**Mayang Vasava**

**Hadoop Developer  
Phone: 848-260-8497**

**PROFESSIONAL SUMMARY:**

* 8+ years of IT experience which includes 5+ years of work experience in Big Data, Hadoop ecosystem related technologies.
* Overall 5 years of experience in application development and design using **Object Oriented Programming, Java /J2EE technologies.**
* Good knowledge of Hadoop Architecture and various components such as **HDFS, Job Tracker, Task Tracker, Namenode and Datanode.**
* Technical expertise in **Big data/Hadoop HDFS, Map Reduce, Apache Hive, Apache Pig, Sqoop, HBase,Flume, Storm, Kafka, Spark, Oozie ,Zookeeper** ,**NoSQL Data bases HBase, Cassandra, MongoDB.**
* Experience in the developing NoSQL database by using **CRUD, Sharding, Indexing and Replication.**
* Experience in good understanding of **Apache Storm-Kafka pipelines.**
* Extensive experience working in **Teradata**, **Oracle**, **Netezza**, **Informatica, SQL Server** and **MySQL** database.
* Good Experience in data loading from **Oracle** and **MYSQL** databases to **HDFS** system using **Sqoop** (Structure Data) and **Flume** (Log Files & XML).
* Designed the data models to be used in data intensive AWS Lambda applications which are aimed to do complex analysis creating analytical reports for end-to-end traceability, lineage, definition of Key Business elements from Aurora.s
* Knowledge on analyzing data interactively using**Apache Spark** and **Apache Zeppelin** .
* Extensive experience in developing **PIG Latin Scripts** and using **Hive Query Language** for data analytics.
* Experienced in writing custom **Hive UDF's** to incorporate business logic with **Hive queries**.
* Good experience in optimizing **Map Reduce algorithms** using **Mappers**, **Reducers**, **combiners** and **partitioners** to deliver the best results for the large datasets.
* Experience in understanding the security requirements for Hadoop and integrate with **Key Distribution Centre**.
* Proficient in **Java, Scala** and **Python.**
* Expertise in Amazon AWS concepts like **EMR** and **EC2** web services which provides fast and efficient processing of Big Data.
* Hands on experience in using BI tools like **Tableau/Pentaho.**
* Detailed understanding of **Software Development Life Cycle (SDLC)** and sound knowledge of project implementation methodologies including **Waterfall and Agile.**
* Involved in design and development of various web and enterprise applications using various technologies like **JSP**, **Servlets**, **Struts**, **Hibernate**, and **spring**, **JDBC**, **JSF**, **XML**, **Java Script**, **HTML**, **AJAX**, **SOAP** and **Amazon Web Services**.
* Experience in constructing pipelines using workflow tools like**Oozie**.
* Experienced in providing real time analytics on big data platforms using **HBase, Cassandra** and **Mongo DB.**
* Hands on experience in application development using core **JAVA, RDBMS, Linux shellscripting** and also developed **UNIX shell** scripts to automate various processes.
* Having Experience on Development applications like Eclipse, RAD etc.
* Expertise in Unit Testing, Integration Testing, System Testing and experience in preparing the Test Cases, Test Scenarios and Test plans.
* Ability to work independently as well as in a team and able to effectively communicate with customers, peers and management at all levels in and outside the organization.

**AREA OF EXPERTISE :**

|  |  |
| --- | --- |
| **Languages** | **R,C, C++, Java(JSP,Servlets,JavaBeansJDBC,XML), Shell Scripting** |
| **Big Data Ecosystem** | **Hadoop, MapReduce, YARN, Pig, Hive, HBase, Flume, Sqoop, Impala, Oozie, Zookeeper, Spark, Ambari, Mahout, MongoDB, Cassandra, Avro, Parquet, Snappy, Kafka, AWS, GCP** |
| **Databases** | **Oracle, MySQL,PL/SQL,PostgreSQL** |
| **No SQL Databases** | **Cassandra, MongoDB, Hbase, DynamoDB,** |
| **Operating Systems** | **UNIX,Linux,MAC OS, Windows XP, Server 2003, Server 2008** |
| **Development Tools** | **Eclipse 3.3,Ant,Maven,JUNIT.log$J,ETL** |
| **programming Languages** | **HTML5,CSS 3,JAVASCRIPT,AJAX,JQUERY,.NET,Visual Studio 2010** |
| **Network protocols** | **TCP/IP, UDP, HTTP, DNS, DHCP, OSPF, RIP** |
| **Frameworks** | **Struts, Spring, Hibernate, MVC** |

