**FANI TEJA KOLLU**

**Summary:**

* Around 9 years of programming experience involved in all phases of Software Development Life Cycle (SDLC).
* Around 5 Years of Big data related architecture experience developing **Hadoop applications**.
* Experienced in the Hadoop ecosystem components like **Hadoop Map Reduce, Cloudera, Hortonworks, HBase, Oozie, Hive, Sqoop, Pig, Flume, Kafka, Storm, Spark, MongoDB, and Cassandra.**
* Proven Expertise in performing analytics on Big Data using **Map Reduce, Hive and Pig**.
* Implemented POC to migrate map reduce jobs into **Spark RDD transformations using Scala.**
* Developed Apache Spark jobs using Scala in test environment for faster data processing and used Spark SQL for querying.
* Hands on developing and debugging **YARN (MR2)** Jobs to process large Datasets.
* Data Processing: Processed data using **MapReduce and Yarn**. Worked on **Kafka** as a proof of concept for log processing
* Experienced with performing real time analytics on **NoSQLdatabases like HBase and Cassandra.**
* Good knowledge in working with **Impala, Storm and Kafka.**
* Experienced with Dimensional modeling, Data migration, Data cleansing, Data profiling, and ETL Processes features for data warehouses.
* Experience web services, data modeling, content processing, data replication.
* Worked with Oozie workflow engine to schedule time based jobs to perform multiple actions.
* Experienced in importing and exporting data from **RDBMS into HDFS using Sqoop.**
* Familiarity with distributed coordination system Zookeeper.
* Design and develop test plan for test cases based on functional and design.
* Analyzed large amounts of data sets using **Pig scripts and Hive scripts**.
* Experience in data warehousing with **ETL tool Oracle Warehouse Builder (OWB).**
* Hands on experience in working with database like **Oracle, MySQL and PL/SQL.**
* Good working knowledge on processing Batch applications.
* Experienced in developing web-based applications using **Python, Django, PHP, XML, CSS, HTML, Java Script and JQuery.**
* Experience with streaming data using IBM streams processing language.
* Experience in capturing and analyze the data in motionusing info sphere stream.
* Experienced in writing **MapReduce programs and UDFs for both Hive and Pig in Java.**
* Experienced in developing Web Services with Python programming language.
* Involved in Design and Development of technical specifications using **Hadoop Echo System tools**
* Cutting edge experience on Splunk (Log based performance monitoring tool).
* Experience with configuration of Hadoop Ecosystem components: **Hive, HBase, Pig, Sqoop and Flume.**
* Good experience in Hive partitioning, bucketing and perform different types of joins on Hive tables and implementing **Hive SerDe like JSON a Avro.**
* Experience in using different file formats - **Avro, Sequence Files, ORC, JSON and Parquet**.
* Experience in Performance Tuning, Optimization and Customization.
* Experience with Unit Testing Map Reduce programs using **MRUnit, JUnit.**
* Experience in Active Development as well as onsite coordination activities in web-based, client/server and distributed architecture using Java, J2EE which includes Web services, Spring, Struts, Hibernate and JSP/Servlets along with incorporating MVC architecture.
* Good working knowledge on servers like Tomcat, Web Logic 8.0.
* Experience with **python scripting**.
* Ability to work in teams as well as an individual, quick learner and able to meet deadlines.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Big Data Eco System** | HDFS, Map Reduce, Hive, Pig, HBase, Spark, Spark Streaming, Spark SQL, Kafka, Cloudera CDH4, CDH5, Hortonworks, Hadoop Streaming,Splunk, Zookeeper, Oozie, Sqoop, Flume, Impala, Solr, and Ranger. |
| **No SQL** | HBase, MongoDB, Couchbase, Neo4j, Cassandra |
| **Languages** | Java/ J2EE, SQL, Shell Scripting, C/C++, Python, Scala |
| **Web Technologies** | HTML, JavaScript, CSS, XML, Servlets, SOAP, Amazon AWS, Google App Engine |
| **Web/ Application Server** | Apache Tomcat Server, LDAP, JBOSS, IIS |
| **Operating system** | Windows, Macintosh, Linux and Unix |
| **Frameworks** | Springs, MVC, Hibernate, Swings |
| **DBMS / RDBMS:** | Oracle 11g/10g/9i, SQL Server 2012/2008, MySQL |
| **IDE:** | Eclipse, Microsoft Visual Studio (2008,2012), NetBeans, Spring Tool Suits |
| **Version Control:** | SVN, CVS and Rational Clear Case Remote Client, GitHub, Visual Studio |
| **Tools:** | FileZilla, Putty, TOAD SQL Client, MySQL Workbench, ETL, DWH, JUnit, SQL Oracle Developer, WinScp, Tahiti, Cygwin, pentaho |

