UDAY KUMAR VURUKONDA
Data Engineer

+1 (469)-473-6948 | udaykumar777v@gmail.com | linkedin.com/in/uday-kumar7

SUMMARY

- Experienced and results-driven Data Engineer and software Engineer with a proven track record of **over 4 years** of experience software development experiencewith expertise in Big Data, Hadoop Ecosystem, Cloud Engineering, Data Warehousing.
- Sound Experience with AWS services like Amazon EC2, S3, EMR, Amazon RDS, VPC, Amazon Elastic Load Balancing, IAM, AutoScaling, Cloud Front, CloudWatch, and Lambda to trigger resources.
- Experience in building data pipelines using Azure Data Factory, Azure Databricks, and loading data to Azure Data Lake, Azure SQL Database, Azure SQL Data Warehouse to control and grant database access.
- Good experience with Azure services like HDInsight, Stream Analytics, Active Directory, Blob Storage, Cosmos DB, Storage Explorer.
- Strong Hadoop and platform support experience with all the entire suite of tools and services in major Hadoop Distributions – Cloudera, Amazon EMR, Azure HDInsight, and Hortonworks.
- Strong familiarity with GCP components, Google Container Builders, and client libraries, leveraging cloud SDKs for development efficiency.
- Led Proof of Concept (POC) initiatives to assess cloud offerings, including Google Cloud Platform.
- Proficient in handling and ingesting terabytes of Streaming data (Kafka, Spark streaming, Strom), Batch Data, Automation and Scheduling (Oozie, Airflow).
- Profound knowledge in developing production-ready Spark applications using Spark Components like Spark SQL, DataFrames, Datasets, Spark-ML and Spark Streaming.
- Expertise in developing multiple confluent Kafka Producers and Consumers to meet business requirements. Store the stream data toHDFS and process it using Spark.
- Strong working experience with SQL and NoSQL databases (Cosmos DB, MongoDB, HBase, Cassandra), data modeling, tuning, disaster recovery, backup and creating data pipelines.
- Experienced in scripting with Python (PySpark), Scala and Spark-SQL for development, aggregation from various file formats such asXML, JSON, CSV, Parquet.
- Great experience in data analysis using HiveQL, Hive-ACID tables, Pig Latin queries, custom MapReduce programs and achieved improved performance.
- Experience in monitoring document growth and estimating storage size for large MongoDB clusters as part of the data life cycle management.
- Hands-on experience on Ad-hōc queries, Indexing, Replication, Load balancing, Aggregation in MongoDB.
- Expertise in creating Kubernetes cluster with cloud formation templates and PowerShell scripting to automate deployment in a cloud environment.
- Sound knowledge in developing highly scalable and resilient Restful APIs, ETL solutions, and third-party integrations as part of Enterprise Site platform using Informatica.
- Highly involved in all facets of SDLC using Waterfall and Agile Scrum methodologies.
- Involved in migration of the legacy applications to cloud platform using DevOps tools like GitHub, Jenkins, JIRA, Docker, and Slack

WORK EXPERINCE

Data Engineer | *BCBS, USA* *Jun 2023 – Current*

- Mastered a wide array of Azure technologies, including HDInsight, Databricks, Data Lake, Blob Storage, Data Factory, SynapseAnalytics, Azure SQL Database, and SQL Data Warehouse, to streamline cloud operations and data management.
- Designed and automated custom input adapters using Spark, Sqoop, and Oozie, significantly enhancing data ingestion from RDBMSto Azure Data Lake, showcasing a blend of innovation and efficiency in data handling.
- Played a pivotal role in the creation of automated workflows for daily incremental data loads, facilitating seamless data transition from traditional RDBMS systems to more scalable data lake solutions, thus optimizing data availability and reliability.
- Executed comprehensive ETL (Extract, Transform, Load) processes, utilizing Azure Data Factory, T-SQL, Spark SQL, and U-SQLfor Azure Data Lake Analytics, ensuring efficient and scalable data movement and transformation strategies.
- Engineered and deployed data pipelines in Azure Data Factory, targeting enhanced ETL processes from diverse sources such as AzureSQL, Blob storage, and Azure SQL Data Warehouse, thereby ensuring data integrity and accessibility.
- Crafted and deployed JSON scripts in Azure Data Factory, leveraging SQL Activity for efficient data processing and integration, highlighting technical proficiency and advanced deployment capabilities.
- Demonstrated deep understanding of data ingestion and orchestration using Airflow Operators and relevant Python libraries, enhancing data flow and operational efficiency.
- Analyzed SQL scripts and implemented solutions using PySpark, focusing on optimizing data processing and analytics workflows.
- Extensively used Databricks notebooks for interactive analytics, employing Spark APIs to facilitate advanced data analysis and insights.
- Integrated data across various sources including MongoDB, MS SQL, and cloud platforms, using Azure Data Factory, SQL API, andMongo API, to ensure a unified data ecosystem.
- Enabled efficient data visualization and reporting through interactive Power BI dashboards and reports, providing actionable insightsto decision-makers.
- Developed and executed efficient data warehousing strategies using Hive on Azure HDInsight for large-scale data analysis.
- Spearheaded the Continuous Integration and Continuous Deployment (CI/CD) of applications into Azure Cloud, enhancingoperational agility and system reliability.

