**Venkata Metla** **Email**: venkatasandeep86@gmail.com

Cassandra Consultant **Phone**: 929-279-2801

**Professional Overview:**

* Over 8+ years of experience in software design, development, maintenance of enterprise applications using enterprise Java, Hadoop and Cassandra
* Having a good experience in Cassandra database modeling and administration
* Implemented and monitored Multi DC - Multi Rack Cassandra Clusters using **nodetool** and **OpsCenter**
* Experience in migrating data from multiple relational databases to Cassandra and validated data
* Experience in performance tuning and maintenance of Cassandra Database and has a good knowledge in write and read parts along with the internal architecture
* Developed several applications using J2EE/Spring and AngularJS from scratch
* Implemented security, backup/recovery on Cassandra Cluster using nodetool snapshots
* Good understanding of Linux implementation, customization and file recovery
* Managing Cassandra clusters using Datastax Opscenter
* Experience in Cassandra systems backup and recovery
* Excellent CQL and data modeling knowledge from conceptual to physical data models including modern Chebotko diagrams
* Good experience in supporting data analysis projects using **Amazon Web Services** (AWS) cloud
* Experience in Cassandra Cluster using Cassandra stress tool
* Good Knowledge of **Spark** and **SCALA programming** on Cassandra
* Good experience in all the phases of Software Development Life Cycle
* Experience with Disk Management and User Management and permissions and also has proficiency in the shell scripting
* Expertise in developing Web applications using Java, J2EE, Servlets, Spring, PHP, HTML5 and CSS3
* Excellent experience in Agile and scrum methodologies
* Excellent team player, self-starter with effective communication skills

**Technical Skills:**

* **Cassandra:** Cassandra, DataStax, DevCenter and OpsCenter, nodetool, Spark on Cassandra
* **Big data:** Hadoop/HDFS, Spark, Hive, HBase, Pig, Sqoop, Zookeeper, Flume, AWS, YARN and Kafka
* **Scripting:** Python, Unix Shell Scripting
* **Enterprise:**  J2EE, Spring Framework (Spring MVC, Security, Batch, Core) and Hibernate
* **RDBMS:** SQL Server, MySQL, Oracle
* **Operating Systems:** Windows, IOS, Unix
* **Servers:** Apache Tomcat, Web Logic, Web Sphere, JBOSS
* **Tools:**   Eclipse, NetBeans, Maven, Gradle, Ant, ArgoUML
* **Version Control:** Git, SVN and CVS
* **Software Methodologies:** Agile, Scrum, Waterfall

**Professional Experience:**

**Cassandra Developer**  **June 2016- Present**

BCBS Health Plus, Charlotte, NC

Responsibilities:

* Involved in the process of data modelling Cassandra Schema
* Experienced in performance tuning a Cassandra Cluster to optimize writes and reads using the Cassandra-stress tool
* Extensively worked on Java persistence layer in application migration to Cassandra using Spark to load data to and from Cassandra Cluster
* To set the configuration properties like memtable size, replication factors, node addresses, flush times and client storage I have wrote and modified the YAML scripts and optimized the Cassandra cluster by making changes in Cassandra configuration file
* Created User defined types (UDT) to store specialized data structures in Cassandra
* Administered and maintained multi rack Cassandra cluster using OpsCenter based on the use case implemented consistency level for reads and writes
* Involved in a POC using Apache YARN to implement a failsafe distributed data storage and computation system
* Tune databases and provide design changes and support stress tests to proactively fix problems
* Used the in-memory capabilities of Spark to implement text analytics and processing
* Loaded and transformed large sets of structured, semi structured and unstructured data in various formats like text, zip, XML, YAML and JSON.
* Data loading from MongoDB to Cassandra Clusters.
* Experience using DataStax Pig functionality to develop the PIG UDFs for manipulating the data and extracting useful information according to Business Requirements
* Developed Spark Streaming applications for Real Time Processing
* Implemented a distributed messaging queue to integrate with Cassandra using Apache Kafka and ZooKeeper
* Implemented several Scala scripts for analytics on top of Cassandra cluster

***Environment****:* Cassandra, J2EE, Spring MVC, Spring Cassandra, Spring Batch, Spring Security, Apache Kafka, Scala on Spark, Linux

