**Raja Shekhar**

**Professional Summary:**

* Worked around 10+ years in Data Warehouse / Data Marts, ETL Systems in traditional and Big Data environments and gained good knowledge in Data analysis with incremental SQL.
* Databricks Certified Data Engineering Associate programmer.
* Experience in architecting, designing, and implementing complex Extract, Transform, Load (ETL) solutions. Proficient in Informatica Power Center, AWS cloud services, and Kafka for real-time data processing.
* Have hands-on experience in designing, implementing, and optimizing data warehousing solutions on cloud platforms such as Amazon Redshift and Snowflake.
* Proficient in database design, ETL development, query optimization, and performance tuning to deliver scalable and high-performance data warehouse solutions.
* Good hands-on in Azure PaaS services like Azure Data Factory (ADF v2), HDInsight Spark, HDInsight Interactive Query Clusters, ADLS Gen 2 File systems and Azure Data bricks.
* Design ETL systems in my latest project using Databricks Data Lake Architecture with new concepts like DLT tables, Unity Catalog and Delta sharing
* Designed scalable Spark Batch ETL (using Scala) application on Big Data cluster for huge volumes of Data.
* Successfully implemented various optimizations in spark data processing using different techniques like Data salting, executors assignment and broadcasting and performed cluster sizing for Spark based on data volumes.
* Experience in Building common ETL framework using Scala/Spark API and IntelliJ IDE to handle all ETL features like Business Transformations, SCD1/ SCD2 and Dimensional type Jobs and ETL subsystems like Job metadata management, Reusable framework for repeated ETL task .
* Strong problem-solving abilities combined with excellent communication and collaboration skills to effectively work within cross-functional teams and deliver solutions that meet business requirements
* Working As Data Engineer Lead, Designed Data Engineering systems for various clients for different use cases using technologies like AI,ML,Azure PAAS, ADF, Azure Data Bricks, AWS
* Good experience in IICS (Informatica Intelligent Cloud Services), Informatica Cloud Application Integration (CAI), Informatica Cloud Data Integration (CDI), MQ, Azure Monitor.
* Strong experience in migrating other databases to snowflake.
* Experience with snowflake cloud data warehouse for integrating data from multiple source systems which include loading nested JSON formatted data into a snowflake table.
* Experience in using snowflake clones and time travel.
* Participates in the development and improvement of snowflake database applications.
* Hands-on experience with snowflake utilities, Snow SQL, Snow pipe using python.
* Good experience in Big Data technologies with hands-on experience using Hadoop, Kafka, Apache Spark, Hive, Impala.
* Experience on Migrating SQL database to Azure Data Lake, Azure data lake Analytics, Azure SQL Database, Data Bricks and Azure SQL Data warehouse and controlling and granting database access and Migrating On premise databases to Azure Data Lake store using Azure Data factory.
* Experience in Developing Spark applications using Spark - SQL in Databricks for data extraction, transformation and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
* ETL pipelines of data warehousing combination of python and snowflakes snow SQL writing SQL queries against snowflake.
* Good experience with Oracle, MongoDB, Big Query and knowledge of container runtime environments (Docker, Kubernetes) and implementing applications onto Public Cloud (AWS/Azure Platform).
* Good understanding of Spark Architecture including Spark Core, Spark SQL, Data Frames, Spark Streaming, Driver Node, Worker Node, Stages, Executors and Tasks.
* Extensively used ETL tools like Informatica client tools –Mapping designer, Informatica Repository Manager and Informatica Workflow Manager and Informatica BDE
* Expert knowledge in moving data between on prime and Cloud
* Have extensive experience in site reliability engineering
* Good development skills in Informatica Cloud (IICS) in Data Integration and Application Integrations tasks like Mass Ingestion Tasks, Mapping, Task Flows and scheduling these using Autosys.
* Good experience with technical architecture and software design, designing and implementing distributed enterprise applications with Microservices using Java/Python and with the design and implementation of API’s – REST Services, Open (API’s) application programming interface.
* Entrusted as technical lead on multiple projects ranging from small BAU requirement changes to large complex regulatory reporting projects
* Strong leader with experience in training developers and advising technical groups on ETL best practices.
* Good knowledge in Linux and python scripting.
* Good knowledge in enterprise metrics/monitoring with frameworks such as Splunk.
* Have extensive experience in solving the tasks for scalability, performance and stability.
* Extensive experience in writing UNIX shell scripts and automation of the ETL processes using UNIX shell scripting.
* Good knowledge in RDBMS systems like Oracle and Teradata
* Experience in using Automation job Scheduling tool Appworx and Autosys
* Have knowledge on stash and Jenkins for code deployment.
* Worked on Jira tool for reporting issues and updating the status on development tasks.
* Worked in Agile methodology for project execution.
* Worked on numerous Offshore – Onshore model Projects and had very good exposure in client interactions.
* Having extensive experience developing ETL technical documentation, mapping documents and creating Test Scenarios and Test Cases.
* Good Knowledge on Data modeling concepts like Star Schema and Snowflake Schema, Dimensions and Fact tables.
* Have good knowledge in big data stacks.
* Specialized in ETL Performance Optimization to improve the performance of the production load.
* Have good knowledge in real-time, streaming and data pipelines frameworks.
* Experience with cloud computing services, particularly deploying and running services in Azure.
* Good experience with Agile software development frameworks and (CI/CD) continuous integration/continuous deployment and development using Jenkins, GitHub and delivering full deployment lifecycle automation, designing, developing, and maintaining complex applications.
* Have excellent analytical & problem-solving skills in business problems. Excellent team worker. Self-motivated and quick learner. Flexible to a fast growing and changing environment.

