**ANIL KUMAR**

**409-444-5487**

**AGUBBALA90826@GMAIL.COM**

**Professional Summary:**

* Around 8+ years of professional experience in various Software Development positions in core and enterprise software development using Big Data, Java/J2EE and Open Source technologies.
* Having 3+ years of hands-on Experience on Hadoop ecosystem components like Hadoop MapReduce, HDFS, HBase, Oozie, Hive, Sqoop, Pig, Zookeeper, Flume, NiFi and Kafka including their installation and configuration.
* Experience in AWS, Hortonworks and Cloudera Hadoop distributions.
* In depth knowledge of Hadoop architecture and various components such as HDFS, JobTracker, NameNode, DataNode, MapReduce and Yarn concepts.
* Experience in writing Hive Queries for analyzing data in Hive warehouse using Hive Query Language (HQL).
* Handling and further processing schema oriented and non-schema oriented data using Pig.
* Designed and developed SQOOP scripts for datasets transfer between Hadoop and RDBMS.
* Experience in extending Hive and Pig core functionality by writing custom UDFs.
* Hands on experience in extending the core functionalities of HIVE using UDF, UDAF and UDTF.
* Experience and good at Data modeling with Hive.
* Experience in using NIFI processor groups, processors and concepts on process flow management.
* Implemented Data Warehousing Methodologies for ETL using Informatica Designer, Repository Manager, Workflow Manager, Workflow Monitor, Repository Server Administration Console.
* Knowledge of job workflow scheduling and monitoring tools like Oozie and Zookeeper.
* Experience in using Flume to collect weblogs.
* Actively involved in successfully migration project without affecting Ab Initio applications data and processing with through testing.
* Developed MapReduce jobs to automate transfer the data from HBase.
* Handling different file formats on Parquet, Proto Buffer, Avro, Sequence file, JSON, XML and Flat file.
* Experience working on Kafka cluster. Also have experience in working on Spark and Spark streaming.
* Good Knowledge in creating event processing data pipelines using Kafka and Spark Streaming.
* Configured and maintained different topologies in Storm cluster and deployed them on regular basis.
* Experienced in Apache Spark for implementing advanced procedures like text analytics and processing using the in-memory computing capabilities written in Scala.
* Experience in using Hcatalog for Hive and Pig.
* Involved in the ETL process using Ab Initio tool to setup a data extraction from several databases.
* Loaded data in elastic search from datalake using SPARK/Hive.
* Experienced in NoSql databases such as Hbase, MongoDb and Cassandra.
* Experienced in developing Shell scripts and Python scripts for system management.
* Involved in data modeling and sharing and replication strategies in MongoDB.
* Experience in creating custom Lucene/Solr Query components.
* Utilized Kafka for loading streaming data and performed initial processing, real time analysis using Storm.
* Experience in developing distributed Web applications and Enterprise applications using Java/ J2EE technologies (Core Java (JDK 6+).
* Got experience in working on Scala with Spark.
* Excellent programming skills with experience in Java, C, SQL and Python Programming.

**Technical Expertise:**

|  |  |
| --- | --- |
| **Languages** | Core Java, Servlets, Hibernate, Spring, Struts, Scala, Python |
| **Enterprise Java** | JSP, JSTL, Servlets, JPA, JTA, EJB, JSF, REST and SOAP-Web Service, Jackson, JAX-B, JAX-RPC, JAX-WS, JAX-RS, Jersey, Apache AXIS, Apache CXF, Jersey, JMS. |
| **Hadoop Core Services** | HDFS, MapReduce, Spark, Yarn Hadoop Distribution Hortonworks, Cloudera. |
| **Frameworks/Libraries** | Spring Framework, Spring IOC, Spring AOP, Spring MVC, Spring Boot, Spring DAO (Spring JDBC, Spring ORM), Hibernate, Struts, iBATIS, Apache Camel, Maven, Ant, Jenkins. |
| **Web Technologies** | HTML, CSS, JavaScript, AJAX, jQuery, JSON, XML, XSD, XSLT, Bootstrap, Angular JS. |
| **Tools** | JUnit, Log4J, Mockito, Maven, Firebug, JIRA, Selenium, Microsoft SQL Studio, Eclipse, NetBeans, IntelliJ. |
| **Application Server** | Web Logic, Web Sphere, JBoss, Tomcat Operating Systems UNIX, Windows, LINUX. |
| **Databases** | Hbase, Cassandra, MongoDB, Oracle, MySQL, SQL Server |
| **Hadoop Data Services** | Hive, Pig, Impala, Sqoop, Flume, NiFi, Kafka, Storm, Solr Hadoop Operational Services Zookeeper, Oozie |
| **Version Control** | SVN, Git, CVS. |
| **Design Patterns** | MVC, Singleton, Factory. |
| **Methodologies** | Agile, Scrum, RUP, TDD, OOAD, SDLC |

