## **Sunil Kumar Mail: sunkumar8711@gmail.com**

## **Hadoop Developer Phone: 240-702-6436**

**TECHNICAL SUMMARY**

* **7+** years of Professional experience in IT Software and Services in design, development and testing various applications in telecom and insurance domains.
* Expertise in **Hadoop ecosystem- HDFS, YARN**, **Pig, HBase, Spark** and **Hive** for data analysis, **Sqoop** for data migration, **Flume** for data ingestion, **Oozie** for scheduling and **Zookeeper** for coordinating cluster resources.
* Involved in designing **Hive schemas**, using performance tuning techniques like partitioning, bucketing.
* Optimized **HiveQL/ pig scripts** by using execution engine like **Tez**, **Spark**.
* Expertise in implementing **HBase schemas** with optimized **Row- key design** to avoid **Hot-spotting**.
* Good understanding on **Spark SQL, Spark Transformation Engine and Spark Streaming**.
* Experience working with **Scala**.
* Loaded streaming log data from various webservers into HDFS using **Flume**.
* Exposed **HBase** **tables** to web applications with **REST** web services.
* Used **Apache Solr** to create full text searches for more than 9 million rows.
* Experience in integrating Hive server with visualization tools like **Tableau, Qlikview, Informatica** using **ODBC** driver.
* Experience in building **Pig scripts** to extract, transform and load different file formats- **JSON, TXT, XML** data onto **HDFS, HBase, and Hive** for data processing.
* Used **SFTP** to transfer the files to server.
* Knowledge on working with Batch processing and Real time systems using various Open source technologies like Solr, hadoop, NoSQL DB’s, Storm, kafka etc.
* Analyzed the data using **Hive queries** and running **Pig scripts** to study customer behavior.
* Developed **Pig UDF'S** to pre-process the data for analysis.
* Good understanding of Java **Object Oriented Concepts** and development of multi-tier enterprise web applications.-
* Good Understanding in complete SDLC and STLC.
* Strong problem solving skills, good communication, interpersonal skills and a good team player
* Have the motivation to take independent responsibility as well as ability to contribute and be a productive team member.
* Experience with Operating Systems like Windows, Linux, and Macintosh.
* Excellent debugging and troubleshooting skills.
* Self-motivated with a strong desire to learn and an Effective Team Player.

**TECHNICAL SKILLS:**

* **Hadoop Ecosystem:** HDFS, YARN, Map Reduce, Apache Spark, Pig, Hive, HBase, Sqoop, Flume, Solr, Elastic Search, Scala, Oozie
* **Cluster:** IBM Big Insights 4.1, Hortonworks, Cloudera CDH4 & CDH5
* **Languages :** Core Java, J2EE, C, JSP, EAI, Shell Script
* **Web Technologies:** HTML5, CSS3, JavaScript, jQuery, XML, XHTML.
* **Servers:** Putty, WebSphere, WebLogic, JBoss, Apache Tomcat.
* **Database:** MySQL, Oracle, PL-SQL, NO-SQL(HBase)
* **IDE’s:** Eclipse Mars.1

**PROFESSIONAL EXPERIENCE:**

**TIAA – Financial Services, Fairfax, VA Mar 2016 – Present**

**Hadoop Developer**

**Description:** Teachers Insurance & Annuity Associate – Financial Services is a leading retirement provider for people who work in the academic, research, medical and cultural fields, provides services to over 3.9 million customer who are active and retired employees in participating at more than 15,000 institutions. It professionally and personally delivers financial solutions with every opportunity. The scope of the project was to moving their existing oracle endeca servers to BDA Servers to implement demographic and delta transactions to perform analytics on customer’s related information such as loans, adjustments, balances, distributions, transfers, contributions and forfeiture.

