**Manoj Kumar**

**949-619-7484**

**n.manojkumar0819@gmail.com**

**PROFESSIONAL SUMMARY:**

* 8+ years of experience in IT industry with extensive experience in Java, J2ee and Big data technologies.
* 4 +years working of exclusive experience on Big Data technologies and Hadoop stack
* Strong experience working with HDFS, MapReduce, Spark. Hive, Pig, Sqoop, Flume, Kafka, Oozie and HBase.
* Good understanding of distributed systems, HDFS architecture, internal working details of MapReduce and Spark processing frameworks.
* More than two years of hands on experience using Spark framework with Scala.
* Good exposure to performance tuning hive queries, map-reduce jobs, spark jobs.
* Expertise in Inbound and Outbound (importing/exporting) data form/to traditional RDBMS using SQOOP.
* Tuned PIG and HIVE scripts by understanding the joins, group and aggregation between them.
* Extensively worked on HiveQL, join operations, writing custom UDF's and having good experience in optimizing Hive Queries.
* Worked on various Hadoop Distributions (Cloudera, Hortonworks, Amazon AWS) to implement and make use of those.
* Participated in design, development and system migration of high performance metadata driven data pipeline with Kafka and Hive/Presto on Qubole, providing data export capability through API and UI.
* Experience data processing like collecting, aggregating, moving from various sources using Apache Flume and Kafka.
* Hands on experience in NOSQL databases and SQL databases.

**TECHNICAL SKILLS:**

**Big Data Ecosystem:** Hadoop, Spark, Scala, MapReduce, HDFS, Hive, Pig, Sqoop. Flume, Kafka , HBase

**Java Technologies:** JSP, Servlets, Junit, Spring Hibernate

**Database Technologies:** MySQL, SQL server , Oracle, MS Access

**Programming Languages:** Scala, Python, Java and Linux shell scripting

**Operating Systems:** Windows, LINUX

**PROFESSIONAL EXPERIENCE:**

**Big Data Engineer July 2016 to Present**

**Knipper, Lakewood, NJ**

**Responsibilities:**

* Involved in requirements gathering and building data lake on top of hdfs.
* Worked on Go-cd (ci/cd tool) to deploy application and have experience with Munin frame work for Bigdata Testing.
* Involved in writing udfs in hive.
* Worked extensively on AWS Components such as Airflow, Elastic Map Reduce(EMR), Athena, Snow-Flake.
* Developed SQOOP scripts to migrate data from Oracle to Big data Environment.
* Extensively worked with Avro and Parquet files and converted the data from either format Parsed Semi Structured JSON data and converted to Parquet using Data Frames in Spark.
* Developed a Python Script to load the CSV files into the S3 buckets and created AWS S3 buckets, performed folder management in each bucket, managed logs and objects within each bucket.
* Created Hive DDL on Parquet and Avro data files residing in both HDFS and S3 bucket
* Created Airflow Scheduling scripts in Python to automate the process of sqooping wide range of data sets.
* Involved in file movements between HDFS and AWS S3 and extensively worked with S3 bucket in AWS
* Created data partitions on large data sets in S3 and DDL on partitioned data.
* Converted all Hadoop jobs to run in EMR by configuring the cluster according to the data size.
* Extensively used Stash Git-Bucket for Code Control
* Monitor and Troubleshoot Hadoop jobs using Yarn Resource Manager and EMR job logs using Genie and kibana.
* Created data pipeline for different events of ingestion, aggregation and load consumer response data in AWS S3 bucket into Hive external tables in HDFS location to serve as feed for tableau dashboards.
* Worked with different file formats like Json, AVRO and parquet and compression techniques like snappy.
* Developed python code for different tasks, dependencies, SLA watcher and time sensor for each job for workflow management and automation using Airflow tool.
* Developed shell scripts for dynamic partitions adding to hive stage table, verifying Json schema change of source files, and verifying duplicate files in source location.
* Converted Hive queries into Spark transformations using Spark RDDs.
* Exploring with the Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frame, Pair RDD's, Spark YARN.
* Worked with importing metadata into Hive using Python and migrated existing tables and applications to work on AWS cloud(S3).
* Worked extensively with importing metadata into Hive and migrated existing tables and applications to work on Hive and AWS cloud and making the data available in Athena and Snowflake.
* Imported the data from different sources like AWS S3, Local file system into Spark RDD.   
     
