**Mohd Yahiya**

**Hadoop/Java Developer**

**Experience Overview**

Strategic, multi-lingual, business professional known for being energetic, analytical problem solver with project management in education and work experience along with exceptional work ethic, strong leadership abilities and competent working in any environment.

* **Over 8 years** of diverse experience in the field of Information Technology, which includes experience in development, and Implementation of various applications.
* Experience in dealing with **Apache Hadoop** components like **HDFS, MapReduce, HIVE, HBase, PIG, SQOOP, OOZIE, and Flume Big Data and Big Data Analytics**.
* Well versed in Core **Java, J2EE, RESTful Services, Web Services (SOA), JDBC, Swings, MySQL**,and **DB2.**
* Expert in installing, configuring and using eco system components like **Hadoop, MapReduce, HDFS, HBase, Zoo Keeper, Hive, Sqoop and Pig.**
* Experience in pulling data from **Amazon S3 cloud to HDFS**.
* Experience in HBase bulk load process, created **HFiles**using MapReduce and then loaded data to **HBase** tables using **completebulkload** tool.
* Experience in importing data to **SAS environment from HDFS.**
* Expertise in Creating **Hive Internal/External Tables/Views** using shared Meta store, writing scripts in **HiveQL** also data transformation & file processing using **Pig Latin** Scripts.
* Experienced in writing **Map Reduce** programs &**UDFs** for both **Hive**&**Pig** in Java.
* Expertise in Data load management, importing & exporting data using **SQOOP & FLUME**.
* Expertise in **NoSQL Database HBase** and plugging them to Hadoop eco system, **Hive&HBase Integration**.
* Having good knowledge in **Benchmarking & Performance Tuning of cluster.**
* Experienced in **commissioning & decommissioning of DataNode**
* Extensive experience in developing applications using **Core Java and multi-threading.**
* Knowledge in **Object Oriented Analysis and Design (OOAD)** using **UML** (Structural and behavioralmodeling diagrams) including Use Cases, Activity, Sequence, Class, Entity relationship diagrams, etc...
* Defining **Test Cases**, analyzing bugs, interaction with team members in fixing errors, **Unit testing and User Acceptance Testing (UAT)**.
* Knowledge about **Splunk**architecture and various components (Search Head, indexer, forwarder, deployment server).
* Excellent hands-on experience in **Unit testing, Integration Testing and Functionality testing**.
* Flexible and always open to learn new technologies at fast pace.
* Apart from developing on the Hadoop ecosystem, also have good experience in installing and configuring of the **Cloudera’s** distribution (CDH 3,4 and 5), **Hortonworks distribution** ( HDP2.1 and 2.2) and **IBM BigInsights**(2.1.2 and 3.0.1**)**.

**Skills in Review**

|  |  |
| --- | --- |
| **Hadoop ecosystem** | HDFS, MapReduce, Sqoop, Hive, PIG, HBase, Zookeper, Flume, Oozie. |
| **Programming Languages** | Java 1.4, C++, C,PIG, PL-SQL,RPGIV/RPGLE, CLLE and COBOL. |
| **IDE’s & Utilities** | Eclipse and JCreator, NetBeans. |
| **Frameworks** | MVC, Struts, Hibernate, And Spring. |
| **RDBMS** | MySQL, MS-SQL server, DB2, Oracle 11g/10g/9i. |
| **Web Dev. Technologies** | HTML, XML, JavaScript, AJAX, SOAP, WSDL. |
| **NoSQL** | HBase and Cassandra |
| **Operating Systems** | Linux, OS/400, MacOS, WINDOWS 98/00/NT/XP. |

**Professional Experience**

|  |  |  |
| --- | --- | --- |
| **Projects** | **Role** | **Location** |
| **[1]**American Express | Hadoop Developer | Phoenix, AZ |
| **[2]**Wal-Mart | Hadoop Developer | Bentonville, AR |
| **[3]**Convergys Info Management | Java/Hadoop Developer | Hyderabad, India |
| **[4]**HSBC Software Development | Java Developer | Hyderabad, India |
| **[5]** CSSI | Java Developer | Hyderabad, India |

