**Sandeep Shankar Murugesan**

Phone: 781-333-9598

Email: [msandeep.hadoop@gmail.com](mailto:msandeep.hadoop@gmail.com)

**Summary**

* Over **7 years** of IT experience which includes close to **3 years** of work experience in **Big Data, Hadoop** **ecosystem** related technologies.
* E**xcellent** understanding**/**knowledge of **Hadoop Ecosystem** including **HDFS**, **MapReduce**, **Hive**, **Pig, Storm, Kafka, YARN, HBase**, **Oozie**, **ZooKeeper**, **Flume** and **Sqoop** based **Big Data** Platforms
* Expertise in design and implementation of Big Data solutions in Banking, Retail and E-commerce domains
* Experienced with **NoSQL** databases like **HBase, MongoDB** and **Cassandra**
* Comprehensive experience in building Web-based applications using **J2EE** Frame works like **Spring, Hibernate, EJB, Struts** and **JMS**
* Excellent ability to use analytical tools to mine data and evaluate the underlying patterns
* Assisted in **Cluster maintenance**, **Cluster Monitoring** and **Troubleshooting**, Managing and Reviewing data backups and log files
* Good working Experience in client-side development with HTML, XHTML, CSS, JavaScript, JQuery, JSON and AJAX
* Good understanding of application servers WebLogic, WebSphere and XML methodologies (XML, XSL, XSD) including Web Services like SOAP and REST.

**MapReduce**

* Hands-on experience in developing **MapReduce** programs using **Apache Hadoop** for analyzing the **Big Data**
* Expertise in optimizing traffic across network using **Combiners**, joining multiple schema datasets using **Joins** and organizing data using **Partitioners** and **Buckets**
* Experience in writing **Custom Counters** for analysing data and testing using **MRUnit** framework
* Experienced in writing complex MapReduce programs that work with different file formats like **Text, Sequence, Xml and Avro**
* Expertise in composing **MapReduce Pipelines** with many user-defined functions using **Apache Crunch**

**PIG**

* Expertise in writing ad-hoc MapReduce programs using **Pig Scripts**
* Used **Pig** as ETL tool to do transformations, event joins, filter and some pre-aggregations
* Implemented business logic by writing **Pig Latin UDFs** in Java and used various UDFs from Piggybanks and other sources

**HIVE**

* Expertise in Hive Query Language(**HiveQL**), Hive Security and debugging Hive issues
* Responsible for performing extensive data validation using **HIVE Dynamic Partitioning** and **Bucketing**
* Experience in developing custom **UDFs** for Pig and **Hive** to incorporate methods and functionality of **Python/Java** into Pig Latin and HQL (Hive QL)
* Worked on different set of tables like **External** Tables and **Managed** Tables
* Experience with working different **Hive SerDe's** that handle file formats like avro, xml
* Analyzed the data by performing **Hive queries** and used **HIVE UDF’s** for complex querying

**NoSQL**

* Expert database engineer, **NoSQL** and relational data modeling
* Responsible for building scalable distributed data solutions using **MongoDb**, **Cassandra,** and **Neo4J**
* Expertise in **HBase** Cluster Setup, Configurations, HBase Implementation and **HBase Client API**
* Worked on importing data into HBase using **HBase Shell** and **HBase Client API**
* Expertise in performing large-scale web crawling with **Apache Nutch** using a **Hadoop/HBase** cluster

**Java/J2EE**

* Expertise in several J2EE technologies like JDBC, Servlets, JSP,Struts, Spring, Hibernate, JPA, JSF, EJB, JMS, JAX-WS, SOAP, JQuery, AJAX, XML, JSON, HTML5/HTML, XHTML, Maven, and Ant
* Expert knowledge over **J2EE** **Design Patterns** like MVC Architecture, Front Controller, Session Facade, Business Delegate and Data Access Object for building J2EE Applications
* Thorough knowledge on **JAX-WS** to access the external **Web Services**, get the xml response and convert it back to java objects
* Experience in using **Jenkins** for Continuous Integration and **Sonar jobs** for java code quality
* Extensive experience in developing Internet and Intranet related applications using **J2EE**, **Servlets**, **JSP**, **Jboss**, **WebLogic**, **Tomcat**, and **Struts** Frame Work

**SQL, Script & Oracle Database**

* Extensive experience with database DB2, Oracle9i/10g/11g (Database Design, and SQL Queries)
* Good experience in SQL, PL/SQL, Perl Scripting, Shell Scripting, Partitioning, Data modeling, OLAP, Logical and Physical Database Design, Backup and Recovery procedures
* Good knowledge of database connectivity (JDBC) for databases like Oracle, DB2, SQL Server, MySQL, MS Access, Netezza.

