**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 4+ year’s hands-on experience in Big-data/Hadoop Technologies.
* 4+ years of extensive experience in, Database development, ETL Tools, Data Analytics.

**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 analyzing data using HBase and custom Map Reduce programs in Java.
* Good understanding of NoSQL databases and hands on experience in writing applications on NoSQL databases like HBase.
* Worked on NoSQL database HBase for storing huge amounts of web logs.
* 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 HBase and its Integration with Hive.
* Used HBase 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 DataStage.
* Experience in developing the complex SQL queries, unions and multiple table joins and experience with views.
* Extensive experience in DataStage using Components like DataStage Designer, DataStage Manager, DataStage Director, DataStage Administrator, Information server and Parallel Extender. Designed Mapping documents ETL architecture documents and specifications.
* Extensive experience in Reverse engineering and Complete Compare Data models using data modelling tools from disparate systems and analyzing the same.
* Extensively experience in SQL, PL/SQL Stored Procedures, Functions and Packages to implement business logic.
* Involved in High level design and specification documents.
* Experience in integration of various data sources like Oracle, DB2 UDB, Teradata, SQL Server, XML and Flat files into Data Warehouse.
* Experience in using cloud components and connectors to make API calls for accessing data from cloud storage Amazon S3 in Talend open Studio.
* Expertise in creating mappings in TALEND using tMap, tJoin, tReplicate, tParallelize,
* Created Talend ETL jobs to receive attachment files from pop e-mail using tPop, tFileList and tFileInputMail and then loaded data from attachments into database and achieved the files.
* 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, Hue, Sqoop, Storm, Kafka, Oozie, SparkSql, SparkR, Flume, Zookeeper, Cassandra, Spark,Hortonworks. |
| **Databases** | Oracle, MySQL, SQL Server, DB2, Familiar with NoSQL (HBase, Cassandra, MongoDB) |
| **ETL** | DataStage 8.5, 8.1, DataStage Server, Talend 5.5/5.0. |
| **Scripting & Query Languages** | UNIX Shell scripting, SQL and PL/SQL. |
| **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 June 2016 – Till Date Role: Big Data/Hadoop Developer**

**Responsibilities:**

* Involved in Analysis, Design, System architectural design, Process interfaces design, design documentation.
* Responsible for developing prototypes of the selected solutions and implementing complex big data projects with a focus on collecting, parsing, managing, analyzing and visualizing large sets of data using multiple platforms.
* Understand how to apply technologies to solve bigdata problems and to develop innovative big data solutions.
* Developed various data loading strategies and performed various transformations for analyzing the datasets by using Hortonworks Distribution for Hadoop ecosystem.
* Worked extensively with importing metadata into Hive and migrated existing tables and applications to work on Hive and Spark.
* Involved in setting up Name node, Resource Manager, HBase Master and Oozie high availability.
* Importing and exporting data from different databases like DB2, Teradata into HDFS and HBASE using Sqoop.
* Created HBase tables to load large sets of structured, semi-structured and unstructured data coming from UNIX, NoSQL and a variety of portfolios.
* Responsible for Hive-HBase integration by defining external tables in hive and pointing the HBase as data store for better performance and lower I/O.
* Worked on HBase Java API to populate operational HBase table with Key value. Worked with Hadoop eco system covering HDFS, HBase and YARN .
* Loaded cache data into HBase using Sqoop.
* Created lots of external tables on Hive pointed to HBase tables.
* Analyzed HBase data in Hive by creating external partitioned and bucketed tables.
* Implemented scripts in Spark to import and export data from HBase, Teradata, Hadoop.
* Developed RDD's/Data Frames in Spark using Scala and applied several transformation logics to load data from Hadoop Data Lake to HBase.
* 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 HBase using Sqoop with Kafka and Spark Streaming API.
* Apply Spark Streaming to receive data from Kafka to do the continually data cleaning and aggregating, then store the data in HBase.
* Used Spark for interactive queries, processing of streaming data and integration with HBase 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 ,HBase and SQL/Teradata.
* Implementing Sqoop jobs, PIG and Hive scripts for data ingestion from relational databases to compare with historical data.
* 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 HBase system.
* Implemented a proof of concept (Poc's) using Kafka, Strom, HBase for processing streaming data.
* 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 HBase
* 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.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala.
* Used Zookeeper for various types of centralized configurations.

**Environment*:*** Hadoop, Spark core, Scala, Spark, Hortonworks distribution, Python, Spark SQL, Spark Streaming, Hive, HBase, Pig, Sqoop, Kafka, Zookeeper, Java 8, and UNIX Shell Scripting.

