**Chandra Lekkala**

**+1 321-615-4012**

[**chan.lekkala@gmail.com**](mailto:chan.lekkala@gmail.com)

**PROFESSIONAL SUMMARY:**

* Having 8+ years of total IT Experience, including 5+ year’s hands-on experience in Big-data/Hadoop Technologies.
* 3+ years of extensive experience in J2EE Technologies, Database development, Data Analytics.
* Very good experience in the Application Development and Maintenance of SDLC projects using various technologies such as Python, Java/J2EE, JavaScript, Data Structures and UNIX shell scripting.

**BIG DATA EXPERIENCE**

* Expertise in all components of Hadoop Ecosystem- Hive, Hue, Pig, Sqoop, HBase, Impala, Flume, Zookeeper, Oozie, and Apache Spark.
* Having sound experience in Big Data Hadoop Ecosystems experience in Ingestion, storage, querying, processing and analysis of big data.
* Excellent understanding/knowledge of Hadoop architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, YARN and MapReduce programming paradigm.
* Experience in importing and exporting data from different RDBMS like MySQL, Oracle and SQL Server into HDFS and Hive using Sqoop.
* Experience in developing custom MapReduce programs using Apache Hadoop to perform Data Transformation and analysis as per requirement.
* Experience in writing PIG scripts and Hive Queries for processing and analyzing large volumes of data.
* Experience in creating PIG and HIVE UDFs using java in order to analyze data sets.
* Experience in collecting the log data from different sources like (web servers and social media) using Flume and storing on HDFS to perform MapReduce jobs.
* Experience in Spark Streaming in order to ingest data from multiple data sources into HDFS.
* Good Hands on Experience in Stream Processing including Storm and Spark Streaming.
* Experience in Design & Development, tuning and maintenance of NoSQL databases such as MongoDB, HBase, Cassandra and its Integration with Hive.
* Used Cassandra in accordance with PIG/Hive as and when required for real time low latency queries.
* Developed multiple POCs using Python Spark API and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL.
* Very good knowledge in advanced Big-Data technologies like Spark Ecosystem (Spark SQL, SparkR and Spark Streaming), Kafka and Predictive analytics.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Spark SQL and Scala.
* Worked on reading multiple data formats on HDFS using Python Spark API.

**RELATIONAL DATABASES & ETL**

* Involved in database design, creating Tables, Views, Stored Procedures, Functions, Triggers and Indexes.
* Strong experience in Data Warehousing and ETL using Data stage.
* Experience in developing the complex SQL queries, unions and multiple table joins and experience with views.
* Experience with various Business Intelligence tools and SQL databases.
* Extensively worked on data extraction, Transformation and loading (ETL) data from various sources like Oracle, SQL Server and flat files.
* Experience in preparing and executing unit test plan and unit test cases after software development.
* Ability to work effectively with associates at all levels within the organization.

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Programming Languages** | SQL, C, C++, Java, Core Java, Java 8, J2EE, Python, Scala, Pig Latin, HiveQL and Unix shell scripting |
| **Big Data Technologies** | Hadoop, HDFS, MapReduce, Hive, Pig, HBase, Impala, Hue, Sqoop, Storm, Talend, Kafka, Oozie, SparkSql, MlLib, SparkR, Flume, Zookeeper, Cassandra, Spark, **IBM BigInsights,** Cloudera CDH4, and Hortonworks. |
| **Databases** | Oracle, MySQL, SQL Server, DB2, Familiar with NoSQL (HBase, Cassandra, MongoDB) |
| **Scripting & Query Languages** | UNIX Shell scripting, SQL and PL/SQL. |
| **Web Technologies** | JSP, Servlets, JavaBeans, JDBC, AWT, Swing, JSF, XML, CSS, HTML, XHTML, JavaScript, AJAX. |
| **Hadoop Paradigms** | MapReduce, YARN, In-memory computing, High Availability, Real-time Streaming. |
| **Operating Systems** | Windows, UNIX, Linux distributions (Centos, Ubuntu), Mac OS |
| **Other Tools** | Eclipse, Tableau 9.0, JUnit, QTP, JIRA, QC (Quality Center) |

**Trainings & Certifications:**

* Edureka certified Apache Spark and Scala developer.
* Successfully earned certification on Apache Spark and Apache Hadoop with IBM Big Data University.
* Successfully earned certification on Big data analysis with Apache Spark and Distributed Machine Learning with Apache Spark with edX BerkeleyX Sponsored by Databricks.
* Attended Webinars on Spark, conducted by Databricks.

