Spark/Hadoop Developer Spark/Hadoop Developer Spark/Hadoop Developer - CVS Health Marlborough, MA Work Experience Spark/Hadoop Developer CVS Health - Woonsocket, RI October 2016 to Present - Assisted in different data Modeling and Data Warehouse design and development. - Used Sqoop to connect to the Sql Server and move the pivoted data to Hive tables and stored in different file formats like Avro, SerDe, Text etc. - Using Spark API over Hortonworks Hadoop YARN to perform analytics on data in Hive. - Analyzing of large volumes of structured data using SparkSQL. - Enhanced and optimized product Spark code to aggregate, group and run data mining tasks using the Spark framework. - Created internal and external Hive tables and defined static and dynamic partitions for optimized performance. - Creating Hive tables to import large data sets from various relational databases using Sqoop and export the analyzed data back for visualization and report (Tableau) generation by the BI team. - Wrote Python scripts to parse JSON documents and load the data in database. - Worked on Apache spark writing python applications to convert txt, xls files and parse data into JSON format. - Developed the Shell, Perl and Python Scripts (Linux& UNIX) to execute jobs. - Exploring with the Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frame, Pair RDD's, Spark YARN. - Loading data into Spark RDD and do in memory data Computation to generate the Output response. - Loading the data to HBASE by using bulk load and HBASE API. - Used PySpark to expose spark API to Python. - Developed Spark code and Spark-SQL for faster testing and processing of data. Experiencing in Oozie a workflow scheduler to manage Hadoop jobs by Direct Acyclic Graph (DAG) of actions with control flows. - Loading data into Spark RDD and do in memory data Computation to generate the Output response. - Developed Spark code using Python for faster processing of data on Hive. - Worked on Integrating Talend and SSIS with Hadoop and performed ETL operations. - Developed MapReduce jobs in Python for data cleaning and data processing. - Used spark cluster to manipulate RDDS (Resilient Distributed Datasets) and also used concepts of RDD partitions. - Connecting MySQL database through spark driver. - Develop applications that use MongoDB database and Pymong. - Wrote Python modules to extract/load asset data from the MySQL source database. - Developed and maintain several batch

jobs to run automatically depending on business requirements. - Involved in cluster maintenance, commissioning & decommissioning Data nodes, Troubleshooting, Manage and review data backups, Manage and review Hadoop log files.   Environment: Python, Hortonworks, Hadoop, HDFS, MapReduce, YARN, Spark, Pig, Hive, Sqoop, Kafka, HBase, Bedrock, Oozie, Flume, Tableau, SQL Scripting, Linux Shell Scripting, Sublime. Hadoop Developer Eliza Corporation - Danvers, MA July 2015 to June 2016 - Installed and configured Apache Hadoop, Hive and Pig environment on AWS. - Handling Hadoop cluster setup involved in start to end process of installation, configuration and monitoring. - Build servers using AWS: Importing volumes, launching EC2, RDS, creating security groups, auto-scaling, load balancers (ELBs) in the defined virtual private connection. - Migrating physical Linux/Windows servers to cloud (AWS) and testing. - Performed Export and import of data into simple storage service (S3). - Development, Acceptance, Integration, and Production AWS Endpoints. - Developed complex MapReduce streaming jobs using Java language that are implemented Using Hive and Pig. - Loading data into Spark RDD and do in memory data Computation to generate the Output response. - Optimized MapReduce Jobs to use HDFS efficiently by using various compression mechanisms. - Collected and aggregated large amounts of web log data from different sources such as web servers, mobile using Apache Flume and stored the data into HDFS for analysis. - Handled importing of data from various data sources, performed transformations using Hive, MapReduce, loaded data into HDFS and Extracted the data from MySQL into HDFS using Sqoop. - Automating backups by the shell for Linux to transfer data in S3 bucket. - Designed and crafted High-Performance Clusters in AWS along with AWS Red shift. - Experience in AWS EC2, configuring the servers for Auto scaling and Elastic load balancing. - AWS CLI Auto Scaling and Cloud Watch Monitoring creation and update. - Develop Hive scripts for end user / analyst requirements to perform ad hoc analysis. - Plan, deploy, monitor, and maintain Amazon AWS cloud infrastructure consisting of multiple EC2 nodes and VMware Vm's as required in the environment. - Expertise in AWS data migration between different database platforms like SQL Server to Amazon Aurora using RDS tool. - Optimized MapReduce Jobs to use HDFS efficiently by using various compression mechanisms. - Loading Data into HBase using Bulk Load

