJ, Pycharm, Visual Studio</i>
Version Control	<i>GitHub, SVN, GitLab. Azure DevOps Git Repo</i>
Environment	<i>UNIX, IBM AIX 5.3, AIX 6.1, Linux, Windows Xp/7</i>
Others	<i>SQL Plus, SQL*Loader, TOAD, Autosys 4.5, Tivoli Workload Scheduler, Airflow and HPQC, Jira, Zen hub, Tableau, QuickSight, Terraform , Pulumi, Monto Carlo, Airflow, Dbt, NiFi.</i>

Professional Experience

RIVIAN

Sr. Data Engineer

Jan 2024 - On going

Rivian is an American electric vehicle manufacturer, project is to collate data from various data sources and provide to the end users for analytics use cases, processing of raw and unstructured data coming from sensors, API's and logs to process and visualize business metrics, support commercial, Charging, Facilities and remarking Analytics.

Roles and Responsibilities :

- Creation and grooming of sprint user stories received from functional downstream system owners.
- Design of streaming and batch pipelines to propagate data through bronze, silver and gold layers using Data bricks.
- Ensuring of data quality and data accuracy before giving datasets to downstream system owners.
- Drafting design and architectural data flow diagrams to support individuals to implement ETL pipelines.
- Orchestration of data pipelines in airflow and configuring Manto Carlo to ensure availability and accuracy of data.
- Working with cross functional domains commercial, re marketing, PDS, supply chain and Finance teams to deliver required datasets for analysis.
- Configured unity catalog and data governance activities, creation of tables in all layers and providing access to groups and users on objects.
- Building robust end to end data pipelines using AWS redshift, AWS glue, Python, Spark, S3, Airflow and terraform.
- Promoting code using CI/CD pipeline, resolving Git conflicts and deployments.
- Preparing of drafts on daily, weekly and quarterly status on key performance metrics and providing updates to leadership team.

Technical Environment: AWS Glue, S3, Data bricks, Redshift, Athena, Dynamo Db, cloud watch logs, Secrets, Manager, IAM, Airflow, Python, Pyspark, GitLab, Sql, Unix, Db viewer, Confluence, Jira, Tableau, Airflow

Sr Data Engineer

Bank of America initiated a project to make Smooth transition of GRDS to GRRS(global regulatory reporting system) in eliminating of unwanted over head components in the system and making system robust enough to apply more functionality and improve to handle future upcoming requirements. And applying technology to provide lineage and tracking data points from targeted reporting system to source layer to give confidence to end users in addressing various queries.

Roles and Responsibilities :

- Understanding of current state architecture of the system and design to modernize system.
- Liaising with stake holders and creation of mapping documents from source to target.
- Building data ingestion engine, bring source data from multiple heterogeneous source systems, building strategy to apply transformation to load data into target system.
- Validation of data accuracy between new modernized system with respect to existing system's data and root cause analysis to fix the issues in new system.
- Data quality and validation checks to impose to ensure to send finite data to global regulatory reporting system (GRRS)
- Creation of checkpoints and logging inputs to database to support the lineage tracking of granular level of target reporting layer records to source and giving confidence in addressing various insightful questions.
- Orchestration of data pipelines, managing and supporting existing data pipelines to process data within SLA times.
- Co-ordinating with offshore team, managing of deliverables, peer review of data pipelines code.
- Documentation of low level designs and managing of release of code into SIT, UAT and production environment.

Environment: AWS, Glue, Hadoop, Spark, Scala, Sql, Data bricks, Git, Confluence, Jira, Mongo Db, Nifi, Unix.

AWS**San Jose****March 22 - June 23****Senior Data Engineer**

Amazon Fraud Detector (AFD) is a fully managed service that provides scalable, easy-to-use real-time fraud detection for e-commerce and online services. AFD evaluates online activities and account sign ups or online orders using a combination of machine learning(ML) and rules to assess the risk that the activity is fraudulent. AFD provides tailored-made fraud detection ML models that learn from their historical fraud data and from past attempts to defraud amazon. Project is to build a Data Lake to get insights on revenue, usage and operational activities of AFD service. Includes to collect data from billing & metering , cloud watch logs and model performance metrics from model version Dynamo DB tables and transform unstructured data to the meaningful structured data for analysis and proactively reaching customers to resolve their blockers while training ML models for fraud detection to improve the customer experience.

