Divya Sree Velugula

divyasreevelugula@my.unt.edu +1 (682)-252-8817 www.linkedin.com/in/divya-velugula

Big Data Developer, Data Engineer, Data Analyst

Expertise:

Big Data Technologies:	**Hadoop, HDFS, SQOOP, HIVE, DWH, MapReduce, Apache Spark, PIG, FLUME, StreamSets.**
Skills:	**SQL, PLSQL, Data Visualization(TABLEAU)**
Scripting Languages:	**UNIX Bash/Shell Scripting, Python.**
Web Technologies:	**JavaScript, HTML5, CSS3, JQuery.**
Databases:	**MSSQL Server 2016, Oracle 11g, MySQL**
Version Control:	**Git**
Others:	**Office 365 Suite, JIRA, Confluence, RESTful**
Cloud & Servers:	**Cloudera, Hortonworks, VMWare**
Developer Tools:	**SQL Tuner, NiFi, HUE,**

Familiar with:

High Level Languages:	**C, Java**

Education:

- **Master of Science in Data Science, University of North Texas** **GPA-3.7**
  Graduation- summer 2020
- **Bachelor of Technology, Jawaharlal Nehru Technological University** **GPA-3.8**
  Graduated- spring 2016

Work Experience:

Research Assistant: 6 months, UNT Data Science, Denton, TX 76201, USA (07/2019-Present)

Roles and Responsibilities:

- **Custom application programming and portal development.**
- **Migration and manipulation of data from legacy systems, HR and Student.**

Big Data Developer: 2+ years, TechMahindra, Hyderabad, India

Roles and Responsibilities: **(08/2016-12/2018)**

- **Cloudera was used as a distribution platform for Hadoop.**
- **Created HIVE Tables on top of datasets which are in staging layer.**
- **Used Various Hive performance optimization techniques.**
- **Created Impala Tables as the target tables to load data into the integration layer.**
- **Extensively used Cloudera Stream-sets to transfer Raw data files into HDFS foundation layer.**
- **Used Parquet file format in all target tables in Impala for better performance.**
- **Used Spark core and Spark SQL for data transformations.**
- **Involved in Building data frames and RDD's using Spark SQL and Spark Core**

Projects:

- **Twitter-Data-Analysis:**
  Collected Tweets from Twitter to store and analyze the data using Apache Hadoop, Apache Spark. For Visualization, Tableau is used.
- **Predicting Movie Ratings using Multimodal data(Research Project):**
  Movie ratings were predicted prior to the release from the posters available. Image recognition is used for the images and NLP for text processing from the posters. Various ML algorithms were applied and analyzed the results.

Certifications:

- **IBM certifications on HIVE and PIG from Big data university.**

Strengths:

A systematic, organized, adaptive, hardworking and dedicated team player with an analytical bent of mind determined to become part of a growth-oriented organization.