

Siddhartha Reddy Yedmala

Data Engineer

(417) 413-3777 | Siddharthaa9898@gmail.com | Manchester, NH

PROFILE SUMMARY

- Experienced Data Engineer with 4+ years of expertise in designing and implementing robust data pipelines, ETL processes, and data integration solutions using SSIS, Informatica, and other leading ETL tools.
- Proficient in Cloud Platforms with hands-on experience in both AWS (EC2, S3, Redshift, Glue, Athena, Kinesis, Lambda) and Azure (Azure Data Factory, Azure Data Bricks, Azure Functions, Azure Synapse Analytics, Blob Storage, Cosmos DB, Microsoft Fabric) for scalable data storage, processing, and analytics.
- Advanced SQL Skills including writing complex queries, creating, and managing stored procedures, and optimizing SQL performance in Microsoft SQL Server and Teradata SQL environments.
- Big Data Technologies expertise in the Hadoop ecosystem, including Spark, HDFS, Hive, Pig, Sqoop, Flume, Kafka, Oozie, and Pyspark, for processing and analyzing large-scale datasets.
- Data Visualization and Reporting using tools like Tableau, Power BI, Amazon QuickSight, and Advanced Excel to create interactive dashboards and reports that drive business insights.
- Machine Learning Integration by leveraging Python libraries and tools to develop and deploy machine learning models, enhancing data-driven decision-making capabilities.
- Strong Programming Skills in Python, C#, and JavaScript, enabling automation of data workflows and development of custom data solutions.
- Interacted with lead and UI developers to enhance data integration and visualization, leading to a 15% improvement in user interface responsiveness and data accuracy.
- Engaged with stakeholders by integrating Power BI dashboards to visualize analytics, enhancing reporting capabilities and supporting strategic decision-making processes.
- Collaborated with cross-functional teams to support cloud data integration projects, resulting in a unified customer view that enhanced marketing campaign success.
- Proven track record of working in Agile/Scrum environments, effectively collaborating with cross-functional teams to deliver high-quality data solutions and meet project deadlines.
- ETL Development and Data Transformation ensuring seamless data extraction, transformation, and loading processes to support business intelligence and analytics.

TECHNICAL SKILLS

Programming Languages: C#, Python, JavaScript, Scala, MATLAB

Database | ETL: Microsoft SQL Server, Teradata SQL, SSIS, Informatica, Cassandra, MongoDB, MySQL

Data Visualization: Tableau, Power BI, Pandas, Advanced Excel, Amazon Quick Sight

Big Data: Spark, HDFS, Map reduce, Yarn, Hive, Pig, Sqoop, Flume, Kafka, Oozie, Nifi, HBase, Pyspark

Cloud Platforms: Amazon Web Services (EC2, EMR, S3, Redshift, Lambda, Glue, Athena, Kinesis), Azure (Azure Data Factory, Azure Data bricks, Azure Functions, Azure Synapse Analytics, HDInsight, Blob Storage Cosmos DB).

Functional Skills: Data Analysis, Data Pipelining, Data Visualization, Data Streaming, Data Integration, Data Transformation

WORK HISTORY

Credit Suisse, MA | Data Engineer

Aug 2022 - Present

- Served as a Data Engineer, responsible for Data modelling, Data migration and Designing ETL pipelines for both cloud and on-premises environments.
- Developed Spark applications for data validation, cleansing, transformation, and aggregation.

