

Data Engineer Data Engineer Data Engineer - RDOPS Wayne, PA Have around 7 years of work experience into development. Actively looking for opportunities in Big Data industry as a (Big Data Developer/Data Engineer/Tableau Developer). Extensive experience with different phases of project life cycle (Analysis, Design, Implementation, Testing, and debugging new and existing client-server based applications). Extensive experience as SQL Server Developer in utilizing triggers, cursors, functions and stored procedures. Strong technical knowledge in T-SQL including ETL, Microsoft Analysis Services, Performance tuning, Reporting, Designing logical/physical databases and Troubleshooting. Expertise in Database development for OLTP (Batch Processing, Online Processing), OLAP, ETL, Data warehousing, Data mining, DBMS and Data Modeling. Experience in creating Views, Constraints, Triggers, joins, Functions, Batch Scripts and numerous stored procedure consumed at various stages of migration/cleansing process, reports or data validation. Proficient in writing complex T-SQL (DDL, DML), Ranking Functions, TOP(n), PIVOT, XML PATH, SQL programming, physical, logical database design and SQL Performance tuning. Excellent knowledge in ETL validation, Data Model Validations and troubleshoot migration issues. Experience in Query tuning and performance-tuning using Clustered and Non-Clustered indexes on tables for faster query retrieval. Extensively worked on Normalized and de-normalized database design and RDBMS concepts to design Relational Database model as per Business requirement. Experience in developing Dashboard, Parameterized Reports, linked reports, Sub reports, Create complex Reports (ex: drill down reports using different objects), drill through reports using SSRS. Experience in Migrating databases from SQL 2008 R2 to SQL 2014 and Extraction & Migration of Data from heterogeneous data sources like MS Access, Teradata, and DB2 etc. to MS SQL Server.

High experience/comfort level interpreting ETL Data Mapping documents, transformation logic, reusable logic, etc. Worked on administration tasks such as creating logins, user mapping, batch jobs, backup and Recovery, Maintenance tasks, table management, upgrades, creating databases/File groups/files/Transaction logs. Experience in using Data Modeling Tools Microsoft Visio and Erwin, designing and developing periodic and on-demand reports using Crystal Reports and Report Builder. Ability to comprehend the requirements and formulate solutions in a deadline

oriented setting. Excellent track record as a team player with effective communication skills. Strong writing and documentation skills for the management and development. Authorized to work in the US for any employer

Work Experience

Data Engineer RDOPS July 2018 to Present

Responsibilities: Working on "LIA" - a Coordination of Benefits solution with intelligent analytics for Medicare and Medicaid claims. Clearing off the Protected Health Information (PHI), cleaning data from EDI file formats (e.g. 837,834), flat files, client databases, and load into the MS SQL database.

Used Sqoop import to move data from MSSQL to HDFS in formats (parquet/ avro/ textfile), with compression codecs (gzip/ snappy/uncompressed) on incremental append with splits on for boundary conditions. Changing the number of mappers, delimiters of fields and lines, writing custom queries in Sqoop. Developed maven projects as KIE sessions, with dependencies and plugins for the running environments. Implemented business rules for Medicare/Medicaid insurance programs. Run Spark jobs using maven jar files and required files and rules are passed as arguments. This filters recoverable claims and gets saved as text files. Launching an EMR cluster (Release version 5.x) with Hive, Presto, Pig, Zoo Keeper, HBase, Sqoop, Spark. Creating steps for running custom maven jars and storing output files to S3 bucket security groups restrictions of authorized systems. Performed DDL operations in HQL on AWS EMR. Worked with TBLPROPERTIES, bucketing, SerDe formats, vectorization, compression sizes and types, indexing, clustering & partitioning, storage formats for tables. Worked on external tables, skewed tables, managed tables, temporary tables, transaction tables in Hive. Export data from SQL tables using Sqoop to Hive on AWS EMR cluster and applying DML operations. Use AWS Presto for querying, by connecting to Hive metastore which was used to create schema. Create AWS S3 buckets and pass path as arguments in spark jobs. Store output flat files into output bucket in the AWS S3. Use Tableau for joins, blends and unions depending on the cardinality. Connecting data from different sources, custom SQL queries, creating pivot tables, stored procedures. Using live and/or extract type connections. Developed a data pipeline for streaming applications to get live data with kafka and process it in pyspark/scala and store in required format in HDFS. Write pyspark/ scala streaming scripts as required. Implemented Oozie workflows to automate the data processing of