**Cassandra Developer January 2015- May 2016.**

Allstate Insurance, Chicago, IL

**Responsibilities:**

* Responsible for the build out, day-to-day managing and supporting Cassandra clusters
* Configure backup, alerts, repairs and monitoring of Cassandra clusters using Opscenter
* Involved in business requirement gathering and proof of concept creation
* Created data models from scratch in CQL for customer data and optimized the data models
* Involved in Hardware installation and capacity planning for cluster setup
* Exported the analyzed data to relational databases using Sqoop for visualization and to generate reports
* Created data-models for customer data using the Cassandra Query Language
* Tuned and recorded performance of Cassandra clusters by altering the JVM parameters
* Changed garbage collection cycles to place them in tune with backups to mitigate disk contention
* Used collections like lists, sets and maps to create data models highly optimized for reads and writes
* Created User defined types to store specialized data structures in Cassandra
* Designed and developed a full-stack application for data analytics on top of Cassandra/Spark cluster with which people can run jobs and see results
* Queried and analyzed data from Datastax Cassandra for quick searching, sorting and grouping
* Used the Spark - Cassandra Connector to load data to and from Cassandra
* Involved in migrating data from MongoDB to Cassandra
* Worked on the optimization of the read path and write path of Cassandra by running many tests
* Created the upgrade plans for DSE upgrades
* Administered, monitored and maintained multi data-center Cassandra cluster using OpsCenter and Nagios in production

***Environment****:* Cassandra, Apache Hadoop, MapReduce, Scala on Spark, Linux, J2EE, Spring Framework, AngularJS, REST services

**Cassandra Developer April 2013- December 2014.**

Visa Inc - Austin, TX

Responsibilities:

• Implemented the multi datacenter cluster consisting of Cassandra rings   
• Worked with developers for choosing the right Compaction Strategies and consistency levels   
• Optimized the Cassandra cluster by making changes in Cassandra configuration file and Linux OS configurations

• Documented and demonstrated on various ways to securely connect to the cluster   
• Designed and implemented a strategy to securely move production data to Development for testing purposes using stable loader   
• Designed and implemented a strategy to upgrade the DataStax enterprise cluster   
• Worked on major and minor upgrades of cluster, Knowledge on applying updates and maintenance patches for the existing clusters    
• Involved in Data modelling design for various use cases   
• Designed, automated the process of installation and configuration of secure DataStax Enterprise Cassandra using puppet

• Administered Cassandra cluster using Datastax OpsCenter and monitored CPU usage, memory usage and health of nodes in the cluster

• Used DataStax OpsCenter and nodetool utility to monitor the cluster  
• Involved in commission and decommission of the nodes in the existing Cassandra cluster   
• Evaluated, benchmarked and tuned data model by running endurance tests using JMeter, CassandraStress Tool and OpsCenter   
• Worked on tuning Bloom filters and configured compaction strategy based on the use case   
• Performed importing data from various sources to the Cassandra cluster using Java/Python APIs   
• Configured node-to-node and client-to-node encryptions by generating and distributing keystore and truststore files to all nodes   
• Involved in requirements gathering and Capacity planning for multi datacenter (four) Cassandra cluster.   
• Created testing environment with OpenStack VM technologies   
• Configured internode communication between Cassandra nodes and client using SSL encryption   
• Evaluated Data model by running endurance tests using JMeter and OpsCenter   
• Configured authorization to Cassandra cluster using Password Authenticator and Kerberos

**Environment:** Cassandra 2.2, Java, Python, BASH Scripting, Apache Hadoop 2.2.0, Apache Kafka, MapReduce, Spark, OpenStack, JMeter, OpsCenter, DevCenter, Linux, XML

**Hadoop Consultant February 2012- March 2013**

JLL - Chicago, IL

Responsibilities:

• Responsible for building scalable distributed data solutions using Hadoop

• Installed and configured Apache Hadoop to test the maintenance of log files in Hadoop cluster   
• Installed and configured Hive, Pig, Sqoop, and Oozie on the Hadoop cluster   
• Installed Oozie Workflow engine to run multiple Hive and Pig Jobs   
• Setup and benchmarked Hadoop/HBase clusters for internal use   
• Extracted data from databases like SQL Server and Oracle 9g into HDFS for processing using Pig and Hive

• Developed Java MapReduce programs for the analysis of sample log file stored in cluster   
• Developed Simple to complex Map/Reduce Jobs using Hive and Pig   
• Developed Map Reduce Programs for data analysis and data cleaning   
• Stored and retrieved data from data-warehouses using Amazon Redshift    
• Performed optimization on Pig scripts and Hive queries to increase efficiency

• Performed statistical analysis using Splunk   
• Developed PIG Latin scripts for the analysis of semi structured data   
• Used Hive and created Hive tables and involved in data loading and writing Hive UDFs   
• Used Sqoop to import data into HDFS and Hive from other data systems   
• Generated aggregations and groups and visualizations using Tableau   
• Continuous monitoring and managing the Hadoop cluster using Cloudera Manager   
• Migration of ETL processes from Oracle to Hive to test the easy data manipulation   
• Conducted some unit testing for the development team within the sandbox environment   
• Developed Hive queries to process the data for visualizing and reporting

• Used Visualization tools such as Power view for excel, Tableau for visualizing and generating reports