**[1]American Express**| Hadoop Developer

**Phoenix, AZ| March 2014 – Till Date**

**[Project]Prospect Marketing & Digital–Big Data Capabilities:** Merchant Financing from American Express is a unique approach to small business lending. Available exclusively to American Express® Card-accepting businesses, it's a simple way to boost cash flow — all on terms that are simple and easy to understand. The project was to design an internal web portal called Maverick to discover/explore the prospects.Maverick serves as an internal web portal for all the statistical analysis teams at American Express. This particular portal has much functionality like discovering a prospect, knowing people, knowing business, understanding the digital data. Maverick is developed for ETL processing such as extracting data from different sources (legacy systems, Amazon S3), transforming it as per business use and then loading of transformed data to HBase.

**Roles and Responsibilities:**

* Worked on analyzing**Hadoop cluster** and different big data analytic tools including **Pig, HBase and Sqoop.**
* Automated the data pull from **Amazon S3 to HDFS,** where it downloads the file whenever there is new file present on **Amazon S3 bucket.**
* Responsible for HBase bulk load process, created **HFiles**using MapReduce and then loaded data to **HBase** tables using **completebulkload** tool.
* Imported data to **SAS environment from HDFS.**
* Developed **Java MapReduce** jobs to transform data.
* Implemented **Map-Side Join and Reduce-Side Join** in Java MapReduce.
* Responsible for building scalable distributed data solutions using **Hadoop**
* Managing and scheduling **Jobs on a Hadoop cluster.**
* Involved in loading data from UNIX file system to **HDFS.**
* Created **HBase** tables to store variable data formats of PII data coming from different portfolios.
* Implemented best income logic using **Pig scripts.**
* Implemented test scripts to support test driven development and continuous integration.
* Responsible to manage data coming from different sources.
* Installed and configured Hive and also written **Hive UDFs.**
* Created **Hive External tables** with **partitions** on daily ingesting files.
* Experienced on loading and transforming of large sets of structured, semi structured and unstructured data.
* Cluster coordination services through **Zookeeper.**
* Experience in managing and reviewing **Hadoop log files.**
* Involved in setting up storm and **Kafka** cluster in AWS environment, monitoring and troubleshooting cluster
* Documented the data flow form Application >**Kafka**>**Spark**> HDFS > Hive tables.
* Used **Spark** API over Hadoop YARN to perform analytics on data
* Improved the performance and optimization of the existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frame, Pair RDD's, Spark YARN
* Developed **Spark** code using Scala and Spark-SQL/Streaming for faster testing and processing of data
* Exported the analyzed data to the relational databases using **Sqoop** for visualizing and generating report.
* Design and implement solutions with **Amazon Web Services (AWS)** cloud technologies such as S3, EC2 and Dynamo DB.

**ENVIRONMENT:**Apache Hadoop, HDFS, Hive, Map Reduce, Java, Hive, Pig, PowerPivot, Sqoop, Flume, MapR, Oozie, Oracle, MySQL, Amazon S3.

**[2] Wal-Mart**| Hadoop Developer

**Bentonville, AR| Feb 2013 – Feb 2014**

**[Project]**Wal-Mart, like other Retail organizations, has a legacy system and researchers need to access to the most refined form of legible data to make business decisions. Wal-Mart’sMAP- Member Analytics Platform project is making use of the data from different sources to do data analytics to increase their customer base and revolutionize the retail market and hold its position in the competitive market. Any electronically generated data can be ingested and stored in the HDFS which will be used for analytic to aid in lowest price sales of products and quality customer care at the lowest possible cost and an environment to enable retail researchers to examine data.

