**Prudhvi Puppala**

**Hadoop Developer**

**Email:pdkr523@gmail.com**

**Mobile No: 361-720-7750**

**PROFESSIONAL SUMMARY**

* Over **7+ years** of professional experience in IT, including 3.5 years of work experience in Hadoop Eco system.
* Ample experience on Big Data Analytics with hands on experience in writing Map Reduce jobs on Hadoop Ecosystem including **pig and Hive**
* Excellent knowledge on Hadoop architecture; as in HDFS, Job Tracker, Task Tracker, Name Node, Data Node and Map Reduce programming paradigm
* Hands-on Experience in creating complex SQL Queries and SQL tuning, writing **PL/SQL** blocks like stored procedures, Functions, Cursors, Index, triggers and packages .
* Work experience in major components of Hadoop Ecosystem like **Flume, Hbase, ZooKeeper, Oozie Hive, Sqoop, PIG** and **YARN.**
* Hands-on experience in writing Pig Latin scripts, working with grunt shells and job scheduling with **Oozie.**
* Experience in analyzing data using **Hive QL** ,**flume Pig Latin**, and custom Map Reduce programs in Java .
* Extensive experience in developing components using JDBC**, Java, Oracle, XML** and **UNIX** Shell Scripting.
* Involved in converting **Hive/SQL** queries into **Spark**  transformations using **Spark  RDDs, Python** and **Scala.**
* Experience in extending Hive and Pig core functionality by writing custom **UDF**s like **UDAFs** and **UDTFs.**
* Implemented **SOAP** and **Restful** Web Services.
* Experience in **NoSQL** database like **HBASE**. Knowledge of job workflow scheduling and monitoring tools like O**ozie** and **Zookeeper.**
* Hands-on experience with related/complementary open source software platforms and languages (e.g. **Java, Linux, UNIX, Python**).
* Hands-on experience with **data** warehouse modeling and **ETL** design.
* Very good understanding on NOSQL databases like **MongoDB** and HBase .
* Experience in Object Oriented Analysis and Design, Java/J2EE technologies including HTML, XML.
* Good technical Skills in Oracle 11i, SQL Server, ETL Development using Informatica tool.
* Involved in developing solutions to analyze large data sets efficiently.
* Extensive experience in **Agile** methodology.
* Proficient in preparing System Design Document and Installation Guide.
* Experience in working with the Different file formats like **TEXTFILE, AVRO** and **JSON.**
* Ability to blend technical expertise with strong Conceptual, Business and Analytical skills to provide quality solutions and result-oriented problem solving technique and leadership skills.
* Good experience in writing **PIG** scripts and **Hive** Queries for processing and analyzing large volumes of data.
* Experience in Hadoop administration activities such as installation and configuration of clusters using Apache and Cloud era.
* Extensive experience in ETL (Informatica) development using Oracle, Flat Files as both Sources and Targets
* Experience in extracting data from both **Relational systems** and **Flat Files.**
* Experience in deploying **Hadoop** cluster on Public and Private Cloud Environment like Amazon Web Services **(AWS**) EC2.
* Ability to perform at a high level, meet deadlines, adaptable to ever changing priorities.
* Excellent work ethics, self-motivated, quick learner and team oriented. Continually provided value added services to the clients through thoughtful experience and excellent communication skills.

|  |  |
| --- | --- |
| **Big Data Technologies** | HDFS, Hive, Map Reduce, HBaes, Pig, Sqoop, Flume, Oozie, Zookeeper, YARN, Spark, Kafka, and Storm. |
| **Scripting Languages** | Shell, Python, Scala. |
| **Languages** | C, C++, Java, SQL, PL/SQL, Python, Scala, PIG Latin, HiveQL, Unix Shell Scripting. |
| **Front End Technologies** | HTML, XHTML, CSS, XML, JavaScript, AJAX, Servlets, JSP. |
| **Java Frameworks** | MVC, Apache Struts2.0, Spring and Hibernate. |
| **Web Services** | SOAP(JAX-WS), WSDL, Apache CXF, Apache Axis SOA, Restful(JAX-RS), JMS,AWS. |
| **Application Servers** | Apache Tomcat 5.5/6.0/7, WebLogic Server 8x/9x/10x, WebSphere 5.1/6.0, JBoss. |
| **Databases** | Oracle 9i/10g/11g, IBM DB2, MySQL, MS SQL Server. |
| **NoSQL Databases** | HBase, MongoDB, Cassandra. |
| **IDE** | Eclipse, NetBeans. |
| **Operating Systems** | Linux, UNIX, Mac, Windows 7/8/10. |
| **Reporting Tools** | Tableau, Talend. |

**Education**

* Bachelors of Technology in Electrical Engineering, JNTU, Anantapur, India.

