ANUDEEP CHITIPIRALLA

[anudeepc7r@gmail.com](mailto:anudeepc7r@gmail.com)

(913) 433- 6233

* Comprehensive experience of 7 years, with over 4 years in Hadoop and Scala(spark) development and Administration experience along with 3+ of experience in Java/J2EE enterprise application design, development and maintenance.
* Extensive experience implementing Big Data solutions using various distributions of Hadoop and its ecosystem tools.
* Experience in installing, configuring and monitoring HDFS clusters (on premise & cloud).
* In depth understanding of MapReduce programs to scrub, sort, filter, join and query data
* Planning, deployment, and tuning of SQL (SQL Server, MySQL) and NoSQL databases.
* Implemented innovative solutions using various Hadoop ecosystem tools like Pig, Hive, Impala, Sqoop, Flume, Kafka, HBase, ElasticSearch, Zookeeper, Couchbase, Storm, Solr, Cassandra and Spark.
* Experience developing PigLatin and HiveQL scripts for Data Analysis and ETL purposes and also extended the default functionality by writing User Defined Functions (UDF s) for data specific processing.
* Practical understanding of the Data modeling (Dimensional & Relational) concepts like Star-Schema Modeling, Snowflake Schema Modeling, Fact and Dimension tables.
* Experience with migrating data to and from RDBMS and unstructured sources into HDFS using Sqoop& Flume.
* Well-versed database development knowledge using SQL data types, Joins, Views, Transactions, Large Objects and Performance tuning.
* Good knowledge of Data warehousing concepts and ETL and Teradata.
* Experience writing Shell scripts in Linux OS and integrating them with other solutions.
* Intensive work experience in developing enterprise solutions using Java, J2EE, Servlets, JSP, JDBC, Struts, spring, Hibernate, JavaBeans, JSF, MVC.
* Strong experience in Extraction, Transformation and Loading (ETL) data from various sources into Data Warehouses and Data Marts using Informatica Power Center (Repository Manager, Designer, Workflow Manager, Workflow Monitor, Metadata Manger), Power Exchange, Power Connect as ETL tool on Oracle, DB2 and SQL Server Databases.
* Fluent with the core Java concepts like I/O. Multi-threading, Exceptions, RegEx. Collections, Data-structures and Serialization.
* Excellent problem-solving analytical, communication, presentation and interpersonal skills that help me be a core member of any team.
* Experience mentoring and working with offshore and distributed teams.

**Areas of Expertise**

**Technologies:**

**Big Data Ecosystems:** Hadoop ecosystem, Apache Yarn (Hadoop version 2), Apache Drill, Apache Spark and its ecosystem, Storm

**Programming Languages:** Java, Scala

**Scripting Languages:** JSP & Servlets, PHP, JavaScript, XML, HTML, Python and Bash

**SQL Databases:** Oracle, Hive, Pig, PrestoDB, Impala, SparkQL.

**NoSQL:** HBase, Cassandra, Neo4J, MongoDB

**UNIX Tools:** Yum, RPM

**ETL Tools:**Informatica Power Center 9.1/8.6/8.5/8.1/7.1 (Designer, Workflow Manager, Workflow Monitor, Repository manager and Informatica Server)

**Tools:** IntelliJ IDEA, Eclipse, R Studio, JDeveloper, JProbe, CVS, Ant, MS Visual Studio, MATLAB

**Platforms:** Mac OSX, UNIX, Linux, Solaris, Windows

**Automation tools:** Chef, Puppet

**Methodologies:** Agile, UML, Design Patterns

**Cloud:**

**Amazon Web Services:** EC2, Amazon Elastic Map Reduce, Amazon S3  
**Google Cloud Platform:** Big Query, App Engine, Computer Engine

**Data:**

Climate Data

Plant Breeding

Stock Market Analysis

Financial Services

E-Commerce

OnStar data

Logs and Click Events data

**Work Experience**

**Big Data Cloud Analytics Developer**

**Monsanto, St Louis, MO June 2016 to Present**

Produce more. Conserve more. Improve lives. That’s Monsanto’s vision for a better world. Achieving this vision demands revolutionizing agriculture through technology and analytics is central to attaining this transformation.  The global analytics team is a cutting-edge group providing recommendations and solutions to accelerate and optimize Monsanto’s product development.  Developing and implementing a full end-to-end Cloud solution for our analytics backend including Infrastructure Design, Platform, Automation, and Storage Architecture. Integrating the big data infrastructure in the AWS cloud.

