

Bhuvan Potluri

Ontario

bhuvanpotluride2_s6r@indeedemail.com

438-725-6769

- 4+ years of IT experience in Business Requirement analysis, design, development and implementation of Data warehousing and Business Intelligence applications using ETL.
- Experience in design, development, and Implementation of Big data applications using Hadoop ecosystem frameworks and tools like HDFS, MapReduce, Yarn, Pig, Hive, Sqoop, Spark, Scala, Storm HBase, Kafka, Flume, Nifi, Impala, Oozie, Zookeeper, Airflow etc.
- Strong technical experience in SDLC including requirement gathering, Design, Analysis, Coding, Testing, Documentation and Deployment.
- Worked on several projects like ACA, HCSIS, OMAP and IES, CWIS projects which required conversion from Legacy systems to give a one stop solution for public services catered in diversified areas like senior benefits, health and financial assistance, child and family welfare, transport etc.
- Understanding and working knowledge of Informatica CDC (Change Data Capture).
- Good knowledge of Relational Databases like SQL Server, Oracle, DB2.
- Experienced in Logical and Physical Database design & development, Normalization and Data modeling using Erwin and SQL Server Enterprise manager.
- Experience with slowly changing dimension methodology and slowly growing targets methodologies.
- Experience in multidimensional modeling and RDBMS concepts Data mart, OLAP, OLTP using Star Schema and Snow-Flake Schema.
- Good understanding in writing Test plans, Test cases, Unit testing, System testing, Integration testing and Functional testing.
- Strong experience in Informatica Data Quality (IDQ), Power Center, Data Cleansing, Data profiling, Data quality measurement and Data validation processing
- Used Unix Shell scripts to write pmcmd commands to automate workflows.
- Experience with different scheduling tools to accommodate requisites.
- Expertise in developing streaming applications in Scala using Kafka and Spark Structured Streaming.
- Expertise in developing and tuning Spark applications using various optimizations techniques for executor tuning, memory management, garbage collection, Serialization assuring the optimal performance of applications by following best practices in the industry.
- Hands on experience in importing and exporting data into HDFS and Hive using Sqoop.
- Exposure on usage of NoSQL databases column oriented HBase and Cassandra.
- Hands-on experience in working with AmazonWebServices (AWS) cloud and its services like EC2, S3, AthenaRDS, VPC, IAM, ElasticLoadBalancing, Glue, Lambda, RedShift, Auto Scaling, Cloud Front, Cloud Watch, and other services of the AWS family.
- Experience in ingesting, processing, analyzing of structured and unstructured data on Hadoop clusters.
- Expertise in developing spring based Microservices
- Experienced in performing code reviews, involved closely in smoke testing sessions, retrospective sessions.
- Have good exposure with the star, snowflake schema, data modelling and work with different data warehouse projects.
- Strong experience in the design and development of relational database concepts with multiple RDBMS databases including Oracle10g, MySQL Server & PL/SQL.

- Trouble-shooting production incidents requiring detailed analysis of issues on web and desktop applications, Autosys batch jobs, and databases.
- Strong troubleshooting and production support skills and interaction abilities with end users.
- Experience in working with various SDLC methodologies like Waterfall, Agile Scrum, and TDD for developing and delivering applications.

