Mani
Sr. Hadoop Developer

Email: vmaniv08@gmail.com   Contact: (615) 813-1551

PROFESSIONAL SUMMARY:

- 8+ years of overall software development experience on **Big Data** Technologies, **Hadoop** Eco system and **Java/J2EE Technologies** with experience programming in **Java, Scala, Python and SQL**
- 4+ years of strong hands-on experience on Hadoop Ecosystem including **Spark, Map-Reduce, HIVE, Pig, HDFS, YARN, HBase, Oozie, Kafka, Sqoop, Flume**.
- Experience in **architecting**, **designing**, and building **distributed** software systems.
- **Scala** and **Java**, Created frameworks for processing data pipelines through **Spark**
- Wrote **python** scripts to parse XML documents and load the data in database.
- Deep knowledge of **troubleshooting** and tuning **Spark** applications and **Hive** scripts to achieve optimal performance.
- Used **Sqoop** to import data into HDFS / Hive from RDBMS and exporting data back to HDFS or HIVE from RDBMS.
- Worked with **real-time data** processing and streaming techniques using **Spark streaming**, **Storm** and **Kafka**.
- Experience developing **Kafka producers and Kafka Consumers** for streaming millions of events per second on streaming data
- Significant experience writing custom **UDF's** in **Hive** and custom **Input Formats** in **MapReduce**.
- Knowledge of job workflow scheduling and monitoring tools like **Oozie**.
- Strong experience **productionalizing** end to end data pipelines on hadoop platform.
- Expertise in Database Design, Creation and Management of Schemas, writing Stored Procedures, Functions, DDL and DML SQL queries and writing complex queries for Oracle
- Experience working with NoSQL database technologies, including MongoDB, Cassandra and HBase.
- Good experience is designing and implementing end to end Data Security and Governance within Hadoop Platform using **Kerberos**.
- Strong experience with **UNIX shell scripts** and commands.
- Experience in using various Hadoop Distributions like **Cloudera, Hortonworks and Amazon EMR.**
- Strong hands-on development experience with **Java, J2EE (Servlets, JSP, Java Beans, EJB, JDBC, JMS, Web Services)** and related technologies.
- Work with the **team** to help understand **requirements**, evaluate new features, architecture and help **drive decisions.**
- Excellent interpersonal, communication, problem solving and analytical skills with ability to make independent decisions
- Experience successfully delivering applications using **agile** methodologies including extreme programming, **SCRUM** and Test-Driven Development (**TDD**).
- Experience in **Object Oriented Analysis, Design, and Programming** of distributed web-based applications.

- Extensive experience in developing standalone **multithreaded** applications.

- Configured and developed web applications in **Spring** and employed spring **MVC** architecture and **Inversion of Control.**
- Experience in building, deploying and integrating applications in Application Servers with **ANT**, **Maven** and **Gradle.**
- Significant application development experience with **REST Web Services, SOAP,** WSDL**, and **XML**.

TECHNICAL SKILLS:

Big Data Technologies	Hadoop, HDFS, Map Reduce, Hive, Pig, HBase, Impala, Sqoop, Flume, NoSQL (HBase, Cassandra), Spark, Kafka, Zookeeper, Oozie, Hue, Cloudera Manager, Amazon AWS, Hortonwork clusters
Java/J2EE & Web Technologies	J2EE, JMS, JSF, JDBC, Servlets, HTML, CSS, XML, XHTML, AJAX, Angular JS, JavaScript
Languages	C, C++, Core Java, Shell Scripting, PL/SQL, Python, Pig Latin
Scripting Languages	JavaScript and UNIX Shell Scripting, Python
Operating systems	Windows, Linux and Unix
DBMS / RDBMS	Oracle, Talend ETL, Microsoft SQL Server 2012/2008, MySQL, DB2 and NoSQL, Teradata SQL, MongoDB, Cassandra, HBase
IDE and Build Tools	Eclipse, NetBeans, MS Visual Studio, Ant, Maven, JIRA, Confluence
Version Control	Git, SVN, CVS
Web Services	RESTful, SOAP
Web Servers	Web Logic, Web Sphere, Apache Tomcat,

PROFESSIONAL EXPERIENCE:

Client	:	**TMNAS**	**Sep 2016 – Present**
Location	:	**Bala Cynwyd, PA**	
Role	:	**Sr. Hadoop Developer**	

Project Description: *TMNA offers the security of nearby expertise, enhanced by the diversity and power of one of the world's most respected insurance groups. Tokio Marine's companies offer access to leading commercial insurance solutions spanning the property and casualty landscape including professional liability, workers' compensation and property coverage.. The project deals with analyzing clickstream data of users who are visiting the company websites and applications to derive useful insights that help in optimizing future promotions and advertising*.