**Responsibilities**

.

* Designed & developed workflows to automate Migration using Python.
* Designed & implemented database cloning using Python.
* Involved in ETL, Data Integration and Migration.
* Worked on Talend to run ETL jobs on the data in HDFS.
* Involved in creating Hive Tables, loading the data and writing Hive Queries that will run internally in a map reduce run internally in a map reduce way.
* Imported data using Sqoop to load data from Oracle to HDFS on a regular basis.
* Developed core search component using Apache Solr.
* Customized Apache Solr to handle fallback searching.
* Used Flume to collect the logs data with error messages across the cluster.
* Integrated big data infrastructure in the AWS cloud.
* Built services, deployed models, algorithms, performed model training and provided tools to make our infrastructure more accessible to all our data scientists.
* Integrated R models into a Scala Project.
* Helped in generating Log files.

**Environment**: Scala, Python, Kafka, Apache Spark, OpenCPU, ASReml, IntelliJ IDEA, R Studio, Git, AWS, Google Cloud, Azure

**Hadoop Developer**

**Dow Jones & Company, Atlanta, GA March 2015 to June 2016**

The company was best known for the publication of the Dow Jones Industrial Average and related market statistics, Dow Jones Newswire and a number of financial publications. In 2010 the Dow Jones Indexes subsidiary was sold to the CME Group and the company focused on financial news publications, including its flagship publication The Wall Street Journal and providing financial news and information tools to financial companies.

**Responsibilities**

* Responsible for developing efficient MapReduce on Hortonworks cloud programs for more than 20 years' worth of claim data to detect and separate fraudulent claims.
* Worked on Hortonworks Hadoop.
* Participated in development/implementation of Hortonworks Hadoop environment.
* Uploaded and processed more than 30 terabytes of data from various structured and unstructured sources into HDFS (Hortonworks cloud) using Sqoop and Flume.
* Played a key-role is setting up a 40 node Hadoop cluster utilizing Apache Spark by working closely with the Hadoop Administration team.
* Worked with the advanced analytics team to design fraud detection algorithms and then developed
* MapReduce programs to efficiently run the algorithm on the huge datasets.
* Developed Scala programs to perform data scrubbing for unstructured data.
* Responsible for designing and managing the Sqoop jobs that uploaded the data from Oracle to HDFS and Hive.
* Well versed in installing, configuring & managing of Hadoop Environment including Cloudera Manager for Cloudera Hadoop Distribution (CDH) and Ambari for Hortonworks Data Platform.
* Experience with job workflow scheduler like Oozie.
* Designed and Maintained Oozie workflows to manage the flow of jobs in the cluster.
* Played a key role in installation and configuration of the various Hadoop ecosystem tools such as Solr, Kafka, Storm, Pig, HBase and Cassandra.

**Environment**: Java, Hadoop, Hive, Pig, Hortonworks, Apache Solr, Sqoop, Flume, Druid, HBase, Kafka, Apache Spark, Storm, Oracle 10g, Teradata, Cassandra

**Hadoop Developer**

**AMERICAN EXPRESS, Phoenix, AZ August 2014 to March 2015**

American Express Company is a multinational financial services corporation best known for its credit card, charge card, and travelerscheque businesses. As many corporate companies enrolled to this card services, transaction data sets are basically huge. The log files generated by the system for these credit transactions by customers are maintained in Hadoop cluster for further analysis and operations

