## BHAVANA T

[bhavana4101@gmail.com](mailto:bhavana4101@gmail.com)

## 913 - 730 - 0714

**Senior Data Engineer**

[LinkedIn](https://www.linkedin.com/in/bhavana-t-56ab56275/)

# PROFESSIONAL SUMMARY

* Over 9+ years of experience in Information Technology which includes Hands on experience in Hadoop ecosystem including Spark, Kafka, HBase, MapReduce, Python, Scala, Pig, Impala, Sqoop, Oozie, Flume, Storm, big data technologies and worked on Spark SQL, Spark Streaming and using Core Spark API to explore Spark features to build data pipelines.
* Experience in building bigdata solutions using Lambda Architecture using Cloudera distribution of Hadoop, Twitter Storm, Trident, MapReduce, Cascading, HIVE, PIG and Sqoop.
* Strong Knowledge and experience on implementing Big Data in Amazon Elastic MapReduce (Amazon EMR) for processing, managing Hadoop framework dynamically scalable Amazon EC2 instances, Lambda, SNS, SQS, AWS Glue, S3, RDS and Redshift.
* end-to-end implementation project experience on data processing pipeline using Hadoop (CDP) ecosystem, Data Lake, and Data Warehouse.
* Strong experience with AWS Cloud on data integration with Databricks, Apache Spark, Airflow, EMR, Glue, Kafka, Kinesis and Lambda in S3, Redshift, RDS, MongoDB/DynamoDB ecosystems.
* Responsible for the design and development of data related solutions (databases, ETL systems, code, scripts, data models, reports, documentation) and supporting various products. Extensive hands-on experience developing database queries, ETL, performance tuning and database unit tests in a relational database environment.
* Having experience in developing a data pipeline using Kafka to store data into HDFS and Experience in building Real-time Data Pipelines with Kafka Connect and Spark Streaming and developing end to end data processing pipelines that begin with receiving data using distributed messaging systems Kafka for persisting data into Hive.
* Experienced in development methodologies like Agile/Scrum.
* Excellent Experience in Designing, Developing, Documenting, Testing of ETL jobs and mappings in Server and Parallel jobs using Data Stage to populate tables in Data Warehouse and Data marts.
* Expertise in transforming business requirements into analytical models, designing algorithms, building models, developing Data Mining, Data Acquisition, Data Preparation, Data Manipulation, Feature Engineering, Machine Learning Algorithms, Validation and Visualization and reporting solutions that scales across massive volume of structured and unstructured Data.
* Strong knowledge in working with ETL methods for data extraction, transformation, and loading in corporate- wide ETL Solutions and Data Warehouse tools for reporting and data analysis
* Experience with different ETL tool environments like SSIS, Informatica, and reporting tool environments like SQL Server Reporting Services, and Business Objects.
* Having good experience on all flavors of Hadoop (Cloudera, Hortonworks, and MapR etc.) and Hands on experience in AVRO and Parquet file format, Dynamic Partitions, Bucketing for best Practice and Performance improvement.
* Expert in designing Server jobs using various types of stages like Sequential file, ODBC, Hashed file, Aggregator, Transformer, Sort, Link Partitioner and Link Collector.
* Proficiency in Big Data Practices and Technologies like HDFS, MapReduce, Hive, Pig, HBase, Sqoop, Oozie, Flume, Spark, Kafka.
* Develop data pipeline to synchronize and aggregate data from various data sources. Build the query data platform and provide guidelines for partner team to use aggregated data. Publish standard data sets & schema.
