**Akash Pandey Current Address:**

Wipro Limited Flat No. 415, 5th Avenue, B-Block

✆: +91 9616063775 Gaur city 1, sector-4, Greater Noida

🖂: akashpandey882@gmail.com Uttar Pradesh-201309

**CAREER OBJECTIVE**

Seeking a position to utilize my skills and abilities in the Information Technology industry that offers professional growth while being resourceful, innovative and flexible.

**PROFILE SUMMARY**

* Working as Big Data professional
* Comprehensive Knowledge of **Big Data Hadoop**
* Knowledge of **Hadoop** components like **HDFS**, **Hive**, **Flume**, **MR**, **HBase, ELK Stack, NiFi, Sqoop, Solr, Nutch**
* **Installed** and **configured** Hadoop cluster in Test and Production environments.
* Knowledge on setup **single-node** Hadoop Cluster as well as **multi-node** Hadoop Cluster.
* Moving data from **Local** to **HDFS** and vice-versa
* Developed multiple MapReduce jobs in java for data segregation and preprocessing like **wordcount**
* Collecting and aggregating large amounts of log data using **PIG** and **Hive** and staging data in HDFS for further analysis.
* Involved in writing the **Hive** scripts to reduce the search result
* Analyzing/Transforming data with Hive and Pig.
* Implemented Commissioning and Decommissioning of new nodes to existing cluster.
* Very good understanding of **Partitions, Bucketing** concepts in Hive and designed both **Managed and External** tables in Hive to optimize performance.
* Installed and configured **Elasticsearch** **(ELK** **Stack)** and **Beats** for storing, parsing, indexing and visualization of data
* Installed and configured **NiFi** for automated and managed flow of information between systems
* Installed and configured both **Solr & Nutch** for gathering data using website crawling
* Involved in writing the Pig scripts to reduce the job execution time.
* Knowledge on **SPARK** and **SCALA**
* Experience in IDE’s for development of project **(Eclipse)**
* Knowledge of **My-SQL**
* Exceptional ability to learn new concepts
* Comprehensive Knowledge on **Linux** andCan install operating systems like **Ubuntu, CentOS**
* Good in understanding, analysing and streamlining **Test Requirements**
* Worked on **Putty, VmWare, VirtualBox, WinScp**
* Effective in working independently and collaboratively in teams
* Commitment to co-operative teamwork and involved in delivering high quality product to client
* Strong and dedicated team player to work on projects
* Ability to learn new technologies and methodologies quickly

**TECHNICAL SKILLS**

* Operating System **:** **Unix /Linux (Ubuntu, Centos & Redhat), Windows-XP/7**
* Big Data eco System **:** **HDFS, FLUME, PIG, HIVE, HBASE, Sqoop, SOLR, Nutch,**

**ELK Stack, Nifi, Ambari**

* Programming Languages **: C, Java, PHP**
* Web Technologies **:** **HTML**
* Database **: MySQL**
* NOSQL Databases : **Mongo DB, HBase**
* IDE’s : **Eclipse**
* Methodologies : **Agile, Water Fall**
* Backups : **Raid Levels**
* Testing Tools : **BugZilla, JIRA**
* Other Tools **: Putty, VmWare, VirtualBox, WinScp**

**ORGANISATIONAL EXPERIENCE**

**Company name**: Wipro Limited, India

**Designation:** Big Data Engineer

**Duration:** Sep 2017 to present

**Client:** National Informatics centre HQ (NIC HQ)

**PROJECT DETAILS**

**PROJECT #1:**

**Project Name** : Stream

**Client** :National Informatics centre HQ (NIC HQ), India

**Environment** : Intel Dual Core Processors, Hadoop 2.7, Java 1.8.64, Ubuntu 16.04

**Software** : HDFS, PIG, HIVE, Map Reduce, SQOOP, SOLR, NUTCH, ELASTICSEARCH,

LOGSTASH, KIBANA, Ambari, JAVA, MYSQL

**Role** :Big Data Engineer

**Description:**

This Project is all about the re-hosting of the current existing data of the organization into Hadoop platform. Previously the organization was using mysql DB for storing their complex data.

