**Kumar Yuvarajulu**

**(213) 536-9936**

[**kumar.yuvaraj0329@gmail.com**](mailto:kumar.yuvaraj0329@gmail.com)

**Summary:**

* **7+ years of IT experience** in Architecture, Analysis, design, development, Testing, implementation, maintenance and support with experience in developing strategic methods for deployingBIG DATA technologies to efficiently solve Big Data processing requirement.
* 4 years of Experience on **BIG DATA** using **HADOOP** framework and related technologies such as **HDFS, HBASE, Map Reduce, HIVE, PIG, FLUME, OOZIE, POSTGRES, SQOOP, TALEND, IMPALA** and **ZOOKEEPER**.
* Around 2 years of experience on apache **SPARK STORM** and **KAFKA**.
* Experience in data analysis using **HIVE, PIG LATIN, HBASE** and custom Map Reduce programs in **JAVA**.
* Good Experience with Cloudera and Horton works distributions working with **machine learning frameworks** such as Tensor flow, **MLLib,** scikit-learn.
* Experience in working with **FLUME, SHELL SCRIPTING** to load the log data from multiple sources directly into **HDFS**.
* Worked on data load from various sources i.e., **Oracle, MySQL, DB2, MS SQL Server, Cassandra, Nifi, MongoDB, Hadoop** using **Sqoop** and **PYTHON SCRIPT**.
* Excellent understanding /knowledge on Hadoop (Gen-1 and Gen-2) and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, Resource Manager (YARN).
* Experienced with **SPARK SREAMING API** to ingest data into **SPARK ENGINE** from **KAFKA**.
* Developed analytical components using **SCALA, SPARK, STORM** and **SPARK STREAM**.
* Excellent experience on **Python** Scripting.
* Excellent understanding and knowledge of **NOSQL** database **HBASE** and **CASSANDRA**.
* Experience with **AWS, AZURE, EMR** and **S3**.
* Worked extensively with Dimensional **MODELING (OLAP), DATA MIGRATION, DATA CLEANSING, DATA PROFILING**, and **ETL** Processes features for data warehouses.
* Implemented Hadoop based data warehouses, **INTEGRATED HADOOP** with **ENTERPRISE DATA WAREHOUSE** systems. Extensive experience in ETL Testing**,** Data Ingestion, In-Stream data processing, **BATCH ANALYTICS** and Data **PERSISTENCE STRATEGY**.
* Experienced **in Testin**g in **HADOOP and HIVE.**
* Experience with creating the **TABLEAU** dashboards with relational and multi-dimensional databases including Oracle, MySQL and HIVE, gathering and manipulating data from various sources.
* Design and document **RESTFUL/HTTP**, SOAP APIs, including JSON data formats and API versioning strategy.
* Installed and configured **JENKINS FOR AUTOMATING** Deployments and providing automation solution.
* Experience in designing both time driven and data driven automated workflows using Oozie and **Ambari**.
* Experience working with **JDK 1.7, JAVA, J2EE, JDBC, ODBC, JSP, JAVA ECLIPSE, JAVA BEANS, EJB, SERVLETS, MS SQL SERVER**.
* Experience in J2EE technologies like **Struts, JSP/Servlets,** and **spring**.
* Good Exposure on scripting languages like **JAVASCRIPT, ANGULAR JS, JQUERY** and **XML**.
* Created a **JAVADOC TEMPLATE** for engineers to use to develop API documentation.
* Expert in **JAVA 1.8 LAMBDAS, STREAMS,** Type annotations.
* Experience in all stages of SDLC (Agile, Waterfall), writing Technical Design document, Development, Testing and Implementation of Enterprise level Data mart and Data warehouses.
* Extensive experience working IN **ORACLE, DB2, SQL SERVER** and **My SQL** database.
* Ability to work in high-pressure environments delivering to and managing stakeholder expectations
* Application of structured methods to: Project Scoping and Planning, risks, issues, schedules and deliverables.
* Strong analytical and Problem-solving skills.
* Good Inter personnel skills and ability to work as part of a team. Exceptional ability to learn and master new technologies and to deliver outputs in short deadlines.

