**SRI**

**Email:sri.cheru5@gmail.com**

**Phone: 512-6454-156**

**PROFESSIONAL SUMMARY:**

* Overall **8+ years** of experience in IT industry, which includes **3+years** of experience in **Hadoop.**
* **4+ years** of experience in **Hadoop Ecosystem** providing and implementing solutions for **Big Data** Applications with excellent knowledge of Hadoop architecture.
* Strong experience in building, maintaining multiple **Hadoop clusters** of different sizes and configuration and setting up the rack topology for large clusters.
* Hands on experience in various Hadoop infrastructures such as Map Reduce, Hive, PIG, SQOOP, Flume and HBase.
* Good understanding of Hadoop architecture and Hands-on experience with Hadoop components such as **Job Tracker**, **Task Tracker**, **Name Node**, **Data Node** and **MapReduce** concepts and **HDFS** Framework.
* Expertise in optimizing **MapReduce algorithms** using **Mappers**, **Reducers**, **Combiners** and **partitioners** to deliver the best results for the large datasets.
* Experience in using **Spark API** over Hadoop Map Reduce to perform analytics on data.
* Good hands on experience in creating the RDD's, DF's for the required input data and performed the data transformations using **Spark Scala.**
* Experience in converting Hive queries into Spark transformations using **Spark** RDDs and Scala.
* Expertise in writing **MapReduce** jobs using **Java native code**, **Pig, Hive** for various business use cases.
* Good experience in data loading from **Oracle** and **MYSQL** databases to **HDFS** system using **Sqoop** (Structure Data) and **Flume** (Log Files & XML).
* Experience in developing customized **UDF's** in java to extend **Hive** and **Pig Latin** functionality.
* Experience with **NoSQL** databases like **HBase**as well as other ecosystems like **ZooKeeper**, **Oozie**.
* Extensive experience in middle-tier development using **J2EE** technologies like **JDBC, Servlets, JSP.**
* Expertise in **RDBMS** like **Oracle, MS SQL Server, MySQL** and**DB2**.
* Capable of processing large sets of **structured**, **semi-structured** and **unstructured** data and supporting systems application architecture.
* Knowledge on installing, configuring, monitoring, and maintaining **HDFS**, **Yarn**, **HBase**, **Flume**, **Sqoop**, **Oozie, Pig, Hive**.
* Installation of various **Hadoop Ecosystems** and **Hadoop Daemons**.
* Experience with web-based UI development using **jQuery UI**, **jQuery,CSS, HTML, HTML5, XHTML** and **JavaScript.**

**TECHNICAL SKILLS:**

**Hadoop Tools:** Big Data(Hadoop Framework) HDFS, Map Reduce,Pig, Hive,HBase, Sqoop,

Flume,Oozie, Zookeeper, Spark, Scala

**Web Technologies:** JSP, Servlets, JavaBeans, JDBC, XML, HTML, VMWare, Amazon EC2, JavaScript,

Hibernate, Spring, Struts

**Languages:** Java, C, HTML, CSS, XML,Pig Latin

**Database:** Oracle 10g, My SQL, HBase, NO SQL (Cassandra)

**Tools:** Eclipse, Toad, My SQL

**Markup Languages:** HTML, XML, DHTML

**Operating Systems:** Linux, Windows, Ubuntu, Centos

**PROFESSIONAL EXPERIENCE:**

**University Hospitals, Cleveland, OH**

**Nov 2014 –Till Date**

**Role: Hadoop Developer**

**Description:** University Hospitals is one of the nation’s leading health care systems, providing high-quality, patient-centered medical care at locations throughout Northeast Ohio.University Hospitals offers the largest network of primary care physicians, outpatient centers and hospitals.The purpose of the project is to create Enterprise Data Hub, so that various business units can use data from Hadoop to do Data Analytics. The data will be stored in Hadoop file system and processed using Map/Reduce jobs.

**Responsibilities:**

* Analyzing **Hadoop cluster** and different **Big Data** analytic tools including **Pig, Hive, HBase** and **Sqoop.**
* Creating multiple **MapReduce** jobs in **Pig** and **Hive** for data cleaning and pre-processing.
* Successfully loading files to Hive and **HDFS** from Oracle, SQL Server using **SQOOP**.
* Writing**Hive jobs** to parse the logs and structure them in tabular format to facilitate effective querying on the log data.
* Extracted and updated the data into HDFS using **Sqoop** import and export command line utility interface.
* Responsible for developing data pipeline using **Flume**, Sqoop and pig to extract the data from weblogs and store in HDFS.
* Creating **Hive tables**, loading with data and writing **Hive queries**.
* Involved in **Spark** for fast processing of data.Defining job flows.
* Using**Hive** to analyze the partitioned data and compute various metrics for reporting.
* Managing and reviewing the **Hadoop log files**.
* Using**Pig** as ETL tool to do Transformations, even joins and some pre-aggregations.
* **Unit testing** and delivered Unit test plans and results documents.
* Exporting data from **HDFS** environment into **RDBMS** using **Sqoop** for report generation and visualization purpose.
* Worked on **Oozie** workflow engine for job scheduling.