**Professional Experience:**

**American Express, Phoenix, AZ Aug’2019- Present**

**Role: Hadoop Developer**

**Project:** **California Consumer Privacy Act.**

**Description:**

The California Consumer Privacy Act (CCPA) of 2018 was passed on June 28th 2018. The law, which will take effect on January 1st 2021, introduces new privacy rights for consumers and will force companies that conduct business in the State of California to implement structural changes to their privacy programs. The new rights given to California consumers are similar to the rights provided in the European Union’s General Data Protection Regulation (GDPR). The CCPA also subjects non-compliant businesses to expensive fines, class-action lawsuits, and injunctions.

**Responsibilities:**

* Understanding the existing process and gathering requirements from business team to convert them into technical requirement.
* Developed data pipelines using custom-built in-house tool (Magellan).
* Designed, developed and implemented data transformation to meet client’s expectation
* Developing HQL queries to retrieve the data and ingest back it into Cornerstone after applying necessary business logics.
* Developed CDC (Change Data Capture) process to handle upcoming incremental data.
* Developed PySpark applications to load JSON data into Hive table from CSV and Json files and create the final file and insert it into cornerstone (HDFS).
* Performing detail Unit Testing: Creation of Unit test Plans/ scenarios, Test data preparation and log the test results.
* Provide the input to Quality Analyst for preparing the test plan for the system and UAT test phase.
* Build and manage teams that collaboratively work well with each other as well as other teams (e.g. architecture, QA teams).
* Deploying the artefacts into production using CI/CD process.
* Providing necessary technical support to team by giving the training Spark and Event engine.

Responsible for deployment related activities and developing UNIX shell scripts for deployment

**Client: PCG ( Public Consulting Group), New York, NY Jul’17– Jul’2019**

**Role: Sr. Hadoop/Big Data Developer**

**Description:**    
PCG ( Public Consulting Group) helps primarily public sector health, education, and human services organizations make measurable improvements to their performance and processes.

**Responsibilities:**

* Worked with highly unstructured and semi structured data of 90 TB in size (270 TB with replication factor of 3)
* Developing scripts and Batch Job to schedule various Hadoop programs.
* Used **Pig** as ETL tool to do transformations, event joins, filter & some pre-aggregations.
* Collecting and aggregating large amounts of log data using **Apache Flume** and staging data in HDFS for further analysis.
* Used **Spark stream** processing to get data into in-memory, implemented **RDD** transformations, actions to process as units
* Created/modified **UDF** and **UDAFs** for Hive.
* Populated **HDFS** and **Cassandra** with huge amounts of data using **Apache Kafka.**
* Used DML statements to perform different operations on **Hive Tables**.

• Developed Hive queries for creating foundation tables from stage data .

* Adjusting the minimum share of maps and reducers for all the queues.
* Analyzed the data by performing Hive queries and running **Pig scripts** to study customer behavior.
* Managed **Amazon Web Services** (AWS) EC2 with **Puppet**.
* Working with **Apache Crunch** library to write, test and run HADOOP MapReduce pipeline jobs.
* Efficiently put and fetched data to/from HBase by writing Map/Reduce job in **Java/Python.**
* Cluster coordination services through **Zookeeper**.
* Creating **Hive tables**, **dynamic partitions**, **buckets for sampling**, and working on them using **Hive QL**.
* Experienced on loading and transforming of large sets of **semi structured data** using **Pig Latin operations**.
* Extracted the data from Teradata into HDFS using **Sqoop.**
* Data Visualization using T**ableau** for reporting from Hive Tables.
* Worked in using Sequence files, RCFile, AVRO and HAR file formats.