and Non-bulk load. - Experiencing in Oozie a workflow scheduler to manage Hadoop jobs by Direct Acyclic Graph (DAG) of actions with control flows - Handled importing of data from various data sources, performed transformations using Hive, MapReduce, loaded data into HDFS and Extracted the data from MySQL into HDFS using Sqoop. - Created Hive tables, partitions and loaded the data to analyze using HiveQL queries. - Configuring EC2 instances in VPC network & managing security through IAM and Monitoring servers' health through Cloud Watch. - Development, Acceptance, Integration, and Production AWS Endpoints. - Automating backups by the shell for Linux to transfer data in S3 bucket. - Designed and crafted High-Performance Clusters in AWS along with AWS Red shift.   Environment: Hadoop, Python, HDFS, Amazon Web Services (AWS), YARN, Spark, Pig, Hive, Sqoop, Oozie, Flume, Linux Shell Scripting, Spark-SQL, MongoDB. Data Engineer (Hadoop) ABS Pvt ltd - Mumbai, Maharashtra May 2014 to March 2015 - Exported data to a MySQL from HDFS using Sqoop and NFS mount approach. - Moved data from HDFS to Cassandra using Map Reduce and Bulk Output Format class. - Developed Map Reduce programs for applying business rules on the data. - Developed and executed hive queries for demoralizing the data. - Works with ETL workflow, analysis of Big Data and loaded them into Hadoop cluster. - Installed and configured Hadoop Cluster for development and testing environment. - Implemented Fair scheduler on the Job tracker to share the resources of the cluster for the map reduces jobs given by the users. - Automated the workflow using shell scripts. - Performance tuning of the Hive queries, written by other developer. - Prototype various applications that utilize modern Big Data tools.   Environment: Java, Python, Hadoop, Map Reduce, HDFS, Hive, Pig, Sqoop, Cloudera, MongoDB, HUE, Linux. Java/ Python Developer GBS Technologies January 2012 to March 2014 - Responsible for understanding the scope of the project and requirement gathering. - Review and analyze the design and implementation of software components/applications and outline the development process strategies. - Used Spring JDBC to write some DAO classes to interact with the database to access account information. - Created entire application using Python, Django, MySQL and Linux. - Involved in creation of Test Cases for JUnit Testing. - Wrote python scripts to parse XML documents and load the data in database. - Designed and configured database and back end

applications and programs. - Used Oracle as Database, used Toad for queries execution, and involved in writing SQL scripts, PL/SQL code for procedures and functions. - Used CVS, Perforce as configuration management tool for code versioning and release. - Developed application using Eclipse and used build and deploy tool as Maven. Environment: Python, Java1.5, J2EE [Servlets, JSP], XML, spring 3.0, Design Patterns, Log4j, CVS, Maven, Eclipse and Oracle 11g. Education Master of Science in Computer Science Rivier University Skills LINUX (4 years), MYSQL (4 years), PYTHON (4 years), JAVA (3 years), SQL (3 years) Additional Information Areas of Expertise: Hadoop Ecosystems: HDFS, YARN, Spark, Map Reduce, Hive, Pig, Zookeeper, Sqoop, Oozie, Bedrock, Flume, Kafka, Impala, Ambari, HUE, MongoDB, HBase Amazon Web Services: Redshift, EMR, EC2, S3, RDS, Cloud Search, Data Pipeline, Lambda Languages: Python, Java, HiveQL, Pig Latin, Advanced PL/SQL, SQL Databases: Oracle, MS-SQL Server, MySQL, MS-Access, PostgreSQL, NoSQL, Teradata Tools: Eclipse, Net Beans, SQL Developer, IntelliJ, MS Visual Studio, Sublime BI Tools: Tableau, Qlikview, Jinfonet JReport Hadoop Platforms: Hortonworks, Amazon Web services (AWS), Cloudera Platforms: Linux Red Hat/Ubuntu/Centos, Windows, Macintosh, Linux, Solaris

Name: Regina Young

Email: antonio43@example.com

Phone: +1-920-784-0076x085