**Environment:** Hadoop, HDFS, MapReduce, Pig, Hive, Sqoop, HBase, Oozie, Zookeeper, Java, Spark, Scala.

**First Data, Austin, TX**

**July 2013 – Sep 2014**

**Role: Hadoop Developer**

**Description:** First Data Corporation is a global payment technology solutions company The company's star interbank network offers PIN-secured debit acceptance at ATM and retail locations.First Data has a global footprint of 6 million merchants, which indeed the largest in the payments industry. The objective of this project is to centralize the source of historic data existing in different sources. It provides a comprehensive picture of customer’s behavior.

**Responsibilities:**

* Worked on importing data from various sources and performed transformations using **MapReduce**, **Hive** to load data into **HDFS.**
* Continuous monitoring and managed the **Hadoopcluster** using **Cloudera Manager**. Worked on **Oozie** workflow to run multiple jobs.
* Building scalable distributed data solutions using **Hadoop.** Written various **Hive** and **Pig scripts.**
* Created **HBase** tables to store variable data formats coming from different portfolios.
* Performed real time analytics on **HBase** using **Java API** and **Rest API.**
* Implemented **HBase Co-processors** to notify Support team when inserting data into **HBase Tables**.
* Configured **Sqoop jobs** to import data from RDBMS into HDFS using **Oozie workflows**.
* Worked on setting up **Pig**, **Hive** and **HBase** on multiple nodes and developed using Pig, Hive, HBase and MapReduce
* Worked on compression mechanisms to optimize **MapReduce** Jobs.
* Analyzed the customer behavior by performing click stream analysis and to ingest the data used F**lume**.
* Solved small file problem using **Sequence files** processing in **Map Reduce**.
* Implemented business logic by writing **UDF's** in Java and used various UDF's from Piggybanks and other sources.
* Exported the analyzed data to the relational databases using **Sqoop** for visualization and to generate reports for the BI team.

**Environment**: Horton works, Map Reduce, HBase, HDFS, Hive, Pig, Java, SQL, Cloudera Manager, Sqoop, Flume, Oozie, Java (JDK 1.6), Eclipse.

**JC Penny Plano, Tx.**

**May 2012– June 2013**

**Role: Hadoop Developer**

Description: JC Penny is one of America's largest department store, catalog, and e-commerce Retailers. The objective of the project Enterprise Contract Ordering System (ECOS) is to support the Enterprise Merchandise Action Team (MAT) and the Enterprise Merchandising & Planning processes defined in the overall Integrated Planning work streams. ECOS will do this by providing a single application for the MAT to Create and manage Contracts, Purchase Orders, and Allocation Instructions regardless of channel (Store Channel and Direct Channel).   
   
Responsibilities:   
• Installed and configured Hadoop, Developed multiple MapReduce jobs in Java for data cleaning and preprocessing.   
• Installed and configured Pig for ETL jobs.   
• Troubleshooting the cluster by reviewing Hadoop LOG files.   
• Imported data using Sqoop from Tera data using Tera data connector.   
• Used Oozie to orchestrate the workflow.   
• Creating Hive tables and working on them for data analysis in order to meet the business requirements.   
• Got good experience with NoSQL database.   
• Designed and implemented MapReduce-based large-scale parallel relation-learning system.   
• Installed and benchmarked Hadoop/HBase clusters for internal use.   
• Written HBASE Client program in Java and web services.   
• Model, serialize, and manipulate data in multiple forms (xml).   
• Supported postproduction enhancements.   
• Experience with data model concepts-star schema dimensional modeling Relational design (ER).   
• Involved in JUnit testing of the application using JUnit framework.   
   
Environment: Hadoop, MapReduce, HDFS, Hive, Pig, HBase, Sqoop,MySQL ,Java 6, Eclipse.

**Bayer, Pittsburgh, PA**

**Role: Hadoop Developer**

**Oct 2011 – May 2012**

**Description:** Bayer HealthCare is among the world's foremost innovators in the field of pharmaceutical and medical products. The project was about Consumer Care where the Dietary products and Supplements data of the past years areanalyzed. The solution is based on the open source Big Data software Hadoop. The data will be stored in Hadoop file system and Processed using MapReduce jobs.