* Experience with source control systems such as Git, Bitbucket, and Jenkins. Working experience in CI/CD deployments
* Good knowledge in RDBMS concepts (Oracle 112c/1g, MS SQL Server 2012) and strong SQL, query writing skills (by using Erwin & SQL Developer tools), Stored Procedures and Triggers and Experience in writing Complex SQL Queries involving multiple tables inner and outer joins.
* Excellent knowledge with Unit Testing, Regression Testing, Integration Testing, User Acceptance Testing, Production implementation and Maintenance.
* Excellent performance in building, publishing customized interactive reports and dashboards with customized parameters including producing tables, graphs, listings using various procedures and tools such as Tableau and user-filters using Tableau.
* Working knowledge of Azure cloud components (HDInsight, Databricks, Data Lake, Blob Storage, Data Factory, Storage Explorer, SQL DB, SQL DWH, Cosmos DB).
* Experienced in building data pipelines using Azure Data Factory, Azure Databricks, and loading data to Azure Data Lake, Azure SQL Database, Azure SQL Data Warehouse, and controlling database access.
* Extensive experience with Azure services like HDInsight, Stream Analytics, Active Directory, Blob Storage, Cosmos DB, and Storage Explorer.
* Expertise in MS SQL Server Database and Performance tuning experience, Batch and distributed computing using ETL/ELT (Spark /SQL Server DWH)
* Integrate Power BI reports into other applications using embedded analytics like Power BI service (SaaS), or by API automation. Also, one must be experienced in developing custom visuals for Power BI.
* Self-starter with a passion and curiosity for solving unstructured data problems and ability to manipulate and optimize large data sets.
* Experience in developing Map Reduce Programs using Apache Hadoop for analyzing the big data as per the requirement. Practical understanding of the Data modeling (Dimensional & Relational) concepts like Star-Schema Modeling, Snowflake Schema Modeling, Fact and Dimension tables.
* Ability to review technical deliverables, mentor and drive technical teams to deliver quality products.
* Demonstrated ability to communicate and gather requirements, partner with Enterprise Architects, Business Users, Analysts, and development teams to deliver rapid iterations of complex solutions.
* Flexible and versatile to adapt to any new environment with a strong desire to keep pace with latest technologies.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| Big Data Technologies | Hadoop, MapReduce, Spark, HDFS, Sqoop, YARN, Oozie, Hive, Impala, Zookeeper, Apache Flume, Apache Airflow, Cloudera, HBase |
| Programming Languages | Python, PL/SQL, SQL, Scala, C, C#, C++, T-SQL, Power Shell Scripting, JavaScript |
| Cloud Services | Azure Data Lake Storage Gen 2, Azure Data Factory, Blob storage, Azure SQL DB, Databricks, Azure Event Hubs, AWS RDS, Amazon SQS, Amazon S3, AWS EMR, Lambda, AWS SNS. |
| Databases | MySQL, SQL Server, Oracle, MS Access, Teradata, and Snowflake |
| NoSQL Data Bases | MongoDB, Cassandra DB, HBase |
| Development Strategies | Agile, Lean Agile, Pair Programming, Waterfall and Test Driven Development. |
| Visualization & ETL tools | Tableau, Informatica, Talend, SSIS, and SSRS |
| Version Control & Containerization tools | Jenkins, Git, and SVN |
| Operating Systems | Unix, Linux, Windows, Mac OS |
| Monitoring tool | Apache Airflow, Control-M |