**Professional Experience:**

**Client: Citizens Bank- Providence, Rhode IslandMarch '2017 –Apr '2018**

**Title: Scala Developer/Java Developer**

**Responsibilities:**

* Developed simple and complex MapReduce programs in Java for Data Analysis on different data formats.
* Developed MapReduce programs that filter bad and un-necessary records and find out unique records based on different criteria.
* Developed Secondary sorting implementation to get sorted values at reduce side to improve MapReduce performance.
* Implemented Customwritables, Input Format, Record Reader, Output Format, and Record Writer for MapReduce computations to handle custom business requirements.   
  Implemented MapReduce programs to classify data records into different classifications based on different type of records.
* Created FanIn and FanOut multiplexing flows with Flume.
* Experience with creating ETL jobs to load JSON data and server data into MongoDB and transformed MongoDB into the Data Warehouse.
* Experience in developing and designing POCs using Scala and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL/Teradata.
* Created Ab Initio graphs that transfer data from various sources like Oracle, flat files and CSV files to the Teradata database and flat files.
* Worked on Sequence files, RC files, Map side joins, bucketing, partitioning for hive performance enhancement and storage improvement.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala.
* Implemented Daily Oozie coordination jobs that automate parallel tasks of loading the data into HDFS and pre-processing with Pig using Oozie co-coordinator jobs.
* Performed advanced procedures like text analytics and processing, using the in-memory computing capabilities of Spark using Scala.
* Responsible for performing extensive data summarization using Hive.
* Importing the data into Spark from Kafka Consumer group using Spark Streaming APIs.   
  Developed Pig UDF's to pre-process the data for analysis using Java or Python.
* Worked with SQOOP import and export functionalities to handle large data set transfer between Oracle database and HDFS.
* Derived modeled the Facts, Dimensions, Aggregated facts in Ab Initio from data warehouse star schema for creating billing.
* Involved in migrating Hive queries into Spark transformations using Data frames, Spark SQL, SQL Context, and Scala.
* Developed a data pipeline using Kafka and Storm to store data into HDFS.
* Worked with JSON for data exchange between client and server.
* Extensively used Spring & Hibernate Frameworks and implemented MVC architecture.   
  Worked on Spring RESTful for dependency injection.
* Developed and retrieved No-SQL data using Mongo DB using DAO's.
* Implementation of Business logic layer for MangoDB services.
* Implemented test scripts to support test driven development and continuous integration.
* Involved in story-driven agile development methodology and actively participated in daily scrum meetings.

**Environment**: Hadoop, CDH4, Map Reduce, HDFS, Pig, Hive, Impala, Oozie, Java, Kafka, Storm, Linux, Maven, Oracle 11g/10g, SVN, MongoDB, Informatica.