**PROFESSIONAL EXPERIENCE**

**NBC Universal New York, NY Jan 2016-present Role: Hadoop/Spark Developer**

**Responsibilities:**

* Installed and configured Apache Hadoop clusters using YARN for application development and Apache toolkits like Apache Hive, Apache Pig, Cassandra, Apache Spark, Zookeeper, Flume, Kafka and Sqoop.
* Worked extensively with importing metadata into Hive and migrated existing tables and applications to work on Hive and Spark.
* Involved in converting Hive QL queries into Spark transformations using Spark RDDs, and Scala.
* Developed Spark code using Scala and Spark-SQL/Streaming for faster testing and processing of data.
* Responsible for loading Data pipelines from web servers and Teradata using Sqoop with Kafka and Spark Streaming API.
* Used Spark for interactive queries, processing of streaming data and integration with popular NoSQL database for huge volume of data.
* Performance tuning the Spark jobs by changing the configuration properties and using broadcast variables
* Worked on performing transformations & actions on RDDs and Spark Streaming data
* Responsible in handling Streaming data from web server console logs
* Developed multiple POCs using Spark and deployed on the Yarn cluster, compared the performance of Spark, with Hive and SQL/Teradata.
* Implementing Sqoop jobs, PIG and Hive scripts for data ingestion from relational databases to compare with historical data.
* Implemented scripts in Spark to import and export data from Cassandra, Teradata, Hadoop.
* Developed RDD's/Data Frames in Spark using Scala and Python and applied several transformation logics to load data from Hadoop Data Lake to Cassandra DB.
* Used IMPALA to write and query the Hadoop data in HDFS or NoSql Cassandra and pull the data from Hive tables.
* Worked on various file formats and compressions Text, Avro, Parquet file formats, snappy, bz2, gzip compression.
* Used Spark API over Hadoop YARN to perform analytics on data in Hive.
* Created tasks for incremental load into staging tables, and schedule them to run.
* Used Apache Kafka as messaging system to load log data, data from UI applications into HDFS system.
* Implemented Hive optimized joins to gather data from different sources and run ad-hoc queries on top of them.
* Wrote Hive Generic UDFs to perform business logic operations at record level.
* Developed workflow in Oozie to automate the tasks of loading the data into HDFS and pre-processing with Pig and used Zookeeper to coordinate the clusters.
* Analyzed large amounts of data sets to determine optimal way to aggregate and report on it.
* Troubleshooting the cluster by reviewing Hadoop LOG files.
* Worked on reading multiple data formats on HDFS using Python Spark API.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs, Python and Scala.
* Used Zookeeper for various types of centralized configurations.

**Environment:** Hadoop 2.3.0, Spark core, Scala, Spark, Cassandra, Python, Spark Sql, Spark Streaming, Hive, Impala, Pig, Sqoop, Kafka, Zookeeper, Control-M, Java 8, and UNIX Shell Scripting.

**Vanguard Inc., Charlotte, NC July 2014- Dev 2015 Role: Big Data Engineer**

**Responsibilities:**

* Developed data pipeline using Spark, Hive 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 Aggregations on the fly to build the common learner data model and persists the data in HDFS.
* Hands on experience in designing, developing, and maintaining software solutions in Hadoop cluster.
* Exploring with the Spark improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark SQL, Data Frame, Spark Yarn.
* 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
* Experienced with Spark Streaming to ingest data into Spark Engine.
* Designed the ETL runs performance tracking sheet in different phases of the project and shared with Production team.
* Involved in converting Hive/SQL queries into Spark Transformations using Spark RDDs and Scala.
* Involved in using SQOOP for importing and exporting data between RDBMS and HDFS.
* Used Hive to analyze the Partitioned and Bucketed data and compute various metrics for reporting.
* Involved in developing Hive DDLS to create, alter and drop Hive tables.
* Involved in loading data from Linux file system to HDFS.
* Created PIG scripts to load, transform and store the data from various sources into HIVE metastore.
* 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 indata warehousing and Business Intelligence systems.
* Responsible for System performance management, Systems change / configuration management and Business requirements management.
* Primary contributor in designing, coding, testing, debugging, documenting and supporting all types of applications consistent with established specifications and business requirements to deliver business value. Identify and design most efficient and cost effective solution through research and evaluation of alternatives.
* Demonstrated Hadoop practices and broad knowledge of technical solutions, design patterns, and code for medium/complex applications deployed in Hadoop production.
* Ingested semi structured data using Flume and transformed it using Pig.

**Environment:** Spark, Hive, Pig, Spark SQL, Spark Streaming, Hbase, Sqoop, Python, Kafka, Cloudera, Scala, Scala IDE(Eclipse), Maven, HDFS.