In this project, we will process the datasets coming with the thousands of volume of records of data from various sources through websites, software etc. The purpose of publishing this dataset is to figure out the users depending up on various constraints like type of website, depending upon IP location, Geo-location, country name, hosted by, status of the website, etc.

Specifically, in this project we will move different log files to hadoop environment which contains tb’s of data, then we will store that data in **Hive** by sorting that data through unix and linux commands and will find and convert the log’s data IP’s into their Geo-locations and sorting them with country names with the help of **Map** **Reduce** and store this data into HIVE tables again for further processing, we also remove the duplicate records and separate them depending on constraints. We will look at the metadata files which contain tables about IP records depending upon geo location and their country. Hive and **Pig** queries are written to search and process the stored data in hive and pig according to the need.

After that the data is sent to **ELK** **stack** so that the data can further be processed. Here, through **Beats** the data is being collected and then sent to **Logstash** where data is aggregated and processed, further the processed data is sent to **elasticsearch** where the data is indexed and after indexing it is stored in elasticsearch itself. After indexing with the help of **Kibana** the data is being visualized to the users and is displayed for an easy view through web console. After processing, indexing and storing data this data is sent to **NiFi** where nifi automates the flow of data between multiple system in a managed way

**Roles and Responsibilities:**

* Completely involved in the requirement analysis phase.
* Analyzing the requirement to setup a single node and multi node cluster
* Analyzing the metadata as per the requirements
* Moved all crawl data flat, log/text files generated from various sources to HDFS for further processing, i.e. loaded the data in HDFS from local system.
* Written the Map Reduce program for word count
* Written Map Reduce code that will take input as log files and parse the logs and structure them in tabular format to facilitate effective querying on the log data
* Written Map Reduce code for the job.
* Written the PIG scripts to process the HDFS data.
* Created Hive tables to store the processed results in a tabular format.
* Created External Hive Table on top of parsed data.
* Setup Hive with MySQL as a Remote Metastore
* Created hive table with schema and loaded the data using Sqoop.
* Writing queries in hive to map the data resided in HDFS.
* Developed the sqoop scripts in order to make the interaction between Pig and MySQL Database.
* Involved in configuring SPARK,SCALA,SQOOP,SOLR,NUTCH,ELK AND AMBARI
* Involved in working of ELASTICSEARCH,LOGSTASH,KIBANA,FILEBEAT
* Involved in gathering the requirements, designing, development and testing
* Writing CLI commands using HDFS.
* Created two different users (hduser for performing hdfs operations and map red user for performing map reduce operations only)
* Ensured NFS is configured for Name Node
* Setting Password less hadoop
* Setting up cron job to delete hadoop logs/local old job files/cluster temp files

**EDUCATION SUMMARY**

|  |  |  |  |
| --- | --- | --- | --- |
| Course | Board/University | Year of Passing | Percentage |
| M.Tech (CS) | AKTU, Lucknow | 2016 | 70.075% |
| B.Tech (CSE) | GBTU, Lucknow | 2013 | 68.82% |

**TRAININGS**

* *Successfully**trained**on* ***Linux*** *from* ***zoom technology***
* *Attended seminar on* ***Developing Web Applications*** *Using* ***ASP.NET*** *held by* ***NIIT, New Delhi***
* *Attended seminar of 1-week on* ***ETHICAL HACKING*** *held in my college by* ***IIT Kanpur***

**PERSONAL INFORMATION**

Father’s Name : Mr. Om Prakash Pandey

Mother’s Name : Late Mrs. Chanda Pandey

Date Of Birth : 28 Jan 1991.

Marital Status : Single

Permanent Address : Flat No. #415, 5th Avenue, B-Block, Gaur city sector 4, Greater Noida, Uttar Pradesh - 201309

**PASSPORT DETAILS**

Passport# : L7049455

Issue date : 27/01/14

Validity till : 26/01/24