**PROFESSIONAL EXPERIENCE**

**Fidelity Investments, Two-Destiny Way, TX February 2015 to Present**

**Sr Hadoop Developer**

At Fidelity, our goal is to make financial expertise broadly accessible and effective in helping people live the lives they want. We do this by focusing on a diverse set of customers: - from 23 million people investing their life savings, to 20,000 businesses managing their employee benefits to 10,000 advisors needing innovative technology to invest their clients’ money. We offer investment management, retirement planning, portfolio guidance, brokerage, and many other financial products.

**Responsibilities:**

* Assisted in the development of high quality analytical reports of weather and GIS data.
* Used Talend to generate optimized code to load, transform, enrich, and cleanse data inside Hadoop.
* Moved Relational Database data using Sqoop as ETL tool into the Data Lake,Hive Dynamic partition tables.
* Build Hadoop clickstream workflows using Apache Hive and Pig for extraction, transformation and loading of data.
* Imported unstructured data like logs from different web servers to HDFS using Flume and developed MapReduce jobs for log analysis, recommendations and analytics.
* Involved in real-time data processing using Storm.
* Expertise in real-time analytics, machine learning and continuous monitoring of operations using Storm.
* Converted and loaded local data files into HDFS through the UNIX shell.
* Developed Pig Latin scripts to extract the data from the web server output files to load into HDFS.
* Created HBase tables to store variable data formats coming from different portfolios.
* Worked with NoSQL database Hive, Hbase to create tables and store data.
* Worked with Apache Nifi to Develop Custom Processors for the purpose of processing and disturbing data among cloud systems
* Worked on Apache Nifi to Uncompress and move json files from local to HDFS
* Implemented HBase Co-processors to notify Support team when inserting data into HBase Tables.
* Used Kafka along with HBase to render data streaming.
* DevelopedCassandra data model to match the business requirements
* Involved in Administration of Cassandra cluster along with Hadoop, Pig and Hive.
* Analyzed the customer behavior by performing click stream analysis and to ingest the data used flume.
* Developed spark scripts by using scala shell as per requirements.
* Created Views from Hive Tables on top of data residing in Data Lake.
* Build advanced ETL logic on clickstream, log and tax data based on complex technical and business requirements.
* Developed Spark code using Scala and Spark-SQL for faster testing and data processing.
* In data exploration stage used hive and impala o get some insights about the customer   
  data.
* Evaluated Spark's performance vs impala on transactional data.
* Involved in converting Hive/SQL queries into Spark transformations using Spark RDD, scala.
* Involved in configuring batch job to perform ingestion of the source files in to the Data lake
* Wrote and executed various MYSQL database queries from Python using Python-MySQL connector and MySQL dB package.
* Worked on Github to check-in and checkout source code.
* Wrote scripts in Python for extracting data from HTML file.
* Worked with NoSQL Cassandra to store, retrieve, and update and manage all the details for Ethernet provisioning and customer order tracking.
* Implemented Spark for fast interactive data analysis of datasets loaded in RDD.
* Analyzed the data by performing Hive queries (HiveQL), ran Pig scripts, Spark SQL and Spark streaming.
* Developed tools using Python, Shell scripting, XML to automate some of the menial tasks
* Used Pig for aggregation, cleaning and incremental ETL functions and developing UDFs for filtering.
* Wrote scripts in Python for extracting data from HTML file.
* Working experience on maintaining MySQL databases creation and setting up the users and maintain the backup of cluster metadata databases with Cron jobs.
* Redshift data model and Tableau Server configurations to provide guaranteed response time for reports.
* Implemented AWS provides a variety of computing and networking services to meet the needs of applications.
* Migrated an existing on-premises application to AWS.
* Developed Pig UDFs to specifically preprocess and filter data sets for analysis.
* Used Teradata database management system to manage the warehousing operations and parallel processing.
* Validated data sets graphically with Excel and did touch ups in Photoshop.
* Designed & scheduled workflows for updating system reports using Oozie.
* Ability to working with team environment and solved problems.

