Nithin Bobbiligama

DATA ENGINEER

Columbus, OH.

(740)909-0444 | nithinbobbili007@gmail.com | https://www.linkedin.com/in/nithin-bobbili07/

PROFESSIONAL SUMMARY

- Cloud-Native Data Engineer with 6 years of experience in designing and delivering scalable data pipelines across Azure and AWS, with strong hands-on expertise in ADF, AWS Glue and IBM DataStage.
- Skilled in Python, SQL and Shell scripting, building reusable ingestion frameworks and metadata-driven transformations to automate and streamline batch and real-time data workflows.
- Skillful in Big Data and streaming technologies including Apache Spark, PySpark, Kafka, Flink, Kinesis, Hive and Sqoop for high-volume data processing and real-time analytics.
- Advanced knowledge of modern data platforms such as Snowflake, Synapse, Redshift, Cosmos DB, MongoDB and MySQL, managing both structured and semi-structured data at scale.
- Delivered actionable insights through dashboards built with Power BI, Tableau, Looker and Quick Sight, collaborated within Agile/Scrum/Kanban environments to ensure iterative and high-quality delivery.

TECHNICAL SKILLS:

- **Programming Languages:** Python, Scala, Java, SQL.
- **Databases & Warehousing:** Azure Synapse, Snowflake, Redshift, MySQL, PostgreSQL, SQL Server, Azure SQL DB, DynamoDB, Cosmos DB, MongoDB, HBase, DB2oracle, Teradata.
- **Big Data & ETL:** Apache Spark, PySpark, Hadoop, Hive, Pig, HBase, Sqoop, IBM DataStage, Spark SQL, MapReduce, Apache Beam, Flink, Airflow, Oozie.
- **Cloud Platforms & Services:** Azure (ADF, Databricks, Functions, Logic Apps, Key Vault, Vms), AWS (EMR, Glue, Lambda, EC2, S3, RDS, Redshift, Kinesis, IAM, VPC, Elasticsearch, CloudWatch).
- **CI/CD & Version Control:** Azure DevOps, Jenkins, GitHub Actions, Git, GitHub, GitLab, Bitbucket.
- **Containerization & Orchestration**: Docker, Kubernetes, AWS EKS.
- **Streaming Technologies:** Apache Kafka, AWS Kinesis, Apache Flink.
- **Data Governance & Quality:** Custom Data Quality Frameworks, RCA, Governance Policies.
- **Methodologies:** Agile, Kanban, SDLC.

WORK EXPERIENCE:

Nationwide, Columbus, Ohio, USA. **Aug 2023 – Present**

Azure Data Engineer

- Designed and optimized ETL pipelines using Azure Data Factory, Synapse and ODS to accelerate data delivery and support 15+ real-time reporting use cases.
- Built scalable event-driven architectures with Azure Functions, Logic Apps and Kafka, reducing manual efforts and enabling near real-time integrations.
- Engineered high-volume data workflows in Databricks and Spark (PySpark), reducing batch processing time and boosting analytics performance.
- Automated ingestion from diverse sources (HDFS, SQL, flat files) using IBM DataStage and Python-based frameworks, cutting data latency by 35% and improving onboarding.
- Delivered 15+ Power BI dashboards and implemented CI/CD pipelines via Azure DevOps and Git, achieving faster deployments and fewer ad-hoc reporting requests.
- Integrated real-time monitoring and alerting using Azure Monitor, Log Analytics and Application Insights, improving system observability and reducing issue resolution time by 40%.

Humana, Dublin, Ohio, USA **Feb 2022 – Dec 2022**

Data Engineer

- Built and optimized real-time data pipelines using Apache Spark, PySpark and AWS EMR to analyze provider performance, enabling faster decision-making across healthcare operations.
- Reduced processing time and infrastructure costs by 30% by tuning Redshift, implementing auto-scaling on EMR and designing scalable storage with MySQL and DynamoDB.
- Streamlined ingestion and live analytics by integrating AWS Kinesis with RESTful APIs, supporting low-latency dashboards and improving provider insight visibility.
- Delivered high-performance ETL workflows using IBM DataStage, Snowflake and Spark SQL, enhanced CI/CD automation with Jenkins and GitHub, while contributing to Agile teams.
- Improved data quality and validation by implementing robust error handling, schema checks and audit frameworks, ensuring reliability across multiple healthcare data sources.

Experian, Hyderabad, India **Oct 2019 – Jun 2021**

Data Engineer

- Developed scalable ETL pipelines using Sqoop, Python and Shell scripts to ingest data from SQL Server, MySQL and flat files into HDFS, supporting batch data workflows.
- Reduced ETL effort by 40% through automation of multi-source integration with IBM DataStage and AWS Glue and optimized data transformation using Apache Spark and PySpark.
- Built real-time streaming pipelines with AWS Kinesis, Kafka and Redshift, improving SLA adherence and delivering near-instant insights for business operations.
- Managed diverse data models using MySQL, MongoDB and HBase and enabled real-time analytics via Tableau, Snowflake and S3 dashboards.
- Led performance tuning and job orchestration efforts, ensuring scalable, fault-tolerant data pipelines using Apache Airflow, Oozie and CI/CD tools for automated deployment.

Novartis, Hyderabad, India **Mar 2018 - Sep 2019**

Program Analyst

- Automated data extraction and transformation using Python, improving data accuracy and reducing manual intervention across MySQL and MongoDB sources.
- Built scalable ETL pipelines with Sqoop, Hive, Pig and Oozie on Hadoop and recommended Apache Airflow to modernize workflow orchestration.
- Developed a custom Python GUI tool for hardware simulation, enabling early QA testing without physical devices and accelerating test cycles.
- Optimized Spark SQL performance and supported interactive data visualizations using JavaScript, AJAX and JSON for business insights.
- Designed and scheduled reporting pipelines, integrating structured and semi-structured data and ensuring consistent delivery through automated workflows.

EDUCATION:

- University of Dayton
  Master of Science in Computer Science.

CERTIFICATIONS:

- Azure Associate Data Engineer.