**Shelly Bansal**

**s.bansal3010@gmail.com**

**(301) 820-5693**

**Professional Summary**

* **Dedicated Data engineer and analyst with over 11 years of experience** in data analysis, reporting, and visualization. Proven track record of driving data-driven decision-making and delivering actionable insights. Seeking a challenging role to leverage my analytical skills and contribute to the success of a dynamic organization
* Working knowledge of migrating SQL databases to Azure data lake, Azure data lake analytics, Azure SQL Database, Databricks, and Azure SQL Data warehouse, and controlling and granting database access and migrating on-premises databases to Azure Data Lake storage using Azure Data factory.
* Worked on various projects related to Data Modeling, System/Data Analysis, Design, and Development both for OLTP and Data Warehouse applications.
* Experience with AWS services like S3, Athena, Redshift Spectrum, Redshift, EMR, Glue, Data pipeline, step functions, cloud watch, SNS, and Cloud formation.
* Experience with Snowflake cloud data warehouse and AWS S3 bucket for integrating data from multiple source systems, including loading nested JSON formatted data into Snowflake table.
* Working knowledge of ETL (Extraction, Transformation, and Loading) data from various sources into Data Warehouses, as well as data processing (such as collecting, aggregating, and moving data from various sources) using Apache Flume, Kafka, Power BI, and Microsoft SSIS.
* Extensive knowledge on Amazon Web Services (AWS) EC2, S3, Elastic Map Reduce (EMR) and on Snowflake, Redshift, Identity and Access Management (IAM).
* Built pipelines in Azure Data Factory to transfer data between locations and created ETL pipelines in and out of the data warehouse using a combination of Python and Snowflake's Snow SQL.
* Automated OpenStack and AWS deployment using Cloud Formation, Ansible, Chef, and Terraform.
* Experience using Fivetran SQL to build and optimize data models within a data warehouse, including creating tables, indexes, and views.
* Expert in importing and exporting data between HDFS and Relational Systems such as MySQL and Teradata using Sqoop.
* Experience building ETL pipelines to load data into Data Vault models from various source systems.
* Experience with ETL tools such as Informatica, Talend, DellBoomi or DataStage, for extracting data from various source systems and loading it into the EDW.
* Developed Python scripts for data modeling and import/export, and extensive experience in deploying, managing, and developing MongoDB clusters. Wrote Script for performing DML operations with MongoDB.
* Proficient in designing and developing ETL processes using Control-M and other ETL tools, such as Informatica to integrate data from multiple sources and support downstream analytics and reporting.
* Expert on creating SQL objects like Tables, Complex Stored Procedures, Triggers, Views, Indexes, and User Defined Functions to facilitate efficient data manipulation and consistent data storage in Azure Synapse.

**Skills**

• Data Analysis

• Data Modeling

• SQL/ MySQL

• ETL/ELT pipeline

• Data Visualization (Tableau)

• Data Warehousing

• Git, GitHub, Bitbucket

• Python, Java

• Informatica

• Atlassian JIRA

• Report Generation

• S3/Lambda/Redshift

• Blob/Cosmos/Azure DB

• Big Query/Cloud Shell

**Professional Experience**

**TransAmerica – TX, USA**

**Senior Lead Data Engineer Remote**

**05/2023 - Current**

• Developed and maintained end-to-end operations of ETL data pipelines and worked with large data sets in Azure Data Factory and Azure Synapse environments.

• Leveraged existing Apache Airflow DAGs to facilitate loading of data into target systems, streamlining data ingestion and transformation processes.

• Debugged and troubleshooted issues within Apache Airflow DAGs to ensure smooth data loading operations, identifying and resolving errors or inconsistencies in process

• Expertise in utilizing AWS S3 as scalable object storage solution, implementing data lake architectures and optimizing data storage and retrieval for enhanced performance.

• Implemented AWS Athena for interactive querying of data stored in S3, enabling ad-hoc analysis and exploration of data with SQL-like syntax, reducing need for data preprocessing.

• Designed and optimized ETL processes using AWS Glue and Spark, enabling efficient extraction, transformation, and loading of data from diverse sources into target data warehouses.

• Demonstrated expertise in utilizing Jenkins as a key tool for deploying code to various environments, including development, testing, staging, and production.

• Proficient in utilizing BitBucket as a version control system for managing source code repositories, including creating branches, merging code, and resolving conflicts.

**Cigna, CT - Senior Data Engineer - Azure, Remote**

**05/2022 - 04/2023**

• Developed and maintained end-to-end operations of ETL data pipelines and worked with large data sets in Azure Data Factory.

• Researched and implemented various components like pipeline, activity, mapping data flows, data sets, linked services, triggers and control flow.

• Performed extensive debugging, data validation, error handling mechanism, transformation types and data clean up analysis within large datasets.

• Collaborated with DevOps Engineers to developed automated CI/CD and Test driven development pipeline using Azure as per client standards.

• Experience in developing Spark applications using PySpark and Spark- SQL in Databricks for data extraction, transformation, and aggregation.