**Roles and Responsibilities:**

* Extracted files from **DB2** through **Kettle** and placed in **HDFS** and processed.
* Analyzed large data sets by running **Hivequeries** and **Pig scripts**.
* Worked with the Data Science team to gather requirements for various data mining projects.
* Involved in creating **Hive tables**, and loading and analyzing data using **hive queries**.
* Developed Simple to **complex Map ReduceJobs** using **Hive** and **Pig**.
* Involved in running Hadoop jobs for processing millions of records of text data.
* Worked with application teams to install operating system, **Hadoopupdates**, **patches**, **version**upgrades as required.
* Building **Hadoop-based big dataenterprise** platforms coding in python.
* Developed multiple **Map Reducejobs** in **java** for data cleaning and pre-processing.
* Involved in unit testing using **MRunit** for **Map Reduce jobs**.
* Involved in loading data from **LINUX** file system to **HDFS**.
* Responsible for managing data from multiple sources.
* Experienced in running **Hadoop** **streamingjobs** to process **terabytes** of **xml** format data.
* Load and transform large sets of structured, semi-structured data.
* Responsible to manage data coming from different sources.
* Involved in standardizing **Splunk** forwarder deployment, configuration and maintenance across UNIX and windows platforms.
* Helping application teams in on-boarding**Splunk** and creating dashboards/alerts/reports.
* Responsible in exporting analyzed data to relational databases using **Sqoop**.
* Created and maintained **Technical documentation** for launching **HADOOP** Clusters and for executing **Hivequeries** and **Pig Scripts**.
* Experience of **Github** and use of Git bash for code submission in **Github** repository.
* Wrote content explaining installation, configuration, and administration of core**Hortonworks** Data Platform (HDP) Hadoop components (**YARN**, HDFS) and other Hadoop components (Hive, Pig, **Tez,** etc.). You can see my recent work at docs.**Hortonworks**.com (YARN, HDFS, System Administration, Security, Reference Guides, etc).
* Data migration from **Teradata** to **Cassandra** using Teradata Fexport& Cassandra loader.
* Installing, configuring **Cassandra** on **AWS** platform.

**ENVIRONMENT:**Hadoop, HDFS, Pig, Hive, Map Reduce, Sqoop, Python, LINUX, MRUnit and Cassandra.

**[3]Convergys Information Management**| Java/Hadoop Developer

**Hyderabad, India | August 2012 – January 2013**

**[Project]**The purpose of the project is to store terabytes of different departmental information generated by ICOMS and extract significant information out of it. The data will be stored in Hadoop file system and processed using Map-Reduce or Pig jobs. Which intern includes getting into the Excel, Csv and PDF data from the client team, Process the data to obtain respective department related information with respect to their domains, Extract various reports out of the processed information and Export the information for further processing. Also moving the data to the various data sources and then further present in the UI. This project is for the reorganizing the existing system which is running in Oracle DB to a new cloud solution technology called Hadoop which can able to process large date sets in order to meet the client requirements with the increasing data volume from other departments.

**Roles and Responsibilities:**

* Installed and configured **Hadoop MapReduce, HDFS**, Developed multiple **MapReduce jobs** in java for data cleaning and preprocessing.
* Wrote reference architecture documents describing **Hortonworks** platform integration with other platforms (Informatica) and analytic applications (Excel, Tableau).
* Wrote sandbox tutorials that demonstrated the functionality of the **Hortonworks** platform; HDFS (Hadoop), Pig, Hive, HBase, and HCatalog. Topics included how to install and configure the**Hortonworks** ODBC drivers, refining data with the**Hortonworks** platform, and analyzing data with 3rd-party BI tools (Excel, Tableau).
* Installed virtual machines on Windows and Mac using Oracle VirtualBox and VMware. Installed and documented**Hortonworks** platform features on Windows 7, Windows Server 2012, and other operating systems.
* Managed delivery pipeline, tuned OS params, throttled **SOA** application, connection pool,   
  Mediator Worker Threads, Audit flow, JVM Heap garbage collection, security patches, bug fixes.
* Developed server-side services using Java, spring, Web Services (SOAP, Restful, WSDL, JAXB, JAX-RPC), **SOA** (Service oriented architecture).
* Experience in installing, configuring and using Hadoop ecosystem components.
* Experience in administration, in4stalling, upgrading and managing **CDH3, Pig, Hive &HBase**
* Importing and exporting data into **HDFS and Hive** using **Sqoop and Flume.**
* Experienced in defining job flows.
* Knowledge in performance **troubleshooting and tuning** Hadoop clusters.
* Experienced in managing and reviewing Hadoop log files.
* Participated in development/implementation of Cloudera Hadoop environment.
* Responsible for the system user interface creation.
* Developed and implemented GUI of the system (**Java 2.0, Swing**)
* Developed server-side business logic software modules using  **JDBC**
* Proficient in database development MySQL
* Associated in development in System management and Out patience modules using **JAVA, Swing**
* Developed Application based on J2EE using Hibernate, Spring, JSP frameworks and SOAP/REST web services.
* Experience in preparing Unit Test cases prepared.
* Participated code review implemented by peers.