**Work Experience**

**Molina HealthCare Inc, Long Beach, CA** March 2015 to Present

**Hadoop Developer**

**Responsibilities:**

* Proactively monitored systems and services, architecture design and implementation of Hadoop deployment, configuration management, backup, and disaster recovery systems and procedures.
* Imported Bulk Data into **HBase** Using **Map Reduce** programs.
* Involved in Analyzing system failures, identifying root causes, and recommended course of actions.
* Analyzed the web log data using the **HiveQL** to extract number of unique visitors per day, page views, visit duration, most purchased product on website.
* Used **HIVE** join queries to join multiple tables of a source system and load them into Elastic Search Tables.
* Involved in enhancing the speed performance using Apache **Spark** .
* Worked with **NoSQL** database **HBase** to create tables and store data**.**
* Real time streaming the data using **Spark** and **Kafka.**
* The **HUE** is used for data validation and running **hive** queries.
* Analyzed the data by performing **Hive** queries and running Pig scripts to know user behavior.
* Used Flume to collect, aggregate, and store the web log data from different sources like web servers, mobile and network devices and pushed to **HDFS.**
* Loaded the dataset into **Hive** for **ETL** Operation.
* Migrated **ETL jobs to Pig scripts** do Transformations, even joins and some pre-aggregations before storing the data onto **HDFS**.
* Developed Spark code using **Scala** and **Spark-SQL**/Streaming for faster testing and processing of data.
* Developed new mappings for complex **ETL** operations such as CDC (change data capture) & Surrogate Key Generation by utilizing **Informatica** designer..
* Used **Spark** for predictive analysis and tumor simulator.
* Used **Kafka**   integration to ingest the log data directly into **HDFS.**
* Worked on different kind of **custom filters** and handled pre-defined filters **on HBase data using API**
* Involved in creating Hive tables, loading with data and writing Hive queries which will run internally in map reduce way
* Used Pig as **ETL**  tool to do Transformations, even joins and some pre-aggregations before storing the data onto HDFS
* Load the data into **Spark RDD** and do in memory data Computation to generate the Output response.
* Moving data from DB2, Oracle Exa-data to HDFS and vice-versa using **SQOOP.**
* Knowledge on creating and monitoring Hadoop clusters VM, Horton works Data Platform 2.1 & 2.2, CDH4, and CDH 5.4 Cloudera Manager on **CentOS**, **Ubuntu** etc.
* Pre-Processing large data sets in parallel across the **Hadoop** cluster.
* Responsible for loading data from **UNIX** file system to **HDFS**.
* Experienced in running Hadoop streaming jobs to process terabytes of **CSV** format.

**Environment: Hadoop, Map Reduce, Yarn, Hive, Pig, HBase, Oozie, Sqoop, Flume, Oracle 11g, Core Java Cloudera HDFS, Eclipse, CentOS, Ubuntu.**