**Environment:** Hadoop, HDFS, Apache Crunch,Map Reduce, Hive, Flume, Sqoop, Zookeeper, Kafka, Storm, Cassandra , Spark, Puppet, Storm, Linux**.**

**Client: The Travelers, Inc., New York, NY March‘15– Jun‘17**

**Role: Big Data Developer**

**Description:**

The Travelers Companies, Inc. is an American [insurance](https://en.wikipedia.org/wiki/Insurance) company. It is the second largest writer of U.S. commercial property casualty insurance and the third largest writer of U.S. personal insurance through independent agents. Travelers is incorporated in Minnesota, with headquarters in [New York City](https://en.wikipedia.org/wiki/New_York_City) .

**Responsibilities:**

* Have real-time experience of **Kafka-Storm** on **HDP 2.2** platform for real time analysis.
* Created **PoC** to store Server Log data in **MongoDB** to identify System Alert Metrics
* Implemented Hadoop framework to capture user navigation across the application to validate the user interface and provide analytic feedback/result to the UI team
* Loaded data into the cluster from dynamically generated files using Flume and from relational database management systems using **Sqoop**.
* Performed analysis on the unused user navigation data by loading into **HDFS** and writing **MapReduce jobs.** The analysis provided inputs to the new APM front end developers and lucent team.
* Wrote **MapReduce** jobs using **Java API and Pig Latin.**
* Loaded the data from **Teradata to HDFS** using Teradata Hadoop connectors.
* Wrote **Pig scripts** to run **ETL jobs** on the data in **HDFS** and further do **testing**.
* Used **Hive** to do analysis on the data and identify different correlations.
* Imported data using **Sqoop** to load data from **MySQL** to **HDFS** and **Hive** on regular basis.
* Written Hive queries for data analysis to meet the business requirements.
* Involved in collecting, aggregating and moving data from servers to HDFS using **Apache Flume**.
* Involved in using **Oozie** for defining amd scheduling jobs to manage apache Hadoop jobs by **Directed Acyclic graph (DAG**) of actions with control flows.
* Involved in creating **Hive tables** and working on them using **HiveQL and** perform data analysis using **Hive** and **Pig.**
* Automatically Importing data regular basis using **sqoop** to into the Hive partition by using **apache Oozie.**
* Supported Map Reduce Programs those are running on the cluster**.**
* Weekly meetings with technical collaborators and active participation in code review sessions with senior and junior developers.
* Used **Qlikview and D3 ss**for visualization of query required by BI team
* Continuous monitoring and managing the Hadoop cluster through Cloudera Manager

**Environment:** Hadoop, MapReduce, HDFS, Pig, Hive, HBase, Flume, ZooKeeper, Cloudera Manager,Oozie, Java (jdk1.6), MySQL, SQL, Windows NT, Linux

**Client: Toms HQ, Los Angeles, CA                            Jan‘14 - Feb‘15**

**Role: Hadoop Developer**

Toms is a profit company based in California. The company designs and sells shoes and eyewear. When Toms sells one pair of shoes, a new pair of shoes is given to person in need and when toms sells a pair of eyewear, part of the profit is used to restore the eyesight for people. Toms believe it’s One for One.