**Education:**

* Electronics and Communications Engineering from Anna University, India.

**Technical Skills:**

|  |  |
| --- | --- |
| **Technology** | Hadoop Ecosystem/J2SE/J2EE/JDK1.7,1.8 / Data base. |
| **Operating Systems** | Windows, LINUX (Ubuntu, Cent OS), UNIX |
| **DBMS/Databases** | DB2, My SQL, PL/SQL |
| **Programming Languages** | C, C++, Core Java, XML, JSP/Servlets, Struts, Spring, HTML, JavaScript, jQuery, Web services, Xml |
| **Big Data Ecosystem** | HDFS, Map Reducing, Oozie, Hive, Pig, Sqoop, Flume, splunk, Zookeeper, spark, Python Kafka and Hbase. |
| **Methodologies** | Agile, Water Fall |
| **NOSQL Databases** | Hbase |
| **Version Control Tools** | SVN, CVS |
| **ETL Tools** | IBM data stage 8.1, Informatica |

**Professional Experience:**

**First Republic Bank, San Francisco, CA May 2020 to Till Date**

**Role: Big Data Developer**

**Responsibilities:**

* Developed data pipeline using **Spark, Hive, Pig, python, Impala and HBase** to ingest customer behavioral data and financial histories into Hadoop cluster for analysis.
* Responsible for implementing a generic framework to handle different data collection methodologies from the client primary data sources, validate transform using spark and load into S3
* Collected data using **Spark Streaming** from **AWS S3** bucket in near-real-time and performs necessary **Transformations** and Aggregation on the fly to build the common learner data model and persists the data in HDFS.
* Create dashboards on snowflake cost model, usage in **QlikView**.
* Created program in **python** to handle PL/SQL functions like cursors and loops which are not supported by snowflake
* Explored the usage of **Spark** for improving the performance and optimization of the existing algorithms in Hadoop using **Spark Context, Spark SQL** and **Spark Yarn**.
* Developed Spark Code using Scala and **Spark-SQL/Streaming** for faster testing and processing of data.
* Involved in converting **Hive/SQL** queries into Spark Transformations using **Spark RDDs** and **Scala**.
* Worked on the **Spark SQL** and Spark Streaming modules of **Spark** and used Scala and Python to write code for all Spark use cases.
* Exploring with the Spark improving the performance and optimization of the existing algorithms in Hadoop using **Spark-Context, Spark-SQL, Data Frame** and **Pair RDD's**.
* Migrated historical data to S3 and developed a reliable mechanism for processing the incremental updates.
* Used **Oozie** workflow engine to manage independent Hadoop jobs and to automate several types of Hadoop such as java **Map Reduce**, **Hive** and **Sqoop** as well as system specific jobs
* Create dashboards on snowflake cost model, usage in **QlikView**.
* Used to monitor and debug Hadoop jobs/applications running in production.
* Worked on providing user support and application support on Hadoop infrastructure.
* Worked on evaluating, comparing different tools for test data management with Hadoop
* Helping testing team **on Hadoop Application Testing**.

**Environment:** Spark, Hive, Pig, Spark SQL, Spark Streaming, HBase, Sqoop, Kafka, AWS EC2, S3, Cloudera, Scala IDE (Eclipse), Scala, Linux Shell Scripting, HDFS