**Responsibilities:**

* Application installation of **Hadoop, Hive, Map Reduce &Sqoop.**
* Analyzing the log files using Hive queries and writing **UDF's** for **Hive queries.**
* Creating **Hive tabular partitions** and **buckets.**
* Moving all log files generated from various sources to **HDFS** for further processing through **Flume.**
* Used **R programming** language to graphically analyse the data.Used **Tableau** to represent data.
* **HDFS** support and maintenance and Adding/Removing a Node, Data Rebalancing.
* Developed **MapReduce** application using Hadoop, MapReduce programming and **HBase.**
* Developed the **Pig scripts** and **Hive Reports.**
* Responsible for developing data pipeline using **Flume**, Sqoop and pig to extract the data from weblogs and store in HDFS.
* Writing **Mapreduce** programs and tested using **MRUnit.**
* Analyzed the advantage of introduction of **Cassandra NoSQL** database into the present system.
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.
* Participate in requirement gathering and analysis phase of the project in documenting the business requirements by conducting workshops/meetings with various business users.

**Environment:** Hadoop, Java-MapReduce, Hive, Pig, Cassandra, HBase, Ubuntu, Eclipse,Sqoop.

**Essent Guaranty Insurance, Radnor, PA Apr 2010 – Oct 2011**

**Role: Java Developer**

**Description:** Essent Guarantyis a private mortgage insurer, approved by Fannie Mae and Freddie Mac. Focused on managing risk and dedicated to serving as a strong and fair counterparty. This is an internal project Reference Data Service and Maintenance (RDSM) for providing services to the customers. The project designs and develops service interfaces, for retrieving reference data from common user interface for agents and online customers to get the Insurance quotes.

**Responsibilities:**

* Coded **Servlets** on the server side, which gets the requests from the client and processes the same by interacting the Oracle database.
* GUI development using HTML Forms and Frames and validating the data with **JavaScript**.
* Used **JDBC** to connect to the backend database and developed stored procedures.
* Developed code to handle web requests involving Request Handlers, Business Objects, and Data Access Objects.
* Developed **Java Servlets** to control and maintain the session state and handle user requests.
* Carrying out actual development of JSP pages including the use of JSP custom tags and other methods of Java Beam presentation and all HTML and graphically oriented aspects of the site's user interface.
* Used XML to transfer data universally among different data sources.
* Involved in unit testing and documentation.

**Environment:** J2EE (Servlets, JSP, Java Scripts, Java Beans, JDBC 2.0, XML technologies), HTML, JavaScript, CSS, XML, Oracle 9i, PL/SQL, Apache Tomcat, JBuilder, UNIX.

**Cigna, Windsor, CT**

**Oct 2008 – Mar 2010**

**Role: Jr. Java Developer**

Description: Greater Health care project maintains a website quote for the people looking for health insurance and provides the various offers available and suggests a plan that suits them better based on the various inputs from them. 

**Responsibilities:**

* Analyzed all the test cases based on the requirements gathered and documented for unit testing aswell as for integration testing.
* Designed the user interface required for the portal with all the components for selection of plan.
* Provided the design using Restful Web Services to populate the individual details of plans available for the customers to pick.
* Programmed functionality for all the components in the user interface interacting with the database using MySQL Server.
* Developed various Controller classes and business logic using the spring libraries which interact with the middle tier to perform the business operations.
* Responsible to develop the custom tools as per the client needs.
* Tested the application by programming the test cases using JUnit for both units testing and Integration testing and bug tracking for the entire application.

**Environment:** Core Java, JSP, MySQL, SOAP, JUnit, Eclipse, HTML, JavaScript, XML.

**Junior Java Developer**

**May 2008 –Oct 2008**

**Andhra Bank, Hyd, India.**

**Description:** Banking sector that is used to provide loans and other banking sections such as personal, saving. It is one of the leading governments banking sector that is user friendly with all the subsystems integrated in it.   
   
**Responsibilities:**• Developed programs for accessing the database using JDBC thin driver to execute queries, Prepared statements, Stored Procedures and to manipulate the data in the database   
• Involved in resolving the Design issues and other project related problems.   
• Developed back-end logic with Core Java using technologies including Collection Framework, Multi-Threading.   
• Developed the credit check module using Servlets and JSP.   
• CSS is used to style Web pages with user friendly guidelines.   
• Used Asynchronous JavaScript Technology and XML (AJAX) for critical front-end validations to reduce the burden on the server side.   
• Involved in various phases of Software Development Life Cycle (SDLC).   
• Front-end is designed by using HTML, CSS, JSP, Servlets, JSTL, Ajax and Struts.   
• Deployed web components, presentation components and business components in Web Logic Application Server.   
• Performed client side validations using JavaScript.   
• Extensively worked on Java server side programming using Servlets, and JSP.   
• Involved in peer code reviews and performed integration testing of the modules.   
   
**Environment:** Core Java, JDK, J2EE, Servlets, JSP, MVC, Java Beans, SQL, PL/SQL, JDBC, Web Logic, Log4J, Apache ANT, JavaScript, CSS, HTML, Stored Procedures, Windows 2000/NT/98 OS.

**EDUCATION:**

* **Bachelors in Information Technology**