  **Environment:** Spark, AWS, EC2 ,EMR, Hive, SQL Workbench, GenieLogs, Kibana, Sqoop ,Spark SQL, Spark Streaming, Scala, Python

**Big Data Developer January 2015 to June 2016**

**Computech Corporation, Detroit, MI**

**Responsibilities:**

* Integrated Kafka with Spark Streaming for real time data processing
* Experience in writing Spark applications for Data validation, cleansing, transformations and custom aggregations.
* Imported data from different sources into Spark RDD for processing.
* Developed custom aggregate functions using Spark SQL and performed interactive querying.
* Worked on installing cluster, commissioning & decommissioning of Data node, Name node high availability, capacity planning, and slots configuration.
* Developed Spark applications for the entire batch processing by using Scala.
* Automatically scale-up the EMR instances based on the data.
* Run and Schedule the Spark script in EMR pipes.
* Utilized spark data frame and spark sql extensively for all the processing
* Experience in managing and reviewing Hadoop log files.
* Experience in hive partitioning, bucketing and perform joins on hive tables and utilizing hive SerDes like REGEX, JSON and AVRO.
* Exported the analyzed data to the relational databases using Sqoop and to generate reports for the BI team.
* Executed tasks for upgrading cluster on the staging platform before doing it on production cluster.
* Perform maintenance, monitoring, deployments, and upgrades across infrastructure that supports all our Hadoop clusters.
* Installed and configured various components of Hadoop ecosystem.
* Optimized HIVE analytics SQL queries, created tables/views, written custom UDFs and Hive based exception processing.
* Involved in transforming the relational database to legacy labels to HDFS, and HBASE tables using Sqoop and vice versa.
* Replaced default Derby metadata storage system for Hive with MySQL system.
* Supported in setting up QA environment and updating configurations for implementing scripts with Pig.
* Configured Fair Scheduler to provide fair resources to all the applications across the cluster.   
     
  **Environment:** Hadoop (cloudera stack), Hue, Spark, Kafka, HBase, HDFS, Hive, Pig, Sqoop, Oracle

**Hadoop Developer June 2012 to November 2014  
Nationwide Mutual Insurance company, Columbus, OH**

**Responsibilities:**

* Experience on AWS-EMR, Spark installation, HDFS and MapReduce Architecture.
* Participated in Hadoop Deployment and infrastructure scaling.
* Involved in creating Hive tables, and loading and analyzing data using hive queries.
* Developed Simple to complex Map Reduce Jobs using Hive and Pig.
* Developed workflow in Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig.
* Parsed high-level design spec to simple ETL coding and mapping standards.
* Maintained warehouse metadata, naming standards and warehouse standards for future application development.
* Worked with Linux systems and RDBMS database on a regular basis to ingest data using Sqoop.
* Implemented Kafka consumers to move data from Kafka partitions into Cassandra for near real-time analysis.
* Involved in Hadoop cluster task like adding and removing nodes.
* Managed and reviewed Hadoop log files and loaded log data into HDFS using Sqoop.
* Created and maintained Technical documentation for launching HADOOP Clusters and for executing Hive queries, Pig Scripts, Sqoop jobs.   
     