**WORK EXPERIENCE**

**Role: Sr. AWS Data Engineer**

**Client: JP Morgan Chase & CO, Jersey City, NJ**

**Duration: September 2023 – Present**

**Responsibilities:**

* Developed Spark applications using Pyspark and Spark-SQL for data extraction, transformation, and aggregation from multiple file formats.
* Extensive experience in working with AWS cloud Platform (EC2, S3, EMR, Redshift, Lambda and Glue).
* Worked on building the data pipelines (ELT/ETL Scripts), extracting the data from different sources (MySQL, AWS S3 files), transforming, and loading the data to the Data Warehouse (AWS Redshift)
* Working knowledge of Spark RDD, Data Frame API, Data set API, Data Source API, Spark SQL and Spark Streaming.
* Created SQL scripts for daily extracts, ad-hoc requests, reporting and analyzing large data sets from S3 using AWS Athena, Hive and Spark SQL.
* Creation of ETL, built data pipelines using spark SQL, Pyspark, AWS Athena and AWS Glue.
* Writing to Glue metadata catalog which in turn enables us to query the refined data from Athena achieving a serverless querying environment.
* Developed Spark Applications by using Python and Implemented Apache Spark data processing Project to handle data from various RDBMS and Streaming sources.
* Worked with the Spark for improving performance and optimization of the existing algorithms in Hadoop.
* Using Spark Context, Spark-SQL, Spark MLlib, Data Frame, Pair RDD and Spark YARN.
* Used Spark Streaming APIs to perform transformations and actions on the fly for building common.
* Learner data model which gets the data from Kafka in real time and persist it to Cassandra.
* Performed API calls using the python scripting. Performed reads and writes to S3 using Botto3 libraries.
* Developed Kafka consumer API in python for consuming data from Kafka topics.
* Consumed Extensible Markup Language (XML) messages using Kafka and processed the XML file using Spark Streaming to capture User Interface (UI) updates.
* Performed Raw data ingestion into S3 from kinesis firehouse which would trigger a lambda function and pit refined data into another S3 bucket and write to SQS queue as aurora topics.
* Developed Preprocessing job using Spark Data frames to flatten JSON documents to flat file.
* Load D-Stream data into Spark RDD and do in memory data Computation to generate output response.
* Experienced in writing live Real-time Processing and core jobs using Spark Streaming with Kafka as a Data pipeline system.
* Migrated an existing on-premises application to AWS. Used AWS services like EC2 and S3 for data sets processing and storage.
* Experienced in Maintaining the Hadoop cluster on AWS EMR.
* Hands on expertise with AWS Databases such as RDS(Aurora), Redshift, DynamoDB and Elastic Cache (Memcached & Redis).
* Loaded data into S3 buckets using AWS Glue and PySpark. Involved in filtering data stored in S3 buckets using Elasticsearch and loaded data into Hive external tables.
* Configured Snow pipe to pull the data from S3 buckets into Snowflakes table.
* Stored incoming data in the Snowflakes staging area.
* Created numerous ODI interfaces and load into Snowflake DB.
* Worked on Amazon Redshift for shifting all Data warehouses into one Data warehouse.
* Good understanding of Cassandra architecture, replication strategy, gossip, snitches etc.
* Designed columnar families in Cassandra and Ingested data from RDBMS, performed data transformations, and then exported the transformed data to Cassandra as per the business requirement.
* Used the Spark Data Cassandra Connector to load data to and from Cassandra.
* Worked from Scratch in Configurations of Kafka such as Mangers and Brokers.
* Experienced in creating data-models for Clients transactional logs, analyzed the data from Cassandra.
* Tables for quick searching, sorting, and grouping using the Cassandra Query Language.
* Tested the cluster performance using Cassandra-stress tool to measure and improve the Read/Writes.
* Used Hive QL to analyze the partitioned and bucketed data, Executed Hive queries on Parquet tables.
* Stored in Hive to perform data analysis to meet the business specification logic.
* Used Apache Kafka to aggregate web log data from multiple servers and make them available in
* Downstream systems for Data analysis and engineering type of roles.
* Worked in Implementing Kafka Security and Boosting its performance.
* Experience in using Avro, Parquet, RCFile and JSON file formats, developed UDF in Hive.
* Developed Custom UDF in Python and used UDFs for sorting and preparing the data.
* Worked on Custom Loaders and Storage Classes in PIG to work on several data formats like JSON, XML, CSV, and generated Bags for processing using pig etc.
* Developed custom Jenkins jobs/pipelines that contained Bash shell scripts utilizing the AWS CLI to automate infrastructure provisioning.
* Written several Map Reduce Jobs using Pyspark, NumPy and used Jenkins for Continuous integration.
* Setting up and worked on Kerberos authentication principals to establish secure network communication.
* On cluster and testing of HDFS, Hive, Pig and MapReduce to access cluster for new users.
* Continuous monitoring and managing the Hadoop cluster through Cloudera Manager.

**Environment:** Python, Flask, NumPy, Pandas, SQL, MySQL, Cassandra, API, AWS EMR, Spark, AWS Kinesis, AWS Redshift, AWS EC2, AWS S3, AWS Beanstalk, AWS Lambda, AWS data pipeline, AWS cloud-watch, Docker, Shell scripts, Agile Methodologies.

