**Ajay Kumar**

**Professional Summary:**

* 8years of overall IT experience in various sectors, which includes hands on experience on Big Data Analytics, and Development.
* Having **good** experience in Bigdata related technologies like **Hadoop frameworks**, **Map Reduce**, **Hive**, **HBase**, **PIG**, **Sqoop**, **Spark**, **Kafka**, **Flume**, **ZooKeeper**, **Oozie**, and **Storm**.
* Excellent knowledge on Hadoop ecosystems such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node and Map Reduce programming paradigm.
* Experienced in writing complex MapReduce programs that work with different file formats like **Text**, **Sequence**, **Xml, JSON** and **Avro.**
* Having working experience on Cloudera Data Platform using VMware Player, Cent OS 6 Linux environment.
* Strong experience on Hadoop distributions like Cloudera, and HortonWorks.
* Good knowledge of No-SQL databases **Cassandra, MongoDB** and **HBase.**
* Expertise in Database Design, Creation and Management of Schemas, writing Stored Procedures, Functions, DDL, DML **SQL** queries.
* Worked on **HBase** to load and retrieve data for real time processing using.
* Very good experience of Partitions, Bucketing concepts in **Hive** and designed both Managed and External tables in Hive to optimize performance.
* Good working experience using **Sqoop** to import data into HDFS or Hive from RDBMS and exporting data back to HDFS or HIVE from RDBMS.
* Extending **HIVE** and **PIG** core functionality by using custom User Defined Function’s **(UDF)**, User Defined Table-Generating Functions **(UDTF)** and User Defined Aggregating Functions **(UDAF)** for Hive and Pig.
* Experience in converting SQL queries into Spark Transformations using Spark RDDs, Data Frames and Scala, and performed map-side joins on RDD's.
* Experienced in designing, built, and deploying a multitude applications utilizing almost all of the AWS stack (Including EC2, Route53, S3, RDS, DynamoDB, SQS, SNS,SWF, IAM and EMR), focusing on high-availability, fault tolerance, and auto-scaling.
* Hands on experience with Amazon web services, created EC2 (Elastic Compute Cloud) cluster instances, setup data buckets on S3 (Simple Storage Service), set EMR (Elastic MapReduce) with Hive scripts to process big data.
* Created a multi terabyte database in multiload
* Have implemented terabyte parallel transporter(tpt) in data ware house
* Worked with Big Data distributions like Hortonworks (Hortonworks 1.7 and 2.1) with Ambari.
* Worked with Big Data distributions like Cloudera (CDH 3 and 4) with Cloudera Manager.
* Worked in ETL tools like **Talend, Pentaho** to simplify Map Reduce jobs from the front end. Also used to develop reports based on data in Hive.
* Worked with BI tools like Tableau for report creation and further analysis from the front end.
* Connecting Hive using Tableau and generating Bar chart etc based on business requirement.
* Extensive knowledge in using SQL queries for backend database analysis.
* Involved in unit testing of **Map Reduce** programs using Apache **MRunit**.
* Worked on Amazon Web Services and EC2
* Excellent Java development skills using J2EE, J2SE, Servlets, JSP, Spring, Hibernate, JDBC.
* Experience in creating Reusable Transformations (Joiner, Sorter, Aggregator, Expression, Lookup, Router, Filter, Update Strategy, Sequence Generator, Normalizer and Rank) and Mappings using Informatica Designer and processing tasks using Workflow Manager to move data from multiple sources into targets.
* Doing query performance to increase speed.
* Experience in database development using SQL and PL/SQL and experience working on databases like Oracle 9i/10g/11g, Informix, and SQL Server.
* Experience working with build tools like **Maven**, **Ant, SBT**.
* Experienced in both **Waterfall** and **Agile Development** (**SCRUM**) methodologies
* Strong Problem Solving and Analytical skills and abilities to make Balanced & Independent Decisions.
* Experience in developing service components using **JDBC**.