  **Environment:**  Hadoop(Hortonworks stack), HDFS, Oozie, Pig, Hive, MapReduce, Sqoop, Cassandra, Linux.

**Hadoop Developer November 2010 to May 2012**

**US Bank, Denver, CO**

**Responsibilities:**

* Worked on analyzing, writing Hadoop MapReduce jobs using JavaAPI, Pig and Hive.
* Responsible for building scalable distributed data solutions using Hadoop.
* Involved in loading data from edge node to HDFS using shell scripting.
* Created HBase tables to store variable data formats of PII data coming from different portfolios.
* Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the BI team.
* Worked with using different kind of compression techniques to save data and optimize data transfer over network using LZO, Snappy, and Bzip etc.
* Analyze large and critical datasets using Cloudera, HDFS, HBase, MapReduce, Hive, HiveUDF, Pig, Sqoop, Zookeeper, &Spark.
* Developed custom aggregate functions using Spark-SQL and performed interactive querying.
* Used Scoop to store the data into HBase and Hive.
* Worked on installing cluster, commissioning & decommissioning of DataNode, NameNode high availability, capacity planning, and slots configuration.
* Creating Hive tables, dynamic partitions, buckets for sampling, and working on them using HiveQL.
* Used Pig to parse the data and Store in Avro format.
* Stored the data in tabular formats using Hive tables and Hive Serdes.
* Collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis.
* Worked with NoSQL databases like HBase for creating HBase tables to load large sets of semi structured data coming from various sources.
* Implemented a script to transmit information from Oracle to HBase using Sqoop.
* Implemented MapReduce programs to handle semi/unstructured data like XML, JSON, and sequence files for log files.
* Fine-tuned Pig queries for better performance.
* Involved in writing the shell scripts for exporting log files to Hadoop cluster through automated process.
* Installed Oozie workflow engine to run multiple Hive and pig jobs.
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.   
     
  **Environment:** Hadoop, MapReduce, HDFS, Yarn, Sqoop, Oozie, Pig, Hive, HBase,Java, Eclipse, UNIX shell scripting, python, Horton works.

**Java Developer January 2010 to October 2010**

**GE Healthcare, Richmond, TX**

**Responsibilities:**

* Effectively interacted with team members and business users for requirements gathering.
* Involved in analysis, design, and implementation phases of the software development lifecycle (SDLC).
* Implementation of spring core J2EE patterns like MVC, Dependency Injection (DI), and Inversion of Control (IOC).
* Implemented REST Web Services with Jersey API to deal with customer requests.
* Developed test cases using J Unit and used Log4j as the logging framework.
* Worked with HQL and Criteria API from retrieving the data elements from database.
* Developed user interface using HTML, Spring Tags, JavaScript, JQuery, and CSS.
* Developed the application using Eclipse IDE and worked under Agile Environment.
* Design and implementation of front end web pages using CSS, JSP, HTML, java Script Ajax and, Struts
* Utilized Eclipse IDE as improvement environment to plan, create and convey Spring segments on Web Logic

**Environment:** Java, J2EE, HTML, JavaScript, CSS, J Query, Spring 3.0, JNDI, Hibernate 3.0, Java Mail, Web Services, REST, Oracle 10g, J Unit, Log4j, Eclipse, Web logic 10.3.

**Java Developer June 2008 to November 2009**

**ICICI bank, Noida, India**

**Responsibilities:**

* Involved in various phases of Software Development Life Cycle (SDLC) such as requirements gathering, analysis, design and development.
* Involved in overall performance improvement by modifying third party open source tools like FCK Editor.
* Developed Controllers for request handling using spring framework.
* Involved in Command controllers, handler mappings and View Resolvers.
* Designed and developed application components and architectural proof of concepts using Java, EJB, JSF, Struts, and AJAX.
* Participated in Enterprise Integration experience web services
* Configured JMS, MQ, EJB and Hibernate on Web sphere and JBoss
* Focused on Declarative transaction management
* Developed XML files for mapping requests to controllers
* Extensively used Java Collection framework and Exception handling.

**Environment:** Core Java, XML, Servlets, Hibernate Criteria API, Web service, WSDL, XML,UML, EJB, Java script, JQuery, Hibernate, SQL, CVS, Agile, JUnit.