**NAME: VINEETH**

**EMAIL: Vineeth3936@gmail.com**

**CONTACT: +13048533083**

**Summary:**

* Over **8+** years of IT experience in software Development and 4 years in Big Data Technologies and Analytical Solutions with 4+ years of hands-on experience in development and design of Java and related frameworks, Full stack web development.
* Over 4 years’ experience as Hadoop developer with good knowledge of Hadoop framework, Hadoop Distributed file system and Parallel processing implementation, **Hadoop Ecosystems HDFS, Map Reduce, Hive, Pig, Python, HBase, Sqoop, Hue, Oozie, Impala, Spark.**
* Excellent understanding / knowledge of Hadoop architecture and various components such as **HDFS, Job\ Tracker, Task Tracker, Name Node, Data Node and Map Reduce programming paradigm**.
* Experienced in handling different file formats like **Text file, Avro data files, Sequence files, Xml and Json files.**
* Extensively worked on **Spark Core, Numeric RDD's, Pair RDD's, Data Frames, and Caching for developing Spark applications**
* Experience and Expertise in **ETL, Data analysis and designing data warehouse strategies**. Good knowledge in using apache **NiFi** to automate the data movement between different **Hadoop system**s.
* Extensive experience writing custom Map Reduce programs for data processing and UDFs for both **Hive** and **Pig** in Java. Extensively worked on **MRV1** and **MRV2** Hadoop architectures.
* Strong experience in analyzing large amounts of data sets writing **PySparkscripts** and **Hive queries**.
* Experience in using Python's packages like **xlrd, numpy, pandas, scipy, scikit-learn and IDEs - PyCharm, Spyder, Anaconda, Jupyter, IPython.**
* Extensive experience in working with structured data using Hive QL, join operations, writing custom UDF's and experienced in optimizing Hive Queries.
* Extensive experiences in working with semi/unstructured data by implementing complex map reduce programs using design patterns.
* Experience in importing and exporting data using Sqoop from HDFS to Relational Database.
* Experience in Apache Flume for collecting, aggregating and moving huge chunks of data from various sources such as webserver, telnet sources etc.
* Experience with **Oozie** Workflow Engine in running workflow jobs with actions that run **Hadoop, MapReduce, Hive, Spark jobs**.
* Involved in moving all log files generated from various sources to HDFS and Spark for further processing.
* Experienced with the Spark improving the performance and optimization of the existing algorithms in Hadoop using **Spark Context, Spark-SQL, Data Frame, Pair RDD's and Datasets**.
* Excellent understanding and knowledge of **NOSQL** databases like **MongoDB, HBase, and Cassandra.**
* Experience in implementing Kerberos authentication protocol in Hadoop for data security.
* Experience in Dimensional modelling, logical modelling and Physical data modelling. Experienced with code versioning and dependency management systems such as **Git, SVT, and Maven**.
* Experience with Testing Map Reduce programs using **MRUnit, Junit, ANT, Maven**.
* Adequate knowledge and working experience in **Agile & Waterfall methodologies**.
* Great team player and quick learner with effective communication, motivation, and organizational skills combined with attention to details and business improvements. In depth knowledge of Apache **Cassandra** and **Datastax** Enterprise **Cassandra**.
* In depth knowledge in**Cassandra** read, writes paths and internal architecture.
* Experience in upgrading and migrating various versions of Mongo database on different platforms.
* Developed enhancements to **MongoDB** architecture to improve performance and scalability.
* Familiar with **MongoDB** clusters, Java scripting to load unstructured data into sharding environment.
* Experience in Importing data from various sources to the Cassandra cluster using Java API's
* Experience in data modeling of Cassandra
* Experienced in installing, configuring and monitoring the **Datastax Cassandra Cluster**, DevCenter and OpsCenter.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| Big Data | **Hadoop, Hive, Sqoop, Pig, Puppet, Ambary, HBase, MongoDB, Cassandra, Power Pivot, Defamer, Pentaho, Spark, Flume, SolrCloud ,Flume , Impala** |
| Operating Systems | **Windows, Ubuntu, Red Hat Linux, Linux, UNIX** |
| Project Management | **Plan View, MS-Project** |
| Programming & Scripting Languages | **Java, SQL, Unix Shell Scripting, C ,Python#** |
| Modelling Tools | **UML, Rational Rose** |
| IDE/GUI | **Eclipse** |
| Framework | **Struts, Hibernate** |
| Database | **MS-SQL, Oracle, MS-Access** |
| Middleware | **Web Sphere, TIBCO** |
| ETL | **Informatica, Pentaho** |
| Business Intelligence | **OBIEE, Business Objects** |
| Testing | **Quality Centre, Win Runner, Load Runner, QTP** |

**WORK EXPERIENCE:**

**Senior Hadoop Developer**

**NEC Corporation of America**, **Irving, TX Jan 2018-Present**

**Description**:Lithium Response - LR (Formally known as Lithium Social Web - LSW) allows **NEC Corporation ofAmerica** to manage Conversations on Social networks where**NEC Corporation of America** has presence (such as Facebook, Twitter etc.) at scale and increase the quality of member's social experience.

