****

**Neha Khatri**

**Senior Hadoop Developer**

**SUMMARY**:

Senior Hadoop Developer with 6+ years of experience building scalable, distributed data solutions with 35GB+ of data and driving business improvements with innovative Hadoop and Ecosystem .

**PROFILE : Senior Hadoop Developer**

“Data are becoming the new raw material of business.”

* Have 6.4 years of total experience in IT industry and 4 years of Big Data Experience.
* Developed multiple MapReduce jobs in Java for data cleaning and processing.
* Performed code/design analysis, strategy development and project planning.
* Having good insight of HDFS Design, NameNode, DataNode.
* Experience in loading data from Linux file system to HDFS,Creating Hive tables, writing hive queries that will internally run as Map Reduce job.
* Implemented Partitioning, Dynamic Partitioning, and Bucketing in HIVE.
* Familiar with data architecture including Machine learning, Spark streaming, Kafka.
* Received praise from users, shareholders and analysts for developing a highly interactive and intuitive Web Application using JavaScript,JQuery, AJAX, Rest API,RDBMS,HTML & CSS techniques.
* Cloudera Certified Developer for Apache Hadoop,Version 5.

**AREAS OF EXPERTISE:**

* **Big Data Ecosystems:** Hadoop, MapReduce,HDFS, Hive, Linux
* **Programming Languages:** Java
* **Scripting Languages:** JavaScript,Jquery, HTML
* **Databases:**MySql, HDFS
* **Tools:** Eclipse ,Putty, Winscp,Maven,Github
* **Platforms:** Windows(2000/XP), Linux ( Centos & Ubuntu)
* **Application Servers:** Apache Tomcat 7
* **Cloud Platform :** Amazon Web Services(AWS)

**PROFESSIONAL EXPERIENCE:**

**Project #1**

**Rani Technologies Inc.**

**Module 1 :- Streaming**

**Module 2:- Skillskaart.com:**

Developed Internet Web Application for online handshaking between mentee(learner) & mentor based on common skills of interests & locality. Anyone who wants to learn new things be it academic(subjects like Maths, Science,Arts) or non academic (Sports,cooking,Music), Can simply register & use their profile recommendations as per their criteria in order to start classes offline.

**Features Developed:**

1. Implemented GeoCoding & ReverseGeoCoding API to locate address of user on the hit of app URL in browser.
2. Integrated Google & Facebook Signup Functionality along with username/password enabled login features.
3. Designed & Developed homepage & Profile Page once user logins.
4. Reset Password Feature
5. Made Application Secure from http to https using AWS Workmail.
6. Developed responsive version compatible with Mobile, Tablet & Desktop/Laptop**.**
7. **Hosted application multiple times on AWS to make it available live .Configured Elastic-Beanstalk, Route 53,Load Balancer,RDS,AWS workmail etc.**
8. **Made & Implemented algorithms to provide close recommendation of matched profiles to user using Rest API.**

**Project #2**

**Client-American-Express**

**CDSS-Addition & deletion of variables Controlled by MCT.**

April 2016 to May. 2017

**Description:**

In CDSS ,generic software that takes the names of the variables that need to be computed from a file along with the specifics like the stats operation details ,the limits to be applied and the rules for the variables . I got that list of variables distributed in the hadoop cluster and subsequently churned the data from hive sources on hadoop cluster to get the statistical computation achieved for those variables in a distributed paradigm using map reduce . The code was generic enough to handle multiple file types and various stats operations .

* Design algorithm to implement hadoop Mapreduce jobs to perform data profiling & Stats calculations (such as Mean/min/max etc. using descriptive Statics jar)

of 35 GB data generated on per day basis.

* Handling of edge node & run time exception (without going on jobtracker)& push notification to business partners.
* Developed Internet version(Web Application) of the same RCM -To configure, Run ,reporting,analysis job.

Technologies used for above project :Java , MapReduce ,HDFS, Restful Webservices , Tomcat , Maven

**Project #3**

**Teradata & XML Migration into Hadoop(TXMH)**

May 2014 to Dec 2015

**Description:**

In TXMHproject ,data from various sources (like XML file, flat file, Teradata etc.) are migrated into Hadoop . Sqoop was used to transfer data from Teradata to Hadoop and Java map reduce programs to extract data from XML files and store them in Hadoop.

