**SINDHU DATLA**

[Sindh137.datla@gmail.com](mailto:Sindh137.datla@gmail.com)

224-432-3344

**HADOOP DEVELOPER**

**Professional Summary:**

* 7 years of professional **IT** experience in administration, analysis, design, architecture, development, testing and implementation of **Hadoop** and deploying various applications on **Object Oriented Programming**.
* Experience in deploying versions of **Hadoop 1.0 and Hadoop 2.0 (YARN).**
* Experience in Installing, Configuring and Deploying Hadoop clusters with different distributions like **Apache Hadoop** and **Cloudera Manager**.
* Strong knowledge/understanding of Hadoop architecture and various concepts like HDFS, **JOB TRACKER, TASK TRACKER, NAME NODE, DATA NODE AND MAP REDUCE**.
* Hands on experience with major components in Hadoop Ecosystem including **Spark-stream, Hive, YARN, Impala, HBase, Kafka, Pig, Sqoop, Flume, Oozie, Zookeeper.**
* Good understanding and hands on experience on **Apache Nifi.**
* Analyzed or transformed stored data by writing **MapReduce** jobs based on business requirements.
* Good understanding **in sequence files, RC files, Combiners, Counters, Dynamic partitions, bucketing** for performance improvement.
* Worked on multiple map reduce jobs to perform **Data Cleaning** and **Preprocessing.**
* Experienced on loading large sets of structured, semi structured and unstructured data and also performed importing and exporting data into **HDFS** and **Hive** using **Sqoop.**
* Hands on experience in working with **Flume** to load the log data from multiple sources directly into **HDFS.**
* Had 2 years of working experience on building, deploying applications in **Azure** environment.
* Proficient in developing web based applications and client server distributed architecture applications in **Java/J2EE** technologies using **Object Oriented Methodology**.
* Excellent understanding and knowledge of **NOSQL** databases like **MongoDB, HBase**, and **Cassandra**.
* Experience in analyzing data using **Hive, Pig Latin** and custom **MapReduce** programs in Java.
* Extensive experience with databases such as **DB2, Oracle 9i, Oracle 10g, Oracle 11g, MySQL, SQL Server and MS Access**.
* Expertise in Web technologies using **Core Java, J2EE, Servlets, EJB, JSP, JDBC, Java Beans, and Design Patterns.**
* Experience in developing J2EE applications using various other Open Source tools, **Persistence frame work (Hibernate)** and implementing **JPA (Java Persistence API).**
* Able to assess business rules, collaborate with stakeholders and perform source-to-target data mapping, design and review.
* Assisted in designing, development and architecture of **Hadoop** and **HBase** systems.
* Loading the data from the different Data sources like (Teradata and DB2) into **HDFS** using **Sqoop** and load into partitioned **Hive** tables.
* Conducted detailed analysis of system and application architecture components as per **functional** requirements.
* Strong Knowledge on full Software Development life cycle -Software analysis, design, architecture, development and maintenance and good knowledge on **Agile, Scrum and Waterfall model technologies.**
* Experienced with **Java Multithreaded programming** to develop multithreaded modules and applications.
* Good knowledge of all aspects of Testing such as **Unit, Regression, Agile, White-box, Black-box.**
* Strong debugging and critical thinking ability with fabulous understanding of **frameworks** advancement in methodologies and strategies.
* Experience in **debugging** and **troubleshooting** the bugs and resolving the issues.
* Good understanding of **distributed systems and parallel processing** architectures.
* Excellent track record of leading end to end implementation and integrating projects to successful completion under tight deadlines.
* Learning of new technologies that will allow for a more effective design and implementation.
* Intellectual capacity to grasp new situations quickly and problem solving skills.

**Technical Skills:**

|  |  |
| --- | --- |
| Hadoop Ecosystems | Hadoop 1.x/2.x(Yarn), HDFS, Spark, MapReduce, Sqoop, Impala, Hive, HBase, Flume, Kafka, Oozie, Pig, Cassandra, , Eclipse, Cloudera-desktop and SVN |
| Databases | Oracle 11g/10g, MySQL, MS SQL Server, IBM DB2 9.x |
| NoSQL Databases | MongoDB, HBase, Cassandra |
| Programming Languages | Java, Python, C, C++, SQL, PL/SQL |
| Front-end Technologies | XML, HTML5, CSS, JavaScript, PHP, JSP, Servlets |
| Operating Systems | Linux, UNIX, Ubuntu 13.X, Mac OSX, Windows NT / 98 /2000/ XP / Vista, Windows 7, Windows8, Windows 10. |
| Networks | HTTP, HTTPS, FTP, UDP, TCP/TP |
| Methodologies | Agile, UML, Design Patterns (Core Java and J2EE), Waterfall, Scrum |
| Frameworks | MVC, Struts 2/1, JavaScript frameworks. |
| IDE | Eclipse, NetBeans |

**Professional Experience:**

**Client: Motiva, Houston, TX. *July 2015– Till Present***

**Hadoop Developer**

**Description:**Motiva operates multiple lube hydro processing units using patented unit designs to upgrade heavy gas oils into high quality API Group II lubricant base oils and specialty products. It has a systematic approach to health, safety, security and environmental management. Also, manages these matters as critical business activities, and sets standards and targets for improvement.