**Responsibilities:**

* Configured Hadoop components including Hive, Pig, HBase, Spark, Sqoop, Oozie and Hue in the client environment.
* Stored Solr indexes in HDFS.
* Index documents in HDFS using Solr Hadoop connectors.
* Responsible to manage data coming from different sources and involved in HDFS maintenance and loading of structured and unstructured data.
* Developed MapReduce programs to parse the raw data, populate staging tables and store the refined data in partitioned tables in HDFS.
* Implemented knowledge of various java, J2EE and EAI patterns.
* Worked on the backend using Scala and Spark to perform several aggregation logics.
* Created Hive queries that helped market analysts spot emerging trends by comparing fresh data with HDFS reference tables and historical metrics.
* Enabled speedy reviews and first mover advantages by defining the job flow in Oozie to automate data loading into the Hadoop Distributed File System and PIG to pre-process the data.
* Designed HBase schema to avoid Hotspotting and exposed the data from HBase tables to REST API on UI.
* Developed Pig scripts to transform raw data from several data sources into forming baseline data and loaded the data into HBase tables.
* Involved in creating POCs to ingest and process streaming data using Spark and HDFS.
* Used Flume to collect, aggregate, and store the log data from different web servers.
* Developed Shell Scripts to automate the batch processing and processed the daily jobs through Maestro scheduler.
* Provided design recommendations and thought leadership to sponsors/stakeholders that improved review processes and resolved technical problems.
* Co-ordinate with the offshore team and cross-functional teams to ensure that applications are properly tested, configured, and deployed.
* Used Tableau for visualizing and to generate reports.

**Cluster Size: 45 Nodes**

**Environment:** **CDH** 5.7.0(***Cent OS***): Apache Hadoop 2.7.1, MapReduce, HBase 1.1.2, Pig 0.15.0, Sqoop 1.4.6, Oozie 4.2.0, Java 8, Autosys, Hive 1.2.1, Impala, ZooKeeper 3.4.6, Oracle 11g, PL/SQL, SQL Developer 4.0 , UNIX. **Rest API,** Web Services REST, SQL, ANT, Shell Script, JAVA, J2EE

**GEICO – Insurance, IBM India Feb 2015 – Aug 2015**

**Hadoop Developer**

**Description**: GEICO Insurance is one of the top American Insurance players with 15 Million customers and 5-10 % yearly growth. Pulling data from various structured and unstructured data sources into Hadoop platform and standardized all data through a series of master data management processes to reach client goal to become efficient and effective within different geographic location in attracting more customers and improve the user experience.

**Responsibilities:**

* Participated in brainstorming sessions on finalizing the data ingestion requirements and design.
* Worked on SQOOP to import data from various relational data sources.
* Working with Flume in bringing click stream data from front facing application logs.
* Worked on strategizing SQOOP jobs to parallelize data loads from source systems.
* Participated in providing inputs for design of the ingestion patterns.
* Participated in strategizing loads without impacting front facing applications.
* Worked on design on Hive data store to store the data from various data sources.
* Developed MapReduce and MRUnit jobs to operate on streaming data.
* Involved in providing inputs to analyst team for functional testing.
* Worked with source system load testing teams to perform loads while ingestion jobs are in progress.
* Worked on performing data standardization using PIG scripts.
* Worked on building analytical data stores for data science team’s model development.
* Worked on design and development of Oozie works flows to perform orchestration of PIG and HIVE jobs.
* Worked on performance tuning of HIVE queries with partitioning and bucketing process.
* Worked with Ambari UI to configure alerts for Hadoop eco system components.
* Participated in tuning various components in Hadoop Eco System.

**Environment:** Hadoop 2.2 (Horton works) – PIG, MRUnit, Map Reduce, Hive, Hive UD TEZ, HDFS, Apache Sqoop, Oozie, HBase, JUnit, Zookeeper, maven, Hadoop Data Lake with Linux-Cent OS.

**Sprint- Telecom, IBM Jan 2014 – Feb 2015**

**Hadoop Developer**

**Description**: GPRS collection workflow will fetch all the GGSN & SGSN files generated in switch periodically. Mediation Zone GPRS process workflow used to process all the files and distribute records to various billing downstream applications. Project required to handle bottleneck of processing huge files from SGSN/GGSN switch. Hadoop's MapReduce used to split the huge file and map the SGSN/GGSN records to internal format. GPRS process workflow will further process the outputted files and distribute the files to downstream for further billing. Pig used for fine-tuning the GGSN/SGSN records. Sqoop used to push GPRS session details into database which will be used during process workflow. Objective to redesign GPRS Workflow.

