

# Jessica Claire

100 Montgomery St. 10th Floor (555) 432-1000 resumesample@example.com

## SUMMARY

- Having over 7 years of experience as a Big Data and cloud Engineer with expertise in designing data-intensive applications using the Hadoop Ecosystem, Big Data, Cloud Data Engineering, Data Warehouse / Data Mart, Data Visualization, and Reporting. Experience in implementing E2E solutions on Big Data using the Hadoop framework, executed, and designed big data solutions on multiple distribution systems like Cloudera (CDH3 & CDH4), and Hortonworks. Expertise in Big Data processing using Hadoop, Hadoop Ecosystem (Map Reduce, Pig, Spark, Scala, Hive, Sqoop, Flume and HBase, Cassandra, Mongo DB, Kafka Framework, Zookeeper and Oozie, Storm) implementation, maintenance, ETL and Big Data analysis operations. Adept in programming languages like Scala, and Apache Spark, including Big Data technologies like Hadoop, Hive, HBase, Sqoop, Oozie, and Zookeeper. Worked on analyzing data using SQL, Hive, and Spark SQL for Data Mining, Data Cleansing. Extensively used Python Libraries PySpark, Pytest, Pymongo, Oracle, PyExcel, Boto3, Psycopg, embedPy, NumPy, Pandas, Star base, Tabulae, matplotlib, and Beautiful Soup. Experience in developing data pipelines using AWS services including EC2, S3, Redshift, Glue, Lambda functions, Step functions, CloudWatch, SNS, DynamoDB, and SQS. Designed, Developed and implemented ETL processes using IICS Data integration. Experience in collecting real-time streaming data and creating the pipeline for raw data from different sources using Kafka and storing data into HDFS and NoSQL using Spark. Have good knowledge and experience in Google Cloud Platform (GCP). Extensive experience with SQL in designing database objects such as Tables, functions, procedures, triggers, Views, and CTEs. Experience in Migrating SQL database 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 store using Azure Data Factory. Experience in creating Impala views on hive tables for fast access to data. Experienced in Erwin data Modeling iterations (Schematic/ Logical/ Physical) Forward and reverse-engineering processes. Proficient knowledge and hands-on experience in writing shell scripts in Linux. Experience in developing MapReduce jobs for data cleaning and data manipulation as required for the business. Used hive extensively to perform various data analytics required by business teams. Good experience in Tableau for Data Visualization and analysis on large data sets, drawing various conclusions. Experience in writing code in Python to manipulate data for data loads, extracts, statistical analysis, modeling, and data munging. Utilized Kubernetes and Docker for the runtime environment for the CI/CD system to build, test, and deploy. Experience in working on creating and running docker images with multiple microservices. Experience with ETL tools like Informatica, DataStage, and Snowflake. Well-experienced in Normalization and De-Normalization techniques for optimum performance in relational and dimensional database environments.

## SKILLS

- Created Snowflake Schemas by normalizing the dimension tables as appropriate and creating a Sub Dimension named Demographic as a subset to the Customer Dimension.
- Cloud Ecosystem:
- AWS
- Azure
- GCP
- Languages:
- Python
- Scala
- Spark
- SQL
- Databases:
- MySQL
- Oracle
- Cassandra
- MongoDB
- Tools & IDE:
- Git
- IntelliJ
- Visual Studio Code
- Jupyter Notebook
- PyCharm
- Office Suite
- Data Analytics
- Warehouse Models
- SAN Technologies
- SQL Transactional Replications
- Deep Learning
- Quality Analysis
- Data Mining
- Critical Thinking
- Team Building
- Team Management
- Data Management
- Data Mining
- Data Visualization and Presentations
- Data Analysis
- Amazon Web Services
- Project Management
- Data Storage and Retrieval
- Big Data Technologies
- Hadoop Programming

