**Yeswanth P**

Phone: 302-990-2580 | yeswanthp28@gmail.com

**Professional Summary**

* Around **9** years of professional experience in IT field and **6** years of experience in **Hadoop** ecosystem for **large**-**scale** **data** processing both in **real** time and **batch**
* Good knowledge on **Spark**, **HDFS, MRv1, MRv2 (YARN)**, **HIVE, HBase, Sqoop, Oozie, Pig, Hue and Kafka**
* Extensive knowledge in programming with **Resilient Distributed Datasets (RDDs) and DataFrames** in Spark
* Strong experience **troubleshooting** spark applications and various **performance** considerations to take for efficient memory handling of long running spark applications
* Experience integration of **Kafka** with **Spark** for real time data processing
* Developed custom **Kafka** **producer** and **consumer** for different publishing and subscribing to Kafka topics
* Good experience working with **Spark-streaming** API
* Experience in importing data from data warehouses such as Teradata and Oracle into the HDFS and exporting the processed data back into databases using **Sqoop**
* Implemented **Flume** for collecting, aggregating and moving large amount of server logs and streaming data to HDFS.
* Experienced in working with spark eco system using Spark SQL and Scala queries on different formats like Text file, Avro Parquet files.
* Expertise in working with **Hive** and using **partitioning**, **bucketing**, writing and **optimizing** HiveQL queries
* Strong experience developing custom **UDF’s** and using custom **SerDe’s** in Hive
* Excellent understanding of Hadoop architecture and hands-on experience in writing MapReduce programs
* Experience working with AWS **EMR** clusters, **Cloudera** and **Hortonworks** Distributions
* Hands on experience in installing, configuring and deploying Hadoop distributions in cloud environments (**Amazon Web Services**)
* Good knowledge on NoSQL data stores like **Cassandra, MongoDB** and **HBase**
* Good knowledge in analyzing data using **Pig Latin** and **Hive QL**
* Experience in cluster capacity planning, deploying, performance tuning, administering and monitoring Hadoop ecosystem
* Good knowledge on software development life cycle (SDLC)
* Experience in developing MVC architecture using **Servlets, JSP, Struts Framework, Hibernate Framework and Spring Framework**
* Experience in using Java IDE tools like IBMWebSphere Studio Application Developer **(WSAD), Rational Application Developer, Eclipse** and familiar with other IDE's like **Net Beans, JBuilder, and JDeveloper**
* Good knowledge of Data Modeling, Data Warehousing concepts, OLAP Concepts, Star Schema, Snowflake Schema, and Entity-Relationship Diagrams
* Extensive work experience in ETL processes consisting of data sourcing, data transformation, mapping and loading of data from multiple source systems into Data Warehouse using Informatica Power Center.
* Proficient using version control tools like **SVN** and **GIT**
* Ability to work on multiple tasks, strong interpersonal, analytical and presentation skills
* Excellent knowledge in **Agile software development methodology**

**Technical Skills**

|  |  |
| --- | --- |
| Hadoop Ecosystem | HDFS, Hive, Pig, Flume, Oozie, Zookeeper, HBase, Sqoop, Kafka, Storm, Elastic Search, Spark |
| Languages | C, C++, Java, Python, Scala, PIG Latin, Linux, Shell Scripting, Hive QL |
| Operating System | Linux, Windows 7 |
| Databases | MySQL, Oracle, SQL Server |
| NoSQL | Cassandra, MongoDB, HBase |
| Web Development | HTML, CSS, JavaScript, R Shiny, Servlets, Struts, JSP, Hibernate, Spring |
| Build tools | Ant, Maven |
| Version Control | SVN and GIT |
| Cluster Management and Monitoring | Cloudera Manager, Ambari |

**Work Experience:**

***InterContinental Hotels Group, Atlanta, GA November 2016 – Present***

**Hadoop/Spark Developer**

***Responsibilities:***

* Worked on building end-to-end **Scala** based **Spark** applications for **cleansing**, **auditing** and **transforming** raw data feeds from multiple report suites
* Developed custom **UDF’s** in **Scala** to be used in **spark**-**SQL**
* Improved the performance of existing **spark** applications
* Extensively worked on **HIVE**, created numerous internal and external tables as part of the requirements
* Written custom **UDF’s** in HIVE according to business requirements
* Hands on experience in data loading techniques like **Sqoop.**
* Involved in loading data from UNIX file system to HDFS.
* Experience in using **Oozie** for workflow design and **Oozie** coordinator for scheduling workflows
* Used Spark **DataFrames**, **Spark**-**SQL** and **Spark** **MLLib** extensively
* Created **batch** and **real** time **pipelines** using Spark as the main processing framework
* Integrated **Kafka** with **Spark** **Streaming** for real time data processing
* Work closely with business and transforming business requirements to technical requirements
* Hands on experience in working with **AWS** Cloud Services like **EMR**, **S3** and **Redshift**
* Been part of Design Reviews & Daily Project Scrums
* Successfully loaded files to **S3** from Teradata, and loaded from **S3** to **Redshift**
* Knowledge in installing and configuring various services on the Hadoop cluster
* Worked on exporting data into relational database using **Sqoop** for making it available for visualization for the BI team
* Weekly meetings with technical collaborators and active participation in **code** **review** sessions with senior and junior developers using **Agile**/scrum methodologies