**Philips HealthCare, Bangalore, India. May 2013- June 2014 Role: Hadoop Developer**

**Responsibilities:**

* Involved in all phases of development activities from requirements collection to production support.
* Detailed understanding of current system and find out the different sources of data.
* Involved in analyzing system failures, identifying root causes, and recommended course of actions.
* Wrote the shell scripts to monitor the health check of Hadoop daemon services and respond accordingly to any warning or failure conditions.
* Managing and scheduling Jobs on a Hadoop cluster.
* Designed a data warehouse using Hive.
* Deployed Hadoop Cluster in all modes.
* Developed Pig Latin scripts to extract the data from the web server output files to load into HDFS.
* Developed the Pig UDF'S to pre-process the data for analysis.
* Develop Hive queries for the analysts.
* Involved in Cluster setup
* Performed Batch processing of logs from various data sources using MapReduce.
* Defined UDFs using PIG and Hive to capture customer behavior
* Design and implement map reduce jobs to support distributed processing using java, Hive and Apache Pig.
* Create Hive external tables on the map reduce output before partitioning, bucketing is applied on top of it.
* Developed HIVE scripts for analyst requirements for analysis.
* Hands-on experience in using Hive partitioning, bucketing and execute different types of joins on Hive tables and implementing Hive SerDes like JSON and Avro.
* Worked on Developing custom MapReduce programs and User Defined Functions (UDFs) in Hive to transform the large volumes of data with respect to business requirement.
* Maintenance of data importing scripts using Hive and Map reduce jobs
* Developed and maintain several batch jobs to run automatically depending on business requirements
* Import and export data between the environments like MySQL, HDFS and Unit testing and Deploying for internal usage monitoring performance of solution

**Environment:** Apache Hadoop, Hive, PIG, HDFS, Java Map-Reduce, Core Java, Python, Maven, GIT, Jenkins, UNIX, MYSQL, Eclipse, Oozie, Sqoop, Flume and Cloudera Distribution, Oracle, Teradata and MySQL.

**L&T Technologies, Mumbai, India May 2011- April 2013 Java Developer**

**Responsibilities:**

* Reviewed the functional, non-functional requirements and high level designs.
* Worked with Agile Methodologies.
* Developed application using Struts MVC, Spring Dependency Injection and iBATIS as the ORM tool.
* Worked extensively in the backend Oracle database.
* Created logging system for development stage using Apache Commons Logging.
* Developed web pages using HTML, CSS, JavaScript, Angular Js and Free marker.
* Produced SOAP Web Service using Apache Axis. Generated WSDL files for the Service that served XML data.
* For logging and other functions used Spring AOP and Spring IOC Framework to implement them.
* Soap UI is used for web services testing.
* Used multithreading in programming to improve overall performance.
* Developed Multi-Threaded/Concurrent applications and Unix Shell Scripting.
* Implemented Collections for managing the dynamic data and other Core Java APIs.
* Worked on Garbage Collection for memory issues to improve the performance.
* Created functions and stored procedures in Oracle SQL/PL-SQL.
* Worked with XPath when need to identify objects in the XML.
* Developed lookup tables and views to speed up queries.
* Provide critical support to current order management systems being used as and when required.
* Used Jenkins as build management tool. Selenium was used to keep a track of bugs.
* Worked in Quality analysis and testing.

**Environment:** Java, EJB, Hibernate, Struts, Free Marker, Selenium, Oracle SQL, PL/SQL, XML, XPath, JUnit, Eclipse IDE, HTML, CSS, JavaScript, SOAP, JAX-B, JAX-RPC, JAX-WS, JAX-RS, Apache Tomcat, Apache Commons Logging, SOAP, Multi-Threaded/Concurrent , Collections , Stored procedures.

**ICICI Bank, Mumbai, India Feb 2010-Apr 2011 Role: Software Developer**

**Responsibilities**

* Core member of the one of the main product of the company
* The core project was developed using Core Java, Multithreading and JDBC
* Analyzed and understood the business requirements.
* Involved in various SDLC phases of the project.
* Connecting to MySQL database using JDBC
* Developed the presentation tier using Struts Framework.
* Used Hibernate as ORM and also involved in configuring hibernate to access database for various modules.
* Designed the database tables, and wrote the respective mapping files in hibernate and POJO classes.
* Used log4j for logging for the application.
* Build the Ant Scripts for building and deploying the application.
* Automated the settlement process with the application and Banks and sends the reports to the client.
* Developed the loyalty program applications to enhance the client business.
* Developed a protocol between the client and server using XML and build the parse to communicate between both ends.
* Developed a router application for the ICICI Bank to route the transactions.
* Implementation of PGP in Java to transfer and receive the confidential files between our servers and clients.
* Developed test cases and performed unit testing using JUnit.
* Used CVS to maintain the versions of the source code of the project.
* Implemented many file parses and XML parses in the application to communicate between the server and client.
* Developed the UI using HTML, JSP, jQuery, AJAX, JavaScript.
* Certification of the product with various Banks and meet the standards of the industry.

**Environment:** Core Java/J2EE, Multithreading, JDBC, Struts, Java Mail, JDBC, MySQL, MSSQL, Ant, Apache Tomcat, WSDL, SOAP, XML, Log4j, Hibernate, JSP, Ajax, jQuery, SFTP, PGP, Linux, jPOS.