Responsibilities:

- Involved in story-driven **agile** development methodology and actively participated in daily scrum meetings.
- Ingested terabytes of click stream data from external systems like FTP Servers and S3 buckets into HDFS using custom **Input Adaptors**.
- Implemented end-to-end pipelines for performing user behavioral analytics to identify user-browsing patterns and provide rich experience and personalization to the visitors.
- Developed **Kafka producers** for streaming real-time clickstream events from external Rest services into topics.
- Used HDFS File System API to connect to FTP Server and HDFS. S3 AWS SDK for connecting to S3 buckets.
- Written Scala based **Spark applications** for performing various data transformations, denormalization, and other custom processing.
- Implemented data pipeline using **Spark, Hive, Sqoop and Kafka** to ingest customer behavioral data into Hadoop platform to perform user behavioral analytics.
- Created a multi-threaded Java application running on edge node for pulling the raw clickstream data from **FTP servers** and **AWS S3** buckets.
- Developed **Spark streaming** jobs using Scala for real time processing.
- Involved in creating external **Hive tables** from the files stored in the HDFS.
- Optimized the Hive tables utilizing improvement techniques like partitions and bucketing to give better execution Hive QL queries.
- Used **Spark-SQL** to read data from hive tables, and perform various transformations like changing date format and breaking complex columns.
- Wrote spark application to load the transformed data back into the Hive tables using **parquet** format.
- Used **Oozie** Scheduler system to automate the pipeline workflow to exact data on a timely manner.
- Implemented installation and configuration of multi-node cluster on the cloud using **Amazon Web Services (AWS)** on **EC2**.
- Worked on data visualization and analytics with research scientist and business stake holders.

Environment: Hadoop 2.x, Spark, Scala, Hive, Pig, Sqoop, Oozie, Kafka, Cloudera Manager, Storm, ZooKeeper, HBase, Impala, YARN, Cassandra, JIRA, MySQL, Kerberos, Amazon AWS, Shell Scripting, SBT, Git, Maven.

Client	:	Davita Inc	Jan 2015 - Sep 2016
Location	:	Nashville, Tennessee	
Role	:	Sr.Hadoop Developer	

Project Description: *Davita is one of the largest kidney dialysis companies in world. The idea of the project is to ingest data from different multiple sources to Hadoop Data Lake, perform transformations on it according to business requirements and exporting the data to external systems. The system is a scalable BI platform that can adapt to the speed of the business by providing relevant, accessible, timely, connected, and accurate data.*

Responsibilities:
- Involved in gathering and analyzing business requirements and designing Data Lake as per the requirements.
- Built distributed, scalable, and reliable data pipelines that ingest and process data at scale using **Hive** and **MapReduce**.
- Developed **MapReduce** jobs in **Java** for cleansing the data and preprocessing.
- Loaded transactional data from Teradata using **Sqoop** and create **Hive Tables**.
- Extensively used **Sqoop** for efficiently transferring bulk data between HDFS and relational databases.
- Worked on automation of delta feeds from **Teradata** using **Sqoop** and from **FTP Servers to Hive**.
- Worked on various performance optimizations like using distributed cache for small datasets, **Partition**, **Bucketing** in hive and Map Side joins.
- Created components like **Hive UDFs** for missing functionality in HIVE for analytics.
- Used **IMPALA** to analyze the data present in Hive tables.
- Handled **Avro** and **JSON** data in Hive using **Hive SerDe.**
- Worked with different **compression codecs** like **GZIP, SNAPPY and BZIP2** in MapReduce, Pig and Hive for better performance.
- Analyzed the data by performing the Hive queries using **Hive QL** to study the customer behavior.
- Wrote **python** scripts to parse XML documents and load the data in database.
- Generate auto mails by using **Python** scripts.
- Implemented the recurring workflows using **Oozie** to automate the scheduling flow.
- Worked with application teams to install OS level updates and version upgrades for Hadoop cluster environments.
- Participated in design and code reviews.

Environment: HDFS, Hadoop, Pig, Hive, HBase, Sqoop, Talend, Flume, Map Reduce, Podium Data, Oozie, Java 6/7, Oracle 10g, YARN, UNIX Shell Scripting, SOAP, REST services, Oracle 10g, Maven, Agile Methodology, JIRA.

Client	:	**NASBA**	**Aug 2012 - Dec 2014**
Location	:	**Nashville, TN**	
Role	:	**Hadoop Developer**	

Project Description: *National Association of State Boards of Accountancy enhances the effectiveness and advance the common interests of the Boards of Accountancy. Existing ETL platform is overloaded with data coming from variety of sources and as data is growing day by day, it is not able to perform well and cost of managing the Relational database servers are going up. So, the data is migrated from multiple sources to Hadoop Data Lake and transformations are performed on it according to business requirements and the processed data is exported to external systems.*

Responsibilities:
- Analysed business requirements and created/updated Software Requirements and design documents

- Imported the data from relational databases to **Hadoop** cluster by using **Sqoop**.
- Provided batch processing solution to certain unstructured and large volume of data by using **Hadoop Map Reduce** framework.