## EXPERIENCE

### DATA ENGINEER

02/2021 to CURRENT

#### Verizon Communications | Boston, MA

- Participated in all phases of software development, including requirements gathering and business analysis
- Designed and developed complex data pipelines using Sqoop, Spark, and Hive
- Developed Python and Scala code for data processing and analytics using inbuilt libraries
- Used various spark Transformations including mapToPair, filter, flatMap, groupByKey, sortByKey, join, cogroup, union, repetition, coalesce, distinct, intersection, map Partitions, map Partitions with Index and Actions for cleansing the input data
- Designed data models for AWS Lambda applications and analytical reports
- Created scripts to read CSV, JSON, and parquet files from S3 buckets in Python and run SQL operations then load into AWS S3, DynamoDB, and Snowflake and used AWS Glue with the crawler
- Designed the Staging and Operational Data Storage ODS environment for the enterprise data warehouse (Snowflake) including the Dimension and fact table design using Kimball's Star Schema approach
- Unit tested the data between Redshift and Snowflake
- Developed data transition programs from DynamoDB to AWS Redshift (ETL Process) using AWS Lambda by creating functions in Python for certain events based on use cases
- Performed review of system specifications related to DataStage ETL and related applications and created functions in AWS Lambda for event-driven processing
- Built a full-service catalog system using Elasticsearch, Logstash, Kibana, Kinesis, and CloudWatch with the effective use of MapReduce
- Developed Airflow dags to orchestrate sequential and parallel ETL Jobs
- Experience in moving data between GCP and Azure using Azure Data Factor
- Installed and configured Apache Airflow for workflow management and created workflows in Python
- Worked on Pyspark script for data encryption using hashing algorithms concepts on client-specified columns
- Implemented a Continuous Delivery pipeline with Docker and GitHub
- Devised PL/SQL Stored Procedures, Functions, Triggers, Views, and packages
- Made use of Indexing, Aggregation, and Materialized views to optimize query performance
- Developed ER diagrams and modeled transactional databases and data warehouses using ER/ Studio and Power Designer
- Developed Databricks Notebooks to generate the HIVE create statements from the data and load the data into the table
- Experience in developing spark applications using Spark-SQL in Databricks for data extraction, transformation, and aggregation from multiple file formats for analyzing and transforming the data to uncover insights into customer usage patterns
- Build data pipelines in airflow in GCP for ETL-related jobs using different airflow operators
- Developed PySpark code that is used to compare data between HDFS and S3
- Wrote reports using Tableau Desktop to extract data for analysis using filters based on the business use case
- Environment: Spark, Scala, AWS, ETL, GCP, Hadoop, HIVE, Tabulae, Python, Snowflake, HDFS, Hive, MapReduce, PySpark, Pig, Docker, GitHub, Apache Spark, Teradata, JSON, Databricks, PostgreSQL, MongoDB, SQL, Agile, and Windows.

### DATA ENGINEER

09/2018 to 01/2021

#### Verizon Communications | Denver, CO

- Utilized AWS services with a focus on big data architect /analytics/enterprise Data warehouse and business intelligence solutions to ensure optimal architecture, scalability, flexibility, availability, and performance, and to provide meaningful and valuable information for better decision-making
- Developed Scala scripts, and UDFs using both data frames/SQL and RDD in Spark for data aggregation, queries, and writing back into the S3 bucket
- Experience in data cleansing and data mining
- Wrote, compiled, and executed programs as necessary using Apache Spark in Scala to perform ETL jobs with ingested data
- Wrote Spark applications programs for data validation, cleansing, transformation, and custom aggregation and used Spark engine, and Spark SQL for data analysis and provided to the data scientists for further analysis
- Prepared scripts to automate the ingestion process using Python and Scala as needed through various sources such as API, AWS S3, Teradata, and Snowflake
- Designed and Developed Spark workflows using Scala for data pull from AWS S3 bucket and Snowflake applying transformations on it
- Designed and implemented ETL pipelines between various Relational data Bases to the Data Warehouse using Apache Airflow
- Developed Custom ETL Solution, Batch processing, and Real-Time data ingestion pipeline to move data in
- And out of Hadoop using Python and shell Script
- Experience in GCP Dataproc, GCS, Cloud functions, and Big Query
- Worked on Data Extraction, aggregations, and consolidation of Adobe data within AWS Glue using PySpark
- Implemented Spark RDD transformations to Map business analysis and applied actions on top of transformations
- Installed and configured Apache Airflow and automated resulting scripts to ensure daily execution in production
- Created DAG to use the Email Operator, Bash Operator, and spark Livy operator to execute in EC2
- Created scripts to read CSV, JSON, and parquet files from S3 buckets in Python and load them into AWS S3, DynamoDB, and Snowflake
- Realtime data from the source were ingested as file streams to the SPARK streaming platform and data was saved in HDFS and HIVE through GCP
- Implemented AWS Lambda functions to run scripts in response to events in Amazon DynamoDB table or S3 bucket or to HTTP requests using Amazon API gateway
- Worked on Snowflake Schemas and Data Warehousing and processed batch and streaming data load pipeline using Snow Pipe and Matillion from data lake Confidential AWS S3 bucket
- Google Cloud Platform GCP (Google Cloud Storage, Big Query, Big Table, Cloud SQL, Pub/Sub) Lead SQL
- Data..
- Profile structured, unstructured, and semi-structured data across various sources to identify patterns in data and Implement data quality metrics using necessary queries or Python scripts based on the source
- Environment: AWS, MapReduce, GCP, Snowflake, ETL, Tabulae, Pig, Spark, Scala, Hive, Kafka, Testing framework Cucumber, Python, Airflow, JSON, Parquet, CSV, Code cloud, Git