**Exela Technologies, Irving, TX Jan 2019 to Apr 2020**

**Role: Sr Hadoop/Spark Developer**

**Responsibilities:**

* Developed data pipeline using **Spark, Hive, Pig, python, Impala and HBase** to ingest customer behavioral data and financial histories into Hadoop cluster for analysis.
* Collected data using **Spark Streaming** from **AWS S3** bucket in near-real-time and performs necessary **Transformations** and Aggregation on the fly to build the common learner data model and persists the data in HDFS.
* Involved in Data Warehouse concepts, testing methodologies.
* Configure Flume to ingest log file data into **HDFS**.
* Used **Pig** to do transformations, event joins and some **Pre-Aggregations** before storing the data onto HDFS.
* Involved in using Sqoop for importing and exporting data between **RDBMS** and **HDFS and Impala**.
* Used Hive to analyze the **Partitioned** and **Bucketed** data and compute various metrics for reporting.
* Performed hive performance tuning aspects like **Map join, cost based optimization** and **column level statistics**.
* Involve in working with **big data Testing** analysis and Hadoop and hive as well.
* Create notebooks to load xml files in Azure SQL datwarehouse using **Azure databricks**.
* Build pipelines in**Azure data factory**to move data from on prem to Azure SQL Datawarehouse, from Amazon S3 buckets to Azure blob storage.
* Involved in developing **Hive DDLS** to create, alter and drop Hive tables.
* Involved in loading data from **Linux file** system to HDFS.
* Installed and configured Hive and also written Hive **UDFs**.
* Involved in creating **Hive tables**, loading with data and writing **hive queries** which will run internally in map reduce way.
* Experience in working with **NOSQL database and Postgres** like **HBase.**

**Environment:** Spark, Hive, Pig, Spark SQL, Spark Streaming, HBase, Sqoop, Kafka, AWS EC2, S3, Cloudera, Scala IDE (Eclipse), Scala, Linux Shell Scripting, HDFS.

**Sutter Health, Sacramento, CA Mar 2018 to Dec 2018**

**Role: Hadoop Developer**

**Responsibilities:**

* Developed data pipeline using **Flume, Sqoop, PIG and Java Mapreduce** to ingest customer behavioral data and financial histories into HDFS for analysis.
* Involved in Sqoop, HDFS Put or Copy from Local to ingest data and Map Reduce jobs.
* Used **PIG** to do transformations, event joins, filter boot traffic and **Some Pre-Aggregations** before storing the data onto HDFS.
* Extensive experience in ETL Data Ingestion, In-Stream data processing, **Batch Analytics** and **Data Persistence Strategy**.
* Implemented Hadoop based data warehouses, **Integrated Hadoop** with **Enterprise Data Warehouse** systems. Working on **modeling process on OLAP.**
* Worked on Designing and Developing ETL Workflows using Java for processing data in HDFS/Hbase using Oozie.
* Involved in developing PIG UDFS for the needed functionality that is not out of the box available from Apache Pig.
* Expertise with the tools in Hadoop Ecosystem including **PIG, HIVE, HDFS, MAP REDUCE, SQOOP, KAFKA, YARN, OOZIE, AND ZOOKEEPER**. Hadoop architecture and its components.
* Involved in integration of Hadoop cluster with spark engine to perform **Batch** and **Graphx** operations.
* Developed **Kafka Producer** and consumers, HBase clients, SPARK and Hadoop Map Reduce jobs along with components on HDFS, Hive.
* Used Hive to analyze the **partitioned** and bucketed data and compute various metrics for reporting.
* Created action **filters, parameters** and calculated sets for preparing dashboards and worksheets in **Tableau**.
* Experience in Converting csv files to .tde files using TABLEAU extract API.
* Involved in developing **HIVE DDLS** to create, alter and drop Hive tables and storm.
* Create scalable and high-performance web services for data tracking.
* Involved in loading data from **UNIX file** system to HDFS. Installed and configured Hive and also written Hive UDFs and Cluster coordination services through Zoo Keeper.
* Involved in creating **Hive tables**, loading with **data and writing hive queries** which will run internally in map reduce way.
* Experienced in managing Hadoop Cluster using **Cloudera Manager Tool**.
* Involved in using **Hcatalog** to access Hive table metadata from Map Reduce or Pig code.
* Computed various metrics using Java Map Reduce to calculate metrics that define user experience.
* Responsible for developing data pipeline using **Flume, Sqoop, Postgres** and **PIG** to extract the data from weblogs and store in HDFS.
* Extracted and updated the data into **Monod Using Mongo** import and export command line utility interface.
* Involved in developing Shell scripts to orchestrate execution of all other scripts (Pig, Hive, and Map Reduce) and move the data files within and outside of HDFS.
* Involved in Hadoop testing, Developed Unit test cases using Junit, Easy Mock and MRUnit testing frameworks.