**Environment:** CDH, AWS, HDFS, Hive, Spark, Spark Streaming, Spark-SQL, Spark MLLib, Oozie, Unix, Kafka, Sqoop, Java, Scala, Git, Shell Scripting, Eclipse

***Citi Bank, Edison, NJ May 2015 – October 2016***

**Hadoop/Spark Developer**

***Responsibilities:***

* Involved in loading and transforming large sets of **Structured** and **Semi-Structured** data and analyzed them by running **Hive** queries and **Pig** scripts
* Developed **MapReduce** (YARN) programs to **cleanse** the data in **HDFS** obtained from multiple data sources to make it suitable for ingestion into **Hive** for analysis
* Designed and developed **Spark** jobs to process data coming in different file formats like **XML**, **CSV**, and **JSON**
* Experience in **hive** **partitioning**, **bucketing** and perform **joins** on hive tables and utilizing hive **SerDes** like REGEX, JSON and AVRO.
* Migrated existing python programs to Spark using Scala.
* Executed tasks for upgrading cluster on the staging platform before doing it on production cluster
* Performed maintenance, monitoring, deployments, and upgrades across infrastructure that supports all our Hadoop clusters.
* Configured deployed and maintained multi-node Dev and Test Kafka Clusters.
* Optimized **HIVE** analytics SQL queries, created tables/views, written custom **UDFs** and Hive based exception processing.
* Migrating the coding from Hive to Apache Spark and Scala using Spark SQL, RDD.
* Involved in transforming the large data from **Teradata** to HDFS and vice versa using Sqoop incremental imports
* Stored and processed data by using low level Java API’s to ingest data directly to HBase
* Exported the analyzed data to the relational databases using **Sqoop** and to generate reports for the BI team
* Experience in managing and reviewing **Hadoop**log files
* Supported in setting up **QA** environment and updating configurations for implementing scripts with Pig
* Installed and configured various components of **Hadoop** ecosystem
* Replaced default Derby metadata storage system for Hive with MySQL system

**Environment:** CDH 5, Pig, Hive, Map Reduce, YARN, Sqoop, Spark, Scala, Kafka, HBase, Teradata, SQL Server, Shell Scripting

***Xceed Technologies, Washington, DC December 2013 – April 2015***

**Sr. Hadoop Developer**

***Responsibilities:***

* Developed custom **MapReduce** jobs to perform data **transformations**
* Created few **Hive** **UDF's** to parse complex columns
* Involved in End-to-End implementation of **ETL** logic
* Developed bash scripts to bring the log files from **FTP** server and then processing it to load into Hive tables
* Moved data from HDFS to **HBase** using **Map** **Reduce** and **BulkOutputFormat** class
* Worked collaboratively with all levels of business stakeholders to architect, implement and test Big Data based analytical solution from disparate sources.
* Developed SPARK CODE using SCALA and Spark-SQL/Streaming for faster testing and processing of data.
* Written automated **HBase** test cases for data quality checks using **HBase** command line tools
* Developed **Oozie** workflows for daily incremental loads, which gets data from Teradata and then imported into hive tables
* Implemented various **Hive** queries for analysis and call them from java client engine to run on different nodes
* Loading the data from the different Data sources like (**Teradata**, DB2, Oracle and flat files) into **HDFS** using **Sqoop** and load into Hive tables, which are partitioned
* Used dashboard tools like **Tableau** connecting directly to Hive
* Created **HBase** tables and column families to store the user event data

**Environment:** Hadoop, Map Reduce, HDFS, Hive, Sqoop, Oozie, Scala, Spark, Kafka, HBase, Java, Linux, Maven, Teradata, SVN