**Responsibilities:**

* Created **HBase** tables to load large sets of semi-structured and unstructured data coming from **UNIX, NoSQL** and a variety of portfolios.
* Experience in Bench Marking and performance tuning of Hadoop cluster namely CPU, memory, I/O, mapred & yarn configuration.
* Experience in ingesting Structured, unstructured and log data to Hadoop HDFS and **Netezza & Greenplum** using **Spark & Sqoop and Informatica**.
* Experience inhandling data cleansing, data profiling, data lineage and denormalization, & aggregation of big data.
* Analyzing/Transforming data with **Hive** and **Pig.**
* Worked with different Hive file formats like RC file, Sequence file, ORC file format and Parquet.
* Experience with load Balancers on the **AWS.**
* Worked on google cloud platform (GCP) services like compute engine, cloud load balancing, cloud storage, cloud SQL, stack driver monitoring and cloud deployment manager.
* Setup GCP Firewall rules to allow or deny traffic to and from the VM's instances based on specified configuration and used GCP cloud CDN (content delivery network) to deliver content from GCP cache locations drastically improving user experience and latency.
* Took **Splunk** tools used for log aggregation and implemented log data analysis.
* Automated the process for extraction of data from warehouses and weblogs into HIVE tables by developing workflows and coordinator jobs in **Oozie**.
* Exported the analyzed data to the relational databases using **Sqoop** for visualization and to generate reports for the BI team.
* Involved in writing optimized **Pig** Script along with involved in developing and testing PIG Latin Scripts.
* Working knowledge in writing Pig's Load and Store functions.
* Developed job flows to automate the workflow for **PIG** and **HIVE** jobs.
* Responsible for writing Hive Queries for analyzing data using **Hive QueryLanguage (HQL).**
* Tested and reported defects in an **Agile** Methodology perspective.
* Managed and reviewed Hadoop log files.
* Created **R function** and **Spark stream** to pull customer sentiment data from Twitter.
* Experience in using **Pentaho Data Integration** tool for data integration, OLAP analysis and ETL process.
* Experienced in converting sssss operations to Hadoop system using **Pig Latin** operations, transformations and functions.

**Environment:** Hadoop, YARN,HDFS, Map Reduce, Hive, Oozie, HiveQL, Netezza, Informatica, HBase, Pig, MySQL, NoSQL, Spark Sqoop, Pentaho

**Client: Sonata Software Limited, India Mar ‘11 – Dec‘ 13**

**Role: Java Developer**

**Responsibilities:**

* Involved in Analysis, design and coding on Java and J2EE Environment.
* Implemented struts **MVC framework**.
* Designed, developed and implemented the business logic required for Security presentation controller.
* Set up the deployment environment on **Web Logic Developed** system preferences UI screens using JSP and HTML.
* Developed UI screens using Swing components like **JLabel, JTable, JScrollPane, JButtons, JTextFields**, etc.
* Used JDBC to connect to Oracle database and get the results that are required.
* Designed asynchronous messaging using Java Message Service (JMS).
* Consumed web services through **SOAP protocol**.
* Developed web Components using **JSP, Servlets** and Server-side components using **EJB** under **J2EE Environment**.
* Designing **JSP using Java Beans**.
* Implemented Struts framework 2.0 (Action and Controller classes) for dispatching request to appropriate class
* Design and implementation of front-end web pages using **CSS, DHTML, JavaScript, JSP, HTML, XHTML, JSTL, Ajax** and **Struts Tag Library**.
* Designed table structure and coded scripts to create tables, indexes, views, sequence, synonyms and database triggers.
* Involved in writing Database procedures, Triggers, PL/SQL statements for data retrieving.
* Developed using Web 2.0 to interact with other users and changing the contents of websites.
* Implemented AOP and IOC concept using UI Spring 2.0 Framework.
* Developed using Transaction Management API of **Spring 2.0** and coordinates transactions for Java objects
* Generated WSDL files using AXIS2 tool.
* Developed using CVS as a version controlling tool for managing the module developments.
* Configured and Tested Application on the **IBM Web Sphere App**. Server
* Used Hibernate ORM tools which automate the mapping between SQL databases and objects in Java.
* Developed using XML XPDL, BPEL and XML parsers like DOM, SAX.
* Developed using XSLT to convert XML documents into XHTML and PDF documents.
* Written **JUnit test** cases for Business Objects, and prepared code documentation for future reference and upgrades.
* Deployed applications using **WebSphere Application** Server and Used IDE RAD (Rational Application Developer).

**Environment:** Java, J2EE, JSP, Servlets, MVC, Hibernate, Spring 3.0, Web Services, Maven 3.2.x, Eclipse, SOAP, WSDL, Eclipse, jQuery, Java Script, Swings, Oracle, REST API, PL/SQL, Oracle 11g, UNIX.