**Responsibilities:**  
• Provide design recommendations and thought leadership to sponsors /stakeholders that improved review processes and resolved technical problems  
• Developed **ETL** Applications using **HIVE, IMPALA & SQOOP** and Automated using **Oozie**  
• workflows and Shell scripts with error handling Systems and scheduled using Autosys.  
• Built **Sqoop jobs** to import massive amounts of data from relational databases - **Teradata & Netezza**   
• back-populate on **Hadoop** platform.  
• Working on creating a common workflow to convert from **EBCDIC** format to **ASCII** from the Mainframe sources to a delimited file in the **Avro** format to **HDFS**.  
• Worked on **Avro** and **Parquet** File Formats with **snappy compression**.  
• Developing scripts to perform business transformations on the data using **Hive** and **PIG.**  
• Creating Impala views on top of Hive tables for faster access to analyze data through **HUE/TOAD**.  
• Connected Impala with different BI tools like **TOAD and Sql** Assistant to help modeling team to run the different RISK models. Responsible for data integrity checks, data validation checks and data certification checks in order to provide the quality data for the downstream applications.

• Worked on building Data-pipe line using Talend-Big data as an **ETL** tool, which involves custom Java code that validates the data in the pipe line.

• Involved in building validation framework around the data that is present in Enterprise Data Platform.

• Developed efficient bulk loading techniques with the huge sets of data to the **Hbase & Phoenix** tables more efficiently achieving industrial range bench mark.

• Worked on bringing unstructured data (Image file and Email data) to HDFS and automate the process in bringing the business value by making the claims process in much efficient way than the traditional systems.

• Worked on cleaning and parsing the different file formats such as **Parquet, Avro, Csv and XLSb** files.

• Used **Cassandra CQL** with **Java API's** to retrieve data from Cassandra tables.  
• Worked on analyzing and examining customer behavioral data using Cassandra.  
• Created partitioned tables in Hive, mentored analyst and SQA team for writing Hive Queries.  
• Developed **Pig** Latin scripts to extract the data from the web server output files to load into HDFS.  
• Involved in build/deploy applications using Maven and integrated with **CI/CD** server Jenkins.  
• Involved in agile methodologies, daily scrum meetings, Spring planning's

• Developed **PIGUDF'S** for manipulating the data as per the business requirements and worked on developing custom PIG Loaders.  
• Used **Gradle** for building and testing project.  
• Used **GIT** for version control  
**Environment**: CDH 4, CDH5, HDFS, Map Reduce, Hive, Pig, Sqoop, Oozie, Impala, Spark, Kafka, Teradata, Linux, Java, Eclipse, SQL Assistant, TOAD.

**Senior Hadoop Developer:**

**ENVISION EMI** - **Vienna, VA Feb 2017-Dec 2017**

**Description**: ENVISION EMI is a consumer credit reporting agency. ENVISION EMI collects and aggregates information on over 800 million individual consumers and more than 88 million businesses worldwide. Founded in 1899 and based in Atlanta, Georgia, it is one of the three largest credit agencies along with Experian and Trans Union.

**Responsibilities:**

• Worked on Hadoop eco-systems including **Hive, HBase, Oozie, Pig, Zookeeper, Spark Streaming MCS (MapR Control System)** and so on with MapR distribution.

• Installed and configured **Hadoop Map Reduce, HDFS,** developed multiple Map Reduce jobs in Java for data cleaning and Pre-processing.

• Primarily involved in Data Migration process using Azure by integrating with **Github** repository and **Jenkins**.

• Built code for real time data ingestion using **Java, Map R-Streams (Kafka) and STORM**.

• Involved in various phases of development analysed and developed the system going through Agile Scrum methodology.

• Worked on **Apache Solr** which is used as indexing and search engine.

• Involved in development of Hadoop System and improving multi-node Hadoop Cluster performance.