**Role: Sr. Data Engineer**

**Client: Discover Financials, Chicago, IL**

**Duration: July 2022 – September 2023**

**Responsibilities:**

* Involved in gathering requirements, design, implementation, deployment, testing and maintaining of the applications to meet the organization's needs using SCRUM methodology.
* Followed Agile Software Development Methodology to build the application iteratively and incrementally. Participated in scrum related activities and daily scrum meetings.
* Implemented a 'serverless' architecture using API Gateway, Lambda, and Dynamo DB and deployed AWS Lambda code from Amazon S3 buckets. Created a Lambda function and configured it to receive events from your S3 bucket.
* Developed and implemented scalable data pipelines using Databricks on AWS to ingest, transform, and load large volumes of data into data lakes and data warehouses.
* Designed and optimized data models and schemas for efficient data storage and retrieval in Databricks on AWS.
* Designed the data models to be used in data intensive AWS Lambda applications which are aimed to do complex analysis, creating analytical reports for end-to-end traceability, lineage, definition of Key Business elements from Aurora.
* Writing code that optimizes performance of AWS services used by application teams and provide Code-level application security for clients (IAM roles, credentials, encryption, etc.)
* Used ETL component Sqoop to extract the data from MySQL and load data into HDFS.
* Good hands-on experience with Python API by developing Kafka producer, consumer for writing Avro Schemes.
* Managed Hadoop clusters using Cloudera. Extracted, Transformed, and Loaded (ETL) of data from multiple sources like Flat files, XML files, and Databases.
* Designed infrastructure for AWS application and workflow using Terraform and had done implementation and continuous delivery of AWS infrastructure using Terraform.
* Developed Python scripts to take backup of EBS volumes using AWS Lambda and Cloud Watch.
* Developed and deployed stacks using AWS Cloud Formation Templates (CFT) and AWS Terraform.
* Used Jenkins and pipelines which helped us drive all Microservices builds out to the Docker registry and then deployed to Kubernetes.
* Designed and setup Enterprise Data Lake to provide support for various uses cases including Storing, processing, Analytics and Reporting of voluminous, rapidly changing data by using various AWS Services.
* Used various AWS services including S3, EC2, AWS Glue, Athena, RedShift, EMR, SNS, SQS, DMS, kinesis.
* Extracted data from multiple source systems S3, Redshift, RDS and Created multiple tables/databases in Glue Catalog by creating Glue Crawlers.
* Created AWS Glue crawlers for crawling the source data in S3 and RDS.
* Created multiple Glue ETL jobs in Glue Studio and then processed the data by using different transformations and then loaded into S3, Redshift and RDS.
* Created multiple Recipes in Glue Data Brew and then used in various Glue ETL Jobs.
* Design and Develop ETL Processes in AWS Glue to migrate data from external sources like S3, Parquet/Text Files into AWS Redshift.
* Used AWS glue catalog with crawler to get the data from S3 and perform SQL query operations using AWS Athena.
* Implemented data governance practices and data cataloging using Databricks and AWS services to enhance data discovery and lineage capabilities.
* Conducted capacity planning for Redshift Spectrum clusters, forecasting resource needs based on data growth and query demands.
* Assisted in data migration efforts, extracting, transforming, and loading data from on-premises data warehouses to Redshift Spectrum in the cloud.
* Actively monitored and troubleshooted data pipelines and jobs in Databricks on AWS to ensure data availability and reliability.
* Written PySpark job in AWS Glue to merge data from multiple tables and in Utilizing Crawler to populate AWS Glue data Catalog with metadata table definitions.
* Used AWS Glue for transformations and AWS Lambda to automate the process.
* Used AWS EMR to transform and move large amounts of data into and out of AWS S3.
* Created monitors, alarms, notifications and logs for Lambda functions, Glue Jobs using CloudWatch.
* Performed end- to-end Architecture & implementation assessment of various AWS services like Amazon EMR, Redshift and S3.
* Used AWS EMR to transform and move large amounts of data into and out of other AWS data stores and databases, such as Amazon Simple Storage Service (Amazon S3) and Amazon DynamoDB.
* To analyze the data Vastly used Athena to run multiple queries on processed data from Glue ETL Jobs and then used Quick Sight to generate Reports for Business Intelligence.
* Used AWS EMR to transform and move large amounts of data into and out of AWS S3.
* Used DMS to migrate tables from homogeneous and heterogeneous DBs from On-premises to AWS Cloud.
* Created Kinesis Data streams, Kinesis Data Firehose and Kinesis Data Analytics to capture and process the streaming data and then output into S3, Dynamo DB and Redshift for storage and analyzation.
* Created Lambda functions to run the AWS Glue job based on the AWS S3 events.
* Designed and Developed ETL jobs to extract data from Salesforce replica and load it in data mart in Redshift. Responsible for Designing Logical and Physical data modeling for various data sources on Confidential Redshift. Experienced with event-driven and scheduled AWS Lambda functions to trigger various AWS resources.
* Integrated lambda with SQS and DynamoDB with step functions to iterate through a list of messages and update the status into a DynamoDB table.