**Cigna, Bloomfield, CT** Jan 2013 to March 2015

**Hadoop Developer**

**Responsibilities:**

* Developed **Java Map Reduce programs** for the analysis of sample log file stored in cluster.
* Setting up **Amazon EC2** Servers and installing required database servers and **ETL** tools.
* Good experience with **NoSQL** database.
* Performed Map Reduce jobs using **AWS** Elastic Map Reduce, S3 and EC2 platform
* Created partitioned tables in **Hive**
* Experience with **HIVE** DDLs and Hive Query language (HQLs).
* Troubleshooting the cluster by reviewing **Hadoop** **LOG** files.
* Developed UDF's using **python** in LINUX to extract the **JSON** data with **PIG** script
* Used **Sqoop** to import the data from RDBMS to **Hadoop** **Distributed File System** (HDFS) and later analyzed the imported data using Hadoop Components.
* Worked on Prototyping and Implementation using **Java, Node JS,** AWS with Slacker platform.
* Used **UDF’s** to implement business logic in Hadoop and responsible to manage data coming from different sources.
* Experience with **Hive queries** for data analysis to meet the business requirements.
* Created **Impala** tables for faster operations. Designed **ETL** jobs to identify and remove duplicate records using sort and remove duplicate stage and Generated Keys for the Unique records using Surrogate key Generator Stage.
* Involved in loading data from **UNIX** file system to **HDFS.**
* Involved in creating **Hive** tables, loading with data and writing Hive queries which will run internally in map reduce way.
* Analyzing data using Hadoop components **Hive, Impala and Pig** Coordination with Hadoop Admin’s during deployment to production.
* Worked hands on with ETL process using **Python** and **Java.**
* Created partitioned tables in **Hive**.
* Knowledge with NoSQL databases such as MongoDb, HBase.
* Designed and Modified Database tables and used **HBASE** Queries to insert and fetch data from tables.
* Handled importing data from various data sources, performed transformations using Hive and Map Reduce, streamed using **Flume** and loaded data into **HDFS.**
* Worked with business teams and created **Hive** queries for ad hoc access.
* Worked with Flume to transfer online logs data.
* Collected the logs data from web servers and integrated in to **HDFS** using Flume.
* Developed the **UNIX** shell scripts for creating the reports from Hive data.
* Experience in managing and reviewing **Hadoop log** files.
* Create a complete processing engine, based on Cloudera's distribution, enhanced to performance.
* Involved in loading and transforming large sets of structured, semi structured and unstructured data from relational databases into **HDFS** using **Sqoop** imports.
* Wrote the shell scripts to monitor the health check of Hadoop daemon services and respond accordingly to any warning or failure conditions.

**Environment: AWS, Map Reduce, Hive, Pig, Sqoop, Oozie Scheduler, UNIX, Java 7.0, JSON, Spark, HDFS, YARN, Flume, Ozzie and Zookeeper, MySQL**

**The Home Depot, Austin, TX**  July2012 to Dec 2013

**Java Developer**

**Responsibilities:**

* Used various **Core Java**concepts such as Exception Handling, Collection **API**s to implement various features and enhancements.
* Developed and deployed UI layer logics of sites using **JSP, XML, JavaScript, HTML/DHTML,** and **Ajax**.
* Involved in the Software Development Life Cycle (SDLC): Code, Integrate, Deployment and production phases.
* Experience is using XML technologies such as **JAXP, DOM, SAX, XPath, XSD**, and **JAXB** Parser
* Implemented java Mail API for the email services, **java** FTP API for the file transfer service, ZIP API.
* Experience in using **Apache POI**, Apache Axis API's
* Created tile definitions, Struts-config files, validation files and resource bundles for all modules using Struts framework
* Wrote **DB2 SQL** queries and Stored Procedure for accessing, manipulating and updating DB2 database.
* Designed and implemented UI layer using Struts MVC, Struts Forms, **JSP**, Java  Script, XML, **XHTML, XSL,XSLT** and business logic using Servlets, JSP and J2EE framework.
* Prepared queries using **DB2 control** center and Toad.
* Used JSP for presentation layer, developed high performance object/relational persistence and query service for entire application utilizing Hibernate.
* Designed Asynchronous communication using **Ajax.**
* Worked on creating Packages, Stored Procedures & Functions in **Oracle** using **PL/SQL** and TOAD.
* Intensive experience in using **Java** **1.2, 1.3.**
* Style Sheets worked with various like **Cascading Style Sheets** (CSS).

**Environment: Java, J2EE, JSP, JavaScript, JSF Sun RI, Ajax4JSF, Spring, XML XHTML, Hibernate, Oracle9i, PL/SQL, SOAP Web service, Web Sphere, Oracle, JUnit, SVN**

**Webster Bank, Bristol, CT** Oct2011 to July 2012

**Java Developer**

**Responsibilities:**

* Involved in Unit Testing, User Acceptance **Testing** and **Bug Fixing.**
* Developed the **XML** data object to generate the PDF documents and other reports.
* Developed and deployed UI layer logics of sites using JSP.
* Responsible for overall design & analysis Oracle ADF and integration with existing **J2EE** and Web Services Layer.
* Used JavaScript, **HTML** for creating interactive front-end screens.
* Involved in writing SQL queries, stored procedure and **PL/SQL** for back end.
* Co-ordinate with **QA** for **testing,** Production releases, Application deployment, integration and conducting walk-through code reviews.
* Coordinated with technical teams for preparation of custom interfaces utilizing **ADF** and **API** technologies.
* Used various Core Java concepts such as **Exception Handling**, Collection **APIs** to implement various features and enhancements.
* Involved in different applications testing including Unit testing, **integrated testing** and production testing.
* Developed batch programs to update and modify metadata of large number of documents in File Net Repository using **CE APIs**.
* Created connection through JDBC and used **JDBC** statements to call stored procedures.
* Used **Hibernate3.1** to store the persistence data into the MySql-5 database.
* Involved in using **Log4J** to create log files to debug.
* Developed the application by using the Spring **MVC framework**.
* Deployed the application on **JBoss 5.**
* Created connection through **JDBC** and used JDBC statements to call stored procedures.
* Used **SVN** as version control tools to maintain the code repository