**Technical Skills:**

|  |  |
| --- | --- |
| ETL Tools | * IICS, Informatica Power center 10.X/9.X, Talend, Azure Data Factory
 |
| Amazon Web Services | * S3, Redshift, Athena, Glue, Kinesis, RDS, EMR, Aurora, and EC2
 |
| Hadoop Ecosystem | * Pivotal HD, Hive, Hawq, Pig, Sqoop,spark
 |
| Databases | * Snowflake, Redshift, Oracle, Teradata, DB2, MS-SQL Server
 |
| RDBMS Tools | * SQL Developer, TOAD, Dbeaver
 |
| Scheduling Tools | * Appworx, Control-M, Autosys
 |
| Domain knowledge | * Aviation, Pharmaceutical, Banking and Hospitality
 |

**Project Summary:**

**Client: KPMG, New York, NY Jan 2023 to present**

**Role: Lead Data Engineer**

**Project Description:**

Data pipelines will be built to extract, transform and load the data to Azure, which in turn will be utilized as a source data for multiple downstream systems and processes Data Wrangling, Massaging and Enrichment will be performed in the Data Lake staging area. Here we dealt with our third-party data vendors who provide us with the Marketing and Point of sales data.

**Roles and Responsibilities:**

* Designed and developed a scalable data processing and analytics framework and integrated with Azure cloud services. Migrated on-prem databases and data from third party vendors which is basically marketing data and point of sales data to Azure data Lake, Azure Synapse, Azure Blob, Azure SQL Database, Azure SQL leveraging Azure Data bricks and Azure Data Factory (ADF)
* Extracted, Transformed and Loaded (ETL) data from disparate sources to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL and Azure Data Lake Analytics.
* Worked on the snow SQL and snow pipe.
* Developed a data warehouse model in snowflake for over 100 datasets using whereScape.
* Consulting on the snowflake data platform solution architecture, design, development and deployment focused to bring the data driven cultures across the enterprises.
* Provide guidance to development team working on PySpark as ETL platform
* Developed store procedures/views in snowflake and use in Talend for loading dimensions and facts.
* Evaluate snowflake design considerations for any change in the applications.
* Created ETL/ELT pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform and load data from different sources like Azure SQL, Cassandra, Blob storage, Azure SQL Data warehouse, write-back to and backwards. Built data migration jobs into enterprise Azure cloud and Snowflake.
* Analyzed the SQL scripts and designed it by using PySpark SQL for faster performance.
* Developed Spark applications using PySpark and Spark-SQL for data extraction, transformation, and aggregation from multiple file formats and loaded into the Delta Lake layer.
* Optimization of Hive queries using best practices and right parameters and using technologies like Hadoop, YARN, Python, PySpark.
* Develop framework for converting existing PowerCenter mappings and to PySpark (Python and Spark) Jobs.
* Responsible for estimating the cluster size, monitoring and troubleshooting of the Azure data bricks
* Analyzed and built cloud data solutions using Azure PaaS service to support visualization of data.
* Building the pipelines to copy the data from source to destination in Azure Data Factory (ADF V2)
* Responsible for developing ETL pipelines to meet business use cases by using data flows, Azure Data Factory (ADF), Data Lake, Azure Data warehouse, Azure Synapse.
* Experience in CDL LEVEL 0 APIs-python scripting to bring data into CDL level 0, Performing validation testing with source/ destination teams.
* Developed Databricks Python notebooks to Join, filter, pre-aggregate, and process the files stored in Azure data lake storage.
* Created automated workflows with the help of triggers and Scheduled Jobs in Flows and ADF
* Manage data coming from different sources and loading of structured and unstructured data.
* Worked on different files like CSV, JSON, Flat, Parquet to load the data from source to raw tables.
* Designed and developed Power BI graphical and visualization solutions with business requirement documents and plans for creating interactive dashboards.