**ENVIRONMENT:**Apache Hadoop, HDFS, Hive, Map Reduce, Java, Hive, Pig, Sqoop, Flume, Cloudera CDH3, Oozie, Oracle, MySQL, JAVA, Swings, JSP, JDBC, JUnit.

**[4]HSBCSoftware Development**| Java Developer

**Hyderabad, India | July 2010 – August 2012**

**[Project]CLS** (Continuous Linked Settlement) is an industry initiative to reduce Foreign Exchange (FX) settlement risk. HSBC Bank plc, representing the HSBC group, is one of approximately 70 CLS shareholders and, as a‘Settlement member’, is authorized to submit trade instructions to CLS Bank for settlement. Other HSBC group sites, acting as ‘Submission locations’ can also submit their instructions to CLS, where HSBC is authorized to offer a CLS service to Third Party customers. On settlement date, CLS Bank provides each member bank with a ‘Pay-In Schedule’ detailing its net positions in the 17 CLS-eligible currencies and providing the day’s schedule by which any payments in the ‘short’ currencies must be met.

**Roles and Responsibilities:**

* Implemented the project according to the **Software Development Life Cycle** (SDLC).
* Implemented **JDBC** for mapping an object-oriented domain model to a traditional relational database.
* Created Stored Procedures to manipulate the database and to apply the business logic according to the user’s specifications.
* Developed the **Generic Classes**, which includes the frequently used functionality, so that it can be reusable.
* Exception Management mechanism using Exception Handling Application Blocks to handle the exceptions.
* Designed and developed user interfaces using **JSP, Java script and HTML.**
* Involved in Database design and developing **SQL Queries**, stored procedures on **MySQL**.
* Logging was done through **log4j.**

**ENVIRONMENT:**JAVA, Java Script, HTML, JDBC Drivers, Soap Web Services, UNIX, Shell scripting, MySQL.

**[5] Computer Solutions and Software International**| Java Developer

**Hyderabad, India | January 2008 – July 2010**

**[Project]**Binary Vibes is the leading provider of Mobile Lifecycle Management delivered as a managed-service that automates and simplifies the management of mobile devices and services. This application of Binary Vibes acts as third-party which provides different wireless devices and service plans at lower cost to the corporation. This application designed for employees of corporation which includes modules like Authentication, Search Catalog for wireless devices, service provider and Authorization modules.

**Roles and Responsibilities:**

* Handled 3 modules Login, Register and Update Profile in this application (It uses waterfall methodology).
* Involved in changing in the look and feel (UI) of web pages as per requirement by modifying the JSPs and HTMLs.
* Involved in working on the client side form validation using JavaScript.
* Involved in Code review and design review and documentation of design document.
* Involved in modifying the Java backend objects (EJBs, XMLs, and Servlets).
* Modified the DAOs as part of the enhancements.
* Modified the JDBC connection parameters in the application.
* All Java code development was done through Net Beans.
* Involved in resolving the fixes to the tickets at the time of production deployment.
* Maintained the database by purging the old data and inserting the new data (user information) on daily basis. It was done by running the UNIX scripts.
* Involved in the unit testing and UAT testing for each release.
* Maintained all the standards of GE while doing the code change.

**ENVIRONMENT:** J2EE, EJB1.1, XML, XSLT, WebSphere 4.0, Oracle 8i, JDBC, LDAP, Struts, PL/SQL, Toad, NetBeans, JSP, HTML.

**Education:**

* **Bachelor of Engineering** in Computer Science, Osmania University, Hyderabad,India.