**Environment: Spring MVC, Oracle 11g J2EE, Java, JDBC, Servlets, JSP, XML, Design Patterns, CSS, HTML, JavaScript 1.2, JUnit, Apache Tomcat, My SQL**

**Choice Solutions Ltd, Hyderabad, India** Sept 2010 to Oct 2011

**SQL Developer**

**Responsibilities:**

* Managed Enterprise and Distributed Data Warehouses in Developmental and Production Environments using **SQL** **2005**.
* Optimizing the current **ETL process**, identifying and eliminating bottle necks and decreasing data latency.
* Created synchronized mechanism for data management in **SQL Server.**
* Set the job dependencies of **SQL** jobs using **MSDB** objects Wrote Complex Stored Procedures, Triggers, Views and Queries.
* Created indexes, Constraints and rules on database objects.
* Migrated data from Oracle to **SQL** server database using ODBC connectivity
* Optimized schema, performance, and capacity planning of various data transformation processes related to the reports.
* Export & Import data from Flat file, **CSV file to/from SQL**Server Database using SSIS.
* Monitoring database health, disk space, backup verification
* Using **UNIX** environment for performing the testing
* Created Database and **Database** Objects like Tables, Stored Procedures, Views, Triggers, Rules, Defaults,
* User defined data types and functions.
* Involved in implementation, maintenance, and monitoring Database Back-ups, Database Space
* Management, Error Log monitoring, patch application, etc
* Troubleshoot **SQL** Server performance issues for internal and external customers
* Created Ad-Hoc reports for business users using **SSRS**
* Built the tabs with various charts from show me tool bar like (line chart, pie chart, bar chart).
* Automated data fetch using **UNIX** shell script.
* Writing **PL/SQL** code using the technical and functional specifications.

**Environment: PL/SQL, SQL Developer, Toad, SQL\*Plus, Oracle 10g/11g, Erwin, UNIX.**

**HDFC, HYD, India** July 2008 to August 2010

**UNIX/LINUX Developer**

**Responsibilities:**

* Interacting with various teams such Oracle database, critical infrastructure, **UNIX** architecture, Wintel, IT security, Network/application support outside vendor (Bearing Point)
* Installing, Configuring and Maintaining the **DHCP, DNS, NFS, NIS**, send mail server and **LDAP i**nstallation.
* Management and configuration of **PostgreSQL & MYSQL** database servers.
* Involved in developing custom scripts using **Shell (BASH, KSH)** to automate jobs.
* Merged the generated reports as per request. Formatted data into required format using **UNIX Shell** Scripts**, SED** and **AWK.**
* Responsible to handle the Server Administration and Asterisk Server Installation and Configuration in **Linux**.
* Familiar with Disk Management Utility. Hands on experience in file system creation and file system management.
* Work extensively on HA/Disk Volume Management with **VERITAS** Volume Manager/ SVM and **VCS**.
* Created user roles and groups for securing the resources using local operating System authentication.
* Monitored and fine tuning system and network performance for server environments running Solaris, **Red hat Linux.**
* Troubleshooting of day to day system and user problems.
* Involved in creating new users, groups and setup their home directories and appropriate access restrictions to software in the Active Directory environment and **NIS** domains
* Responsible for modifying and optimizing backup schedules and developing **shell scripts** for it.
* Daily tasks such as taking **data backups**, cleaning junk content and updating software.
* Monitored load and performance on the infrastructure and added capacity as needed.

**Environment: Solaris 10/9/8, AIX 4.3, 5.1/5.3L, Sun E10K/15K/ F25k,  Red-Hat Enterprise Linux  8/ 9/ AS3.0/ AS4.0, Solaris 8 & 9, AIX Sun Enterprise […] Servers, Sun SPARC Enterprise T5220.**