**Environment:** AWS Services, Python, Pyspark, AWS Glue, S3, IAM, EC2, RDS, Redshift, EC2, RDS,Redshift, EC2, CloudWatch, Lambda, Boto3, DynamoDB, Apache, Spark, Kinesis, SNS, Kafka, Database, Datawarehouse, ETL, Athena, Hive, Sqoop, Splunk, Terraforms, Databricks.

**Role: Sr. Data Engineer**

**Client: Anthem, Norfolk, Virginia**

**Duration: June 2020 to July 2022**

**Responsibilities:**

* Involved in importing the real time data to Hadoop using Kafka and implemented the Oozie job for daily data.
* Loaded the data from Teradata to HDFS using Teradata Hadoop connectors.
* Import the data from different sources like HDFS/HBase into Spark RDD
* Developed Spark scripts by using Python shell commands as per the requirement
* Issued SQL queries via Impala to process the data stored in HDFS and HBase.
* Used the Spark - Cassandra Connector to load data to and from Cassandra.
* Used Restful Web Services API to connect with the MapRtable. The connection to Database was developed through restful web services API.
* Involved in developing Hive DDLs to create, alter and drop Hive tables and storm, & Kafka.
* Worked with application teams to install operating system, Hadoop updates, patches, version upgrades as required.
* Experience in data migration from RDBMS to Cassandra. Created data-models for customer data using the Cassandra Query Language.
* Responsible for building scalable distributed data solutions using Hadoop cluster environment with Horton works distribution
* Involved in developing Spark scripts for data analysis in both Python and Scala. Designed and developed various modules of the application with J2EE design architecture.
* Worked with stakeholders to identify data access patterns and optimize data partitioning and clustering to improve query performance in Redshift Spectrum.
* Designed and implemented data retention and archiving strategies using Redshift Spectrum, ensuring optimal storage utilization and data lifecycle management.
* Implemented modules using Core Java APIs, Java collection and integrating the modules.
* Experienced in transferring data from different data sources into HDFS systems using Kafka producers, consumers, and Kafka brokers
* Installed Kibana using salt scripts and built custom dashboards that can visualize aspects of important data stored by Elastic search.
* Used File System Check (FSCK) to check the health of files in HDFS and used Sqoop to import data from SQL server to Cassandra
* Streaming the transactional data to Cassandra using Spark Streaming/Kafka
* Implemented a distributed messaging queue to integrate with Cassandra using Apache Kafka and Zookeeper.
* ConfigMap and Daemon set files to install File beats on Kubernetes PODS to send the log files to Log stash or Elastic search to monitor the different types of logs in Kibana.
* Created Database in Influx DB also worked on Interface, created for Kafka also checked the measurements on Databases.
* Installed Kafka manager for consumer lags and for monitoring Kafka Metrics also this has been used for adding topics, Partitions etc. Successfully Generated consumer group lags from Kafka using their API.
* Ran Log aggregations, website Activity tracking and commit log for distributed system using Apache Kafka.
* Used Kafka as a message cluster to pull/push messages into spark for data ingestion, processing, and storing the resultant data in AWS S3 Buckets.
* Involved in creating Hive tables and loading and analyzing data using hive queries.
* Developed multiple MapReduce jobs in java for data cleaning and preprocessing. Loading data from different source
* (database & files) into Hive using the Talend tool.
* Used Oozie and Zookeeper operational services for coordinating cluster and Scheduling workflows.
* Implemented Flume, Spark, and Spark Streaming framework for real time data processing.

