

<div>Jessica Claire</div> <div>Montgomery Street, San Francisco, CA 94105 ♦ (555) 432-1000 ♦ resumesample@example.com</div>	
SUMMARY	
<ul style="list-style-type: none">Dynamic and motivated IT professional with around 3.9 years of experience as a ETL AWS Developer with expertise in designing data intensive applications using AWS Services, Cloud Data engineering , Data Warehouse / Data Mart, Data Visualization , Reporting , and Data Quality solutions .Profound experience in performing Data Ingestion, Data Processing (Transformations, enrichment, and aggregations).Strong Knowledge on Architecture of Distributed systems and Parallel processing, In-depth understanding of MapReduce programming paradigm and Spark execution framework.Hands-on experience in handling database issues and connections with SQL and NoSQL databases such as MongoDB , HBase , Cassandra , SQL server , and PostgreSQL .Experience in designing and creating RDBMS Tables, Views, User Created Data Types, Indexes, Stored Procedures, Cursors, Triggers and Transactions.Expert in designing ETL data flows using creating mappings/workflows to extract data from SQL Server and Data Migration and Transformation from Oracle/Access/Excel Sheets using SQL Server SSIS .Hands-on experience with Amazon EC2, Amazon S3, Amazon RDS, VPC, IAM, Amazon Elastic Load Balancing, Auto Scaling, CloudWatch, SNS, SES, SQS, Lambda, EMR and other services of the AWS family.Instantiated, created, and maintained CI/CD (continuous integration & deployment) pipelines and apply automation to environments and applications.Worked on various automation tools like GIT, Terraform, Ansible. Experienced in fact dimensional modeling (Star schema, Snowflake schema), transactional modeling and SCD (Slowly changing dimension)Efficient Cloud Engineer with years of experience assembling cloud infrastructure. Utilizes strong managerial skills by negotiating with vendors and coordinating tasks with other IT team members. Implements best practices to create cloud functions, applications and databases.	
SKILLS	
<ul style="list-style-type: none">AWS Services : EC2, ECS, Cloud front, S3, Aurora, Lambda, etcDatabases: Oracle, MySQL, SQL Server, MongoDB, Cassandra, DynamoDB, PostgreSQL, Teradata, Cosmos.Programming: Python, PySpark, Scala, Java, C, C++, Shell script, Perl script, SQLCloud Technologies: AWS, Microsoft AzureFrameworks: Django REST framework, MVC, HortonworksTools: ETL, Informatica IDE, Visual Studio, SQL*Plus, SQL Developer, TOAD, SQL Navigator, Query Analyzer, SQL Server Management Studio, SQL Assistance, Eclipse, PostmanVersioning tools: SVN, Git, GitHubOperating Systems: Windows 10/11, Ubuntu Linux, MacOSNetwork Security: KerberosDatabase Modelling: Dimension Modeling, Snowflake Modeling	
EXPERIENCE	
AWS Data Engineer , 01/2022 - 02/2022 Deloitte – Rosslyn, NY <ul style="list-style-type: none">Designed and setup Enterprise Data Lake to provide support for various uses cases including Analytics, processing, storing and Reporting of voluminous, rapidly changing dataResponsible for maintaining quality reference data in source by performing operations such as cleaning, transformation and ensuring Integrity in a relational environment by working closely with the stakeholders & solution architectDesigned and developed Security Framework to provide fine grained access to objects in AWS S3 using AWS Lambda, DynamoDBSet up and worked on Kerberos authentication principals to establish secure network communication on cluster and testing of HDFS, Hive, Pig and MapReduce to access cluster for new usersPerformed end- to-end Architecture & implementation assessment of various AWS services like Amazon EMR, Redshift, S3Implemented the machine learning algorithms using python to predict the quantity a user might want to order for a specific item so we can automatically suggest using kinesis firehose and S3 data lakeUsed AWS EMR to transform and move large amounts of data into and out of other AWS data stores and databases, such as Amazon Simple Storage Service (Amazon S3) and Amazon DynamoDBUsed Spark SQL for Scala & amp, Python interface that automatically converts RDD case classes to schema RDDImport the data from different sources like HDFS/HBase into Spark RDD and perform computations using PySpark to generate the output responseCreating Lambda functions with Boto3 to deregister unused AMLs in all