Hive was used as a Data Store so that SQL like queries can be executed for analysing data and maintaining versions.

Modules Handled:-

* Extracting data from XML files using XML Parser and store it in HDFS using map reduce code.
* Extracting data from Teradata and load it in Hive using Sqoop.
* Writing Hive queries for creation of databases and tables .
* Analyzing and designing Hive query for loading initial and incremental data and maintain all version of data.

Database : Teradata, mysql

Tools : Eclipse

Technology : MapReduce, Sqoop, Hive

Operating System : Windows 7, Ubuntu

**Project #4**

TCS Internal

ABIM-DESS

Jan 2014-Feb 2014

**Description:**

Extensive hands-on training and enhancing competencies in the field of BigData – Hadoop, HDFS, Map-Reduce, Hive. The POC included loading of data into Hive tables and performed analysis of data using joins and other analytic functions.

**Role:**

* POC Performed
  + - * Moved the data generated (in csv format) by hospitals to hdfs.
      * Loaded the data into hive.
      * Performed partitioning and bucketing on the data.
      * Implemented various joins on the data.
* Transferred data between Hadoop and a relational database using sqoop.
* Hands-on understanding of how to configure a Hadoop Cluster.
* Understand use cases for Hive.
* Understanding how Bucketing and Partitioning is implemented.
* Understanding of different file formats & their performance in hive.
* Understanding of different join techniques and their optimization in hive.
* Implementation of various analytic functions on hive data.

**Project #5**

Non Regression Testing Automation Tool

TCS Internal Product

March 2013 to Dec 2014

Team size 8

**Description:**

The purpose of non-regression testing is to verify whether, after introducing or updating a given software application, the change has had the intended effect.

Process Of automation involves running the JAR files of test\_cases on 3 entities like EURP, NMAR, APAC, which in turns call the AC\_MAIN java script. The primary work of AC\_MAIN

Script is to load the Selenium webdriver and get the connection with Oracle database .Usually the records are picked up from tables like NRTTool\_Dates, Test\_case\_dtls, DP\_main, if field “RUN” is marked as “Y” respective to a particular record. On the basis of the script name in the

DP\_main table,the script written in java code is called by selenium webdriver which in turns loads the web application TCSBANCS.

According to the code written in script the pages are loaded and new records are created ,save, release, authorized, deleted as per the requirement and field “RUN” is set as “D” or “N” according to success/failure of the record .Hence Report is generated likewise in every run of tool and Deviation is calculated.

**Roles & Responsibilities:**

* Involved in Requirement Gathering, Analysis, documentation, review, preparing Test plan, Test data, Test Cases, Test Case Execution.
* Development and Integration of Test cases into Automation Tool.
* Automated 30 corporate actions test cases using Selenium WebDriver and JAVA.
* Involve in Setting up of Environment, before triggering Automation Tool.
* Involved in Automation testing.
* Defect identification and reporting.
* Ensure that tasks are executed with a focus on quality and without any deviations.
* Providing KT-Knowledge transfer sessions for the members joining the team.
* Assigning tasks to the team and providing detailed explanation of the task.

Applications: TCSBancs , JIRA

Automation Tool: Selenium Webdriver.

Database: Oracle.Tool: Putty, Winscp, Eclipse.

**EDUCATION:**

Bachelor of Technology [B.TECH.] in INFORMATION TECHNOLOGY from GECA, Rajasthan Technical University, Rajasthan, India in 2012 with Honors.

**CERTIFICATIONS:**

**Cloudera Certified Developer for Apache Hadoop,Version 5(CCHD-410).**

**Hadoop Developer (INTELLIPAAT)** **2015**

**Hadoop Administration (INTELLIPAAT) 2015**

**PERSONAL DETAILS:**

**Name:** Neha Khatri

**Father’s name**: Mr**.** Kulbhushan

**Date of Birth:** 10th Feb,1991

**Email\_id:** nehakhatri10@gmail.com

**Contact No.:** +91 8800236796