**Environment: Cloudera, Avro, HBase, HDFS, Hive, Pig, Java, SQL, Sqoop, Flume, Oozie, Java (jdk 1.7), Eclipse, Splunk ,YARN, SQL Server, Spark,python,Hortonworks, Zookeeper, SVN, Talend**

**AT&T Dallas, Texas April 2014 to Jan 2015**

**Big Data / Hadoop Developer**

 In AT&T our goal is to provide rich insight into consumer behaviors, preferences and experiences in order to improve the customer experience across a broad range of vertical markets. Analyzes, designs, programs, debugs and modifies software enhancements and/or new products used in local, networked, or Internet-related computer programs. May interact with users to define system requirements and/or necessary modifications.

**Responsibilities:**

* Worked on importing data from various sources and performed transformations using MapReduce, Hive to load data into HDFS.
* 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 Map Reduce.
* Involved in the process of data acquisition, data pre-processing and data exploration of telecommunication project in scala.
* Experience with Map Reduce coding.
* Solved small file problem using Sequence files processing in Map Reduce.
* Written various Hive and Pig scripts.
* Experienced with performing real time analytics on **NoSQLdatabases like HBase and Cassandra**
* Experience in Upgrading cluster, CDH and HDP Cluster.
* Involved in the process of data acquisition, data pre-processing and data exploration of telecommunication project in Scala.
* Used flume, sqoop, hadoop,spark and oozie for building data pipeline
* Created HBase tables to store variable data formats coming from different portfolios.
* Experience in upgrading Hadoop cluster hbase/zookeeper from CDH3 to CDH4.
* Performed real time analytics on HBase using Java API and Rest API.
* Developed MapReduce programs to cleanse the data in HDFS obtained from heterogeneous data sources to make it suitable for ingestion into Hive schema for analysis
* Implemented complex MapReduce programs to perform joins on the Map side using distributed cache.
* Setup flume for different sources to bring the log messages from outside to HDFS.
* Implemented HBase Co-processors to notify Support team when inserting data into HBase Tables.
* Worked on compression mechanisms to optimize MapReduce Jobs.
* Real time experience with analytics and BI.
* Wrote Python scripts to parse XML documents and load the data in database..
* Experienced with working on Avro Data files using Avro Serialization system.
* Implemented business logic by writing UDF's in Java and used various UDF's from Piggybanks and other sources.
* Continuous monitoring and managing the Hadoop cluster using Cloudera Manager.
* Unit tested and tuned SQLs and ETL Code for better performance.
* Monitored the performance and identified performance bottlenecks in ETL code. .
* Exported the analyzed data to the relational databases using Sqoop for visualization and to generate reports for the BI team.

**Environment: Map Reduce, HBase, HDFS, Hive, Pig, Java, SQL, Cloudera Manager, Sqoop, Flume, Zookeeper, YARN, Oozie, Java, Eclipse**

***Barclays, Wilmington, DE Oct’12 – Mar’14***

**Big Data/Hadoop Developer**

Barclays is an international financial services provider engaged in personal banking, credit cards, corporate and investment banking and wealth management with an extensive presence in Europe, the Americas, Africa and Asia. Barclays’ purpose is to help people achieve their ambitions – in the right way. With over 300 years of history and expertise in banking, Barclays operates in over 50 countries and employs approximately 140,000 people.

**Responsibilities:**

* Created Hive tables and working on them using Hive QL.
* Involved in installing Hadoop Ecosystem components.
* Validated Name node, Data node status in a HDFS cluster.
* Importing and exporting data from HDFS to RDBMS and vice-versa using SQOOP.
* Experienced in developing HIVE Queries on different data formats like Text file, CSV file.
* Developed multiple MapReduce jobs in java for data cleaning and preprocessing.
* Collecting and aggregating large amounts of log data using Apache Flume and staging data in HDFS for further analysis
* Installed and configured Hadoop cluster in Test and Production environments
* Performed both major and minor upgrades to the existing CDH cluster
* Code review as per the customer coding standards.
* Testing and providing the valid test data to users as per requirement.
* Weekly meetings with technical collaborators and active participation in code review sessions with senior and junior developers.
* Responsible to manage data coming from different sources.
* Supporting Hbase Architecture Design with the Hadoop Architect team to develop a Database Design in HDFS.
* Involved in HDFS maintenance and loading of structured and unstructured data.
* Wrote Hive queries for data analysis to meet the business requirements.
* Installed and configured Pig and also written Pig Latin scripts.
* Developed UDFs for Pig Data Analysis.
* Involved in managing and reviewing Hadoop log files.
* Developed Scripts and Batch Job to schedule various Hadoop Program.
* Utilized Agile Scrum Methodology to help manage and organize a team of 4 developers with regular code review sessions.
* Handled importing of data from various data sources, performed transformations using Hive, MapReduce.
* Analyzed the data by performing Hive queries and running Pig scripts to know user behavior.
* Continuous monitoring and managing the Hadoop cluster through Cloudera Manager.
* Installed Oozie workflow engine to run multiple Hive and Pig jobs.
* Developed Hive queries to process the data and generate the data cubes for visualizing.