Work Experience

Big Data Engineer / ETL Developer

ADP - Toronto, ON

April 2020 to Present

Canada

Roles & Responsibilities

- Involved in requirements gathering, analysis, design, development, change management and deployment.
 - Assisted Business Analyst with drafting the requirements, implementing design and development of various components of ETL for various applications.
 - Monitor and modify daily & on demand batch jobs that load warehouse tables from various source systems like Oracle, Sql server, Salesforce.
 - Fix sessions that result in undesired load data in lower & higher environments for both Data warehouse and DataMart's.
 - Define match rules, columns for match and merge process.
 - Performed Data Cleansing, Address correction using IDQ.
 - Built profiling, cleansing and validation plans using IDQ.
 - Development on existing mappings that acquired extended scope to accommodate new logistics and perform regressive testing before migrating it to higher environments.
 - Extracted data from heterogeneous sources and performed complex business logic on network data to normalize raw data which can be utilized by BI teams to detect anomalies.
 - Worked with Spark Ecosystem using Scala and Hive Queries on different data formats like Text file.
 - Developed Spark code using Scala and Spark-SQL for faster processing and testing.
 - Using Spark-Streaming APIs to perform transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time
 - Involved in the development of real time streaming applications using Kafka, Hive, PySpark, Apache Flink on distributed Hadoop Cluster.
 - Performs debugging, troubleshooting, modifications and unit testing of integration solutions
 - Perform unit and integration testing and document test strategy and results
 - Work closely with teams from other enterprise software vendors that are being integrated
 - Use innovation to improve operational processes and performance, making sure data is of the highest quality, build and unit test integration components
 - Utilized Apache Spark with Python to develop and execute Big Data Analytics.
 - Developed common Flink module for serializing and deserializing AVRO data by applying schema.
 - Implemented layered architecture for Hadoop to modularize design.
- Developed framework scripts to enable quick development.
- Used Microservice architecture with Spring Boot based services interacting through a combination of REST to build, test and deploy identity Microservices

- Developed Spark scripts using Python on AWS EMR for Data Aggregation, Validation and Adhoc querying.
- Data Extraction, aggregations and consolidation of Adobe data within AWS Glue using PySpark
- Developed CI-CD pipeline to automate build and deploy to Dev, QA, and production environments.
- Supported production jobs and developed several automated processes to handle errors and notifications. Also, tuned performance of slow jobs by improving design and configuration changes of PySpark jobs.
- Created standard report Subscriptions and Data Driven Report Subscriptions.

Environment: Hadoop, Informatica 9.8.1/10.2, IDQ, Salesforce, Embarcadero ER Studio Map Reduce, Spark, Spark MLLib, Kafka, Flink, PIG, Hive, AWS, Tableau, SQL, PostgreSQL, Python, PySpark, SQL Server 2012, T-SQL, CI-CD, Git, XML, Splunk.

Data Engineer

MORGAN STANLEY - Quebec City, QC
February 2019 to March 2020

Canada

Roles & Responsibilities

- Involved in implementation of new statistical algorithms and operators on Hadoop and SQL platforms and utilized optimizations techniques, linear regressions, K-means clustering, Native Bayes and other approaches.
- Developed Spark batch job to automate creation/metadata update of external Hive table created on top of datasets residing in HDFS.
- Developed Data Serialization spark common module for converting Complex objects into sequence bits by using AVRO, PARQUET, JSON, CSV formats.
- Worked on ERModeling, Dimensional Modeling (StarSchema, snowflakeSchema), Data warehousing and OLAP tools.
- Populated HDFS and PostgreSQL with huge amounts of data using Apache Kafka.
- Develop disaster recovery and failover strategies for the data integration environments
- Design, implement and perform administration functions for Enterprise data integration environments
- Involved in migrating RESTful services to Microservices.
- Developed Microservices to handle specific business logic to have easy maintainability and ease of testing
- Designed Batch Audit Process in batch\shell script to monitor each ETL job along with reporting status which includes table name, start and finish time, number of rows loaded, status, etc.
- Developed Spark jobs in PySpark to perform ETL from SQL Server to Hadoop.
- Responsible for continuous monitoring and managing Elastic MapReduce (EMR) cluster through AWS console.
- Design and Develop ETL Processes in AWS Glue to migrate Campaign data from external sources like S3, ORC/Parquet/Text Files into DataBase.
- Worked with EMR, S3 and EC2 services in AWS cloud.
- Migrating servers, databases, and applications from on premise to AWS.
- Designed and implemented data acquisition, ingestion, Management of Hadoop infrastructure and other Analytics tools (Splunk, Tableau).
- Working knowledge of build automation and CI/CD pipelines.
- Developed python scripts to automate data ingestion pipeline for multiple data sources and deployed Apache NIFI in AWS.