**Environment:** Hadoop, Python, HDFS, Hive, Scala, MapReduce, AWS S3 Buckets, Agile, Cassandra, Kafka, AWS EC2, AWS, YARN, Spark, ETL, Teradata, NoSQL, Oozie, Java, Cassandra, Talend, LINUX, Kibana,HBase.

**Role: Data Engineer**

**Client:**  **DaVita, Denver, CO**

**Duration: Jan 2018 – May 2020**

**Responsibilities:**

* Develop deep understanding of the data sources, implement data standards, maintain data quality, and master data management.
* Build Complex distributed systems involving huge amount data handling, collecting metrics building data pipeline, and Analytics.
* Expert in building Databricks notebooks in extracting the data from various source systems like DB2, Teradata and perform data cleansing, data wrangling, data ETL processing and loading to AZURE SQL DB.
* Performed ETL operations in Azure Databricks by connecting to different relational database source systems using job connectors.
* Used Pig as ETL tool to do Transformations, even joins and some pre-aggregations before storing the data on to HDFS.
* Implemented incremental loading strategies using SSIS to efficiently process large volumes of data and improve ETL performance.
* Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL and U-SQL Azure Data Lake Analytics.
* Involved in Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing the data in In Azure Databricks.
* Developing ETL pipelines in and out of data warehouses using a combination of Python and Snowflake Snow SQL Writing SQL queries against Snowflake.
* Developed ETL Processes in AWS Glue to migrate Campaign data from external sources like S3, ORC/Parquet/Text Files into AWS snowflake.
* Developing APIs for Integration Project, Implemented and extensively used the Object-Oriented Programming concepts in C#, Solid principles.
* Developed and optimized SQL queries for complex data manipulations, extractions, and aggregations, ensuring high-performance and responsiveness.
* Created Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform, and load data from different sources like Azure SQL, Blob storage, Azure SQL Data warehouse, write-back tool and backwards.
* Involved in developing Spark applications using PySpark and Spark-SQL for data extraction, transformation, and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
* Created several Databricks Spark jobs with PySpark to perform several tables to table operations.
* Create and maintain optimal data pipeline architecture in cloud Microsoft Azure using Data Factory and Azure Databricks.
* Involved in Architecting and developing the Azure data factory pipelines by creating the datasets and source and destination connections to move the data from oracle database to Azure Data Lake Datalake Store Raw Zone.
* Create Self Service reporting in Azure Data Lake Store Gen2 using an ELT approach  
  Transfer data in logical stages from System of records to raw zone, refined zone and produce zone for easy translation and denormalization.
* Creating Data factory pipelines that can bulk copy multiple tables at once from relational database to Azure Data Lake gen2.
* Create custom logging framework for ELT pipeline logging using Append variables in Data factory.
* Enabling monitoring and azure log analytics to alert support team on usage and stats of the daily runs.
* Involved in Configuring and Upgrading the On Premises Data gateway between various Data sources like SQL Server to Azure Analysis Services and Power BI service.
* Design and develop business intelligence dashboards, analytical reports and data visualizations using power BI by creating multiple measures using DAX expressions for user groups like sales, operations, and finance team needs.
* Utilized Ansible playbook for code pipeline deployment delivered denormalized data for PowerBI consumers for modeling and visualization from the produced layer in Data Lake.
* Experience in data transformations, including calculated columns, maintaining relationships, establishing different metrics, queries, changing values, splitting columns, and grouping by date & column.
* Building contemporary data warehouse, and real-time analytics solutions made possible by native interaction with Azure Active Directory (Azure AD) and other Azure services. Storage, Azure SQL, and Azure DW, as well as data processing in Azure Databricks.
* Proven ability to design and implement data flow from source to target to BI tool, ensuring seamless data integration and availability for analytics and reporting.
* Data was gathered, converted, and loaded from source systems to Azure Data Storage services using a combination of Azure Data Factory, T-SQL, Spark SQL, and U-SQL Azure Data Lake Analytics.
* Worked on integration of Azure Data Factory and other Azure services enables to build modern data warehouse and real-time analytics solutions.
* Deep understanding of ACID transactions, ensuring reliable and consistent data processing and maintaining data integrity in transactional systems.
* Developed code in Azure Databricks for gathering and transforming data from various sources MongoDB, MSSQL.
* Developed pipelines using Data Factory and performed batch processing using Azure Batch Processing.
* Built different application platforms in the cloud by leveraging Azure Databricks.  
  Created tables in Azure SQL Data Warehouse for the reporting and visualization of business-related data.
* Created Spark applications utilizing PySpark and Spark SQL in Databricks to modify and aggregate source data before importing it into Azure Synapse Analytics for reporting.
* Responsible for creating Requirements Documentation for various projects.
* Hands-on experience with Azure Cloud Technologies, including Azure Data Factory, Azure Synapse Analytics, or Azure Blob Storage, demonstrating proficiency in utilizing cloud-based data engineering tools for scalable and flexible data processing.
* Self-starter mentality with a proactive and motivated approach, showing a strong ability to work independently and drive projects to completion with minimal supervision.
* Superior communication skills, strong decision making and organizational skills along with outstanding analytical and problem-solving skills to undertake challenging jobs.
* Able to work well independently and in a team by helping to troubleshoot technology and business-related problems.
* Generate incident reports, change reports, turnovers Summary report weekly basis.