**Environment**: Map Reduce, Yarn, Hive, PIG, Hbase, Oozie, Sqoop, Splunk, Kafka, Oracle 11G, Core Java, Cloudera, Eclipse, SQL, Tableau, Teradata, Unix Shell Scripting.

**Neuros Technology, Chicago, IL Nov 2016 to Feb 2018**

**Role: Java Developer**

**Responsibilities:**

* Involved in requirement analysis, design, coding and implementation.
* Worked within **SOA** architecture with Mule ESB based enterprise framework to build web services.
* Involved in development of the application using Spring Design Pattern, AngularJS Framework.
* Provided development support for creating a website using **Java, Groovy & Grails, framework utilizing agile** methodology and **MySQL database**.
* Developed use **case diagrams**, sequence diagrams and class diagrams using IBM Rational Rose.
* Developed user interface using Struts framework, **HTML/JSP, JSP custom tag libraries** and user validations using **JavaScript**.
* Involved in developing **Restful web services** using Apache CXF tool as wrappers to access existing business services by Mobile channel.
* Extensive use of **Maven** to build and deploy the application into **dev** environment and QA environment and work with the front-end developers in displaying the data.
* Extensive use of Collection Framework features like **Map, Object, List to retrieve the data from Web Service,** manipulate the data to incorporate Business Logic and save the data to Oracle database.
* Maintained **CSS** and **HTML, XSL, XSLT, JavaScript, JSF, Angular JS Bootstrap** for designing Webpages.
* Responsible for managing scope, planning, tracking, change control, aspects of the CORE platform area of ecommerce applications.
* Developed data access objects to encapsulate all database specific code using **JDBC API.**
* Worked on implementation of persistence layer using **spring MongoDB** plug-ins.
* Developed highly productive dynamic web applications by **Groovy on Grails.**
* Involved in Production Support and Maintenance for Application developed in the **Red Hat Linux** Environment.
* Experience in continuous integrated build and testing and deployment using **Hudson**
* Used **Spring Framework** for developing the application and used **JDBC** to map to **Oracle database.**
* Consume **Web Services** using java to retrieve the required information to be populated in the database.
* Worked on Asset Management Module in order to develop services using **Restful Web services.**
* Use of **SOAPUI** to verify **the WSDL** end point URL.
* Extensive use of core java features like multithreading, caching, messaging to develop middleware for the application

**Environment**: Java API, Spring 3.0, Spring MVC, JDBC, Maven, SVN, Servlets, Struts, Amazon WS, RESTful Web Services, Bootstrap, MULE ESB, SOAP, HTML 5, CSS, CSS 3, IBM WebSphere, IBM RSA, Hudson, Rational Rose, Collections, JSP, PL/SQL, MongoDB, Drools, SOAP Web services.

**SOSIO Technologies, India Sep 2014 to Oct 2016**

**Role: Java Developer**

**Responsibilities:**

* Involved in various Software Development Life Cycle (SDLC) phases of the project.
* Developed the application using **Struts Framework**, which is based on Model View Controller design pattern.
* Extensively used **Hibernate** in data access layer to perform database operations.
* Used Spring Framework for Dependency Injection and integrated it with the **Struts Framework** and **Hibernate**
* Developed front end using Struts framework.
* Configured Struts Dyna Action Forms, **Message Resources**, Action Messages, Action Errors, Validation.xml, and Validator-rules.xml.
* Designed and developed front-end using struts framework. Used **JSP, JavaScript, JSTL, EL, Custom Tag libraries** and Validations provided by struts framework.
* Used Web services - WSDL and **SOAP** for getting credit card information from third party.
* Worked on advanced Hibernate associations with multiple levels of Caching, lazy loading.
* Designed various tables required for the project in Oracle 9i database and used Stored Procedures and Triggers in the application.
* Involved in consuming **RESTful Web** services to render the data to the front page.
* Performed unit testing using JUnit framework.
* Co-ordinate with **QA team** in manual and automation testing.
* Coordinated work with **DB team**, **QA team, Business Analysts** and Client Reps to complete the client requirements efficiently.

**Environment:** HTML, JSP, Servlets, JDBC, JavaScript, Tomcat, Eclipse IDE, XML, XSL, Tomcat 5.