**Environment: Java, Hadoop, MapReduce, HDFS, Hive, Pig, Sqoop, Zookeeper, Linux, XML, Eclipse, Cloudera**

**Tech Mahindra, India Jan 2010 to May 2012**

**Java Developer**

Tech Mahindra represents the connected world, offering innovative and customer-centric information technology services and solutions, enabling Enterprises, Associates and the Society to Rise™.

**Responsibilities:**

* Performed requirements analysis and design and used of Rational Unified Process for analysis, design and documentation purposes.
* Interacted with the client to understand the requirements thoroughly in a short span of time.
* Developed Use cases, traceability matrices, design specification and test documents including UATs, training and implementation manuals.
* Involved in design and development of the architecture of the web applications in JSP.
* Developed various Java classes, SQL queries and procedures to retrieve and manipulate the data from backend Oracle database using JDBC.
* Involved in developing programs for parsing the XML documents using XML Parser.
* Involved in tuning and profiling the application for better data transactions and performance used JProbe for the same.
* Designed and developed Enterprise Stateless and Stateful Session beans to communicate with the Container Managed Entity Bean backend services.
* Involved in Unit Testing and Integration Testing.
* Designed and developed various modules of the application with J2EE design architecture and frameworks like Spring MVC architecture and Spring Bean Factory using IOC, AOP concept.
* Followed agile software development with Scrum methodology.
* Wrote application front end with HTML, JSP, JSF, Ajax/JQuery, Spring Web Flow and XHTML.
* Used J Query for UI centric Ajax behavior.
* Implemented JAVA/J2EE design patterns such as Factory, DAO, Session Façade and Singleton.
* Used Hibernate in persistence layer and developed POJO's, Data Access Object (DAO) to handle all Database operations.
* Implemented features like logging, user session validation using Spring-AOP module.
* Developed server-side services using Java, Spring, Web Services (SOAP, WSDL, JAXB, JAX-RPC)
* Worked on Oracle as the backend database.
* Used JMS for messaging.
* Used Log4j to assign, track, report and audit the issues in the application.
* Develop and execute Unit Test plans using J Unit, ensuring that results are documented and reviewed with Quality Assurance teams responsible for integrated testing.
* Worked in deadline driven environment with immediate feature release cycles.

**Environment- Java, Spring, Hibernate, JSP, HTML, CSS, XML, JavaScript, JQuery, JUnit, AJAX, Multi-Threading, Oracle, Web Service - SOAP, WebSphere, MYSQL.**

**HCL Info Systems Ltd Jun’ 08 – Dec’09**

**Java Developer**

**Responsibilities:**

* Performed in various phases of the Software Development Life Cycle (SDLC)
* Developed user interfaces using JSP framework with AJAX, Java Script,HTML,XHTML,and
* CSS
* Performed the design and development of various modules using CBD Navigator Framework
* Deployed J2EE applications in Web sphere application server by building and deploying ear fileusing ANT script
* Created tables, stored procedures in SQL for data manipulation and retrieval.
* Used technologies like JSP, JavaScript and Tiles for Presentation tier.
* Performed the design and development of various modules using CBD Navigator Framework
* Deployed J2EE applications in Web sphere application server by building and deploying ear file using ANT script.
* Created tables, stored procedures in SQL for data manipulation and retrieval.
* Used technologies like JSP, JavaScript and Tiles for Presentation tier.
* CVS tool is used for version control of code and project documents.
* Application Server, UML, JUnit, JTest, Netbeans, Windows 2000.

**Environment:** JSP, Servlets, JDK, JDBC, XML, JavaScript, HTML, Spring MVC, JSF, Oracle 8i, Sun Application Server, UML, JUnit, JTest, Netbeans, Windows 2000.

**EDUCATIONAL QUALIFICATION:**

* Bachelor of Engineering in Electronic and Communication from JNTU, India 2004-2008