**Responsibilities:**

* Worked on importing and exporting data from **Oracle** and **DB2** into **HDFS** using **Sqoop**.
* Developed data pipeline using **Flume** to ingest customer behavioral data and financial histories into **HDFS** for analysis.
* Developed multiple **MapReduce programs** for analyzing the insurance data of the customer and produce summary results from **Hadoop** to downstream systems.
* Scheduled a workflow to import the weekly transactions in the revenue department from **RDBMS** database using **Oozie.**
* Involved in loading data from **LINUX** file system to **HDFS**.
* Automated adding partitions to hive tables and verifying schema change in Json files using Shell scripts.
* Handle escalated Support tickets till closure for **MS Azure** IaaS platform.
* Had experience working onproviding consulting and cloud architecture for premier customers and internal projects running on **MS Azure** platform for high-availability of services with low operational costs.
* Load and transform large sets of structured, semi structured and unstructured data.
* Working on **Kafka** messaging system to prevent the data loss within the cluster and to handle a high volume of data from one end to another.
* Parse JSON files through **Spark** core to extract schema for the production data using **Spark, SQL** and Scala.
* Worked on debugging, performance tuning of **Hive**&**Pig Jobs**.
* Used **Nifi** because of itshighly scalable nature and it can also run on something as simple as a laptop or clustered across many high-performance servers.
* Used **Nifi** to perform real-time integrated data logistics.
* Hands on experience working on **NiFi** in automating the movement of data between disparate data sources inorder to make data ingestion fast.
* **Hive DDLs** to create, alter and drop Hive tables.
* Used **Bit Bucket** as code repository and version control.
* Built wrapper shell scripts to hold this **Oozie** workflow.
* Developed **PIG** Latin scripts to transform the log data files and load into **HDFS**.
* Used **Pig** as **ETL** tool to do transformations, event joins and some pre-aggregations before storing the data onto **HDFS**.
* Developed **Hive UDF** for functions that were not preexisting in Hive like the rank etc.
* Created External **Hive** tables and involved in data loading and writing **Hive UDFs**.
* Experienced in implementing **POC's** to migrate iterative **map reduce** programs into **Spark** transformations using **Scala.**
* Created concurrent access for **hive** tables with shared and exclusive locking that can be enabled in **hive** with the help of **Zookeeper** implementation in the cluster.
* Wrote the shell scripts to monitor the health check of **Hadoop** daemon services and respond accordingly to any warning or failure conditions.
* Developed Unit test cases using MRunit for **Mapreduce** code.
* Hands on experience with NoSQL databases like **Cassandra** for POC (proof of concept) in storing URL's and images.
* Involved in creating **Hadoop** streaming jobs.
* Involved in Installing, Configuring **Hadoop Eco System**, and **Cloudera Manager** using **CDH5** Distribution.

**Environment: Hadoop, MapReduce, Hive, HDFS, PIG, Sqoop, Oozie, Cloudera, Flume, HBase, Zookeeper, CDH3, CDH5, MongoDB, Cassandra, Oracle, NoSQL, Unix, Linux and Kafka.**

**Client: Whataburger, San Antonio, TX.** March 2014– June 2015

**Hadoop Developer**

**Description:** Whataburger is an American privately held regional fast food restaurant chain, which is specialized in hamburgers. It was known for many years for its distinct A framed orange-white striped roofing buildings. And, it is committed to providing equal employment opportunities to all employees and applicants.