• Worked on analysing Hadoop stack and different Bigdata tools including **Pig and Hive, Hbase database and Sqoop.**

• Developed data pipeline using **flume Sqoop** and pig to extract the data from weblogs and store in HDFS

• Used Hive to analyse the partitioned and bucketed data and compute various metrics for reporting.

• Worked with different data sources like **Avro data files, XML files, JSON files, SQL server and Oracle** to load data into **Hive** tables**.**

* Experienced in storing the analyzed results into the Cassandra cluster.
* Experienced in provisioning and managing multi-datacentre **Cassandra** cluster on public cloud environment

Amazon Web Services (AWS) - EC2.

* Imported data from various resources to the **Cassandra**cluster using Java APIs.
* Enabled Security properties like authenticator and authorizer to protect the data from unauthorized users.
* Consistency levels for read and write queries were implemented depending on the use case.
* Optimized the Cassandra cluster by making changes in Cassandra properties and Linux OS configurations.
* Working closely with Datastax to resolve issues on cluster using ticketing mechanism.
* Configured Performance Tuning and Monitoring for **Cassandra** Read and Write processes for fast I/O operations and low latency time.
* Used **J2EE** design patterns like Factory pattern & Singleton Pattern.
* Used **Spark** to create the structured data from large amount of unstructured data from various sources.
* Implemented usage of Amazon EMR for processing Big Data across Hadoop Cluster of virtual servers on **Amazon Elastic Compute Cloud (EC2) and Amazon Simple Storage Service (S3)**
* Performed transformations, cleaning and filtering on imported data using **Hive, Map Reduce, Impala and loaded final data into HDFS.**
* Developed Python scripts to find vulnerabilities with **SQL** Queries by doing **SQL injection**.
* Experienced in designing and developing Spark using Scala to compare the performance of Spark with **Hive and SQL/Oracle**.
* Responsible for coding Map Reduce program, Hive queries, testing and debugging the Map Reduce programs.
* Extracted Real time feed using Spark streaming and convert it to **RDD** and process data into Data Frame and load the data into Cassandra.
* Involved in the process of data acquisition, data pre-processing and data exploration of telecommunication project in Scala.
* Implemented a distributed messaging queue to integrate with Cassandra using Apache Kafka and Zookeeper.
* Specified the cluster size, allocating Resource pool, Distribution of Hadoop by writing the specification texts in **JSON** File format.
* Imported weblogs & unstructured data using the Apache Flume and stores the data in Flume channel.
* Exported event weblogs to HDFS by creating a **HDFS** sink which directly deposits the weblogs in HDFS.
* Used **RESTful** web services with **MVC** for parsing and processing XML data.
* Utilized **XML** and **XSL** Transformation for dynamic web-content and database connectivity.
* Involved in loading data from **UNIX** file system to **HDFS**. Involved in **designing schema, writing CQL's and loading data using Cassandra.**
* Built the automated build and deployment framework using **Jenkins, Maven** etc.

**Environment:**Hadoop, Map Reduce, HDFS, Ambari, Hive, Sqoop, Apache Kafka, Oozie, SQL, Alteryx, Flume, Spark, Cassandra, Scala, Java, AWS, GitHub.

**Hadoop Developer**

**Purdue Pharma, Stamford, CT Mar 2016-Nov 2016**  
**Description**: **Purdue Pharma** isan American multinational medical devices, pharmaceutical and consumer packaged goods manufacturing company founded in 1886. Its common stock is a component of the Dow Jones Industrial Average and the company is listed among the Fortune 500.

**Responsibilities:**

• Involved in Analysis, Design, System architectural design, Process interfaces design, design documentation.

• Responsible for developing prototypes the selected solutions and implementing complex big data projects with a focus on **collecting, parsing, managing, analysing and visualizing large sets of data using multiple platforms.**

• Understand how to apply technologies to solve bigdata problems and to develop innovative big data solutions.

• Developed Spark Applications by using Scala, Java and Implemented Apache Spark data processing project to handle data from various **RDBMS** and Streaming sources.

• Used Spark Streaming **APIs** to perform transformations and actions on the fly for building common learner data model which gets the data from **Kafka** in near real time and persist it to Cassandra.

• Responsible for analysing and cleansing raw data by performing **Hive queries** and running **Pig scripts on data.**

• Developed **Pig Latin scripts to extract** the data from the web server output files to load into HDFS.

• Developed **Simple to complex** Map Reduce Jobs using **Hive** and **Pig.**

• Performed importing data from various sources to the **Cassandra cluster using Sqoop**. Worked on creating data models **for Cassandra** from **Existing Oracle data model**.