Roles and Responsibilities:

- Collaborated with the Product managers, gathered requirements and converted them to a technical driven approach to design and build ETL components using Boto3 SDK, AWS Glue, Spark and Redshift.
- Built robust data pipelines to fetch a variety of data from DynamoDB, Cloud watch logs using Glue and developed dashboards in quick sight to understand the customers behavior of model creation in Amazon fraud detector service.
- Have written design docs to blend and convert data to load into aggregate tables into Redshift and get the security clearance from App sec team.
- Owned end to end implementation of ingesting, transforming, aggregation of data and creation of reporting dashboards.
- Coordinated with Data Scientists, understood data requirements. Extracted Amazon/AWS registration and other market place datasets and provided to Data Science team for Machine learning experiments.
- Built pipelines to send daily email updates on model performance metrics, revenue and usage for the AFD service. Owned to provide updates on project status, summary of ad-hoc analysis about the performance of service to leadership team.
- Performed extensive data analysis, shared insights and findings to management to drive business decisions and initiatives. Handled change requests, understood the application, work flows and existing data pipelines and applied new changes.
- Automated to send notifications to stack holders on various billing metrics.

- Created IAM roles & policies to interact with multiple services to allow to build data pipelines. and taken over admin activities to upgrade, provision and maintain redshift clusters and defined back-up/archival of data strategies in redshift. Handled end to end migration activities to centralized system.

Environment: AWS cloud, Ec2, S3, Glue, Redshift, Athena, Dynamo Db, Cloud Watch Logs, IAM, Python, Pyspark, Airflow, Terraform, CICD, Secrets Manager, Lambda, Quick sight, Glue Crawler.

Deloitte Experience

Apple Inc
Cupertino, CA

Jan '20 – Feb 22

Sr. Big Data Engineer

Apple Inc offers a range of Apple media products such as Apple Music, Arcade, Tv+, News+, Fitness+, iCloud services. With the implementation of this new project to re-vamp existing subscriptions BI system to a new standard and consistent system to serve all products and sourcing to power all Exec reporting/ Hydra, internal & external reports. This helps easy governance of Hydra dashboards/KPI metrics and helps in providing better customer service experience.

Roles and Responsibilities:

- Analysis of subscription business drivers and finding out root cause for variances, fixing gaps in logic to deduce business driver metrics.
- Implementation of Data pipelines to read data from HDFS and write into Teradata tables using Apple cloud infrastructure (ACI) in Spark.
- Design Data flow framework to process AMP data projections; Providing solution to process huge volume of data to meet timelines.
- Performance tuning of spark jobs using re-partition and broadcast variables.
- Implementation of prototypes and validation of KPI metrics.
- Addressing data issues/concerns raised by Internal & External reporting teams, validation and solving the data gaps between GBI projections versus future state AMP projections.
- Automation of data pipelines to build aggregated subscription projections, pumping data to Daily, Weekly & Monthly Dashboards for Data Science and Analytics team at Apple.
- Provided Hybrid solution of multithreaded, scalable and robust data processing pipelines to process data in parallel and configurable to increase threads as data grows near in-future. Successfully launched projection pipelines to refresh full ~12 billions of records daily.
- Achieved to match KPI's variance for various metrics between GBI and AMP projections $\leq \sim 2$ to 5%.

Environment: Data bricks, Python, PySpark, Scala, Tera data, Unix Scripting, Airflow, Tableau, Zen Hub, AWS Cloud, and Gitlab.

HPE
Paul Alto, CA

Feb '19 – Dec '19.

Sr. Big Data Engineer

Client is one of leading provider in servers, storage, networking, consulting and support, and Financial Services, headquartered in USA and have offices all over the world.

This project is for creating an integrated big data and advanced analytics platform that unlocks data insights for business and enhances customer experiences. Existing processes present in EDW are being migrated to Big data platform. Data from different sources are stored in Hadoop cluster as required for different processes is transformed and stored in Hive tales. From here Business Users can fetch reports related to sales, forecast and Inventory.

Roles and Responsibilities:

- Designed data model for Supply chain shipment & Delivery.
- Created framework for data ingestion process to process source data files.
- Developed re-usable data pipelines using Scala & Spark to load data into Hive tables.
- Consumption of data from the Kafka topics to Hive and HDFS using Spark Application.
- Supported model right from reading data from Kafka to load data into ingestion and consumption layer and Scheduling spark jobs.
- Performance tuning of spark data pipelines and peer review of code for deployment.
- Defect Fixing; UAT support, migration of code into higher environments.
- Creation of various documents including High level and low-level Technical Design document.

- Obtain sign-off from stack holders on project deliverables and data models.

Environment: Hadoop, Hive, HDFS, Python, Spark, Scala, Json, Talend, HpQc, GitHub, IntelliJ, AWS Cloud Services.

**Bank of America
Hyderabad, India**

Oct '17 – Jan '19.

Senior ETL consultant

Client is one of the world's largest financial institutions serving individual consumers, small and middle market business and large corporations with a full range of banking, investing, asset management and other financial and risk management products and services.