application regions to reduce the cost for EC2 resourcesImporting & exporting database using SQL Server Integrations Services (SSIS) and Data Transformation Services (DTS Packages)Coded Teradata BTEQ scripts to load, transform data, fix defects like SCD 2 date chaining, cleaning up duplicatesDeveloped reusable framework to be leveraged for future migrations that automates ETL from RDBMS systems to the Data Lake utilizing Spark Data Sources and Hive data objectsConducted Data blending, Data preparation using Alteryx and SQL for Tableau consumption and publishing data sources to Tableau serverDeveloped Kibana Dashboards based on the Log stash data and Integrated different source and target systems into Elasticsearch for near real time log analysis of monitoring End to End transactionsImplemented AWS Step Functions to automate and orchestrate the Amazon SageMaker related tasks such as publishing data to S3, training ML model and deploying it for predictionIntegrated Apache Airflow with AWS to monitor multi-stage ML workflows with the tasks running on Amazon SageMakerEnvironment: AWS EMR, S3, RDS, Redshift, Lambda, Boto3, DynamoDB, Amazon SageMaker, Apache Spark, HBase, Apache Kafka, HIVE, SQOOP, Map Reduce, Snowflake, Apache Pig, Python, SSRS, TableauAssessed organization technology infrastructure and managed cloud migration process.Configured computing, networking and security systems within cloud environment.Implemented cloud policies, managed technology requests and maintained service availability.	
Data Engineer , 01/2016 - 11/2019 Verizon Communications – Annapolis Junction, OR <ul style="list-style-type: none">Worked on Azure Data Factory to integrate data of both on-prem (MY SQL, Cassandra) and cloud (Blob storage, Azure SQL DB) and applied transformations to load back to Azure SynapseManaged, Configured and scheduled resources across the cluster using Azure Kubernetes ServiceMonitored Spark cluster using Log Analytics and Ambari Web UITransitioned log storage from Cassandra to Azure SQL Datawarehouse and improved the query performanceInvolved in developing data ingestion pipelines on Azure HDInsight Spark cluster using Azure Data Factory and Spark SQLAlso Worked with Cosmos DB (SQL API and Mongo API)Develop dashboards and visualizations to help business users analyze data as well as providing data insight to upper management with a focus on Microsoft products like SQL Server Reporting Services (SSRS) and Power BIPerformed the migration of large data sets to Databricks (Spark), create and administer cluster, load data, configure data pipelines, loading data from ADLS Gen2 to Databricks using ADF pipelinesCreated various pipelines to load the data from Azure data lake into Staging SQLDB and followed by to Azure SQL DBCreated Databrick notebooks to streamline and curate the data for various business use cases and also mounted blob storage on DatabrickUtilized Azure Logic Apps to build workflows to schedule and automate batch jobs by integrating apps, ADF pipelines, and other services like HTTP requests, email triggers etcWorked extensively on Azure data factory including data transformations, Integration Runtimes, Azure Key Vaults, Triggers and migrating data factory pipelines to higher environments using ARM TemplatesIngested data in mini-batches and performs RDD transformations on those mini-batches of data by using Spark Streaming to perform streaming analytics in Data bricksEnvironment: Azure SQL DW, Databrick, Azure Synapse, Cosmos DB, ADF, SSRS, Power BI, Azure Data lake, ARM, Azure HDInsight, Blob storage, Apache Spark.Adept in troubleshooting and identifying current issues and providing effective solutions.Managed performance monitoring and tuning while identifying and repairing issues within database realm.Identified key use cases and associated reference architectures for market segments and industry verticals.Designed surveys, opinion polls, and assessment tools to collect data.Tested, validated and reformulated models to foster accurate prediction of outcomes.Created graphs and charts detailing data analysis results.Recommended data analysis tools to address business issues.Developed new functions and applications to conduct analyses.Cleaned and manipulated raw data.Collaborated with solution architects to define database and analytics engagement strategies for operational territories and key accounts.	