**Responsibilities:**

* Designed & developed workflows to automate Database Patching & Rollback using Python.
* Built alert & monitoring scripts for applications & servers using Python & Shell Script.
* Implemented complex business rules in Informatica Power Center by creating re-usable transformations, and robust Mapplets.
* Wrote Map Reduce job using Scala.
* Great understanding of REST architecture style and its application to well performing web sites for global usage.
* Responsible for developing, support and maintenance for the ETL (Extract, Transform and Load) processes using Informatica Power Center 8.5.
* Have written Hive Queries for data analysis to meet the business requirements.
* Creating Hive Tables and working on them using Hive QL.
* Experience in automation and configuration management of Hadoop 2.0 using Chef and Puppet
* Wrote and modified stored procedures to load and modifying of data according to business rule changes.
* Developed a custom file system plugin for Hadoop to access files on data platform.
* The custom file system plugin allows Hadoop Map Reduce programs, HBase, Pig, and Hive to access files directly.
* Extracted feeds from social media sites such as Facebook, Twitter using Python scripts.
* Organized and benchmarked Hadoop/HBase Clusters for internal use.

**Environment**: Cloudera, Informatica, Hadoop, Python, HDFS, HBase, Pig, Hive, MapReduce, Sqoop, Apache Spark, Chef, Puppet, Flume, ETL, REST, Java, PL/SQL, Oracle 11g, Unix/Linux, CDH3, CDH4.

### Hadoop Developer/Hadoop Admin

**TCS, Chennai January 2013 to July 2014**

TCS, with view of Digital re-­imagination through analytics, has architect a group to build competency and in turn develop solutions to customer with latest technologies available in these areas. As part of this program, for handpicked associates across TCS, resources were made available to build "Proof of Concepts" on the trending technologies to solve customer centric problems in areas of ABIM.imagination through analytics, has architect a group to build competency and in turn develop solutions to customer with latest technologies available in these areas. As part of this program, for handpicked associates across Wipro, resources were made available to build "Proof of Concepts" on the trending technologies to solve customer centric problems in areas of ABIM.

**Responsibilities**

* Developed multiple Map Reduce jobs in Java for data cleaning and preprocessing.
* Developed and maintained Hive QL, Pig Latin Scripts, Scala and Map Reduce.
* Worked on the RDBMS system using PL/SQL to create packages, procedures, functions, triggers as per the business requirements.
* Importing and exporting data into HDFS from Oracle Database, and vice versa using Sqoop.
* Experienced in defining job flows.
* Experience with NoSQL database HBase.
* Designed and developed Informatica ETL mappings to extract master and transactional data from heterogeneous data feeds and load.
* Developing scripts and batch jobs to schedule various Hadoop Programs.
* Responsible for architecting Hadoop clusters with CDH3
* Involved in the installation of CDH3 and up-gradation from CDH3 to CDH4
* Developer in Big Data team, worked with Hadoop Cloudera, and its ecosystem.
* Installed and configured Hadoop, Map Reduce, and HDFS.
* Created UNIX shell scripts for Informatica ETL tool to automate sessions.
* Used Hive QL to do analysis on the data and identify different correlations.
* Installed and configured Pig and also written Pig Latin scripts.
* Performed Strategic data analysis and research to support business needs.
* Maintained, designed, modified and constructed tools like UNIX and Oracle database
* Specified and analyzed requirements for scalable database and result reporting
* Created world class web data-mining system
* Installed and Configured the Informatica Client tools.
* Implementation, Configuration and monitoring of ESX server Virtualization.
* Handled the tasks of code development and data analysis
* Designed and developed transformations required for generating machine learning data sets
* Installation of required apache and relatedpackages

**Environment**: Cloudera,Informatica, Hadoop, Hive, PIG, MapReduce, HDFS, UNIX, Java, Shell scripting,SQL, HDFS, CDH4.3, Sqoop, Flume, Oozie, VMware, WebCrawler, Red Hat Enterprise Linux(RHEL).

**Core Java/J2EE Developer**

**Reliance Communications, BangaloreMarch 2011 to November 2012**

Reliance Communications Ltd. is an Indian Internet access and Telecommunications Company headquartered in Navi Mumbai, India. It provides CDMA, GSM mobile services, fixed line broadband and voice services, DTH depending upon the areas of operation.