**Client: Credit Suisse Securities, New York, NY Sep 2019 to Dec 2022**

**Role: ETL Developer**

**Project Description:**

* Opera is the strategic accounting sub-ledger operational data store (ODS) for Finance, providing a standardized Finance Data Control (FDC) for GL postings, Daily P&L and Balance Sheet Controls and FA regulatory/ treasury processes.
* Opera DAL was central to the FSI Roadmaps primary objective to implement a CFO architecture to
* improve operational control throughout the Investment Bank’s F2B trading flow.
* Develop and maintain centralized Financial Data warehouse for a huge Conglomerate. A wide range of source systems like SAP R3, IBM DB2, Flat Files and SQL Server provide data for different companies and modules of the Conglomerate which has to be integrated into a Centralized Data Warehouse.

**Roles and Responsibilities:**

* Worked closely with project Business Analyst, Data Modeler to ensure that end-to-end design meets the business and data requirements.
* Implemented performance tuning techniques to optimize ETL workflows and reduce data processing times.
* Worked as an individual contributor for development activities of hive, Spark Data ingestion programs in Inventory module .
* Designed and implemented a data quality assurance framework using Informatica Data Quality to ensure data consistency and accuracy across multiple data sources.
* Great hands on in using Azure DevOps to manage build  work item tracking, code management and with CI/CD process end-to-end.
* Involved in the participation and contribution in quality assurance walk-through of ETL and Database components and supports the development teams with performance tuning and troubleshooting issues.
* Involved in the root cause analysis of all processes and resolves production issues including data quality issues and performs routine tests on databases and provides support to ETL applications.
* Monitors all business requirements and validates designs and schedules ETL processes and prepares documents for all data flow diagrams.
* Analyzes and interprets complex data on target systems and analyzes and provides resolutions to all data quality issues and coordinates with business analyst to validate all requirements, perform interviews with all users and developers
* Supports critical design reviews, provides insight into the system as well as component level and develops all data warehouse models and prepares reports for all metadata integration into systems and drafts all ETL scripts and prepares the required reports for all end users.

**Client: Hyatt, Chicago, IL Aug-2017 to July -2018**

**Role: Data Engineer**

**Project Description:**

* Its lift and shift data warehouse project. This project will replace the existing on premises data warehouse with cloud systems, Informatica PowerCenter jobs are migrated to IICS and on premises data warehouse system (DB2) upgraded to Snowflake. Coming to Hyatt business, an enterprise data warehouse (EDW) has been built for the revenue technologies of Hyatt. Revenue technologies are the applications supporting the lines of business that result in direct revenue such as reservations, bookings and point of sales. The EDW also supports certain auxiliary areas like email communication with the guests (reminders, thank you mails etc.) and guest satisfaction surveys.

**Roles and Responsibilities:**

* Migrating Objects from PowerCenter to IICS.
* Involved in end-to-end low-level design of Data ingestion for all the Business subject Areas
* Guided the development team in understanding the ETL low level design for all modules
* Worked as individual contributor in development activities of hive, Scala /Spark Databricks and ADF Pipelines
* Designed reusable ETL frameworks to reduce repeated work to benefit and efficiently use the team.
* Involved in Quality assurance meetings along with the project management team and helped in completing the required artifacts.
* Design and Develop design solutions on critical issues which require expedited fix.
* Involved in the Solution incidents, problems, changes and requests with documentation standards.
* Created internal and external stages and transformed data during load.
* Created Snowflake DDL.
* Created IICS Mapping, Mapping tasks and Task flow.
* Created Control Audit jobs.
* Followed the SDLC principles with Implementation Plan, Test Plan, Backout Plan, Impact and Risk analysis.
* Created Snow pipe for continuous load.
* Mentor the engineers by facilitating KT sessions and development peer reviews.
* Cloned Production data for code modifications and testing.
* Time traveled to recover missed data.
* Involved in testing Snowflake to understand the best possible way to use the cloud resources.
* Coordinate and escalate the issues requiring technical help/support with Level 4 Vendors (Informatica, Microsoft)
* Involve in Design of Functional and Technical Document and also document standard operating procedures and knowledge articles for the L1, L2 support team.
* Preparing the UTP/UTR Test cases also Integration Test cases.
* Involved in the process of ServiceNow Incident Management, Problem Management, Change Management process

**Client: Hyatt, Chicago, IL Jan-2016 to July-2017**

**Role: ETL Lead**

**Project Description:**

* The project will replace the existing travel agent (TA) commission process managed in the Opera PMS. Due to the complexity of the existing process in Opera, the project is deemed to be high risk. Hyatt will contract with Onyx Payments, a Dallas based company, to handle the processing of the commission transactions generated through this project. All activity associated with the review of the commission transactions and payment processing is outside the scope of this project.
* Extracts could potentially contain Personal Identity Information (PII) and security measures are required due to the PII being displayed e.g. first and last name for guest, payee name, proxy name, and remit name. No report requirements have been defined at this time.
* Today the commission process is decentralized and processed at each property. Decentralization has led to interpretation that may not be consistent from hotel to hotel. This project, as well as other factors, is going to create better standards and consistency.