**Responsibilities:**

* Worked extensively on **HIVE, SQOOP, PIG** and **PYTHON.**
* Implemented Partitioning, Dynamic Partitions, Buckets in **Hive**.
* Used **AXWAY** to FTP the **OPTIM** files to move to **Hadoop** and create tables on top of the data.
* Scheduled **CRON JOB** to schedule the shell scripts.
* Worked with moving the **VSAM** files in mainframes to the Hadoop.
* Managing and reviewing **Hadoop** log files.
* Complete End to End testing from requirement gathering to data load on fact a dimensional tables through**ETL** process, Data validation in Hadoop **HDFS** cluster and front end Hi- Chart re-port validation against the Data.
* Test planning, Test scope estimation, Test execution and defect management.
* Reported the defect to the defect tracking system and prepared the defect reports.
* The I**NFORMATICA** is used to handle Occurs and Redefines of the VSAM files and loaded the data into Oracle.
* Used Scala to store streaming data to **HDFS** and to implement **Spark** for faster processing of data.
* Had working experience on data management by using the data storage module offered by **Azure.**
* Creating **hive** table on parquet schema using **Scala** and **Spark**. Involved in data migration from one cluster to another cluster.
* Worked on **Talend** for Loading and extracting data from **Oracle** and **SQL**.
* Had experience working on **Kafka** architecture and documentation.
* The **HUE** is used for data validation and running hive queries.
* Experienced in managing **Hadoop** Jobs and logs of all the scripts.
* Developed **PIG** predefined functions to convert the fixed width file to delimited file.
* Migrated Complex map reduce programs into in memory spark processing using transformations and actions.
* Used python to read the **AVRO** files.
* Experience in developing shell scripts to perform the incremental loads.
* Used **HIVE** join queries to join multiple tables of a source system and load them into Elastic Search Tables.
* Developed **Pig** Latin scripts to extract data from the web server output files to load into **HDFS**.
* Involved in loading data from **UNIX file system** to **HDFS**.
* Worked on moving data from MULTIMEMBER’S to **Hadoop**.
* Experienced in creating **hive** tables by getting the source field names and data types.
* Experienced in **Data Validations** and gathering the requirements from the business.

**Environment:** Hadoop, Hive, HDFS, Pig, Sqoop,Oracle10g, SQL, Linux,Kafka, Mainframe,Unix, Spark,Scala, Informatica, Talend and CDH3.

**Client: United Airlines, Elk Grove Village, IL** October 2013 – Febraury2014

**Hadoop Developer**

**Description:** United Airlines has a long and distinguished history as the prominent leader in the low cost carriers segment in airline industry in North America region. It has always relied on internal strength and cultivated distinct expertise to create powerful IT portfolios that drive the company. United Airlines is a largest carrier in USA based on number of passengers carried and operates each day.

**Responsibilities:**

* Used **Sqoop** import from RDBMS.
* Installed and configured **Hive**, remote Hive metastore and Apache **Pig** for both development and production jobs as required and also written **Hive UDF’s**.
* Developed **Hive** scripts to alter the tables and perform required transformations.
* Involved in creating **Hive** tables, loading with data and writing hive queries which will run internally in**mapreduce** way.
* Knowledge in implementing **pig UDF’s** for evaluation and filtering the data.
* Load and transform large sets of structured, semi structured and unstructured data.
* Involved in creating **Hadoop** streaming jobs.
* Involved in Installing, Configuring **Hadoop** Eco System, **and Cloudera** Manager using **CDH4** Distribution.
* Had experience on creating **MS Azure** virtual network.
* Had experience working on Blobs using **MS Azure** storage explorer.
* Worked on importing and exporting data from **Oracle** and **DB2** into **HDFS** using **Sqoop**.
* Developed data pipeline using **Flume** to ingest customer behavioral data and financial histories into **HDFS** for analysis.
* Developed **Mapreduce** programs to cleanse the data in **HDFS** obtained from multiple data sources.
* Involved in creating **Hive** tables as per requirement defined with appropriate static and dynamic partitions.
* Used **Hive** to analyze the data in **HDFS** to identify issues and behavioral patterns.
* Involved in production **Hadoop** cluster set up, administration, maintenance, monitoring and support.
* Back-end **Java** developer for Data Management Platform (DMP) and building RESTful APIs to build and let other groups build dashboards.
* Managed and reviewed **Hadoop** log files to identify issues when job fails.

**Environment:** Hadoop, MapReduce, Hive, HDFS, PIG, Sqoop, Cloudera, CDH4, Oracle, DB2, NoSQL, Java, Unix/Linux, Flume.

**Discover Financial, Chicago, IL**November 2012 – September 2013

**Hadoop Engineer**

**Description:**Discover is a leading direct bank and electronic payment services company. It is mainly focused on the customer and a commitment to be different, to be better and to offer the best service in the business.

**Responsibilities:**

* Actively participated with the development team to meet the specific customer requirements and proposed effective **Hadoop** solutions.
* Designing, building, installing, configuring and supporting **Hadoop**.
* Translate complex functional and technical requirements into detailed design.
* Supported data analyst in running **Pig** and **Hive** queries.
* Implemented High Availability for **Name Node** Replication to avoid single point of failure.
* Used Flume in collecting and aggregating large amounts of streaming data such as log data, events into **HDFS**.
* Implemented commissioning and decommissioning of data nodes, killing the unresponsive task tracker and dealing with blacklisted **task trackers**.
* Managing and developing **HBase**.
* Maintain security and data privacy.