**Technical Skill**

|  |  |
| --- | --- |
| **Hadoop Technologies** | HDFS, YARN, Mesos, MapReduce, Hive, Pig, Sqoop, Flume, Spark, Kafka, Zookeeper, and Oozie. |
| **NO SQL Databases** | HBase, Cassandra, |
| **Languages** | Java, Scala, SQL, PL/SQL, Pig Latin, HiveQL, Unix, Java Script, Shell Scripting, Python. |
| **Java & J2EE Technologies** | Core Java, Servlets, Hibernate, Spring, Struts. |
| **Application Servers** | Web Logic, Web Sphere, JBoss, Tomcat. |
| **Cloud Computing Tools** | Amazon AWS. |
| **Databases** | MySQL, Oracle, DB2 |
| **Operating Systems** | UNIX, Windows, LINUX. |
| **Build Tools** | Jenkins, Maven, ANT, SBT (for Scala). |
| **Business Intelligence Tools** | Tableau, Splunk |
| **Development Tools** | Microsoft SQL Studio, Toad, Eclipse, NetBeans, Intellij Idea. |
| **Development Methodologies** | Agile/Scrum, Waterfall. |

**Work Experience:**

**Client: Health First Jan 2016 to till date**

Role**: Spark, Scala developer**

**Location: Rockledge, FL**

**Description**:-.This Product is turning raw data into information and knowledge that allows clear understanding, enabling better insight and decisions. This is achieved by combining data from multiple disparate sources, straddling OT and IT systems. Seeing real-time and operational data presented on a map with connectivity between assets. Combining calculated KPI’s with site schematics and maps. Capturing and using business intelligence in the form of analytical models, driving them with real-time data. Carrying out what-if analysis to optimize choices in sanity measures based on cost, within supply limits while meeting demand. In short, making a step change in translating data into actionable information and knowledge in line with the Information.

**Responsibilities:-**

* Responsible for design development of Spark SQL Scripts based on Functional Specifications
* Responsible for Spark Streaming configuration based on type of Input Source
* Wrote the Map Reduce jobs to parse the web logs which are stored in HDFS.
* Developed the services to run the Map-Reduce jobs as per the requirement basis.
* Importing and exporting data into HDFS and HIVE using Sqoop.
* Responsible to manage data coming from different sources.
* Developing business logic using Scala.
* Responsible for loading data from UNIX file systems to HDFS. Installed and configured Hive and also written Hive UDFs.
* Experienced with AWS services to smoothly manage application in the cloud and creating or modifying instances.
* Checking the health and utilization of AWS resources using AWS CloudWatch.
* Provisioned AWS S3 buckets for backup of the application and sync these contents with remaining S3 backups by creating entry for AWS S3 SYNC in cron tab.
* Collected the Json data from HTTP Source and developed Spark APIs that helps to do inserts and updates in Hive tables.
* Involved the performance and optimization of the existing algorithms in Hadoop using Spark Context,   
  Spark-SQL, Data Frame, Pair RDD's, Spark YARN.
* Involved in creating Hive Tables, loading with data and writing Hive queries which will invoke and run MapReduce jobs in the backend.
* Writing MapReduce (Hadoop) programs to convert text files into AVRO and loading into Hive (Hadoop) tables
* Implemented the workflows using Apache Oozie framework to automate tasks.
* Worked with NoSQL databases like HBase in creating HBase tables to load large sets of semi structured data coming from various sources.
* Developing design documents considering all possible approaches and identifying best of them.
* Loading Data into HBase using Bulk Load and Non-bulk load.
* Developed scripts and automated data management from end to end and sync up b/w all the clusters.
* Exploring with the Spark improving the performance and optimization of the existing algorithms in Hadoop.
* Import the data from different sources like HDFS/HBase into Spark RDD.
* Experienced with Spark Context, Spark-SQL, Data Frame, Pair RDD's, Spark YARN.
* Import the data from different sources like HDFS/HBase into Spark RDD.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDD, Scala.
* Involved in gathering the requirements, designing, development and testing.
* Developing traits and case classes etc in Scala.
* Data storage n Hbase and connecting to elastic search and solr search.
* Developing reports using Tableau ETL.

**Environment**: Hive, Flume, Java, Maven, Impala, Spark, Oozie, Oracle, Yarn, GitHub, Junit, Tableau, Unix, Cloudera, Flume, Sqoop, HDFS, Tomcat, Java, Scala, Hbase.

