

**PROFESSIONAL SUMMARY:**

- Around 5+ years of hands - on experience in **ETL Informatica development** and **Business Intelligence solutions**, leveraging **Data Warehouse/Data Mart Design, ETL, SQL SERVER, MSBI, Power BI**, and **Azure Data Engineer** technologies.
- Excellent knowledge of **Agile, Scrum, Waterfall**, and other **SDLC methodologies** for system analysis, design, development, testing, and implementation.
- Designed, developed, and deployed **Business Intelligence solutions** utilizing **SSIS, SSRS, SSAS**, and **Power BI**.
- Utilized **Apache Airflow** for **ETL workflow management** and demonstrated expertise in writing **Python scripts** for workflow implementation.
- Hands-on experience with **Azure Data Lake Analytics** for analyzing structured and non-structured data from various sources.
- Extensive experience in developing **OLAP Cubes** using **SQL Server Analysis Services (SSAS)** and defining data sources, views, dimensions, measures, hierarchies, attributes, calculations, and roles using **MDX**.
- Extensive experience in **Dynamic SQL, Records, Arrays**, and **Exception handling**, with expertise in **data extraction, transformation, and loading (ETL)** using tools such as **SQL Server Integration Services (SSIS)**.
- Developed **robust ETL pipelines** in **Azure Data Factory** to move data from on-premises to **Azure SQL Datawarehouse**.
- Experienced in configuring **Spark Streaming** to receive real-time data from **Apache Kafka** and store it to **HDFS**, as well as optimizing Spark efficiency through techniques such as **persist, cache, broadcast**, and efficient joins.
- Experienced with **Hadoop distribution** and ecosystem components such as **HDFS, YARN, MapReduce, Spark, Sqoop, Hive**, and **Kafka**.
- Highly skilled in creating complex **SQL queries, stored procedures, functions, packages, tables, views**, and **triggers** for multiple **relational databases** Highly skilled in working with **NoSQL databases** and their integration, including **Dynamo DB, Cosmo DB, Mongo DB, Cassandra**, and **HBase**.
- Extensive experience in writing **UNIX shell scripts** and automation of the **ETL processes** using **UNIX shell scripting**.
- Experience with **ETL tool Informatica** in designing and developing complex **Mappings, Mapplets, Transformations, Workflows, Worklets**, and scheduling the **Workflows** and **sessions**.
- Experience in **modeling data** using the **Star schema, Snowflake schema, Data Schema**, and **Transactional modeling**.
- Hands on expertise with **AWS services** such as **S3, EC2, SNS, SQS, RDS, EMR, Kinesis, Lambda, Step Functions, Glue, Redshift, DynamoDB, Elasticsearch, Service Catalog, CloudWatch**, and **IAM**.
- Designed and built a **Data Lake** using **Hadoop** and its ecosystem components, with experience in **fetching data** into Hadoop Data Lake from various databases using **Sqoop**.
- Built a **data warehouse** on **SQL Server&Azure Database**, with experience in creating real-time data streaming solutions using **Apache Spark/Spark Streaming&Kafka** and building **Spark Data Frames** using **Python**.
- Experienced with **version control systems** like **Git, GitHub**, and **Bitbucket** for code management and also responsible for defect reporting and tracking in **JIRA**.

**TECHNICAL SKILLS:**

<b>AWS Services</b>	<b>Amazon EC2, Amazon S3, Amazon Simple DB, Amazon MQ, Amazon ECS, Amazon Lambdas, Amazon Sagemaker, Amazon RDS, Amazon Elastic Load Balancing, Elastic Search, Amazon SQS, AWS Identity and access management, AWS</b>
<b>Azure</b>	<b>Azure Data Lake, Data factory, Azure Databricks, Azure SQL database, Azure SQL Datawarehouse, Azure Functions, Azure Synapse, Azure HD Insight, Azure Blob Storage, Azure Event Hub, Azure streaming Analytics.</b>
<b>Big Data Technologies</b>	<b>Spark, Hadoop, HDFS, Hive, Sqoop, Oozie, Pig, MapReduce, Kafka, Kinesis, Zookeeper, Flume, HBase, Yarn, Apache Spark.</b>
<b>Relational Databases</b>	<b>MySQL, Oracle, Microsoft SQL Server, PostgreSQL complex SQL queries, stored procedures, functions, triggers</b>
<b>Business Intelligence</b>	<b>SSIS, SSRS, SSAS, Power BI, Tableau, OLAP Cubes (SQL Server Analysis Services)</b>