- Design and develop Tableau visualizations which include preparing Dashboards using calculations, parameters, calculated fields, groups, sets and hierarchies.

Environment: Hadoop, Map Reduce, Spark, Spark MLLib, Kafka, NiFi, SQL, PIG, Hive, AWS, PostgreSQL, Python, PySpark, Microservices SQL Server 2012, T-SQL, CI-CD, Git, XML, Tableau.

Data Engineer/AWS Engineer

Hexaware - Chennai, Tamil Nadu

August 2017 to December 2018

India

Roles & Responsibilities

- Analyzing Functional Specifications Based on Project Requirement.
- Ingested data from various data sources into Hadoop HDFS/Hive Tables using SQOOP, Flume, Kafka.
- Developing Hive Queries for the user requirement.
- Worked on multiple POCs in Implementing Data Lake for Multiple Data Sources ranging from TeamCenter, SAP, Workday, Machine logs.
- Developed Spark code using Scala and Spark-SQL/Streaming for faster testing and processing of data.
- Creates and owns the strategic roadmap for data integration across the enterprise
- Strong understanding of Enterprise data integration patterns and techniques
- Ability to validate data integration by developing and executing test plans
- Planning, scheduling and implementing Oracle to MS SQL server migrations for in house applications and tools.
- Designed and developed Micro Services business components using Spring
- Worked on Solr Search Engine to index incident reports data and developed dash boards in Banana Reporting tool.
- Integrated Tableau with Hadoop data source for building dashboard to provide various insights on sales of the organization.
- Worked on Spark in building BI reports using Tableau. Tableau was integrated with Spark using Spark-SQL.
- Developed Spark jobs using Scala and Python on top of Yarn/MRv2 for interactive and Batch Analysis.
- Created multi-node Hadoop and Spark clusters in AWS instances to generate terabytes of data and stored it in AWS HDFS.
- Developed work flows in Live Compare to Analyze SAP Data and Reporting.
- Participated in daily scrum meetings and iterative development.
- Search functionality for searching through millions of files of logistics groups.

Environment: Hadoop, Hive, Sqoop, Spark, Kafka, Scala, MS SQL Server PDW, AWS.

SQL Data Analyst

Bluepal Solutions PVT LTD - Hyderabad, Telangana

November 2016 to July 2017

India

Roles & Responsibilities

- Designed and created Data Marts in data warehouse database
- Implementations of MS SQL Server Management studio 2008 to create Complex Stored Procedures and Views using T-SQL.

- Collecting the data from many resources and converting into flat text files with comma delimiter separator and importing the data to the SQL server for data manipulations.
- Responsible for deploying reports to Report Manager and Troubleshooting for any errors during the execution.
- Scheduled the reports to run on daily and weekly basis in Report Manager and emailing them to director and analyst to be reviewed in Excel Sheets.
- Created several reports for claims handling which had to be exported out to PDF formats.
- Analyzed business requirements and provided excellent and efficient solutions

Environment: SQL Server 2008, Microsoft Visual Studio 2008, MS Office, SSRS

Skills

- Big Data Eco-system
- HDFS, MapReduce, Spark, Yarn, Hive, Pig, HBase, Sqoop, Flume, Kafka, Oozie, Zoo-keeper, Impala
- Hadoop Technologies
- Apache Hadoop 1.x, Apache Hadoop 2.x, Cloudera CDH4/CDH5, Hortonworks
- Programming Languages
- Python, Scala, Shell Scripting, HiveQL
- Machine Learning
- Regression, Decision Tree, Clustering, Random Forest, Classification, SVM, NLP
- Operating Systems
- Windows (XP/7/8/10), Linux (Ubuntu, Centos)
- NoSQL Database
- HBase
- Database
- RDBMS, MySQL, Teradata, DB2, Oracle
- Container/Cluster Managers Docker, Kubernetes
- BI Tool
- Tableau, Power BI
- IDE Tools
- Eclipse, Jupyter, Anaconda, Pycharm
- Development Methodologies
- Agile, Waterfall