

Data Scientist / Python Developer Data Scientist / Python Developer Data Scientist / Python Developer McLean, VA Work Experience Data Scientist / Python Developer PenFed Credit Union - McLean, VA October 2018 to Present

Implemented full lifecycle in Data Modeler/ Data Analyst, Data warehouses and DataMart's with Star Schemas, Snowflake Schemas, and SCD& Dimensional Modeling Erwin

Gathering all the data that is required from multiple data sources and creating datasets that will be used in analysis.

Update Python scripts to match training data with our database stored in AWS Cloud Search, so that we would be able to assign each document a response label for further classification.

Analyze data from AWS Redshift database in Python using psycopg2, Pandas and Numpy module

Use AWS Python SDK to download data from DynamoDB.

Use Python MySQL client to download data from AWS RDS MySQL database. Wrote SQL queries to analyze transactional and customer data from MySQL database.

Use Spark-context to analyze dataset and explore data using lambda function.

Develop MapReduce/Spark Python modules for machine learning & predictive analytics in Hadoop on AWS.

Implement end-to-end systems for Data Analytics, Data Automation and integrated with custom visualization tools using R, Python and Hadoop

Create Data Quality Scripts using SQL and Hive to validate successful data load and quality of the data.

Create several types of data visualizations using Python and Tableau.

Perform data analysis by using Hive to retrieve the data from the Hadoop cluster, SQL to retrieve data from Oracle database.

Wrangle data, worked on large datasets (acquired data and cleaned the data), analyzed trends by making visualizations using matplotlib and python.

Oversees the mapping of the integrating process for data into Power BI tool and Reporting Platforms from the data warehouse.

Create stored procedures in various complexities for SSRS report datasets and Power BI dashboards.

Created sub-reports, drilldown-reports, summary reports, and parameterized reports in SSRS

Develop data mapping documentation to establish relationships between source and target tables including transformation processes using SQL.

Tableau is used for analyzing the data to show the trends, variations and density of the data in form of graphs and charts. Tableau was connected to files, relational and big data sources to acquire and process data.

Build and maintain SQL scripts, indexes, and complex

queries for data analysis and extraction. Environment: Python, DynamoDB, MySQL, SQL, AWS RDS, AWS Python SDK, Spark, Lambda function, Map-Reduce, Jupyter Notebook, PowerBI, SSRS, Tableau Data Scientist Morgan Stanley - Baltimore, MD January 2017 to September 2018 Created 3NF business area data modeling with de-normalized physical implementation data and information requirements analysis using ERWIN tool. Participated in all phases of data mining; data collection, data cleaning, developing models, validation, visualization and performed Gap analysis. Built analytical data pipelines to port data in and out of Hadoop/HDFS from structured and unstructured sources and designed and implemented system architecture for Amazon EC2 based cloud-hosted solution for client. Performed data analysis by retrieving the data from the Hadoop cluster. Created Logical and Physical data models with Star and Snowflake schema techniques using Erwin in Datawarehouse as well as in Data Mart. Involved in Data Analysis, Data Validation, Data Cleansing, Data Verification and identifying data mismatch. Resolved the data related issues such as: assessing data quality, data consolidation, evaluating existing data sources. Participated in data modeling discussion and provided inputs on both logical and physical data modelling. Data transformation from various resources, data organization, features extraction from raw and stored. Collaborated with data engineers and operation team to implement the ETL process, wrote and optimized SQL queries to perform data extraction to fit the analytical requirements. Involved in creating various reports like Drill through, drill down and Ad-hoc according to the user requirement using SQL Server Reporting Services (SSRS). Tools: Hadoop, Map-Reduce, Tableau, Python Education Master's Skills Mapreduce, Clustering, Hadoop, Data analysis, Mysql, Sql, Git, Hadoop, Json, Mapreduce, Python, Ggplot2, Matplotlib, Numpy, Pandas, Shiny, Anova, Boosting, Decision trees, Dimensionality reduction

Name: Scott Williams

Email: kristi43@example.com

Phone: 763-878-9903x77968