**Others**

* Experienced with build tool **Maven**, **Ant** and continuous integrations like **Jenkins**
* Experience in Administering, Installation, Configuration, Troubleshooting, Security, Backup, Performance Monitoring and Fine-tuning of **Linux** Red Hat
* Developed Unit test cases using **JUnit**, **Easy Mock** and **MRUnit** testing frameworks
* Experienced in **Agile** **SCRUM**, **RUP** (Rational Unified Process) and **TDD** (Test Driven Development) software development methodologies

**Technology**

Hadoop/Big Data/NoSql Technologies HDFS, MapReduce, Hive, Pig, Sqoop, Flume, Oozie, Avro, Hadoop Streaming, Storm, Spark, Kafka, YARN, Crunch, Zookeeper, HBase, Cassandra, MongoDb, Neo4J

**Programming Languages:** Java, Python, C, Scala, SQL, PL/SQL, Shell Script

**IDE Tools:** Eclipse, Rational Team Concert, NetBeansFramework Hibernate, Spring , Struts, JMS, EJB ,JUnit, MRUnit, JAXB

**Web Technologies:**HTML5, CSS3, JavaScript, JQuery, AJAX, Servlets, JSP,JSON, XML, XHTML, Rest Web Services

**Application Servers:** Jboss, Tomcat, Web Logic, Web Sphere

**Databases:** Oracle 11g/10g/9i, MySQL, DB2, Derby, MS-SQL Server

**Operating Systems:** LINUX,UNIX, Windows

**Build Tools:** Jenkins, Maven, ANT

**Reporting/BI Tools:** Jasper Reports, iReport, Tableau, QlikView

**Professional Experience**

**BCBS Detroit, MI 2014 – Present**

**Sr. Hadoop Developer**

**Responsibilities:**

* Developed simple and complex **MapReduce** programs in **Java** for Data Analysis on different data formats
* Developed **MapReduce** programs that filter bad and unnecessary claim records and find out unique records based on account type
* Processed **semi, unstructured** data using **Map Reduce** programs
* Implemented Daily **Cron jobs** that automate parallel tasks of loading the data into **HDFS** and pre-processing with **Pig** using **Oozie co-ordinator** jobs
* Implemented **custom** **DataTypes**, **InputFormat, RecordReader, OutputFormat, RecordWriter** for **MapReduce** computations
* Successfully migrated Legacy application to Big Data application using Hive/Pig/HBase in Production level
* Transformed date related data into application compatible format by developing apache Pig UDFs
* Developed **MapReduce** pipeline for feature extraction and tested the modules using **MRUnit**
* Optimized **MapReduce** jobs to use HDFS efficiently by using various compression mechanisms
* Creating **Hive** tables, loading with data and writing **Hive queries** which will run internally in MapReduce way
* Responsible for performing extensive data validation using **Hive**
* Populated HDFS and **Cassandra** with huge amounts of data using **Apache Kafka**.
* Exposure to Amazon Web Services – AWS cloud computing (EMR, EC2 and S3 services).
* Created the Load Balancer on **AWS EC2 for unstable cluster**
* Expertise with NoSQL databases like **Hbase, Cassandra, DynamoDB (AWS) and MongoDB**
* Very Good experience on **UNIX shell scripting, Python and WLST.**
* Expertise in designing python scripts to interact with **middleware/back** end services.
* Worked on python scripts to analyze the data of the customer.
* Completely Designed, Developed and Owned the of the automation flow of **Taleo Fusion wiring using Python**.
* In-depth understanding of **Data Structures and Algorithms**. Have hands on experience in writing **MapReduce** jobs in **Java, Pig and Python.**
* Implemented Partitioning, Dynamic Partitions and Bucketing in **Hive** for efficient data access
* Worked on different set of tables like **External Tables** and **Managed Tables**
* Used **Oozie** workflow engine to run multiple **Hive** and **Pig** jobs
* Involved in installing and configuring Hive, Pig, Sqoop, Flume and Oozie on the Hadoop cluster.
* Involved in designing and developing non-trivial ETL processes within Hadoop using tools like **Pig**, **Sqoop**, **Flume**, and **Oozie**
* Used **DML** statements to perform different operations on Hive Tables
* Developed Hive queries for creating foundation tables from stage data
* Used **Pig** as ETL tool to do transformations, event joins, filter and some pre-aggregations
* Analyzed the data by performing Hive queries and running Pig scripts to study customer behavior
* Implemented business logic by writing **Pig UDFs** in Java and used various UDFs from Piggybanks and other sources
* Working with **Apache Crunch** library to write, test and run HADOOP MapReduce pipeline jobs
* Involved in joining and data aggregation using **Apache Crunch**
* Worked with **Sqoop** to export analyzed data from HDFS environment into RDBMS for report generation and visualization purpose
* Involved in writing, testing, and running MapReduce pipelines using **Apache Crunch**
* Queried and analyzed data from Datastax **Cassandra** for quick searching, sorting and grouping
* Developed Mapping document for reporting tools