Client embarked on a risk data aggregation (RDA) program using extensive data controls and information management capabilities. Client needed a partner to accelerate the sourcing of brokerage and trust data into an Authorized Data Source (ADS) in alignment with their Risk Reporting requirements and accelerate Data Lake creation for the benefit of Global Wealth and Investment Management (GWIM) Analytics.

Roles and Responsibilities:

- Migration of RDBMS system in Hive and Data Lake using Spark and Python.
- Kafka been introduced to read streaming data from our real time systems.
- Handled nested files like XML, Json files in Spark
- Creation of Hive Tables to feed the data which is been split according the segments and identifiers.
- Ingestion of COBOL files using Podium to hive tables
- Import data from RDBMS to HDFS system (from Oracle and other RDBS systems) using Sqoop.
- Oozie and Airflow (P.O.C) workflows in DAG for scheduling the ETL and Spark pipelines.
- Responsible for user reports extraction by following the business rules and aggregations.
- Creation of Hive ingestion, extraction, and transformation with various hive performance techniques
- Optimization of the hive and Spark queries and jobs for high performance.
- Performance tuning for spark jobs and migrated MR code functionality to spark code.
- Shell scripting (Unix) to automate file validation and triggering of data pipelines.

Environment: Hadoop, Hive, HDFS, Python, Spark, Scala, Json, Talend, HpQc, GitHub, IntelliJ, AWS Cloud Services.

**QBE Insurance, Health Care
Hyderabad, India.**

May '15 – Sep '17.

Senior ETL Consultant

Client is one of the largest insurers and it provides insurance services mainly to Australia, America, Europe and Asia Pacific Region. The GAEE will allow a single, centrally supported approach to earnings be used by all divisions, which will reduce differences in approach across the group and reduce the need for manual calculations. Earnings will be calculated from GWP/RWP, which are calculated upstream of the GAEE, using earnings profiles that describe the percentage of premium to be earned in each discrete period within the risk period. The profiles will include patterns that replicate risk incidence, including linear annual earnings patterns. Commissions and acquisition costs will be earned in line with the premium where possible, and in line with an appropriate earnings profile otherwise. The output of the engine will be transactions which represent changes in unearned premium or deferred acquisition costs.

Roles and Responsibilities:

- Requirements gathering, translating functional requirements to technical, effort estimation and prioritizing the activities, leading and mentoring the offshore team.
- Developing and tuning the SQL scripts for data extraction and data validation.
- Designing DataStage ETL components and scripts, Performance Tuning of ETL components.
- Writing UNIX shell script to automate loading process, Preparation of migration plan, ETL code migration checklist,
- Documentation of TDD and maintaining issue tracker, resolutions and code changes.
- Status updates, conducting reviews, managing of testing team in functional and technical testing activities,
- Conducting workshops and facilitating KT sessions with external vendor, supporting SIT, UAT phases.

Environment: IBM Info sphere DataStage 11.5/9.1, Oracle 11g/10g, SQL Server, Flat files, UNIX, Control-M.

**C. R. Bard, Inc Health care
Hyderabad, India**

Jun'14 – April '15.

ETL Developer

C. R. Bard, Inc. is a leading multinational developer, manufacturer, and marketer of innovative, life-enhancing medical

technologies in the product fields of Vascular, Urology, Oncology and Surgical Specialty. BARD markets its products and services worldwide to hospitals, individual health care professionals, extended care facilities, and alternate site facilities.

Roles and Responsibilities:

- Requirement analysis and gathering to provide technical and architectural support to the team
- Documented the Purpose of Mapping as to facilitate the personnel to understand the process and incorporate the changes as and when necessary.
- Used DataStage Designer to develop processes for Extracting, Cleansing, Transforming, Integrating, and Loading data into Data warehouse
- Worked on logical and physical design of Data warehouse, identified sources/targets, and analyzed source data for dimensional modeling.
- Implemented SCD Type1 and SCD Type2 using different stages in for building the EDW
- Developed the data warehouse repository using DataStage Manager by importing the source and target database schemas
- Extensively designed, developed and implemented Parallel Extender jobs using Parallel Processing (Pipeline and partition) techniques to improve job performance while working with bulk data sources
- Extensively worked with architects and proposed solutions in building common design approach for building the Job control and error recording tables
- Created Naming Standards Document and Low-Level Design Documents and used them across the Projects.
- Extensively Worked on Job Sequences to Control the Execution of the job flow using various Triggers (Conditional and Unconditional) and Activities like Job Activity, Email Notification, Sequencer, Routine activity and Exec Command Activities
- IBM Info sphere 8.5 and 9.1 were utilized where the project also involves migration to 9.1
- Defect Tracking, unit testing, defect reporting, analyzing results and documentation.

Environment: IBM DataStage 8.5/9.1, Microsoft SQL 2005/2008, IBM DB2, PUTTY, WINSCP and HP Quality Center

**Cooper vision
Hyderabad, India.**

Sep '13 – May'14.