**Client: Shelter Insurance Dec 2014- Jan 2016**

**Role: Hadoop, Spark developer**

**Location: Columbia, MO**

**Description:** The purpose of the project is to store terabytes of different departmental information generated by Shelter Insurance marketing team and extract meaning information out of it. The solution is based on the open source Big Data Hadoop .The data will be stored in Hadoop file system and processed using Map/Reduce jobs. Which intern includes getting the word, excel, PPT and PDF data from the client team, Process the data to obtain respective department related information meaningful information with respect to their domains, Extract various reports out of the processed information and Export the information for further processing. This project is mainly for migrating of the current existing system which is running presently in SQL DB to a new cloud solution technology called Hadoop which can able to process large data sets (i.e. Tera bytes and Peta bytes of data) in order to meet the client requirements with the increasing data volume from other departments.

**Responsibilities:-**

* Consumed the data from Kafka queue using spark.
* Configured different topologies for spark cluster and deployed them on regular basis.
* Load and transform large sets of structured, semi structured and unstructured data.
* Involved in loading data from LINUX file system to HDFS
* Importing and exporting data into HDFS and Hive using Sqoop
* Implemented Partitioning, Dynamic Partitions, Buckets in Hive.
* Configured various property files like core-site.xml, hdfs-site.xml, mapred-site.xml based upon the job requirement.
* Responsible for migrating the code base from Cloudera Platform to Amazon EMR and evaluated Amazon eco systems components like Redshift, Dynamo DB.
* POC on Data Search using Elastic Search
* Along with the Infrastructure team, involved in design and developed Kafka and Storm based data pipeline. This pipeline is also involved in Amazon Web Services EMR, S3 and RDS.
* Launched instances with respect to specific applications.
* Developed and managed cloud VMs with AWS EC2 command line client’s and management console.
* Involved in performing the Linear Regression using Scala API and Spark.
* Installed and configured MapReduce, HIVE and the HDFS; implemented CDH5 Hadoop cluster on CentOS. Assisted with performance tuning, monitoring and troubleshooting.
* Created Map Reduce programs for some refined queries on big data.
* Writing logic using Python
* Involved in setting up of HBase to use HDFS.
* Designed and developed Map Reduce jobs to process data coming in different file formats like XML, CSV, and JSON.
* Developed multiple MapReduce Jobs in java for data cleaning and pre-processing
* Developing scripts to perform business transformations on the data using Hive and PIG
* Used Hive partitioning and bucketing for performance optimization of the hive tables and created around 20000 partitions.
* Created RDD's in Spark technology and extracted data from data warehouse on to the Spark RDD's
* Used Spark with Scala.

**Environment**: Spark, Spark SQL, Spark Streaming, Red Shift, CDH5, Scala, Javascript, Cassandra.

Client: Tangoe Inc Oct 2013 – Nov 2014

Role: Hadoop Developer

Location: Austin, Texas

Description:-Tangoe Inc provides internet protocol, data, voice, and wireless communication solutions. The Company offers data and IP services, internet access, security, information technology solutions and hosting, managed networks, premises equipment, contact centers, conferencing, and voice products. This Project is to migrate the data from existing Data warehouse systems like SAP, PeopleSoft & Oracle ECC to Hadoop and Implement the existing use cases in SAP using Sqoop, Spark Streaming, Hive, Spark, Spark SQL. Create high-security framework to prevent any misuse of the data using HDFS Encryption, Ranger, and Kerberos

**Responsibilities:--**

* Worked on analyzing Hadoop cluster using different big data analytic tools including Pig, Hive and MapReduce.
* Involved in exploring Hadoop, Map Reduce programming and its ecosystems.
* Implementing Map Reduce programs / Algorithms for Organizing the data, For performing Aggregation over the data, Joining different data sets, Filtering the data, Classification, Partitioning.
* Created a multi terabyte database in multiload
* Have implemented terabyte parallel transporter(tpt) in data ware house
* Importing and exporting data into HDFS and Hive using Sqoop.
* Writing UDF (User Defined Functions) in Pig, Hive when needed.
* Developing the Pig scripts for processing data.
* Managing work flow and scheduling for complex map reduce jobs using Apache Oozie.
* Involved in creating Hive tables, loading data &writing hive queries.
* Written Hive queries for data analysis to meet the business requirements.
* Created HBase tables to store various data formats of incoming data from different portfolios.
* Created Pig Latin scripts to sort, group, join and filter the enterprise wise data.
* Automated the History and Purge Process.
* Importing and exporting data into HDFS and Hive using Sqoop
* Implemented Partitioning, Dynamic Partitions, Buckets in Hive.
* Load and transform large sets of structured, semi structured and unstructured data.
* Validating the data using MD5 algorithms.
* Experience in Daily production support to monitor and trouble shoots Hadoop/Hive jobs
* Involved in Configuring core-site.xml and mapred-site.xml according to the multi node cluster environment.
* Implemented Data Integrity and Data Quality checks in Hadoop using Hive and Linux scripts.
* Used AVRO, Parquet file formats for serialization of data.