• Hands on experience in installing, configuring MapR, Hortonworks clusters and installedHadoopecosystem components like **Hadoop Pig, Hive, HBase, Sqoop, Kafka, Oozie, Flume, Zookeeper**.

• Worked in **Spark** and **Scala** for Data Analytics. Handle ETL Framework in Spark for writing data from **HDFS** to **Hive**.

• Used **Scala based** written **framework for ETL**.

• Developed multiple **spark streaming** and **core jobs** with Kafka as a **data pipe-line system**

• Worked and learned a great deal from **AWS** Cloud services like **EC2, S3, and EBS**.

• Migrated an existing on-premises application to AWS. Used AWS services like **EC2** and **S3** for small data sets processing and storage.

• Imported data from **AWSS3** into **Spark RDD**, Performed transformations and actions on **RDD's.**

• Extensively use **Zookeeper** as **job scheduler** for **Spark Jobs**.

• Worked on **Talend** with **Hadoop**. Worked in migrating from **Informatica Talend jobs**.

• Implemented a distributed messaging queue to integrate with Cassandra using **Apache Kafka** and **Zookeeper**.

• Developed **Kafka producer** and **consumer components** for **real time data processing**.

• Worked on physical transformations of data model which involved in **creating Tables, Indexes, Joins, Views and Partitions.**

• Involved in Cassandra Data modelling to create **key spaces** and **tables** in **multi Data Centre DSE Cassandra DB.**

**Environment:** Spark, HDFS, Kafka, Map Reduce (MR1), Pig, Hive, Sqoop, Cassandra, AWS, Talend, Java, Linux Shell Scripting.

**Hadoop Developer:**

**TMG Health** - **Jessup, PA Nov 2014-Sep 2015**

**Description:**TMG Health has a legacy of serving both large and small health plans and Pharmacy Benefit Managers, across the nation. We offer a full range of tried and true solutions and can manage the processes for the complex Dual Eligible population. Our expertise, coupled with a strong commitment to our Clients’ success, positions us as a trusted advisor and partner who can help solve the challenges of today and prepare for those of tomorrow.

**Responsibilities:**

• Performed data transformations like filtering, sorting, and aggregation using Pig

• Creating Sqoop queries to import data from **SQL, Oracle**, and **Teradata to HDFS**

• Created **Hive** tables to push the data to **Mongo DB**.

• Implemented several automated process using Bash scripting to validate large volume for XML files and handling sync up for schema changes between Oracle RDBMS and Hive Tables.  
• Worked on Informatica BDM to pull data from Hadoop Data-Lake into Teradata Data warehouse.  
• Used agile methodologies for sprint planning and deliverables.

• Wrote **complex aggregate queries** in **mongo** for report generation.

• Developed scripts to run scheduled batch cycles using **Oozie** and present data for reports.

• Worked on **a POC** for building a movie recommendation engine based on Fandango ticket sales data using **Scala** and **Spark Machine Learning library**.

• Developed big data ingestion framework to process **multi TB data** including data quality checks, transformation, and stored as efficient storage formats like **parquet** and loaded into **AmazonS3** using **Spark Scala API** and **Spark**.

• Implement automation, traceability, and transparency for every step of the process to build trust in data and streamline data science efforts **using Python, Java, Hadoop streaming, Apache Spark, Spark SQL, Scala, Hive, and Pig.**

• Designed highly efficient data model for optimizing large-scale queries utilizing Hive complex data types and Parquet file format.

• Performed data validation and transformation using Python and Hadoop streaming.

• Developed highly efficient **Pig Java UDFs** utilizing advanced concept like Algebraic and Accumulator interface to populate ADP Benchmarks cube metrics.

• Performed data validation and transformation using Python and Hadoop streaming.

• Developed highly efficient **Pig Java UDFs** utilizing advanced concept like Algebraic and Accumulator interface to populate ADP Benchmarks cube metrics.

• Loading the data from the different Data sources like **(Teradata and DB2)** into **HDFS** using **SQOOP** and load into **Hive tables**, which are partitioned.

• Developed bash scripts to bring the **TLOG** file from **ftp** server and then processing it to load into **hive tables**.

• Automated workflows using shell scripts and **Control-M jobs** to pull data from various databases into Hadoop **DataLake**.

• Involved in story-driven Agile development methodology and actively participated in daily scrum meetings.

• Inserted Overwriting the **HIVE** data with **HBase** data daily to get fresh data every day and used Sqoop to load data from **DB2** into **HBASE** environment

• Involved in converting **Hive/SQL** queries into **Spark transformations** using **Spark RDDs**, **Scala** and have a good experience in using **Spark-Shell** and **Spark Streaming**.