<b>NoSQL Databases</b>	<b>Dynamo DB, Mongo DB, Cassandra, HBase</b>
<b>Programming languages and tools</b>	<b>Python, R, Scala, UNIX, PySpark, SparkSQL</b>
<b>SDLC Methodologies</b>	<b>Agile, Scrum, Safe</b>
<b>Version Control</b>	<b>Git, GitHub, Bitbucket</b>

## **PROFESSIONAL EXPERIENCE:**

**Protein Metrics Cupertino, CA**

**Jan 2023- Present**

**Data Engineer**

- Collaborated with Business Analysts to gather requirements and design a reliable and scalable data pipeline.
- **Pulled** the data from data lake (**HDFS**) and **massaged** the data with various **RDD transformations**. Developed and compared the performance of Spark with **Hive** and **SQL/Oracle** using Scala.
- Converted **Hive/SQL queries** into **Spark transformations** using **Spark RDDs** and **Scala**. Developed **PySpark** and **Spark SQL** code to process data in **Apache Spark** on **Amazon EMR**.
- Developed high-speed data ingestion pipeline using **Scala**, **Apache Spark**, **Akka Streams**, **HDFS**, **Hive**, and **Cassandra**.
- Developed **Python scripts** for data cleaning, analysis, and automating day-to-day activities.
- Developed **shell scripts** and **SQL** for data analysis and quality checks.
- **Model, lift, and shift** custom SQL and transpose **LookML** into **dbt** for materializing incremental views. Performed **Data Analysis** and provided statistical insights using **Python** and **Statistical Algorithms**. Developed dashboards using **Grafana** with TSDB as a data source.
- Developed multiple dashboards and reports in **SPLUNK** to improve the agility of Security Analysts.
- **ETL** performed on data collected from the web or provided by the client, primarily using **Python (Pandas, SQLAlchemy)**.
- Used **FLASK** framework to integrate **Python**, **SQL** or **NoSQL (MongoDB)** databases, and **JavaScript** to create dynamic visualizations in **D3**, **Plotly**, and/or **Leaflet (Mapbox)**.
- Discussed issues/complex code to be resolved and translated them into **Informatica** logic. Prepared **ETL design documents**.
- Used **Informatica Designer** to create complex mappings using different transformations to move data to a Data Warehouse.
- Created continuous integration and continuous delivery (**CI/CD**) pipeline on **AWS** to automate steps in the software delivery process.
- Hands-on experience with **Informatica PowerCenter** and **PowerExchange** in integrating with different applications and relational databases.
- Created **AWS Code Pipeline** to build, test, and deploy code every time there is a code change.
- Worked on **Data modeling**, **Data mapping**, **Data cleansing**, and **Data visualization**.
- Under supervision of Sr. Data Scientist, performed **Data Transformation** method for **Rescaling**.
- Built models using **Python** and **PySpark** to predict probability of attendance for various campaigns and events.
- Implemented **classification algorithms** such as **Logistic Regression**, **K-NN neighbors**, and **Random Forests** to predict Customer churn and Customer interface.
- Performed data visualization and designed dashboards and generated complex reports, including charts, summaries, and graphs to interpret findings to team and stakeholders.

**Environment:** Spark, Spark SQL, Python, PySpark, Databricks, AWS services: EMR, Redshift, EC2, S3, Glue, CloudWatch, CloudTrail, SNS, DynamoDB, StepFunctions, Athena, Snowflake, Snowpipe, Shell scripting, PostgreSQL, Jenkins, Tableau.