ETL Developer

It is one of the world's largest manufacturers of contact lenses. Their products are sold in over 100 countries around the world. From China to Europe, from Africa to the Americas.

The client is seeking assistance in implementing a Business Intelligence ("BI") solution that will enable more efficient and reliable decision making by delivering relevant and accurate information and business insight to all its employees in a timely and consistent fashion. The resultant BI solution will consider disparate information sources across the organization including analysis of any in-flight projects that would be impacted by the BI Strategy and architecture.

Roles and Responsibilities:

- Experience in testing mobile features in different environments, which includes Process of new claims, and payment related features
- Conducted source information profiling using Quality Stage Investigation stage and Information Analyzer
- Developed parallel jobs to load data into XML files using XML input and XML output stages
- Extensively used Quality Stage patterns like CASS, Standardize, Match Frequency, etc. to standardize name, address and consumer information
- Ensured Data Quality and similarity matches using Quality Stage Patterns
- Implemented Surrogate Key Generator Stage, Slowly Changing Dimension Stage etc.
- Worked on the DataStage server in both AIX and UNIX simultaneously using virtual machine.
- Created the Enterprise Data Warehouse Data Migration ETL design document utilizing Info Sphere DataStage and Quality Stage
- Created the Quality Stage match specifications while training the employees in Quality Stage and DataStage concepts and programming
- Created project plan, monitored the progress on all tasks while working closely with the data modelers and data base administrators to create the conceptual, logical and physical design of the Enterprise Data Warehouse.
- Implemented Slowly Changing Dimension (SCD) as per business requirement for saving history using WebSphere DataStage 8.1 for dimension tables in Star Schema
- Used DataStage Director and the runtime engine to schedule running the server jobs, monitoring scheduling and validating its components
- Extensively worked on Quality Stage Methodologies in Information Server 8.5 for Data Cleansing and

Standardizing

Environment: IBM DataStage 8.0.1, IBM AIX 5.2, Oracle 10g, XML files, sequential flat files, TOAD.

IBM

Client: CELCOM, Malaysia
Kolkata, India.

March '10 – Sep '13.

ETL Developer

CELCOM embarks into the journey of proposed strategy, CELCOM requires a review of the current data warehousing, reporting and analytical system to provide business information that is secure, accurate, timely, consistent, integrated appropriately detailed, well organized and easy to obtain so that users through the organization will be better able to perform day-to-day operations, decision making, process improvement and planning. With this background, the Project CELCOM Enterprise Business Intelligence (EBI) has been launched in order to help support the business objectives of CELCOM. **Roles and**

Responsibilities:

- Worked on End-to-end implementation, maintenance, optimizations and enhancement of the application.
- Worked on various phases in the project like project estimation, Requirement Gathering and Analysis, Designing, Coding, User Acceptance Testing to deployment in production, including post implementation support.
- Assisting the Management and Quality team (QMS) by providing various Metrics and data for the project.
- This is important as View360 projects are very process oriented and emphasize on Quality standards.
- Adhering to the Process and creating and posting all the required documents and deliverables like the UTC document, Review Checklist, etc.
- Created a data dictionary to map the business requirements to attributes to design logical data model
- Developed Data validation and error handling jobs.
- Assisting the Project Management team in providing the estimation of projects, so we could plan estimate the effort and time by when we could deliver the best solution.
- Extensively worked with DataStage Shared Containers for Re-using the Business functionality
- Worked on performance tuning to address very critical and challenging issues
- Co-ordinate with SIT and UAT team to fix the Test Problem Reports
- Worked closely with Project lead/Manager, Architects, and Data Modelers, System Analyst to understand the business process and functional requirements.

Environment: DataStage 8.1, Oracle 9i/10g, DB2, flat files, PL/ SQL, Windows 2003 Server, SQL Server.

Standard Chartered Bank

Finance
Hyderabad, India.

Aug '06 – March '10

ETL developer

Implemented and transformed legacy application with new Electronic branch banking system EBBS to ease banking operations, Migrated customer demographic and credit & transactions data from legacy to EBBS application.

Education

Master of Science, Acharya Nagarjuna University, Andhra Pradesh, India

Certifications

- IBM Data Stage V8.1 & V8.5 Certified Professional.
- Db2 9 Fundamentals Certified Professional.
- Certified in Tera data professional 12.
- Certified Vantage Advanced Sql.

Key Achievements

- Received TMT Outstanding award in Deloitte.
- Received multiple Reward and Recognition (R&R) for contributions on projects in Banking and Financial Services.
- Received Excellence & Eminence Award for the implementation of Enterprise data warehouse and contribution of excellent client services in IBM India Pvt Ltd from TAC HEAD.