**Environment:** Apache Hadoop, HDFS, MapReduce, Apache Crunch, Java (jdk1.6), MySQL, DB Visualizer, Linux, Sqoop, kafka, Apache Hive, Apache Pig,Cassandra

**HiltonWorldwide, Memphis,TN Jan 2013 – Feb 2014**

**Hadoop Developer**

**Responsibilities:**

* Installed, configured, and maintained Apache Hadoop clusters for application development and major components of Hadoop Ecosystem: **Hive**, **Pig**, **HBase**, **Sqoop**, **Flume**, **Oozie** and **Zookeeper**
* Implemented six nodes **CDH4 Hadoop Cluster** on CentOS
* Importing and exporting data into HDFS and Hive from different RDBMS using **Sqoop**
* Experienced in defining job flows to run multiple **MapReduce** and **Pig** jobs using **Oozie**
* Importing log files using **Flume** into **HDFS** and load into Hive tables to query data
* Involved in writing APIs to Read **HBase** tables, cleanse data and write to another **HBase** table
* Created multiple **Hive** tables, implemented **Partitioning**, **Dynamic Partitioning** and **Buckets** in **Hive** for efficient data access
* Very Good experience on **UNIX shell scripting, Python and WLST.**
* Completely Designed, Developed and Owned the of the automation flow of Taleo Fusion wiring using Python.
* Responsible for configuring Hadoop clusters with CDH4 on CentOS, managing with **Cloudera Manager**
* Written multiple **MapReduce** programs in Java for data extraction, transformation and aggregation from multiple file formats including XML, JSON, CSV and other compressed file formats
* Experience working with **Apache SOLR** for indexing and querying
* Worked on Kafka for real time data feeds.
* Knowledge on **Zookeeper** internals
* Experienced in running batch processes using **Pig Scripts** and developed **Pig UDFs** for data manipulation according to Business Requirements
* Manage Amazon Web Services **(AWS) EC2** with Puppet
* Experience with Load Balancers on the **AWS**
* Took **Splunk** tools used for log aggregation and implemented log data analysis.
* Experience in Operational Intelligence using **Splunk**.
* Knowledge about **Splunk** architecture and various components (indexer, forwarder, search head, deployment server), Heavy and Universal forwarder, License model.
* Experienced in writing programs using **HBase Client API**
* Involved in loading data into **HBase** using **HBase Shell, HBase Client API**, **Pig** and **Sqoop**
* Written MapReduce program in Python with the **Hadoop streaming API**
* Developed unit test cases for Hadoop MapReduce jobs with **MRUnit**
* Expertise in designing python scripts to interact with **middleware/back end services.**
* Worked on **python scripts** to analyze the data of the customer.
* Independently researched the viability of a **Neo4J** database for existing customer data
* Python developer working on a large, multiple data center deployment of **OpenStack**. Build RESTful APIs around OpenStack.
* Excellent experience in ETL analysis, designing, developing, testing and implementing ETL processes including performance tuning and query optimizing of database
* Experience in using Pentaho Data Integration tool for data integration, OLAP analysis and ETL process
* Created **R** **function** and **Spark** stream to pull customer sentiment data from Twitter.
* Used **R** to cleanse data for analysis and Created Codebook to outline the process for data analysis
* Connected from R to HBase, Oracle and TeraData, for Analytics purpose
* Experience integrating R with **Hadoop** with **RHadoop** for statistical analysis and predictive modelling

**Environment:** Apache Hadoop1.0, Hive, Pig, HBase, Sqoop, kafka, Flume, Java, Linux, OpenStack, MySQL Server5.155, MS SQL Server 2012, SQL, PL/SQL, SQL Server Data Tools, SQL Server Business Intelligence Development Studio (SSAS, SSIS, SSRS), R