**Roles and Responsibilities:**

* Leading a team of 8 and SPOC for Client interactions from Offshore.
* Involved in Requirement gathering, Impact Analysis and Design.
* Running Hadoop streaming jobs to process terabytes of data
* Load and transform large sets of structured, semi structured data.
* New User IDs & New Folder Creations in the repository.
* Granting Additional permissions/privileges to the Repository Users.
* Developed mappings, sessions and workflows in Informatica Power Center.
* Writing UNIX shell scripts and automation of the ETL processes using UNIX shell scripting.
* Work closely with business Analyst to understand the business and With Data modeler to understand the Physical and Logical design of the EDW.
* Designed, developed, and maintained ETL processes using Informatica PowerCenter to extract, transform, and load data from disparate source systems into centralized data warehouses.
* Implemented performance tuning techniques to optimize ETL workflows and reduce data processing times.
* Designed and implemented real-time data streaming solutions using Kafka to support mission-critical business processes and analytics.
* Architected and optimized ETL workflows for performance, scalability, and fault tolerance, achieving significant improvements in data processing efficiency and reliability.
* Collaborated closely with business stakeholders to understand requirements, define data integration strategies, and deliver high-quality solutions that meet business objectives.
* Design and implementation of complex ETL solutions for large-scale data integration projects, leveraging Informatica PowerCenter, AWS cloud services, and Kafka.
* Led a team of ETL developers, providing technical leadership, mentoring, and guidance throughout the software development lifecycle.
* Create and Develop interface using ODI with the help of a mapping sheet.
* Documentation of ETL HLD and TDD for Different stakeholders in the project team.
* Perform Data load from source to STG and then EDW and Marts based on Model created by Data modelers.

**Client: Boehringer-Ingelheim, Hyderabad, India Sep 2013 to Dec 2015**

**Role: ETL Lead**

**Project Description:**

* This project involved developing the Global Reporting Data Store (RDS) on top of Veeva CRM(SFDC) to support their data warehousing needs. RDS is implemented for 70+ countries in Boehringer-Ingelheim, Inc and its ETL consists of Global (core) and Country specific (supplemental) mappings. RDS is used as a source for their respective country's data warehousing and reporting needs like SAS extracts, Vendor Extracts, Global reporting data layer views, Best Address, Sales Force hierarchy, Call Plan History etc.

**Roles and Responsibility:**

* Understanding Requirements of building data marts and ETL design.
* Work closely with business Analyst to understand the business and With Data modeler to understand the Physical and Logical design of the EDW.
* Designed, developed, and maintained ETL processes using Informatica PowerCenter to extract, transform, and load data from disparate source systems into centralized data warehouses.
* Implemented performance tuning techniques to optimize ETL workflows and reduce data processing times.
* Designed and implemented real-time data streaming solutions using Kafka to support mission-critical business processes and analytics.
* Architected and optimized ETL workflows for performance, scalability, and fault tolerance, achieving significant improvements in data processing efficiency and reliability.
* Collaborated closely with business stakeholders to understand requirements, define data integration strategies, and deliver high-quality solutions that meet business objectives.
* Design and implementation of complex ETL solutions for large-scale data integration projects, leveraging Informatica PowerCenter, AWS cloud services, and Kafka.
* Led a team of ETL developers, providing technical leadership, mentoring, and guidance throughout the software development lifecycle.
* Create and Develop interface using ODI with the help of a mapping sheet.
* Documentation of ETL HLD and TDD for Different stakeholders in the project team.
* Perform Data load from source to STG and then EDW and Marts based on Model created by Data modelers.
* Responsible for designing and developing of mappings, mapplets, sessions and workflow for loading the data from source to target database using Informatica Power Center.
* Responsible for Code review.
* Maintained Detailed Design documents.
* Developed deployment plan and deploy functionality incrementally.
* Tuned mappings for improving performance.
* Developed several reusable transformations and mapplets that were used in other mappings.
* Design and development of SCD2 process using both time-stamping and flag method for dimension tables.
* Performed Unit and Integration testing.

**Educational Qualifications:** Master’s in computer science (2018-2019) - State university of New York, Newpaltz.