***Pacific Electric and Gas Company, San Francisco, CA June 2012 – November 2013***

**Hadoop Developer**

***Responsibilities:***

* Developed Java Rest Services as part of functional requirements
* Installed and configured **Hadoop** and responsible for **maintaining** cluster and managing and reviewing Hadoop log files
* Installed and configured Hadoop MapReduce, HDFS and developed multiple MapReduce jobs in Java for data **cleansing** and **preprocessing**
* Importing and exporting data into HDFS and Hive using Sqoop
* Written Hive UDF’s to implement business logic in hive scripts
* Responsible for building scalable distributed data solutions using Hadoop
* Involved in loading data from edge node to HDFS using shell scripting
* Created HBase tables to store variable data formats of PII data coming from different portfolios
* Worked with using different kind of compression techniques to save data and optimize data transfer over network using **LZO**, **Snappy**, **Bzip2** etc.
* Used Sqoop to store the data into HBase and Hive
* Worked on installing cluster, **commissioning** & **decommissioning** of DataNode, NameNode high availability, **capacity** **planning**, and slots configuration
* Creating Hive tables, dynamic partitions, buckets for sampling, and working on them using HiveQL
* Used **Pig** to parse the data and store in **Avro** format
* Stored the data in tabular formats using Hive tables and **Hive** SerDes
* Collecting and aggregating large amounts of log data using Apache **Flume** and staging data in HDFS for further analysis
* Worked with NoSQL databases like **HBase** for creating HBase tables to load large sets of semi structured data coming from various sources
* Implemented a script to transmit information from **Oracle** to HBase using Sqoop
* Implemented **MapReduce** programs to handle semi/unstructured data like XML, JSON, and sequence files for log files
* Fine-tuned **Pig** queries for better performance
* Exported the analyzed data to the relational databases using **Sqoop** for visualization and to generate reports for the BI team

**Environment:** Hadoop, MapReduce, HDFS, Pig, Hive, HBase, Flume, Oracle, Java, J2EE, XML, JUnit, Tomcat 6

***Bank of East Asia, New York, NY November 2010 – May 2012***

**Senior Java/J2EE Developer**

***Responsibilities:***

* Used **HTML**, **JavaScript**, **Servlets**, and **JSP** in the presentation tier. Developed business delegates to communicate with the Enterprise Java Beans (**EJB**) and web services residing on the business tier
* Responsible and active in the analysis, definition, design, implementation, management, and deployment of full software development life cycle of project
* Development of the application that was used to create dynamic **JSPs**, given input from a database table containing information about the contents of the **JSP** being developed
* Implemented Business Logic Components using Spring Core and navigation using **Spring MVC**
* Implemented Persistent entities and Persistent Layers using **Hibernate** API and Interfaces
* Implemented **SOAP** based web Services for Enterprise Logging Application Module
* Responsible for coding **SQL** Statements and Stored procedures for back end communication using **JDBC**
* Developed an API to write **XML** documents from a database and utilized **XML** and **XSL** Transformation for dynamic web-content and database connectivity
* Involved in the development of presentation layer and GUI framework in **JSP**. Client-side validations were done using **JavaScript**
* Developed the presentation layer using **CSS** and **HTML** taken from bootstrap to develop for browsers

**Environment:** HTML, CSS, JavaScript, J2EE, Java, XML, EJB, JAXB, JNDI, SOAP Web Services, Hibernate, Spring Framework, Linux, SQL Server and Oracle

***Mphasis, Bengaluru, India August 2009 – September 2010***

**Java Developer**

***Responsibilities:***

* Participated in requirement gathering and converting the requirements into technical specifications
* Developed UI using **HTML**, **JavaScript**, **JSP**, and developed Business Logic and interfacing components using Business Objects, **XML**, and **JDBC**
* Designed user-interface and checking validations using JavaScript
* Involved in design of **JSP’s** and **Servlets** for navigation among the modules
* Developed various **EJBs** for handling business logic and data manipulations from database
* Managed connectivity using **JDBC** for querying/inserting & data management including triggers and stored procedures
* Developed **SQL** queries and Stored Procedures using **PL/SQL** to retrieve and insert into multiple database schemas
* Developed the **XML** Schema and Web services for the data maintenance and structures
* Written test cases in **JUnit** for unit testing of classes
* Provided Technical support for production environments resolving the issues, analysing the defects, providing and implementing the solution defects
* Built and deployed Java applications into multiple **UNIX** based environments and produced both unit and functional test results along with release notes
* Developed the presentation layer using **CSS** and **HTML** taken from bootstrap to develop for browsers

**Environment:** HTML, CSS, JavaScript, Java, JSP, Servlets, JDBC, EJB, JUnit, PL/SQL, XML, SOAP Web services

**Education**

Bachelor’s in Computer Science Engineering from Koneru Lakshmaiah College of Engineering, Guntur, India