### DATA ENGINEER

01/2017 to 08/2018

#### Verizon Communications | San Jose, CA

- Created a lot of complex stored procedures, triggers, functions, indexes, and views with T-SQL statements in SQL Server Management Studio (SSMS)
- Involved in designing and managing schema objects such as Tables, Views, Indexes, Stored Procedures, and Triggers and maintained referential integrity using SQL Server Management Studio
- Designed an ETL strategy to transfer data from source to landing, staging, and destination in the data warehouse using SSIS and DTS (Data Transformation Service)
- Cleansing and messaging of the data is done on the local database
- Experience in handling Python and spark context when writing PySpark programs for ETL
- Developed SSIS packages to export data from Excel/Access to SQL Server, automated all the SSIS packages, and monitored errors using SQL Server Agent Job
- Interacted with subject matter experts in understanding the business logic, and implemented complex business requirements in the backend using efficient stored procedures and flexible functions
- Deploying Azure Resource Manager JSON Templates from PowerShell worked on Azure suite: Azure SQL Database, Azure Data Lake, Azure Data Factory, Azure SQL Data Warehouse, and Azure Analysis Service
- Architect & implement medium to large-scale BI solutions on Azure using Azure Data Platform services (Azure Data Lake, Data Factory, Data Lake Analytics, Stream Analytics, Azure SQL DW, HDInsight/ Databricks, NoSQL)
- Prepared reports using SSRS (SQL Server Reporting Service) based on cube and local data warehouse to declare discrepancies between user expectations and service efforts involved in scheduling the subscription reports via the subscription report wizard
- Designed and Developed ETL jobs to extract data from Salesforce replica and load it in the data mart in Redshift
- Design, set up, maintain, and Administrator the Azure Analysis Service, Azure SQL Data warehouse, Azure Data Factory, and Azure SQL Data warehouse
- Deployed and manage MLFlow libraries in Azure Databricks workspace and tracked servers outside Databricks
- Worked on SQL Server Integration Services (SSIS), SSAS skills, stored procedures, and triggers
- Used various sources to pull data into Power BI such as SQL Server, Excel, Oracle, SQL Azure, etc, and also created visualizations using Dashboards
- Migrate data from traditional database systems to Azure SQL databases
- Propose architectures that consider cost/spend in Azure and develop recommendations to right-size Azure data infrastructure
- Environment: MS SQL Server Management Studio (SSMS), SQL Server Integration Service (SSIS), SQL Server Reporting Services (SSRS), MS Visual Studio 2012, SQL Server Profiler, CSV, Code cloud, Git

### PYTHON DEVELOPER

04/2014 to 12/2016

#### Cumming Llc | Chicago, IL

- Involved in all phases of SDLC including Requirement Gathering, Design, Analysis and Testing of customer specifications, Development, and Deployment of the Application
- Developed entire frontend and backend modules using Python on Django Web Framework along with views and templates with Django's view controller and templating language to create a user-friendly interface using MVC architecture
- Responsible for creating visually appealing web pages with Bootstrap and HTML5
- Used server-side authentication and cookie-based session management to secure information and allow for the persistence of data
- Used Python and Django to interface with the jQuery UI and manage the storage and deletion of content
- Involved in Python OOD code for quality, logging, monitoring, and debugging code optimization
- File handling for the client data in files like .txt and .xlsx for data retrieval and storing purpose
- Graphical data representation using Python modules like NumPy, SciPy, Pandas, and Pygal and installed using pip command toolkit
- Used MATLAB for implementing algorithms and creation of user interfaces
- Used Pandas API to put the data as time series and tabular format for east timestamp data manipulation and retrieval
- Worked in the development of applications in the UNIX environment
- Responsible for debugging and troubleshooting the web application
- Creating RESTful web services with Django MVT and MySQL
- Implemented Multithreading module and complex networking operations like traceroute, SMTP mail server, and web server Using Python
- Collaborated within a team using an Agile Development workflow and widely accepted collaboration practices using Git
- Made use of Jira to manage workflow and project progress over time
- Environment: Python, Django, JavaScript, MySQL, NumPy, SciPy, Pandas API, PEP, PIP, Jenkins, JSON, Git, JavaScript, AJAX, RESTful webservice, MySQL, PyUnit

## EDUCATION AND TRAINING

### Master of Science | Computer Engineering

12/2022

#### University of Missouri - Kansas City, Kansas City, MO

- 3.580 GPA