**Lee Hecht Harrison LLC, Dec 2011 – Nov 2012**

**Hadoop Developer**

**Woodcliff Lake, NJ**

**Responsibilities:**

* This is an initiative from Citibank to move its disconnected legacy billing & monitoring systems to a consolidated platform
* Consolidating customer data from Lending, Insurance,Trading and Billing systems into warehouse and mart subsequently for business intelligence reporting
* Providing improved revenue capture through leakage elimination, Assessing risk score more accurately in its customer portfolios, having better exposure and to offer each customer better products and advice
* Experience working with **JIRA** for project management, **GIT** for source code management, **JENKINS** for continuous integration and **Crucible** for code reviews
* Transferring and exporting data into **HDFS** and **Hive**, **MySQL** and **DB2** using **Sqoop**
* Implementing **MapReduce** programs to analyze large datasets in warehouse for business intelligence purpose
* Used default MapReduce Input and Output Formats
* Developed **HQL** queries to implement the select, insert, update and operations to the database by creating HQL named queries
* Based on user usage identifying the customer potentiality using MapReduce Programs.
* Using **Oozie** in your environment to schedule Hadoop jobs and would like to call Sqoop from within your existing workflows
* Automatically Importing data regular basis using sqoop to into the **Hive** partition by using apache Oozie
* Clustering customers category based on that providing offers using Apache Hive
* Experienced in managing and reviewing Hadoop log files Load and transform large sets of data
* Grouping,Aggregation and Sorting are done by using **Pig** and **Hive** which are higher level abstractions of MapReduce
* Built and maintained network and production infrastructure, including Cisco firewall, Dell Servers with VMware ESXi 5.5 hypervisor, and all client workstations
* Developed BI Reporting dashboard, queries, reports using Hyperion Interactive reporting (IR)
* Conducted data extraction that may include analyzing, reviewing, modeling based on requirements using higher Level Tools such as Hive and Pig
* Supported MapReduce Programs those are running on the cluster
* Involved in creating Hive tables, loading with data and writing Hive queries
* Created data-models for customer data using the **Cassandra Query Language**
* Ran many performance tests using the Cassandra stress tool in order to measure and improve the read and write performance of the cluster
* Queried and analyzed data from **Datastax Cassandra** for quick searching, sorting and grouping

**Environment:** Apache Hadoop (Cloudera), Java (jdk1.6), Teradata, Redhat Linux, Sqoop, Hive, Cassandra, DB visualizer, Oozie

**Denim Group,**  **Jan 2010 – Nov 2011**

**J2EE Software Developer**

**San Antonio, TX**

**Responsibilities:**

* Application was developed using the Struts **MVC architecture**
* Developed action and form classes based on **Struts framework** to handle the pages
* Developed a web-based reporting for credit monitoring system with **HTML5, XHTML, JSTL**, **custom tags** and Tilesusing **Struts framework**
* Developed Servlets and JSPs based on MVC pattern using Struts framework and Spring Framework
* Developed web-based customer management software using Facelets, Icefaces and JSF
* Implemented **Ajax** Frame works, jQuery tools examples like Auto Completer, Tab Module, and Calendar and Floating windows
* Configured Struts-Config file for form-beans, global forwards, error forwards and action forwards
* Designed and implemented Report Module (using **Jasper Report** framework)
* Developed Message-Driven beans in collaboration with Java Messaging Service (**JMS**)
* Developing Web Services using **Apache Axis 2** to retrieve data from legacy systems
* Developed Servlets, Action classes, Action Form classes and configured the struts-config.xml file
* Used XML parser APIs such as **JAXP** and **JAXB** in the web service's request response data marshalling as well as un-marshalling process
* Developed UI components for email and link sharing of documents and files for a Content Management System using **BackBone.js** and **JQuery**
* Planned and implemented various SQL**, Stored Procedure, and triggers**
* Used **Hibernate** to access My SQL database and implemented of connection pooling
* Developed JavaScript based components using **Ext JS** framework like GRID, Tree Panel with client reports customized according to user requirements
* Performed building and deployment of **WAR, JAR files on test, stage**, and **production** systems in **Apache Tomcat** application server
* Used **ANT** for the build process

**Environment:** J2EE, Java 1.4.2, Servlets, JSP, JDBC, EJB 3, JMS, JQuery, backbone.js, HTML5, JSTL, Icefaces, XML, Spring, Struts, Hibernate, Web Services, Apache Tomcat Server, JSF, EXT JS, JAXB, Jasper Report, JUnit, SOAP, SOAPUI, XML, JavaScript, UML, Apache Axis 2, ANT, SVN, MySQL

**Mind Tree Ltd., May 2007 – Nov 2009**

**Java Developer**

**Bangalore, India**

**Responsibilities:**

* Worked on Requirement analysis, gathered all possible business requirements from end users and business Analysts
* Involved in creation of **UML diagrams** like Class, Activity, and Sequence Diagrams using modelling tools of **IBM Rational Rose**
* Worked with core **Java** code extensively using interfaces and multi-threading techniques
* Involved in production support and documenting the application to provide training and knowledge transfer to the user
* Used **Log4j** for logging mechanism and developed wrapper classes to configure the logs
* Used **JUnit** and Test cases for testing the application modules
* Developed and configured the Java beans using **SpringMVC** framework
* Developed the application using **Rational Team Concert** and worked under **Agile Environment**
* Developed **SQL** stored procedures and prepared statements for updating and accessing data from database
* Also used **C++** to create some libraries used in the application

**Environment:**  
C++, Java, JDBC, Servlets, JSP, Struts, Eclipse, Oracle 9i, Apache Tomcat, CVS, JavaScript, Log4J

**Educational Qualification:** B.E. (ECE)/ Karpagam College/Anna University, 2008