**Environment:** Hadoop, MapReduce, Hive, HDFS, PIG, Cloudera, HBase, CDH3, Oracle, NoSQL and Unix/Linux.

**Client: TransDyne, Hyderabad, India.** March 2011– October 2012

**Java Developer**

**Description:** Developed an Issue tracking system to track and solve Technical Problems and Service Requests for Wipro Employees. This project involved implementing an online database for the users to enter technical problems and request service. Online summary of the various requests and time taken for each were provided to facilitate faster response to future customer requests.

**Responsibilities:**

* Involved in design and development phases of **Software Development Life Cycle** (SDLC).
* Involved in designing **UML Use case diagrams, Class diagrams, and Sequence diagrams** using **Rational Rose.**
* Followed **agile** methodology and **SCRUM** meetings to track, optimize and tailored features to customer needs.
* Developed user interface using **JSP, JSP Tag libraries, and JavaScript** to simplify the complexities of the application.
* Implemented **Model View Controller (MVC)** architecture using Jakarta Struts frameworks at presentation tier.
* Developed a **Dojo** based front end including forms and controls and programmed event handling.
* Developed various Enterprise Java Bean components to fulfill the business functionality.
* Created Action Classes which route submittals to appropriate **EJB** components and render retrieved information.
* Validated all forms using **Struts** validation framework and implemented Tiles framework in the presentation layer.
* Used **Core java** and **object oriented** concepts.
* Extensively used **Hibernate** in data access layer to access and update information in the data-base.
* Used **Spring** Framework for Dependency injection and integrated it with the Struts Framework and **Hibernate**.
* Used **JDBC** to connect to backend databases, **Oracle** and **SQL Server 2005.**
* Proficient in writing **SQL** queries, stored procedures for multiple databases, **Oracle** and **SQL Server 2005**.
* Used multithreading in programming to improve overall performance.
* Identified proper Exception Handling and logging Mechanism.
* Developed various helper classes needed following Core java multi-threaded programming and Collection classes.
* Wrote Stored Procedures using **PL/SQL**. Performed query optimization to achieve faster indexing and making the system more scalable.
* Deployed application on windows using **IBM Web Sphere Application Server**.
* Used Java Messaging Services (JMS) for reliable and asynchronous exchange of important information such as payment status report.
* Used Web Services **- WSDL** and **REST** for getting credit card information from third party and used **SAX** and **DOM XML** parsers for data retrieval.
* Implemented **SOA** architecture with web services using Web Services like **JAX-WS**.
* Used **ANT** scripts to build the application and deployed on Web Sphere Application Server.
* Developed **JUnit** test cases to validate the services.

**Environment**: Core Java, J2EE, Oracle, SQL Server, JSP, Struts, JDK, JavaScript, HTML, CSS, AJAX, Junit, Windows.

**Client: Nucleus Software Exports Ltd, India *October 2009– February2011***

**Software Developer**

**Description:** It is an Indian global IT products and solutions company serving the Banking and Financial Services. This project is to develop new software products for IT solutions and consultancy services and also installing/configuring Linux operating system.

**Responsibilities:**

* Involved in the complete **SDLC** software development life cycle of the application from

requirement analysis to testing.

* Developed different Action Classes for different modules in the application.
* Developed presentation layer using tag libraries in **JSP** pages.
* Implemented web page layout using struts tiles libraries, and performed validations using **Struts** validation framework.
* Developed front-end using **JSP, HTML, JavaScript and CSS**.
* Developed **Unit Test** scripts and executed scripts for both automated and manual testing.
* Involved all phases of testing.
* Installation and configuration of **Linux** for new hardware.
* Created volume groups logical volumes and partitions on the Linux servers and mounted file systems on the created partitions.
* Monitoring and troubleshooting user issues with the networks and systems.
* Improve system performance by working with the development team to analyze, identify and resolve issues quickly.
* Ensured data recover by implementing system and application level backups.
* Performed various configurations which include networking and resolving host names and **SSH** keyless login.
* Managing Disk File Systems, Server Performance, Users Creation and Granting file access permissions.
* Automate administration tasks through the use of scripting.

**Environment:**Java, JSP, JavaScript, Eclipse, DB2, JDK, WebSphere, Linux.

**Educational Background:**

Bachelors of Technology, Jawaharlal Nehru Technological University Kakinada, India.