**Environment:** Azure SQL Database, Azure Data Lake, Azure Data Factory (ADF), Azure SQL Data Warehouse, Azure Analysis Service (AAS), Azure Blob Storage, Azure Search, Azure App Service, Snowflake, SQL, Azure Data Factory (ADF), Azure Synapse, Azure Database Migration Service (DMS), GIT, PySpark, Python, JSON, ETL Tools, SQL Azure, C#.

**Role: Data Analyst/ ETL Developer**

**Client: Delta, Hyderabad, India**

**Duration: January 2015 to November 2017**

**Responsibilities:**

* Coordinated with front-end application developers for implementing database architecture and design.
* Used various transformations in SSIS dataflow, control flow using for loop containers, and fuzzy lookups.
* Develop parameterized reports, caching reports, sub reports and Ad Hoc reports using SSRS.
* Implemented error handling and utilized event handlers for automated notifications using SSIS.
* Write Complex SQL Queries to generate Reports based on the business requirement.
* Redesigned the SSIS packages from the legacy DTS packages.
* Execute SSIS package include a master package which include number of child packages.
* Supporting ETL (Extract Transform and Load) for fetching data from multiple systems to single Data Warehouse.
* Created complex Ad-Hoc reports, Sub reports, linked reports related to State compliance reporting. Used custom code in SSRS for row color, visibility, and masking.
* Developed Full Analysis Cycle Project and created packages for extracting data from OLTP to OLAP. Created Multi-Dimensional Expression (MDX) scripts for OLAP data cubes.
* Involved in Designing, Developing and Testing of the ETL (Extract, Transformation and Load) strategy to populate the data from Heterogeneous data sources (SQL Server, Flat Files, Excel source files, XML files etc.)
* Performed different kinds of transformations like Lookup Transformations, Merge Joins, Derived Columns, Merge Join, Conditional Split, and Data Conversion with Multiple Data Flow tasks.
* Bulk data migration using Bulk Insert from flat files.
* Created Package Configurations, Event Handlers for On Error, On Pre/Post execution. Designed Complex Packages with Error Handling and Package Logging that stores the Logging results in SQL Tables and log files  
  Performance tuning and testing on stored procedures, indexes using performance tool like SQL Profiler and DETA.
* Import/Export and successfully migrated the Server usage MS Access database data to SQL Server using ETL/SSIS. Created dynamic and customized packages to support future changes. Scheduled the same packages by creating the corresponding job tasks.
* Formatted the reports using the global variables and expressions in SSRS, deployed the generated reports on to the Reporting server.

**Environment:** MS SQL Server 2012, Visual Studio 2010, T-SQL, MS Excel, Microsoft SQL Server Integration Services (SSIS), Microsoft SQL Server Reporting Services (SSRS), Rally bug tracking and SVN.

**EDUCATION:**

Bachelors in Information Technology, VNR VJIET- 2014