• Designed, developed and maintained BigData streaming and batch applications using Storm.

• Created **Hive**, **Phoenix**, **HBase tables** and **HBase** integrated **Hive tables** as per the design using **ORC** file format and **Snappy compression**.

• Developed **Oozie** Workflows for daily incremental loads, which gets data from **Teradata** and then imported into **hive tables**.

• **Sqoop jobs, PIG** and **Hive scripts** were created for data ingestion from relational databases to compare with historical data.

• Created **HBase** tables to load large sets of structured, semi-structured and unstructured data coming from **UNIX, NoSQL** and a variety of portfolios.

• Developed pig scripts to transform the data into structured format and it are automated through **Oozie coordinators**.

• Used Splunk to captures, indexes and correlates real-time data in a searchable repository from which it can generate reports and alerts.

**Environment:**Hadoop, HDFS, Spark, Strom, Kafka, Map Reduce, Hive, Pig, Sqoop, Oozie, DB2, Java, Python, Splunk, UNIX Shell Scripting.

**Java Developer**

**SoftValue Solutions,India Nov 2012- April 2014**

**Description:**This application enables clients to access, review and update Insurance coverage's online. The website functionality was extended by providing the customers to Login and perform online transactions like paying bills online and updating their insurance policy. Customer Care website was redesigned to provide customized information for the end users.

**Responsibilities:**  
• Responsible for developing various modules, front-end and back-end components using several design patterns based on client's business requirements.  
• Designed and Developed application modules using **spring** and **Hibernate frameworks**.  
• Designed and developed the front-end with **Swings** and **Spring MVC framework, Tag libraries** and **Custom Tag Libraries** and **development of Presentation Tier** using **JSP pages integratingAJAX, Custom Tag's, JSP Tag Lists, HTML, JavaScript and JQuery**.  
• Used **Hibernate** to develop persistent classes following **ORM principles**.  
• Deployed spring configuration files such as application context, application resources and application files.  
• Used **Java-J2EE** patterns like **Model View Controller (MVC)**, **Business Delegate, Session façade, Service Locator, Data Transfer Objects, Data Access Objects, Singleton and factory patterns**.  
• Used **JUnit** for Testing Java Classes.  
• Used **Waterfall methodology**.  
• Worked with Maven for build scripts and Setup the **Log4J** Logging framework.  
• Involved in the Integration of the Application with other services.  
• Involved in Units integration, bug fixing, and testing with test cases.  
• Fixed the bugs reported in User Testing and deployed the changes to the server.  
• Managing the version control for the deliverables by streamlining and re-basing the development streams of the **SVN**.  
  
**Environment**: Java/JDK, J2EE, spring 2.5, Spring MVC, Hibernate, Eclipse, Tomcat, XML, JSTL, JavaScript, Maven2, Web Services, JQuery, SVN, JUnit, Log4J, Windows, Oracle.

**Java Developer**

**InfoBiz Technologies, India May 2010-Aug 2012**

**Description**: Portfolio analyser intimating System is a new development Project for the users to monitor Client's Portfolio and sets respective alerts on their portfolios. These alerts triggered when the portfolios do not meet the respective targets outlined by the users for their respective client applications. The Client applications are operated from remote locations. To implement this functionalities web services SOAP is used to hit the endpoints of the applications. Spring is used to implement the feasible betterment of fine-grained approach in the application.

**Responsibilities:**

• As a Developer in Java, involved in **System Requirements analysis** and **conceptual design.**

• Experience in using **VSS** as **source control**.

• Developed the application using **User interface** using **Struts, Html** and **JavaScript, CSS**.

• Experience in using **JSON, JavaScript** functions to get dynamic data and **client side** validation.

• Experience in handling designed, Developed and Implemented document transformation applications using **XML Spy tool** and **XSD's**.

• Experience in developing **JDBC** to interact with the database

• Involved in developing applications for workflow using **JSP's**, **spring MVC module, Hibernate, AJAX, JavaScript technologies** using **Apache Tomcat** and integrated it with the actual system to improve the quality of the existing system.

• Experience in initiating with **'Workflow module'** of the Project which is a product intended to automate the workflow related functions of any organization.

• Experience in designing full life cycle development of web applications and modification of existing applications to incorporate new components and **features emphasizing maintainability, reliability, scalability and performance.**

**Environment:** Java, Servlets, JSP, JSTL, Spring, Ajax, Hibernate, Java JDK 1.6, Apache Tomcat 6.0, MySQL Query Browser, Windows XP Professional.