- Involved in loading and transforming large datasets of Structured, Semi-Structured and Unstructured data and analyzed them using Hive queries.
- Performed Extract, Transform and Load operations from Source Systems to Azure Data Storage services leveraging Azure Data Factory, Spark SQL and Azure Data Lake Analytics.
- Executed Data Ingestion to Azure Services like Azure Data Lake, Azure Storage, and Azure SQL DW, processing data in Azure Databricks.
- Developed Pipelines in ADF using Linked Services, Datasets, and Pipelines to Extract, Transform, and load data from various sources like Azure SQL, Blob storage, Azure SQL Data warehouse.
- Created PySpark programs to load data into Hive and MongoDB databases from PySpark Data frames.
- Designed and implemented custom-built input connectors using Spark and Sqoop for the ingestion and analysis of data from RDBMS to Azure Data Lake.
- Designed and optimized complex stored procedures for data processing and analytics, improving query performance by 30%.
- Developed and maintained SQL views to streamline data retrieval for reporting and analysis, reducing query complexity.
- Created interactive dashboards and reports in Power BI, providing actionable insights for business stakeholders.
- Integrated Power BI with various data sources, including SQL Server and Excel, ensuring real-time data availability.
- Utilized Azure Key Vault for secure storage and management of sensitive information, ensuring compliance with security standards.
- Developed and managed ETL pipelines in Azure Data Factory, automating data workflows and reducing manual intervention.
- Designed and created normalized and denormalized tables to support various business applications and reporting needs.

Dsquare Technology, India | Data Engineer

Sep 2020 - Jul 2021

- Configured and optimized Spark jobs for large-scale data processing, utilizing YARN for resource management and Pig for scripting.
- Conducted thorough data analysis and validation to ensure data accuracy and consistency.
- Designed and implemented robust ETL pipelines using Informatica and SSIS to facilitate data extraction, transformation, and loading from multiple sources into Teradata SQL.
- Automated data integration processes to improve efficiency and reduce manual intervention.
- Administered MongoDB and Cassandra databases, ensuring optimal data partitioning and replication strategies for fault tolerance and high availability.
- Developed serverless applications with AWS Lambda and orchestrated data workflows using AWS Glue.
- Leveraged AWS services including S3, Lambda, and Glue to build scalable and secure data storage and processing solutions, enhancing data accessibility and integration across the organization.
- Created dynamic visualizations and dashboards with Power BI and Amazon Quick Sight to deliver critical business insights.
- Integrated Flume and Sqoop for seamless data ingestion and export between Hadoop and traditional data stores.

Dsquare Technology, India | Data Engineer

Jan 2020 - Aug 2020

- Designed and implemented complex ETL processes using SQL Server Integration Services (SSIS) to facilitate data integration and migration across multiple systems, enhancing data consistency and accessibility.
- Utilized SSIS to perform data cleansing and transformation, ensuring high data quality and readiness for analytical applications.
- Maintained comprehensive reporting solutions using SQL Server Reporting Services (SSRS), delivering key insights through interactive dashboards and automated reports
- Implemented Kafka for real-time data streaming and Oozie for job scheduling and workflow management.

- Used JavaScript for developing interactive web-based data visualization tools.
- Used Advanced Excel and Pandas for detailed data analysis and visualization, supporting various business functions.
- Managed Hadoop clusters with components like YARN for resource management and Sqoop for data transfer between Hadoop and RDBMS.
- Streamlined data transformation and integration processes, enhancing data quality and operational efficiency across the enterprise.
- Managed NoSQL databases including Cassandra and MongoDB, optimizing them for high availability and performance.

ACADEMIC PROJECT

Real-time Social Media Sentiment Analysis Platform

- Developed an end-to-end data pipeline using Apache Kafka for real-time data ingestion from platforms such as Twitter and Reddit. Leveraged Apache Spark Streaming for data processing, cleaning, and tokenization.
- Integrated sentiment analysis using NLTK to categorize sentiments as Positive, Negative or Neutral. Stored the processed data in distributed warehouses, like Snowflake.
- Visualized real-time sentiment trends and insights using Tableau dashboards.
- Ensured storage optimization and performance while maintaining compliance with data regulations. Utilized Docker and Kubernetes for scalable application deployment.

EDUCATION

Masters in Business Analytics Engineering	Jan 2022 – Dec 2022
Southern New Hampshire University	
Bachelors in Business Administration	June 2017 – Oct 2020
Osmania University, Hyderabad, India	

CERTIFICATIONS

- Google Cloud Professional Data Engineer
- AWS Certified Big-Data Specialty
- Microsoft Certified: Azure Data Engineer Associate
- Microsoft Fabric Certification