spark jobs and streaming applications. Developed charts, plots, continuous/discrete (Dimensions/ Measures), splitting on delimiters, used splits with data interpreter. Customizing filters, aggregations, colors, sizes, tooltip, label, analytic functions like reference, average, forecast. Created calculated fields, new parameters, KPIs, aliases, text tables with calculations like table/ pane (across/down). Developed views, dashboards, workbooks and projects. Optimizing screen resolutions, layouts, scaling, page orientations. Pushing data to Tableau online and creating user accesses to the workbooks, customizing start pages upon login. Environment: Microsoft SQL, Apache Maven, Tableau, Drools, Spark, AWS S3, AWS Presto, AWS EMR, HIPAA compliance. Hadoop Developer UAASC, Western Illinois University June 2016 to July 2018 Responsibilities: Worked on various projects as a software/data engineer to collect data required for research projects of various departments on fields like consumer products data, historical stock prices data, business products reviews etc. Running spark applications on both local mode on test sample and on clusters in yarn mode in clusters. Used Spark on text file datasets from HDFS/ using python text file reading operations on files in local file system and converting them to RDD applying transformations such as filtering, mapping, union and intersection using lambda operations. Setting up accumulator and broadcast variables using spark context, shuffling number of partitions. Converting data sets into (key, value) pairs and applying actions, performing joins on RDD's, applying coalesce to produce single output files, repartition the RDD data on a partitioner. Saving and loading to/from text/sequence/parquet/avro/json file formats with compression codecs like uncompressed, snappy, gzip. Developed global temporary views, temporary tables on the dataframes and using Spark Sql to query output. Performing aggregations functions using spark sql. Created Hive UDFs and UDAFs in java as maven projects with dependencies in pom.xml. Used Hive UDF/UDAF interface in java, importing jar into Hive environment and added it to class path. Automated OS operations in python scripts for combining files in folders, directory creation, deletions, autorunning scripts. Worked on creating Python 3 APIs for parsing data from a webpage and filter out the data and save in the required format. Used python libraries such as beautiful soup, requests, to mine data from URL, with python multi-processing, spoofing headers, iterating pages and filtering content from html

tags. Created a framework which has many API's to scrapping data from different web pages. Developed Django applications to display analysis results on a web application. Used HTML 5 and created pages to display the results, used Bootstrap and CSS 3. Input fields send HTTP responses, the respective files are queried from the database and sent as HTTP response calls and are injected to the Front-end display. Was responsible for creating Tableau Dashboards for the data collected for the Finance/marketing department. Creating views from the data generated after cleaning as per the business requirements. Environment: Python, Django, MS-SQL, AWS, HTML, CSS, Spark, Tableau. Software Engineer First Republic Bank May 2012 to June 2016 Responsibilities: Worked in an Agile environment for the project, tools used were JIRA, GIT. Following up with the activities on JIRA regarding issues, reports, related documents, coordinating with the development, business and testing teams from various locations. Creating an issue, tracking sub tasks, changing and updating fix versions. Creating Scrum projects, discussing about sprints to be added, standup meetings, burndown charts, summary reports. Dealing with use cases, sequence and activity diagrams as a part of UML diagrams, Software Requirements Specification. Deliver test artifacts adhering to customer's needs. Performed Requirement based testing, following the DO - 178B guidelines, Software Verification, and Validation included Analyzing and correct the bugs identified during Testing, performed White - box and Black - box testing. Modified Condition/ Decision Coverage (MCDC) code coverage functionality check by setting T & F conditions in the code. Used GIT to maintain repository, creating and merging branches, commit changes, checking out, moving and removing files. Creating data models, stored procedures, queries for data analysis and manipulations, views, functions. Maintain, upgrade databases and creating backups in SQL. Develop automated python scripts for repetitive task like delimiters splitting, characters joining, stray values filtering, date and data format conversions, regex operations (like code matching, replacing, pattern matching). Environment: Python development, MS SQL, C. Education MS in Computer Science Western Illinois University Additional Information SKILLS: AWS: Real time working knowledge on AWS services like EMR clusters, EC2, Security Configurations, storing data in S3 Querying the tables using Athena, Presto, HQL. Databases like AWS RedShift, Snowflake, Apache

Cassandra, implementing JDBC connectors to AWS. AWS CodeBuild and AWS CodePipeline for Continuous Integration and Continuous Deployment. Hadoop & ecosystems: Experience on HDFS environment and Big data technologies like Apache Spark, Apache Hive, Sqoop, Apache Kafka, Flume, NoSQL database like Apache HBase, writing scripts in Apache Pig. Working with file formats (text/avro/json/parquet) and file compressions (gzip/snappy/uncompressed). Good understanding of Hadoop components like HDFS, Hadoop MapReduce, and YARN. Implemented batch processing, live streaming using Kafka and spark streaming using pyspark/scala. Automating workflow of jobs using Apache Oozie through cli and GUI. Tableau: Proficient with creating interactive dashboards in Tableau, accessing data from multiple sources. Created custom SQL queries for live/extract connections in Tableau. Worked and maintained Tableau online (such as creating projects, workbooks, views, user access privileges). Experience on calculated fields, KPIs, creating views, charts, plots and trees. Worked on saving in different formats, joins, blends, unions, pivot tables, multi-source connectors. SQL: Creating data models, creating and maintaining data lakes, databases. Creating joins, indexes, clusters, creating constraints, stored procedures, DRL, DDL and DML. Experience of migrating data from traditional databases like MSSQL to AWS RedShift. Maven Projects: Worked on creating Spark jobs, Java Maven projects with Drools engine, MS SQL. Developed stateful and stateless KIE sessions, working with relative dependencies and plugins for pom.xml Python: Developed my own Python Framework, worked on API's and developed web applications on Django. Experience of working with HTML5, CSS3, Bootstrap, DOM. Worked on date format, file formats, string manipulations, file and directory accesses, different file formats, OS and IO level operations, concurrent execution, internet protocols libraries etc. Version Control and SDLC: Knowledge of version control in Git, GitHub, worked in Agile environments. Committing, moving and removing files, checkout, branch and merge.

Name: Crystal Ross

Email: hlopez@example.org

Phone: +1-581-356-3624x5912