**Client: Sammons Financial Group, Sioux Falls, SDMarch '2016 - Feb ‘2017**

**Title: Java/Scala Developer**

**Responsibilities:**

* Involved in creating Data Model (RDF) for the Existing PME Application in Semantic Web.
* Created various Parser programs to extract data from Autosys, Tibco Business Objects, XML, Informatica, Java, and database views using Scala
* Created Ingestor to publish extracting data using Data Model to Cesium environment.
* Configured Jenkins build configuration for this application everyday automatically run.
* Worked in Agile Scrum Methodology.
* Deployed and Developed Java/J2EE based application
* Developed application using Spring MVC and Hibernate as the ORM tool.
* Developed user interface with HTML, CSS, JavaScript, JQuery, Angular Js, AJAX, JSP and JSTL.
* Created SOAP based web service to provide the service of Location Look up for the agents.
* Created a Subscriber-Publisher Messaging Topics using JMS to send and receive messages.
* Used Oracle SQL and performed Tuning to improve its performance.
* Created Daemon threads for certain long running reporting process.
* Performed improvements Garbage collection to cleaning up the idle objects.
* Scheduled a certain tasks using Java Timer Class by using Core Java Thread APIs.
* Implemented several Collection classes/Interfaces to manage the client data.
* Worked on Garbage Collection for memory issues to improve the performance.
* Scheduled a certain tasks using Java Timer Class.
* Interfaced with Oracle back-end Hibernate Framework.
* Implemented Test Driven Development using JUnit.
* Used JAXB to unmarshal XML data into POJO Objects.
* Used Spring AOP and Spring IOC Framework for logging and other functions.
* Used JMS to send and receive messages from the MQ and differentiated messages using Apache Camel.
* Created and consumed SOAP based web services using JAX-RPC and JAX-WS specifications with Apache CXF Implementations of SKELETON and WSDL file.
* Used Jersey Implementations to build REST web services.
* Used Git as Version Control.
* Used MongoDB as data storage and utilized aspects like replica sets.
* Deployed applications on WebSphere during development
* Wrote shell scripts to schedule standalone Java programs.
* Created various message transport mechanism with Apache Camel.
* The message transport mechanism supported various data formats like XML and JSON.
* Used MAVEN as build management tool.

**Environment**: Scala 1.7 Spring 2.5, Hibernate 3.x, JMS, Sybase, Toad 10.5, Linux, XML, Log4j, GitHub, Hudson, Scala 4.4.1, Ivy, Semantic Web (RDF) and SPARQL, Top Briad, Java, Spring, JSP, HTML, CSS, JavaScript, Angular JS, JQuery, AJAX, Multithreading, Collections, Garbage Collection, Oracle, Web Service – SOAP, JAX-B, JAX-RPC, JAX-WS, REST-JAX-RS, WebSphere, Apache Camel.

**Client: Care Innovations. Louisville, KentuckyOct '2014 - Feb‘2016**

**Title: Hadoop Developer**

**Responsibilities:**

* Participate in requirement gathering and analysis phase of the project in documenting the business requirements by conducting workshops/meetings with various business users
* Responsible for building scalable distributed data solutions using Hadoop.
* Used Datastage Tool for extraction, transformation and load (ETL) of data in the data warehouse.
* Involved in writing MapReduce programs using combiners and partitioner.
* Created and optimized Hive queries to analyze the customer frustration level which helped reduce the call traffic and ensured better customer satisfaction
* Involved in creating Hive tables, loading structured data and writing hive queries which will run internally in MapReduce way.
* Used R Studio to do Statistical data analysis
* Developed Validation scripts for Data Ingestion.
* Created PIG scripts to load, transform and store the data from various sources into HIVE metastore.
* Monitored OOZIE workflows and managing the coordinator schedules and rerunning the failed jobs to ensure the data is processed accurately and in a timely manner.
* Worked on setting up the Kafka producer and consumer for transmitting and consuming messages
* Developed and maintained ETL (Extract, Transformation and Loading) mappings to extract the data from multiple source systems like Oracle, SQL server and Flat files and loaded into Teradata.
* Importing and exporting data into HDFS and Hive using Sqoop.
* Experienced in managing and reviewing Hadoop log files.
* Involved in loading data from UNIX file system to HDFS.
* Experienced in running Hadoop streaming jobs to process terabytes of xml format data.
* Load and transform large sets of structured, semi structured and unstructured data.
* Responsible to manage data coming from different sources.
* Supported MapReduce Programs those are running on the cluster.
* Involved in creating new table structures and modifying existing tables and fit into the existing Data Model.
* Created Teradata SQL Queries, creating Tables, and Views by following Teradata best practices.
* Tested the different sources such as Flat files, Mainframe Legacy Flat Files, SQL server 2008 to load into the Teradata data Warehouse.