Software Developer | Ecclesiastes Inc. *Apr 2023 – Jun 2023*

- Engineered and managed the creation of 10 automated, scalable, code-based data pipelines, leveraging Amazon Redshift, Data Grip, Amazon S3, and Glue. These pipelines efficiently processed millions of data points, demonstrating a high level of expertise in cloud-based data architecture and automation.
- Successfully executed query optimizations on AWS, utilizing Amazon RDS and Amazon Redshift, to improve database and data warehouse performance by 32%.
- Utilized AWS Athena to manipulate CSV data files stored in AWS S3, applying Scala queries for proficient data extraction and

transformation, showcasing versatility in data handling methodologies.

- Formulated and implemented Python solutions to extract data from AWS S3 and populate it into SQL Server, directly supporting business team requirements through effective data integration strategies.
- Played a key role in contributing to a Databricks Delta Lake environment on AWS, employing Spark for sophisticated data processing tasks, enhancing data lake utility and performance.
- Conducted ETL operations using Python, Spark SQL, S3, and Redshift, handling substantial volumes of data to derive actionable customer insights, thus driving business intelligence initiatives.
- Orchestrated automated CI/CD pipelines utilizing AWS Code Pipeline, Jenkins, and AWS Code Deploy, significantly enhancing deployment efficiency and operational agility.
- Developed and maintained large-scale data warehouses on AWS using Hive and Redshift, enabling efficient data analysis for complex datasets.
- Authored PySpark scripts to streamline ETL processes, extracting data from S3 with a crawler and generating a data catalog to consolidate metadata.

Full stack Developer | Eficens Systems LLC. *Jan 2023 – Apr 2023*

- Designed and implemented a real-time data pipeline to ingest customer data from various sources at a rate of 10,000 records per second using Apache Spark on AWS.
- Developed custom data processing algorithms in Python to cleanse, transform, and load data into a data lake, enabling real-time customer behavior analysis Reduced data processing time by 30% by optimizing Spark jobs and leveraging efficient data serialization techniques.
- Developed unit tests and integrated CI/CD pipeline to ensure code quality and continuous delivery of the model.
- Designed and developed a microservices architecture for the e-commerce platform using Spring Boot, resulting in a 50% reduction in response times during peak load. Containerized microservices using Docker for rapid deployment and scalability across multiple environments

Software Engineer | Tech Mahindra, India *Apr 2018 – Apr 2021*

- Migrated a critical enterprise application from an on premise server to AWS cloud infrastructure, achieving a 40% reduction in operational costs. Responsible for infrastructure provisioning, configuration management, and security best practices implementation.
- Built scalable real-time Kafka pipeline for efficient high-volume data ingestion & processing.
- Developed scalable data pipelines using AWS Glue for automated data ingestion and transformation, resulting in a 50% reduction in manual data processing efforts and improved data quality for downstream analytics.
- Enhanced data analysis capabilities by implementing advanced statistical models and complex data transformations using Python and SQL, resulting in a 20% improvement in predictive accuracy for business forecasts.
- Utilized AWS EMR (Elastic MapReduce) for distributed data processing and analytics, leveraging Hadoop and Spark clusters to handle largescale data workloads and performing complex transformations.
- Successfully implemented POC (Proof of Concept) in a Development Databases to validate the requirements and benchmarking the ETL loads.
- Implemented partitioning and bucketing in Hive, developing queries to process data and generate data cubes for visualization, enabling faster query performance and efficient data organization.
- Loaded the aggregated data into MongoDB for reporting on the dashboard. Worked on MongoDB schema/document modeling, querying, indexing, and tuning

SKILLS

- **Programming Languages:** SQL, Pig Latin, HiveQL, Scala, Python, Unix Shell Scripting.
- **Databases:** MS-SQL SERVER, Oracle, MS-Access, MySQL, Teradata, PostgreSQL, DB2.
- **Big Data Technologies:** Yarn, MapReduce, Pig, Hive, HBase, Cassandra, Oozie, Apache Spark, Scala, Impala, Kafka.
- **Hadoop Distributions:** Apache Hadoop 2.x/1.x, Cloudera CDP, Hortonworks HDP, Amazon EMR (EMR, EC2, EBS, RDS, S3, Glue, Elasticsearch, Lambda, Kinesis, SQS, DynamoDB, Redshift, ECS) Azure HDInsight (Databricks, Data Lake, Blob Storage, Data Factory, SQL DB, SQL DWH, Cosmos DB, Azure DevOps, Active Directory).
- **NoSQL Database:** Cassandra, MongoDB.
- **Reporting Tools/ETL Tools:** Informatica, Talend, SSIS, SSRS, SSAS, ER Studio, Tableau, Power BI.
- **Methodologies:** Agile/Scrum, Waterfall.
- **Development Tools:** Eclipse, NetBeans, IntelliJ, Hue, Microsoft Office Suite (Word, Excel, PowerPoint, Access)
- **Operating Systems:** Windows, Macintosh, Linux, Ubuntu, Unix.
- **Others:** Machine learning, NLP, Stream Sets, Spring Boot, Jupyter Notebook, Docker, Kubernetes, Jenkins, Jira.

EDUCATION

Master of Science in Computer Science ***Aug 2021 - Dec 2022***
Fitchburg State University, Fitchburg, USA

Bachelor of Technology in Aeronautical Engineering ***Sep 2013 - May 2017***
Jawaharlal Nehru Technological University, India

CERTIFICATIONS

- AWS Certified Developer – Associate LINK
- Python for Data Science and ML LINK
- Tableau 2020 A-Z LINK
- AWS Certified Solutions Architect – Associate LINK