**Intermountain Health Care, Salt Lake City, UT July 2014- June 2016 Role: Hadoop/Spark Developer**

**Responsibilities:**

* Developed data pipeline using Spark, Hive and HBase to ingest customer behavioral data and histories into Hadoop cluster for analysis.
* 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.
* Build scalable distributed data solutions using Hadoop Hortonworks Distribution. Ensure data accuracy and create reports.
* Designed the HBase schemes based on the requirements
* Wrote multiple queries to pull data from HBase .
* Extracted files from NoSQL database HBase through sqoop and placed in HDFS for processing.
* Facilitated access to large data sets utilizing Hive/HBase and custom solutions on Hadoop Ecosystem.
* HBase data migration and validation
* Involved in creating Hive tables, loading structured data and writing hive queries which will run internally in MapReduce way
* Efficiently put and fetched data to/from HBase by writing MapReduce job.
* Developed MapReduce jobs to automate transfer of data from/to HBase.
* 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.
* Used Pattern matching algorithms to recognize the fraudulent customer across different sources and built risk profiles for each customer using Hive and stored the results in HBase.
* 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.
* ETL: Data extraction, managing, aggregation and loading into HBase .
* Involved in developing Hive DDLS to create, alter and drop Hive tables.
* Loaded the data from different source such as HDFS or HBase into Spark RDD and do in memory data computation to generate the output response.
* 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.
* Export filtered data into HBase for fast query
* Experienced in managing and reviewing Hadoop log files.
* Involved in data 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*:*** Hadoop, Spark, Hive, Pig, Spark SQL, Spark Streaming, HBase, Sqoop, Python, Kafka, Hortonworks, Scala, Scala IDE(Eclipse), Maven, HDFS.

**Verizon Communications, Tampa, FL Oct 2012- July 2014   
Talend / ETL Developer**

**Roles & Responsibilities:**

* Closely worked with Data Architects in designing of tables and even involved in modifying technical Specifications.
* Involved in the Huge development of a ETL Process Which Dynamically fetches the Source Data and Inserts the files in to S3.
* Involved in the Development of copying Data from S3 to RedShift using the Talend Process.
* Involved in writing custom copy command query's and used the context extensively and implemented in tRedshift Row Component.
* Extensively used t system component to push large sets of data to s3.
* Developed a Talend Code for S3 Tagging in the Process of Moving data from source to S3
* Involved in Extraction, Transformation and Loading of data.
* Utilized Talend components like tS3Put, tS3Get, tS3File List, tRedshift Row, tRedshiftUnload, tRedshift BulkExec, tflowtoiterate, tSetGlobalVar.
* Integrated Redshift SSO cluster with Talend
* Involved in Integrating IAM Roles in Talend Components.
* Used Components in Talend Like (tMap, Tfilelist, Tjava, Tlogrow, ToracleInput, ToracleOutput, tsendEmail etc, tDie, tConvertType, tFlowMeter, tLogCatcher, tRowGenerator, tRedshiftrow, tsystem)
* Involved in Creating the Talend process which captures the Job Exception and Table Level Stats.
* Involved in Performance Tuning of Data Ingestion Process to S3 in all the Job Servers
* Responsible for the Scheduling and Deployment in TAC.
* Involved in implementing the custom build schema converter to generate create table DDL for redshift.
* Used T parallelize component in running the Jobs in Parallel.
* Created Talend Jobs for Batch Processing of Data. Worked on Multi-Threading in Talend.

**Environment:** Talend 5.5/5.0, Oracle 11g, Teradata SQL Assistant, HDFS, MS SQL Server 2012/2008, PL/SQL, Agile Methodology, TOAD, ERwin, AIX, Shell Scripts, AutoSys, SVN.

**L&T Technologies, Mumbai, India May 2011- April 2012 DataStage Developer**

**Roles & Responsibilities:**

* Worked closely with Business analysts and Business users to understand the requirements and to build technical specifications.
* Involved in all business meetings to understand the existing logic, different member information and agency data to come up with the best IT solution.
* Responsible to create Source to Target (STT) mappings.
* Involved in day to day production support activities.
* Worked on various defects raised by concerned business teams from various entities.
* Developed and supported the Extraction, Transformation and Load process (ETL) for a data warehouse from various data sources using DataStage Designer and to load the target tables using the DataStage Designer.
* Worked on various stages like Transformer, Join, Lookup, Sort, Filter, Change Capture and Apply stages.
* Developed parallel jobs using various Development/debug stages (Peek stage, Row generator stage, Column generator stage, Sample Stage) and processing stages (Aggregator, Change Capture, Change Apply, Filter, Sort & Merge, Funnel, Remove Duplicate Stage.
* Designed job sequencer to run multiple jobs with dependency and email notifications.
* Involved in Unit Testing, SIT and UAT Worked with the users in data validations.
* Extensively worked in improving performance of the jobs by avoiding as many transforms as we can.
* Prepared documentation for unit, integration and end - to - end testing.
* Optimized/Tuned DS jobs for better performance and efficiency
* Responded to customer needs; self-starter and customer service oriented.
* Provided support and guidance by creating Release, Deployment & Operation guide documents.
* Involved in Performance tuning of complex queries.
* Interacted with Teradata DBA team for creation of primary and secondary indexes on Data Warehouse tables.
* Develop Universe model as per the business requirements to use for Web rich client.

**Environment:** IBM Infosphere DataStage 8.5 (Designer, Director, Administrator), Teradata, Oracle 10g, Sequential files and Mainframe files, COBOL and UNIX Shell Scripting.

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

**Responsibilities:**

* Core member of the one of the main products 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.