• Exported data to Tableau and excel with Power view for presentation and refining

• Implemented business logic by writing Pig UDFs in Java and used various UDFs from Piggybanks and other sources

• Implemented test scripts to support test driven development and continuous integration

• Actively participated in daily scrum meetings

Environment: Hadoop, Map Reduce, HDFS, Pig, Hive, Sqoop, Flume, Oozie, Java, Linux, Maven, Zookeeper, Tableau, HBase, Cassandra

**Hadoop Consultant October 2009- November 2011**

Brightstar Corporation, Miami, FL

Responsibilities:   
• Worked on analyzing, writing Hadoop MapReduce jobs using Java API, Pig Latin and Hive   
• Responsible for building scalable distributed data solutions using Hadoop   
• Involved in loading data from edge node to HDFS using shell scripting

• Developed pig scripts for analyzing large data sets in the HDFS

• Knowledge on handling Hive queries using Spark SQL that integrate with Spark environment

• Responsible for creating Hive tables, loading the structured data resulted from MapReduce jobs into the tables and writing hive queries to further analyze the logs to identify issues and behavioral patterns

• Performed extensive Data Mining applications using Hive

• Performed streaming of data into Flume and Kafka and transferred to HDFS for analysis

• Used Sqoop to import the data from RDBMS to Hadoop Distributed File System (HDFS) and later analyzed the imported data using Hadoop Components

• Written multiple MapReduce programs in Java for Data Analysis

• Wrote MapReduce job using Pig Latin and Java API

• Performed performance tuning and troubleshooting of MapReduce jobs by analyzing and reviewing Hadoop log files   
• Created HBase tables to store variable data formats of data coming from different portfolios   
• Implemented a script to transmit information from Oracle to HBase using Sqoop   
• Implemented best income logic using Pig scripts and UDFs   
• Implemented test scripts to support test driven development and continuous integration   
• Worked on tuning the performance using Apache Pig queries  
• Worked with QA team in preparation and review of test cases   
• Involved in loading and transforming large sets of Structured, Semi-Structured and Unstructured data and analyzed them by running Hive queries and Pig scripts   
• Experience in managing and reviewing Hadoop log files   
• Assisted application teams in installing Hadoop updates, operating system, patches and version upgrades when required   
• Used Oozie workflow engine to run multiple Hive and pig jobs automatically   
• Analyzed large amounts of data sets to determine optimal way to aggregate and report on it  
   
Environment: Hadoop, HDFS, Hive, Apache Pig, Sqoop, HBase, Shell Scripting, Ubuntu, Linux Red Hat, Zookeeper

**Java Developer**  **June 2008- August 2009**

Vodafone, India

Responsibilities:

• Involved in System Analysis and Design methodology as well as Object Oriented Design and development using OOA/OOD methodology to capture and model business requirements   
• Proficient in doing Object Oriented Design using UML-Rational Rose   
• Created Technical Design Documentation (TDD) based on the Business Specifications   
• Created JSP pages with Struts Tags and JSTL   
• Developed UI using HTML, JavaScript, CSS and JSP for interactive cross browser functionality and complex user interface   
• Implemented the web-based application following the MVC II architecture using Struts framework   
• Used XML DOM API for parsing XML   
• Developed Scripts for automation of productions tasks using Perl, UNIX scripts   
• Used ANT for compilation and building JAR, WAR and EAR files   
• Used JUnit for the unit testing of various modules   
• Project coordination with other Development teams, System managers and web master and developed good working environment   
• Generated Business Logic using servlets, Session beans and deployed them on Web logic server   
• Created complex SQL queries and stored procedures   
• Used Hibernate ORM framework with spring framework for data persistence and transaction management   
• Provided technical support for production environments resolving the issues, analyzing the defects, providing and implementing the solution defects.   
• Analyzed the banking and existing system requirements and validated them to suit J2EE architecture.   
• Designed the process flow between front-end and server side components   
• Developed web based presentation-using JSP, AJAX using Servlets technologies and implemented using struts framework   
• Designed and developed backend Java components residing on different machines to exchange information and data using JMS    
• Used JavaScript for client-side validation and Struts Validator Framework for form validations   
• Implemented Java/J2EE Design patterns like Business Delegate and Data Transfer Object (DTO), Data Access Object   
• Integrated Spring DAO for data access using Hibernate, used HQL and SQL for querying databases   
• Worked with QA team for testing and resolving defects   
• Used ANT automated build scripts to compile and package the application   
• Used JIRA for bug tracking and project management   
   
Environment: J2EE, JSP, JDBC, Spring Core, Struts, Hibernate, Design Patterns, XML, WebLogic, Apache Axis, ANT, Clear case, JUnit, JavaScript, Web Services, SOAP, XSLT, JIRA, Oracle, PL/SQL Developer and Windows