**Responsibilities:**

* Involved in data loading strategies by using Sqoop for importing and exporting the data from HDFS to Relational Database systems and vice-versa.1
* Involved in design and creating Hive partitioned tables and External tables.
* Involved in writing Pig scripts for data cleansing.
* Developed Pig Latin scripts by using HCatalog to read the data from hdfs and load into the hive tables.
* Developed Hive queries for the data analysts.
* Developed workflows in Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig and MR.
* Using JUnit test extensively written test cases for this system to test the application.
* Implemented logging mechanism using Log4j with the help of Spring AOP frame work.

**Environment:** Hadoop Cloudera CDH3, MapReduce, HIVE, PIG, Sqoop, Oozie, Java, Autosys, Hbase, HIVE, Zookeeper, Maven, Oracle 11g, SQL Developer 4.1.3 , Unix.

**Vodafone Portugal, IBM, India July 2010 – Dec 2013**

**Application Developer**

**Description**: Mediation Device is an in house application by Vodafone Portugal. Mediation device main goal is to feed the Billing (Post and Pre Paid), Interconnection and Reporting systems hiding the complexity and the diversity of the Vodafone network to all Mediation client systems. Digital Route from Sweden has developed Mediation Zone product. Mediation Zone platform supports any combination of data exchange between any systems, in an online and offline mode. Mediation Device / Mediation Zone handles all kinds of traffic including GSM/GPRS/SMS/ MMS/WIFI/ADSL/ IPTV/VPN/Email/LTE/Roaming/Interconnect scenarios.

**Responsibilities:**

* Involved in the review and analysis of the Functional Specifications, and Requirements Clarification Defects etc.
* Involved in the analysis and design of the initiatives.
* Involved in the development of the mediation platform using Java.
* Involved in design and implementation of migration of Mediation legacy platform to mediation zone.
* Implemented CDR batch processing for MSC, IN, ADSL, and GPRS Streams to Object Oriented Programming.
* Involved in writing Junit Test for Unit testing.
* Involved in Regression Testing of SAT and Development Environments.
* Involved in writing the SQL queries and stored procedures.
* Participated in the test case reviews, and manual testing of the enhancements during Release 1.5.
* Involved in fixing the defects during integration testing.
* Build and deployment of the application using Ant script on to dev and testing environments.
* Participated in the code reviews for various initiatives, Performed Static Code Analysis to follow the Best Practices for Performance and Security.

**Environment**: Mediation Zone (Digital Route), Legacy Zone, Core Java 6

**Vodafone Italy, IBM, India Jul 2008– Jun 2010**

**Application Developer**

**Description**: The Vodafone Group's principal activity is providing voice & data communications services. Through its mobile businesses, the Group provides a range of mobile communications services including voice, text messages, picture messages and other data. WebMethods is being used to interface Seibel with ARBOR (B2B integrations) using EDI. The scope of the project included inbound Customer Purchase Orders, Advices, Outbound Order Acknowledgements and Invoices. Project included developing a messaging infrastructure for EAI/B2Bi and online monitoring, testing and reporting tools. Customer Order information maintained in SIEBEL systems was integrated with webMethods Integration Server. All Clients/Dealers were connected through the webMethods Integration Server and webMethods Trading Networks.

**Responsibilities:**

* Interpreting product requirements into test requirements, writing test plans and test cases.
* Modeling conceptual design using **Use Case**, **UML Class** and **Activity diagrams** using **Rational Rose.**
* Wrote requirement specific **SQL** and **PL/SQL** scripts including Stored Procedures, functions, packages and triggers.
* Made enhancements to the application which presented me with the opportunity to go through the entire SDLC.
* Providing daily updates to the on-site team over call and making enhancements.
* Code coverage and Test case presentation.
* System Test Planning & Execution.
* Analyzing the HLD and Business Requirements.
* Preparing TPI in the test planning phase and the execution phase.
* Preparing the Test Cases and Test Execution Plan.
* Defect tracking, reporting and closure. Raised defects in QC.
* Reviewing the test plan document, so that there is no functionality or the requirements missing.
* Used **Web Services** like **SOAP** and **WSDL** to communicate over internet.
* Automated the tests using the integration tool (SOAP-UI).

**Environment**: WebMethods, SOAP-UI, SEIBEL, CITRIX, Putty, IBM Rational - ClearQuest.

**EDUCATION:**

* Bachelor of Technology (Computer Science and Engineering), JNTU, Hyderabad, India