Big Data Engineer / Hadoop Developer , 10/2013 - 12/2015 Cumming Llc – Miami, FL AnsibleDenodoDenodoCloudWatchAvroPySparkPySparkPySparkMLlibDataframeNiFiNiFi <ul style="list-style-type: none">Interacted with business partners, Business Analysts and product owner to understand requirements and build scalable distributed data solutions using Hadoop ecosystemDeveloped Spark Streaming programs to process near real time data from Kafka, and process data with both stateless and state full transformationsWorked with HIVE data warehouse infrastructure-creating tables, data distribution by implementing partitioning and bucketing, writing and optimizing the HQL queriesBuilt and implemented automated procedures to split large files into smaller batches of data to facilitate FTP transfer which reduced 60% of execution timeWorked on developing ETL processes (Data Stage Open Studio) to load data from multiple data sources to HDFS using FLUME and SQOOP, and performed structural modifications using Map Reduce, HIVEDeveloping Spark scripts, UDFS using both Spark DSL and Spark SQL query for data aggregation, querying, and writing data back into RDBMS through SqoopWritten multiple MapReduce Jobs using Java API, Pig and Hive for data extraction, transformation and aggregationAvrom multiple file formats including Parquet, Avro, XML, JSON, CSV, ORCFILE and other compressed file formats Codecs like gZip, Snappy, LzoStrong understanding of Partitioning, bucketing concepts in Hive and designed both Managed and External tables in Hive to optimize performanceDeveloped PIG UDFs for manipulating the data according to Business Requirements and also worked on developing custom PIG LoadersDeveloping ETL pipelines in and out of data warehouse using combination of Python and Snowflakes SnowSQL Writing SQL queries against SnowflakeExperience in report writing using SQL Server Reporting Services (SSRS) and creating various types of reports like drill down, Parameterized, Cascading, Conditional, Table, Matrix, Chart and Sub ReportsUsed DataStax Spark connector which is used to store the data into Cassandra database or get the data from Cassandra databaseWrote oozie scripts and setting up workflow using Apache Oozie workflow engine for managing and scheduling Hadoop jobsWorked on implementation of a log producer in Scala that watches for application logs, transform incremental log and sends them to a Kafka and Zookeeper based log collection platformUsed Hive to analyze data ingested into HBase by using Hive-HBase integration and compute various metrics for reporting on the dashboardTransformed tPySpark using AWS Glue dynamic frames with PySpark; cataloged the transformed the data using Crawlers and scheduled the job and crawler using workflow featureWorked on installing cluster, commissioning & decommissioning of data node, name node recovery, capacity planning, and slots configurationDeveloped data pipeline programs with Spark Scala APIs, data aggregations with Hive, and formatting data (JSON) for visualization, and generatiPySparkEnvironment: AWS, Cassandra, PySpark, Apache Spark, HBase, Apache Kafka, HIVE, SQOOP, FLUME, Apache oozie, Zookeeper, ETL, UDF, Map Reduce, Snowflake, Apache Pig, Python, Java, SSRSOnfidentialDeveloped and implemented Hadoop code while observing coding standards.Optimized and tuned Hadoop environments and modified hardware to meet prescribed performance thresholds.Developed new functions and applications to conduct analyses.Created graphs and charts detailing data analysis results.Tested, validated and reformulated models to foster accurate prediction of outcomes.	
Python Developer , 09/2012 - 10/2013 Fiserv – City, STATE <ul style="list-style-type: none">AWS, S3, EC2, LAMBDA, EBS, IAM, Datadog, CloudTrail, CLI, Ansible, MySQL, Python, Git, Jenkins, DynamoDB, Cloud Watch, Docker, KubernetesLeveraged open communication, collective decision-making and thorough reviews to create performant and scalable systems.Worked with server-side and front-end technologies and leveraged common design patterns to code dynamic and user-friendly systems.Implemented new API routes, architected new ORM structures and refactored code to boost application performance.Harnessed version control tools to coordinate project development and individual code submissions.Introduced cloud-based technologies into Python development to expand on-premise deployment options.Wrote clear and clean code for use in projects.Resolved customer issues by establishing workarounds and solutions to debug and create defect fixes.	
EDUCATION AND TRAINING	
Post Graduate : Data Engineering , 02/2022 Purdue University - West Lafayette, IN	
Post Graduate : Data Science And Business Analytics , 09/2021 University of Texas At Austin - Austin, TX	
Bachelor of Arts : Business Administration And Management, 12/2009 California State University - Fullerton, CA	