**PGS Software, India**

**June 2019 - Dec 2021**

**Data Analyst**

- Collaborated with business stakeholders to drive data initiatives, focusing on leveraging data to optimize key performance indicators such as revenue and circulation. Led a team of data professionals specializing in **Analytics & Insight, Data Engineering, and Data Science**.
- Worked closely with the data analytics team to establish user groups for targeted advertisement sales, enhancing revenue streams.
- Conducted extensive Data Analysis on various platforms including **Teradata, Hadoop/HIVE/Oozie/Sqoop**,

**and AWS using SQL, Teradata SQL Assistant, Python, Apache Spark, and SQL Workbench.**

- Implemented data extraction processes utilizing **Flume** to transfer log files data to **HDFS**, ensuring alignment with business requirements. Designed and deployed interactive reports in **Tableau** with Drill Down, Drop-down menus, and Parameterized and Linked reports for enhanced data visualization and analysis.
- Defined and created Cloud Data strategies, including multi-phased implementation plans utilizing **AWS services such as S3**.
- Created and optimized **Snowflake Schemas**, emphasizing normalization of dimension tables for improved data structure.
- Led migration efforts from **Teradata** to **AWS**, automating processes with **Python scripts and SQL**. Utilized AWS migration services like **Kinesis Firehose, AWS Snowball, and S3 Transfer Acceleration** for seamless data transfer.
- Conducted advanced data analysis and complex algorithm development using **Python, R, Hadoop, Spark, Teradata, and Redshift**. Applied statistical analysis methods including time series, regression models, principal component analysis, and multi-variance analysis.
- Designed and implemented Test systems utilizing **Oracle SQL, Oracle, AIX/Unix, and Unix Shell Scripts**. Produced regular reports using **Teradata, Hadoop/HIVE/Oozie/Sqoop, MS Excel, and UNIX**.

**Environment:** Python, R, Hive, Oozie, Amazon AWS (S3, Redshift, EC2, Cloud Watch), MongoDB, Django, MySQL, MS SQL Server, Pig, Hadoop, SQL, SOAP, Rest APIs, Shell Scripting.

**Britenet, India**  
**ETL Developer**

**Mar 2018 - May 2019**

- Responsible for **analysis** of requirements and designing generic and standard **ETL** processes to load data from different source systems.
- Developed objects were thoroughly tested for unit/component testing and meticulously documented test cases for **mappings, sessions, and workflows**.
- Actively participated in daily status meetings, interfacing with the onshore team through mails/calls to ensure module progress and promptly resolve data/code issues.
- Managed the Classification System aspect of the project, focusing on loading data based on predefined conditions.
- Developed complex mappings to extract source data from heterogeneous databases including **SQLServer, Oracle**, and flat files. Applied transformation rules and efficiently loaded data into the **Data Warehouse**.
- Identified and resolved bugs in existing mappings by conducting thorough analysis of data flow and evaluating transformations using the **Debugger tool**.
- Implemented various **Performance Tuning techniques** on Sources, Targets, Mappings, and Workflows to enhance system efficiency.
- Collaborated closely with the Production Control team to schedule **shell scripts, Informatica workflows, and PL/SQL** code in Autosys.
- Conducted comprehensive **Database** testing to ensure integrity including checking **constraints, field size, indexes, and stored procedures**.
- Tracked, reviewed, and analyzed defects to ensure optimal system performance and data accuracy.
- Developed, executed, and maintained appropriate **ETL** development best practices and procedures to streamline workflows and optimize efficiency.
- Assisted in the development of test plans for assigned projects, ensuring thorough testing coverage.
- Monitored and tuned **ETL** processes for performance improvements, proactively identifying, researching, and resolving data warehouse load issues.
- Conducted unit testing of mappings and **SQL** code, ensuring robustness and reliability.
- Developed mappings to effectively load data into slowly changing dimensions, enhancing data accuracy and completeness.
- Played a key role in the performance **tuning of source & target systems, mappings, sessions, and workflows** to meet performance benchmarks.

**Environment:** Informatica PowerCenter, SQL Server, Oracle, Autosys, Shell Scripting, Data Warehouse, PL/SQL.

**Education:**

- Masters in Business Analytics **Lewis University, Chicago, Illinois**
- Bachelor of commerce (Honours) **Osmania University, Hyderabad, India**