**Environment:**Apache Hadoop, Amazon AWS, Hive, R studio, Pig, Oozie, Sqoop , Kafka , Data stage , Teradata , Oracle 11g, Hbase, Toad, MS-SQL Server, Unix Shell Scripting.

**Client: Gapvak, India Jan '2012 - Sep‘2014**

**Title: Java Developer**

**Responsibilities:**

* Involved in Design and Code Construction with OOPs Implementation.
* Involved in Maintenance / Enhancements of the application which includes developing the new reports.
* Developed JSP, Tiles, Validator, Action classes and EJBs for various user transactions.
* Designed Java Servlets and Objects using J2EE standards.
* Implemented the project by using MVC framework.
* Used Struts Validator framework to validate user input.
* Used Struts MVC framework along with JST for developing J2EE based web application.
* Implemented Hibernate to map all the tables from different data sources to make database updating.
* Designed Implementation of GUI using HTML, JSP and Java Script for front-end validations.
* Implemented Singleton Design Pattern by sharing the single service locator object across different Business Delegates.
* Developed the XML Schema for the data maintenance and structures.
* Implemented EJB session beans, messaging technologies like Message Driven Bean, JMS.
* Involved in designing Database Connections using JDBC.
* Involved in design and Development of UI using HTML, JavaScript and CSS.
* Used JSP for presentation layer, entity beans as the model to represent the persistent data and interacted with Oracle database as per MVC Architecture.
* Used JDBC calls in the Enterprise to access Oracle Database.
* Involved in Integration testing and defect fixes.
* Developed Unit Test Scenarios and tested Test Scenarios in Web methods using N-Unit.

**Environment:**Java, J2EE, Junit, JDK 1.5, Spring core, Spring MVC, Hibernate, SOAP/Rest Web services, Ajax, XML, jQuery, HTML, CSS, TCServer, SVN, Maven, Jenkins, Splunk, SFX, MAX, SEED. Oracle

**Client: Wizden Solutions Pvt Ltd, India Aug'2009 - Dec '2011**

**Title: Software Developer**

**Responsibilities:**

* Gathering requirements for data mapping rules from existing system to new system.
* Interacted with designers for preparing the detailed design and UML diagrams, delegated and communicated effectively with the team members and the QA team
* Extracted data from various distributed databases and loaded into Oracle Database on a monthly basis using SQL\*LOADER and external table.
* Created PL/SQL stored procedures and functions in order to validate various new orders, it included deploying the entire packages using exceptions, constraints, auditing tools.
* Created and scheduled numerous shell scripts useful for loading files into Oracle and also involved in automating the file transfer process, report generation process
* Involved in the data extraction and data conversion of fixed format and CSV files using external tables
* Developed processing logic for data validation procedures and functions for verifying and uploading into respective schemas
* Involved with DBAs in creation of tablespaces, users, privileges, indexes, constraints, triggers, synonyms, roles, partitioning etc.
* Wrote anonymous PL/SQL blocks to BULK COLLECT large data and assist in data migration between multiple databases.
* Wrote complex SQL queries using joins, subqueries and correlated subqueries
* Documented test plans, test cases, test scripts, and validations based on design specifications for unit testing, system testing, functional testing, regression testing, prepared test data for testing, error handling and analysis.
* Performance tuning of Data Stage jobs and PL/SQL scripts to handle millions of data.
* Engaged in Unit testing and Peer reviews of codes.

**Environment:** ORACLE 10g/9i, MySQL 5.0, Solaris 9, Toad, Oracle, UNIX.