**Environment:** Solix EDMS, Java 1.6, Hadoop (Map/Reduce, Hive, PIG, Oozie, Sqoop, Flume), Eclipse Helios, Linux/Unix, Talend, Camel.

**Client: HDFC Bank Aug 2011- July 2013**

**Role: Sr. Java Developer**

**Location: Bangalore, India**

Description:-HDFC Bank is a private bank and financial services firm. It follows MVC architecture and connects to Oracle Enterprise database in the backend over a secured layer which handles user interactions. The application allows users to perform online banking and update information like transaction id, Billing / Invoice addresses. It is mainly used as Point Of Sales (POS) application. Each user can be associated with multiple customers and handle return requests too

**Responsibilities:-**

* Developed Web module using Spring MVC, JSP.
* Developing model logic by using Hibernate ORM framework.
* Handle server side validations.
* Involved in Bug fixing.
* Involved in Unit Testing by using Junit.
* Writing Technical Design Document.
* Gathered specifications from the requirements.
* Developed the application using Spring MVC architecture.
* Developed JSP custom tags which support custom User Interfaces.
* Developed front-end pages using JSP, HTML and CSS
* Developed core Java classes for utility classes, business logic, and test cases
* Developed SQL queries using MySQL and established connectivity
* Used Stored Procedures for performing different database operations
* Autowire Java objects using Spring Dependency Injection
* Used Log4j for application logging and debugging
* Spring framework was used in developing the application which uses Model View Controller(MVC) architecture.
* Used Hibernate for interacting with Database
* Developed control classes for processing the request
* Used Exception Handling for handling exceptions
* Prepare Junit test cases for test driven development
* Designed sequence diagrams and use case diagrams for proper implementation
* Used Rational Rose for design and implementation

**Environment**: - JSP, HTML, CSS, JavaScript, MySQL, Spring, Hibernate, MySQL Exception Handling, UML, Rational Rose.

**Client: Flipkart June 2009- July 2011**

**Location: Hyderabad, India**

**Role: Java Developer**

Description:--Flipkart is online shopping cart application used to shop Products. This project mainly useful for Technical people in solving problems and Quality Assessment. The Organization has different classes of employees. Here we are dealing with ESO (Electronic Service Officer), QA (Quality Agent) and Manager only. ESO mainly deals directly with the clients through Chat or E-Mail. QA check the quality of the ESO replied messages on daily basis and gives rating on their performance. Quality Agent selects the files randomly from files created by ESO, manually and verifies them for errors. After verification, QA rates ESO according to their performance like politeness and accuracy of the reply. On demand of Manager QA has to submit their verified reports.

Responsibilities:-

* Involved in Requirements analysis, design, and development and testing
* Developed Custom tags, JSTL to support custom User Interfaces.
* Designed the user interfaces using JSP.
* Designed and Implemented MVC architecture using Struts Framework, Coding involves writing Action Classes/Custom Tag Libraries, JSP.
* Developed Action Forms and Controllers in Struts 1.2 framework. Utilized various Struts features like Tiles, tagged libraries and Declarative Exception Handling via XML for the design.
* Involved in writing and business layer using EJB, BO, DAO and VO.
* Implemented Business processes such as user authentication, Account Transfer using Session EJBs.
* Worked with Oracle Database to create tables, procedures, functions and select statements.
* Used Log4J to capture the log that includes runtime exceptions and developed WAR framework to alert the client and production support in case of application failures.
* Developed the Dao's using SQL and Data Source Object.
* Deployed the application in client's location on Tomcat Server.
* Development carried out under Eclipse Integrated Development Environment (IDE).
* Used JBoss for deploying various components of application
* Used Ant for building Scripts.
* Used JUNIT for testing and check API performance.

**Environment :-**Java1.6, J2EE, Struts, HTML, CSS, JavaScript, Jdbc SQL 2005, ANT, Log4j, JUnit, XML, JSP, JSTL, AJAX, JBoss, ClearCase.

**EDUCATION:**

* Bachelor in Engineering, Osmania University, India (2009).