• Experience in using REST APIs to retrieve analytics data from different data feeds to meet business requirements.

• Expertise in building python notebooks in Azure Databricks to implement transformations as per business needs.

• Worked on Tableau Extracts to send data to Dashboards connected to Databricks and Azure Synapse Analytics (DW).

**Celebal Technologies Remote**

**Data Engineer 03/2021 - 04/2022**

• Developed and maintained interactive dashboards and reports in Tableau, providing real-time insights to support executive decision-making.

• Handled key projects for "Field Service Order KPI Determination” for different business models and different use cases.

• Gathered historical data on field service orders, including order placement timestamps, order details, technician availability, and scheduling-related information.

• Defined a set of KPIs and metrics that will help measure and evaluate the efficiency and effectiveness of the scheduling process. Potential KPIs include:

Order-to-Schedule Time: The time taken from order placement to scheduling.

Scheduling Accuracy: The percentage of orders scheduled without conflicts or rescheduling.

• Utilized data analysis techniques and tools (e.g., SQL, BigQuery, Excel, Python) to process and analyze the collected data.

• Created interactive dashboards and reports using Tableau visualize KPI trends over time.

• Developed a real-time monitoring dashboard to track these KPIs for ongoing orders

• Established a process for ongoing monitoring and reporting of the selected KPIs.

• Conducted ad-hoc data analysis to identify trends, patterns, and anomalies, resulting in actionable recommendations to improve business processes.

• Played a key role in data quality improvement initiatives, identifying data quality issues and implementing data cleansing processes.

• Assisted in the development of predictive models to forecast business outcomes and optimize resource allocation.

• Developed Python scripts for extracting data from web services API's and loading into databases.

• Adept in troubleshooting and identifying current issues and providing effective solutions.

**The Standard - Data Engineer III Portland, Oregon**

**04/2019 - 02/2021**

• Design and build ETL pipelines to automate ingestion of structured and unstructured data using Azure Data Factory.

• Experience in Data Ingestion from various sources to one or more Azure Services – (Azure Data Lake, Azure Storage, Azure SQL, Azure DW) and processing data in Azure Databricks.

• Design and automated DataLab's space (Sandbox environments) in Azure environment using PowerShell Scripts and Python Scripts for business teams.

• Deployed and managed (CI/CD) pipelines to build code using Azure

DevOps. Used Kusto (Azure Data Explorer) Queries to analyze data from ADF monitoring through Log Analytics Workspaces.

• Hands-on experience using PowerShell, Azure Runbooks and Webhooks for automating development and deployments.

• Created YAML pipeline templates(.yml) in pre-production and production environment using Azure DevOps.

• Developed Spark code using Python (PySpark) and Spark-SQL for faster testing and processing of data and Implemented PySpark applications to ingest data into data lake central repository from different legacy sources.

• Worked with Data Governance team (Cross Functionality) to install and configure Collibra infrastructure over Azure environment and helped in migration process from on-premises to cloud. Managed and implemented Sandbox environment with team of 10 members.

**Sisense, NY -Data Engineer (Consultant), Redmond, WA**

**06/2016 - 03/2019**

• 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.

• Developed Spark applications using Scala 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.

• Responsible for estimating cluster size, monitoring, and troubleshooting of HDInsight clusters.

• Used Zeppelin, Jupyter notebooks and Spark-Shell to develop, test and analyze Spark jobs before Scheduling Customized Spark jobs.

• Deployed and tested (CI/CD) our developed code using Visual Studio Team Services (VSTS).

• Expertise in creating HDInsight cluster and Storage Account with End- to-End environment for running jobs.

• Hands-on experience on developing PowerShell Scripts for automation purpose.

• Created Build and Release for multiple projects (modules) in production environment using Visual Studio Team Services (VSTS).

• Experience in using Scala Test Funsuite Framework for developing Unit Tests cases and Integration testing.

• For Log analytics and for better query response used Kusto Explorer(Azure Data Explorer).

• Migrated existing MapReduce programs and Hive Queries into Spark application using Scala.

**GetServ Technologies Noida, India**

**Data Analyst 10/2011 - 03/2016**

• Collaborated with stakeholders across the organization to understand their needs from a data perspective.

• Maintained documentation of all the processes related to Data Analysis.

• Performed quality assurance testing on newly developed reporting solutions.

• Automated manual reporting processes by developing scripts in Python on AWS Lambda.

• Defined KPIs based on collected data sources.

• Built visualizations that enabled users to quickly interpret results from complex analyses.

• Generated reports and obtained data from AWS S3 storage to develop analytics on key performance and operational metrics.

• Worked with internal teams to understand business needs and changing strategies.

• Synthesized BI or trend data in AWS Redshift to support recommendations for action.

• Created or reviewed technical design documentation to drive accuracy of reporting solutions.

**Education And Training**

05/2023

**Training/Certification:**

SQL - MySQL for Data Analytics and Business Intelligence

Udemy

08/2011

**Master in Computer Science**

**Punjab Technical University**

Jalandhar, India