****

**SUMMARY**:

I am a dedicated computer programmer. Computer programming in cloud as well as distributed computing is by now more of a way of life to me . I love making things by which directly / indirectly we are making the life of people better . Have been instrumental in making relevant internet startups as well as have provided data scientists rolled up sanitized data from large big unstructured data so that they can derive intelligent insights out of it . I believe in humbleness .

and learning which helps me to learn new technologies as a developer , as a continuous learner.

I have worked for clients like American Express big data engineering team where my distributed codes in production serves daily processes petabytes of data with a cumulative CPU time much much lower than had it been processed sequentially . In my startup journey , wrote softwares which public liked ( based on alexa ranking ) and had webservices written and benchmarked within 5 mili seconds .

**AREAS OF EXPERTISE:**

* **Big Data Ecosystems:** Hadoop, MapReduce,HDFS, Hive
* **Distributed Queue :** Kafka
* **Programming Languages:** Core Java
* **User interface creation ( mobile compatible ) :** JavaScript,Jquery, HTML
* **Databases:**MySql, HDFS
* **Tools:** Eclipse ,Putty, Winscp,Maven,Github,JenKins,Jira,Rally
* **Platforms:** Windows(2000/XP), Linux ( Centos & Ubuntu)
* **Application Servers:** Apache Tomcat 7
* **Cloud Platform :** Amazon Web Services Compute (Elastic Beanstalk, EC2, Route 53, VPC, AWS RDS,Workmail)
* **J2EE Platform:** Spring, Restful webservice(JAX-RS,Jersey)
* **Certification** : Cloudera Certified Developer for Apache Hadoop,Version 5.

**PROFESSIONAL EXPERIENCE:**

**Project #1 Rani Technology** Nov 2018 – Present

**Module 1 :- EDA Tool**

I have developed a generic software using Mapreduce which performs EDA(exploratory data analysis) to check the integrity of the data like if the data I have got from various sources are statistically matching with the measures provided in the reference file or if some data is missing(when statistics don't match) .

After the confirmation & validation of the data, Customized processing (such as capping of certain variables) is done on the basis of certain rules defined in the Controller file.

The Mapreduce output produced is pushed to Kafka Producer for further processing.

**Module 2:- Skillskaart.com: Rani Technology** June 2017 – Present

Designed,Implemented & 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 ACM(AWS Certification Manager).
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** April 2016 to May. 2017

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

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.

**Project #3 TCS** May 2014 to Dec 2015

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

**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.

**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.

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 hive.Hands-on understanding of how to configure a Hadoop Cluster.U
      * Understand use cases for Hive.

**Project #5** TCS Internal ProductMarch 2013 to Dec 2014

**Non Regression Testing Automation Tool**

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.

**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:**

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

**Contact No.:** +91 8800236796