Python Data Analyst Python Data Analyst Python Data Analyst - Schlumberger Buffalo, NY   Robust professional experience in Data collection, Data Extraction, Data Cleaning, Data Aggregation, Data Mining, Data verification, Data analysis, Reporting, and data warehousing environments.   Working knowledge in google cloud platform (Big Query, Cloud DataProc, Composer/AirFlow).   Skilled experience in Apache Hadoop ecosystem such as HDFS, MapReduce, Hive, Pig, Sqoop, Spark and Presto.   Hands on experience in developing SPARK applications using Spark tools like RDD transformations, Spark core, Spark Streaming (POC) and Spark SQL.   Experience with building data pipelines in python/Pyspark/HiveSQL/Presto/BigQuery and building python DAG in Apache Airflow.   Extensive experience in querying languages using SQL, PL/SQL, T-SQL, HiveSQL, Spark SQL, Presto.   Hands on experience with Python Programming for data analysis (Pandas, Numpy).

  Proficient in Data Analysis with sound knowledge in extraction of data from various database sources like MySQL, MSSQL, Oracle, Teradata and other database systems.   Working knowledge in CI/CD and GIT.   Developed Python jobs to extract and load data into MySQL database and use apache Sqoop to load data into HDFS.   Automate the process of data extraction, transformation and processing using shell scripts.   Used Test driven approach for developing the application and implemented the unit tests using Python unit test framework.   Worked in creating different Visualizations in Power BI using Bar charts, Line charts, Pie charts, Maps, Scatter Plot charts, Heat maps and Table reports.   Extensive experience in various reporting objects like Facts, Attributes, Hierarchies, Transformations, Filters, Prompts, Calculated Fields, Sets, Groups, Parameters in Power BI.   Extensive knowledge in working with SQL Server, ORACLE, PostgreSQL Databases. Experience in integration of various data sources like SQL Server, Oracle, Access DB, DB2 and Flat File, Excel File.   Expertise in writing T-SQL Queries, Dynamic-queries, sub-queries and complex joins for generating Complex Stored Procedures, Triggers, User-defined Functions, Views and Cursors.   Highly proficient in packages to extract, transform and load data (ETL) using SSIS, designed packages which are utilized for tasks and transformations Data Conversion and Pivot tables.   Experience in transforming unstructured data from various data sources using SSIS like Conditional Split, Lookup, Merge Join and Sort.   Experience in debugging the SSIS packages

using breakpoints and checkpoints.   Highly proficient in the use of SQL for developing complex stored procedures, triggers, tables, user profiles, user functions, relational database models and data integrity, SQL joins and query writing.   Widely worked on to improve SQL Query performance, SQL Query analytics and highly complex joins to populate the desired result set.   Extensively worked on Table Design creating Clustered Indexes, Non-Clustered Indexes, Primary key Foreign Key Relationships, Highly complex joins and SQL Scripts for Views and Materialized Views. Authorized to work in the US for any employer Work Experience Python Data Analyst Schlumberger - Houston, TX August 2017 to Present Responsibilities: Used pandas library for data manipulation and numpy for numerical analysis and managed large datasets using pandas data frames and MySQL.   Helping the client partially migrate from MySQL to Hadoop ecosystem.   Experienced in using Hortonworks on-premise Hadoop distribution and have written HiveSQL and Sqoop scripts. Worked briefly in writing pyspark core methods for speeding up HiveSQL queries like non-equi joins.

   Used apache Sqoop import and export and handled data types after moving.   Did POC's by doing bucketing and partitioning on a hive table to understand performance.   Knowledge in various file formats in HDFS like avro, orc, parquet.   Extensive experience in writing python functions and advanced.   Wrote and executed various MYSQL database queries from Python using Python-MySQL connector and MySQL dB package.   Used Collections in Python for manipulating and looping through different user defined objects. Data Analyst/ Engineer JP Morgan Chase - Plano, TX January 2016 to August 2017 Responsibilities: Hands on experience in using Google cloud platform for big query, cloud dataproc and apache airflow services.   Got involved in migrating on prem Hadoop system to using GCP (Google Cloud Platform).   Expertise knowledge in HiveSQL, Presto SQL and Spark SQL for ETL jobs and using the right technology for the job to get done. Migrated previously written cronjobs to airflow/composer in GCP.   Experience in using stack driver service/ dataproc clusters in GCP for accessing logs for debugging.   Have written python DAGs in airflow which orchestrate end to end data pipelines for multiple applications.   Was involved in setting up of apache airflow service in GCP.   Expert level knowledge in using python API for spark and python pandas package for ETL jobs. SQL/ETL Developer American Express - IN October 2013

to July 2014 Responsibilities: Responsible for gathering requirements for the new projects and creating the data flow model of the Business requirement with MS Visio 2013.   Involved in integrating different Database Applications using SSIS packages.   Creating SSIS packages that involves migration from legacy systems to centralized Database.   Designed and tested packages to extract, transform and load data (ETL) using SSIS, Designed packages which are utilized for tasks and transformations such as Execute SQL Task, Mapping the right and required data from source to destination, Data Flow Task, Data Conversion, For each Loop Container.   Validate data before performing ETL transformations and load into SQL server database.   Intensively worked on transformations like Derived column, import column, Export column, Row count, Union all, Copy column, Lookup, Data Conversion with SSIS 2012.   Implemented transformation mapping through BizTalk server with the help of Visual Studio to achieve data transformation from one type to another.   Worked on Deployment of packages into Production servers and managed scheduled SQL jobs.   Created stored procedures, Cursors, Joins in T-SQL.   Implemented Distributed Transaction Control for the SSIS packages and event handlers to keep track of error logs. Also, to trigger emails about the failure using Email task.   Designed and created Report templates, bar graphs and pie charts based on the inspection data. SQL/MSBI Developer Kastech Solutions Pvt Ltd - Hyderabad, Telangana July 2011 to September 2013 Responsibilities: Extensively worked on production support that includes Deployment of SSIS Packages into Development, Production Servers, Creating, Scheduling and Managing the SQL Jobs.   Creating SSIS Package Configurations and maintaining their Tables by editing the values of the variables as per the requirement.   Using SQL Server Integration Services (SSIS) Migrated data from Heterogeneous Sources and legacy system (Oracle, Access, Excel) to centralized SQL Server databases to overcome transformation constraints and limitations.   Extensively worked on Design and development of SSIS (ETL) packages to extract, transform and load data from different Sources like Excel files, flat files to the Data warehouse using different tasks and transformations like File System Task, Data Flow Task, Derived Columns, and Execute SQL Task.   Extensively worked on the packages created including a variety of transformations, for example Slowly Changing Dimensions,

Lookup, Aggregate, Audit, Conditional Split, Merge join, Multi Cast, Pivot, OLE DB Command, Sort, Union All, and Copy Column, Derived Column using SSIS.   Worked on the row count transformation or row count expression and Record Set Destination and event handlers control flow to populate the special Log tables with high level SSIS package execution.   Worked on Execution of the paths depending on the success, failure or Completion of other tasks with precedence constraints and expressions to determine the workflow. Education Master of Science in Computer Science in Computer Science Northwestern Polytechnic University Present

Name: Eric Harris

Email: anthonysmith@example.com

Phone: (804)718-6140