**Responsibilities**:

* Responsible for the implementation of application system with core Java and Spring framework.
* Used Spring framework for dependence injection and integrated it with Hibernate.
* Developed Restful APIs using Spring Cloud.
* Developed the Spring Xml file for database configuration using different spring beans.
* Implemented and Used Web Services with the help of WSDL and SOAP to get updates from the third parties.
* Upgradation of WebLogic servers in development, testing and production environment and applying patch and service packs
* WebLogic Administration, Monitoring and Troubleshooting using Admin Console and JMX.
* Cluster Configuration – Single Cluster, Multiple Clusters and Deployment over clusters.
* WebLogic's HttpProxyServlet and HttpClusterServlet configurations
* Working with WebLogic utilities - WebLogic.Admin, WebLogic.Deployer, ejbc, appc.
* Monitoring WebLogic server health and security.
* Tuning JVM heap size to maximize the number of clients accessing the applications.
* Installation and configuration of WebLogic Integration server 8.1 SP6 9.2 mp2 10.3 on Sun Solaris 8 and Sun Solaris 10
* Involved in implementation of MVC pattern using Angular JS, JSF and Spring Controller.
* Developed specialized search system (using Angular JS, Java Servlets, JUnit).
* Involved in Java and EJB Design Patterns.
* Used Multi-threading to overcome the errors in the process of transactions.
* Used JAXB parser for parsing the valid XML files.
* WebLogic Administration, Monitoring and Troubleshooting using Admin Console
* Managing and Monitoring the JVM Performance by adjusting WebLogic heap size and garbage collection parameters
* Monitoring the WebLogic Server using HP Site scope and SNMP agent based monitoring
* Deploying different applications like WAR files, EJB jars files and enterprise application archives (EAR) on WebLogic using various deployment tools (Console, WebLogic deploy and WLST)
* Knowledge in WebLogic Portals and Jboss
* Used Spring DAO concept in order to interact with database (DB2) using JDBC template.
* Used MVC Framework and integrated Struts Web Module with Java Server Faces (JSF).
* Involved with Master Data Management (MDM) for Customer Data Integration.
* Involved in module testing using JUnit.
* Implemented Hibernate to map all the tables from different data sources to make database updating.
* Maven is used to build and deploy the application.
* Used Spring Framework for Dependency Injection and spring bean wiring
* Used Hibernate 3.0 object relational data mapping framework to persist and retrieve the data from database.
* Implemented Hibernate to map all the tables from different data sources to make database updating.
* Applications are designed using J2EE, JSP, Struts, WSDL, Web Services, JMS

**Environment:** Spring, JUnit, JDBC, MDM, Eclipse, CSS3, HTML5, Multithreading, JQuery, Oracle 11g, DB2, JSF, Hibernate, Angular JS, RESTFUL APIs, SOAP, WebLogic, Web Services, LINUX SCRIPTS Shell Scripting, Jboss, SQL Server 2012, GIT, Maven

**Junior Java Developer**

**HDFC Bank, Mangalore May 2010 to March 2011**

The Housing Development Finance Corporation Limited (HDFC) is one of the oldest and largest financial services firm in India. The aim of this project was to automate the process of Initial Public Offering by designing a GUI- based tool developed in Java Swing to capture the key set of events as per the laid down business specifications. This consisted of generation of IPO Deals, Client Letters and corresponding business functionalities to upload and download such documents within the system.

**Responsibilities:**

* Developed JavaScript behavior code for user interaction.
* Used HTML, JavaScript, and JSP for UI development.
* Used JDBC and managed connectivity, for inserting/querying& data management including stored procedures and triggers.
* Involved in the design and coding of the data capture templates, presentation and component templates.
* Developed an API to write XML documents from database.
* Used JavaScript and designed user-interface and checking validations.
* Part of a team which is responsible for metadata maintenance and synchronization of data from database.

**Environment**: java jdk, java script, JSP, JDBC, HTML, XML.

**Education and Professional Development**

* Bachelor’s in Computer Science and Engineering
* Member, Hadoop Users Group of Atlanta

**References**

Available upon Request