

Data Engineer Data Engineer Data Engineer - Nike Beaverton, OR Over 5+ years of experience in IT industry and have been active in Big Data technologies including Hadoop and Spark. Driven by curiosity, innovative thinking and pleasure of learning developed apt solutions on performance by leveraging emerging technologies and methods. Strong believer in collaboration, teamwork, integrity and interest of the client.

Work Experience Data Engineer Nike - Beaverton, OR August 2018 to Present

This project is an assortment product which brings the analytical solutions to users based on their request data comes from different sources like DB2, Oracle, and Teradata etc. This data will be imported and will undergo several cleansings and value-added processing and then finally views will be created on this data. And this data is consumed by data science team to forecast the demand and supply.

Responsibilities: Ingested the data from various data sources like Teradata into AWS S3 and Snowflake using spark API. Developed ETL frameworks for data using PySpark. Performed fine-tuning of spark applications/jobs to improve the efficiency and overall processing time for the pipelines. Used broadcast variables in Spark, effective & efficient Joins, transformations and other capabilities for data processing. Used Spark-SQL to perform event enrichment and to prepare various levels of user behavioral summaries. Worked with EMR, S3, Athena services in AWS cloud. Implemented the workflows using Airflow scheduler to automate tasks. Created Athena query service for ad-hoc queries, analysis and data discovery.

Experienced working with Spark Core and Spark SQL using Python. Interacted with the infrastructure, network, database, application and BA teams to ensure data quality and availability.

Environment: AWS EMR, Athena, Spark, Python, S3, Airflow, Snowflake, ETL, Teradata, Hive, PySpark, TDCH.

Hadoop Developer Apttus - San Mateo, CA January 2018 to July 2018

Retail Enterprise Credit Risk application calculates Bank's retail data such as credit cards, auto, student and home loans for risk domains including Enterprise Capital Management. Data comes from different system of records such mainly from Teradata. This data will undergo several cleansing and processing and then finally views will be created on this data in Hadoop warehouse. This data will be consumed by BI tools for analyzing and generating reports.

Responsibilities: Developed SQOOP scripts for importing and exporting data into HDFS and Hive. Developing design

documents considering all possible approaches and identifying best of them. Responsible to manage data coming from different sources. Responsible for loading data from UNIX file systems to HDFS. Installed and configured Hive and written Hive UDFs. Involved in creating Hive Tables, loading with data and writing Hive queries. Used Bucketing and Dynamic Partitioning on Hive tables. Import the data from different sources like HDFS/Hive into Spark RDD. Developed Spark SQL scripts using Scala to perform transformations and actions on RDD's in spark for faster data Processing. Experienced with in working with Spark Core and Spark SQL using Scala. Performed data transformations and analytics on large dataset using Spark. Implemented the workflows using Apache Oozie framework to automate tasks. Imported results into visualization BI tool Tableau to create dashboards. Environment: Hadoop, Hive, Spark, Oozie, Teradata, Yarn, Tableau, Unix, Hortonworks, Sqoop, Scala. Software Engineer Accenture Solutions Pvt. Ltd - Hyderabad, Telangana July 2014 to July 2017 Express-Scripts, being the leading company in American Pharmacy Benefit Management company with many claims daily and there is every possibility that some of them are fraudulent claims. The goal of this project is to identify possible fraudulent claims out of the total claims processed daily. Responsibilities: Data analysis and generated reports which helped to improve product quality and decision making. Involved in development of Hadoop System and improving multi-node Hadoop Cluster performance. Experience in importing the data from relational databases such as MySQL to HDFS and exporting the data from HDFS to relational databases using SQOOP. Involved in Big Data Frameworks and tools such as Hadoop, Spark, Hive. Experience in troubleshooting the issues and failed jobs in the Hadoop cluster. Implemented the workflows using Apache Oozie framework to automate the tasks. Communicating with clients to gather the requirements. SQL querying and performance tuning, creating backup tables. Developed Web services component using XML, WSDL and SOAP with DOM parser to transfer and transform data between applications Exposed various capabilities as Web Services using SOAP/WSDL. Used SOAP UI for testing the Restful Web services by sending and SOAP request. Used AJAX framework for server communication and seamless user experience. Mentoring and Training the new recruits. Environment: Hadoop, Hive, SQOOP,

Spark, Oozie, Cloudera Manager, Tableau Education Master of Science in Information Technology
Arkansas Tech University - Russellville, AR Skills Cassandra, Hdfs, Mapreduce, Oozie, Sqoop,
Hbase, Cdh, Db2, Etl, Hadoop, Nosql, Power bi, Teradata, Amazon web services, Apache spark,
Api, Git, Hadoop, Hbase, Hive Additional Information CORE - COMPETENCIES IT experience in
Software Development Life Cycle (Analysis, Design, Development, Testing, Deployment and
Support) using WATERFALL and AGILE methodologies. Experience in Big Data Technologies
using Hadoop Eco System components (Spark, HDFS, MapReduce, Sqoop, Hive) in Retail,
Health-care sector and Financial sectors. Experienced in working with Hadoop distributions
predominantly Amazon EMR, Cloudera (CDH) and knowledge on Hortonworks (HDP).
Experienced in working with cloud services such as EMR, S3, EC2, Athena. Experience in ETL
jobs and developing and managing data pipelines. Working knowledge in creating ETL jobs to
load huge volumes of data into Hadoop Ecosystem and relational databases. Experience in
running Hive scripts, Unix and Linux shell scripting. Created User Defined Functions (UDFs) in
Hive. Flexible with full implementation of spark jobs with PySpark API and Spark Scala API.
Designed Hive queries to perform data analysis, data transfer and table design to load into Hadoop
environment. Proficient in importing / exporting data from RDBMS to HDFS using Sqoop and
TDCH. Experience in using Airflow, Oozie schedulers and Unix scripting to implement cron jobs
that execute different Hadoop actions. Experience in using SQL server, MYSQL and Teradata.
Hands on experience with ORC, and Parquet File formats. Proficient in working with Jira, GIT and
Jenkins Knowledge with NoSQL Databases HBase and Cassandra. Good analytical,
communication, problem solving skills and adore learning new technical and functional skills.
Strong believer in collaboration, teamwork, integrity and interest of the client. AREAS OF
EXPERTISE Big Data Ecosystem HDFS, MapReduce, Hive, YARN, Apache Spark, Airflow, Oozie,
Zookeeper, HUE, Sqoop, TDCH. NoSQL Databases HBase, Cassandra Hadoop Distributions
AWS (Amazon Web Services), Cloudera, Hortonworks Programming Languages Python, Scala,
HiveQL Scripting Languages Shell Scripting, Java Scripting Databases Teradata, Snowflake,
MySQL, DB2 IDE Eclipse, PyCharm, IntelliJ BI Tools Tableau, Power BI Version control tools Git,

GitHub, Bitbucket

Name: Seth Gonzalez

Email: stephenhenderson@example.org

Phone: 210-709-0925