- Developed data pipelines using **Hive scripts** to **transform** data from **Teradata**, **DB2** data sources. These pipelines had customized **UDF'S** to extend the ETL functionality.
- Developed UDF for converting data from Hive table to JSON format as per client requirement.
- Involved in creating tables in **Hive** and writing scripts and queries to load data into **Hive** tables from HDFS.
- Implemented dynamic **partitioning** and **Bucketing** in Hive as part of performance tuning.
- Created custom **UDF'**s in **Pig** and **Hive**.

- Performed various **transformations** on data like changing date patterns, converting to other time zones etc.

- Designed and developed **PIG Latin** Scripts to process data in a batch to perform trend analysis.
- Automated Sqoop, hive and pig jobs using **Oozie** scheduling.
- Storing, processing and analyzing huge data-set for getting valuable insights from them.
- Created various aggregated **datasets** for easy and faster reporting using **Tableau**.

Environment: HDFS, Map Reduce, Hive, Sqoop, Pig, HBase, Oozie, CDH distribution, Java, Eclipse, Shell Scripts, Tableau, Windows, Linux.

Client	:	Copart Inc	Oct 2010 - Aug 2012
Location	:	Dallas, TX	
Role	:	Java Developer	

Project Description: *Copart makes it easy for Members to find, bid and win the vehicles that they are looking for. Members can choose from* classics*, early and late model* cars *and* trucks*,* industrial vehicles *and more. These internal applications are used for content management through which users can request a new site, which is used to store their project related documents.*

Responsibilities:
- Developed the J2EE application based on the Service Oriented Architecture by employing SOAP and other tools for data exchanges and updates.
- Worked in all the modules of the application which involved **front-end presentation logic** - developed using **Spring MVC, JSP, JSTL** and **JavaScript**, Business objects - developed using POJOs and data access layer - using Hibernate framework.
- Designed the **GUI** of the application using **JavaScript**, **HTML**, CSS, **Servlets**, and **JSP**.
- Involved in writing **AJAX** scripts for the requests to process quickly.
- Used **Dependency Injection** feature and AOP features of **Spring framework** to handle exceptions.
- Involved in writing Hibernate Query Language (HQL) for persistence layer.
- Implemented persistence layer using **Hibernate** that uses the **POJOs** to represent the persistence database.
- Used **JDBC** to connect to backend databases, Oracle and SQL Server 2005.

- Proficient in writing **SQL queries**, stored **procedures** for multiple databases, **Oracle** and SQL Server.
- Wrote backend jobs based on Core Java & Oracle Data Base to be run daily/weekly.
- Used **Restful API** and **SOAP** web services for internal and external consumption.
- Used **Core Java** concepts like Collections, Garbage Collection, **Multithreading**, OOPs concepts and APIs to do encryption and compression of incoming request to provide security.
- Written and implemented test scripts to support **Test driven development (TDD)** and continuous integration.

Environment: Java, JSP, HTML, CSS, Ubuntu Operating System, JavaScript, AJAX, Servlets, Struts, Hibernate, EJB (Session Beans), Log4J, WebSphere, JNDI, Oracle, Windows XP, LINUX, ANT, Eclipse.

Client	:	Aricent	Nov 2008 - Sep 2010
Location	:	Hyderabad, India	
Role	:	Java Developer	

Project Description: *Aricent is a global design and engineering company innovating in the digital era. help the world's leading companies solve their most important business and technology innovation challenges - from Customer to Chip. I have worked on developing the Aricent internal applications to automate the business process, store the documents. These internal applications are used for content management through which users can request a new site, which is used to store their project related documents.*

Responsibilities:
- Analyzed user requirements and created Software Requirements and design documents
- Responsible for GUI development using **Java**, **JSP**, **Struts**
- Database design and development
- Created and modified existing database scripts, Tables, Stored Procedures, and Triggers
- Used **XML functions**, Cursors, Mail and Utility packages for Advanced search functionality
- Created data correction and manipulation scripts for Production
- Used **JAXB** for marshalling and un-marshalling of the data
- Created **JUnit** tests for the service layer
- Support for Production issues
- Attending the review meetings for scheduling, implementation and resolving issues in software development cycle

Environment: Java, Struts, Java, Jsp, Servlets, JQuery, Ajax, XML, XSLT, JAXB, FOP, JBoss, Weblogic, Tomcat, SQL server 2005 and MyEclipse

EDUCATION:
- Bachelor of Technology in Computer Science from National Institute of Technology, Rourkela.