Ranjith Kumar D

12135 Goodman St, Overland Park, KS 66213
+1 913-963-8638 | ranjithreddy0702@gmail.com | linkedin.com/in/ranjithh

Education

University of Central Missouri	**Apr 2021 - Sep 2022**
Master of Computer Science	*Warrensburg, MO*
ICFAI Business School	**May. 2017 - Mar 2019**
Master of Business Administration	*Hyderabad, India*

Relevant Coursework

- Data Structures
- Business Analytics
- Big Data
- Data Visualization
- Database Management
- Artificial Intelligence
- Internet Technology
- Systems Programming
- Computer Architecture
- Financial Analytics
- Advanced Excel
- Machine Learning

PROFESSIONAL SUMMARY:

- A Data Engineer and Data Scientist with a solid background in Computer Science with **3.5 years of Experience** working in a fast - paced
- Data-driven setting fostering collaborations to maintain the edge necessary for delivering products and services built on innovative technologies like Artificial Intelligence and Machine Learning.
- Natural knack for diving deep into a problem to find a creative solution with a roll-up-your-sleeve attitude.
- Dynamic and motivated IT professional with around 5+ years of experience as a Big Data Engineer with expertise in designing data-intensive applications using Hadoop
- Ecosystem, Big Data Analytical, **Cloud Data Engineering**, **Data Warehouse** / Data Mart, Data Visualization, Reporting, and Data Quality solutions.
- Strong experience working with Bigdata Hadoop ecosystem components like **HDFS**, **Map Reduce**, **Spark**, **HBase**, **Oozie**,**Hive**, **Sqoop**, **Pig**, **Flume**, **Kafka**, and **Airflow**.
- Expertise in developing production-ready Spark applications utilizing **Spark**-**Core**, **Data Frames**, **Spark-SQL, Spark-ML and Spark-Streaming APIs**
- Good experience working with **AWS Cloud** services like **S3**, **EMR**, **Lambda**, **Redshift, Athena, Glue** etc.,
- Good understanding of **NoSQL** Databases and hands-on work experience in writing applications on No SQL databaseslike **Cassandra** and **Mongo DB**.
- Familiar with data architecture, including data ingestion pipeline design, Hadoop information architecture, **data modeling** and **data mining**, **machine learning** and advanced data processing. Experience optimizing ETL workflows.
- Experienced at creating data regression models, using predictive data modeling, and analyzing data mining algorithms to deliver insights and implement action-oriented
- solutions to complex business problems
- Using machine learning techniques supervised, unsupervised, reinforcement learning and understand the requirement & design for AI/ML use cases
- Strong hands and exposure on experience in Python, (Frameworks/libs: NumPy, SciPy, Pandas, NumPy, Matplotlib/Plotly, scikit-learn, Statsmodels)

- Proficient in Python Scripting and worked in stats function with **NumPy, visualization** using **Matplotlib** and **Pandas** for organizing data.
- Knowledge in installation, configuration, supporting and managing **Hadoop Clusters** using Apache, Cloudera (CDH3,CDH4) distributions and Amazon web services (**AWS**).
- In-depth understanding/knowledge of Hadoop Architecture and various components such as HDFS, MapReduce Programming Paradigm, High Availability and YARN architecture.
- Experienced in working with **Amazon Web Services (AWS)** using **EC2** for computing and **S3** as storage.
- Capable of using AWS utilities such as **EMR**, **S3** and cloud watch to run and monitor Hadoop and spark jobs on Amazon Web Services (AWS).
- Strong knowledge in working with Amazon **EC2** to provide a complete solution for computing, **query processing**, and storage across a wide range of applications.

Technical Skills

Hadoop Components	HDFS, Hue, MapReduce, Hive, Sqoop, Impala, Zookeeper, Flume
Spark Components	Spark RDD, Data Frames, Spark SQL, PySpark, Spark Streaming, Spark ML
Databases	Oracle, Teradata, Microsoft SQL Server, MySQL.
Programming Languages	Java, Python, Scala.
Cloud Services	AWS S3, EMR, Redshift, Glue, Lambda, Step Functions, Athena, GCP Dataproc, Big Query, Nifi, Snowflake
IDE	Eclipse, IntelliJ IDEA, PyCharm
NoSQL Databases	HBase, Cassandra, MongoDB.
Release Management Tools	Jenkins, Maven, SBT, GitHub, Jira
Development methodologies	Agile/Scrum

Experience

Capital One   **May. 2022 – Present**
Big Data AWS Engineer   *Virginia, USA*

- Design and maintain well-structured relational database Schemes. Create automated **ETL pipelines** for a variety of raw manufacturing data sources. Research and develop new data storage architectures for growing data volumes
- Designing a resilient framework using **Amazon Web Services**, which involves reading/streaming data from the producer/Source through the **Apache Kafka** process to ingest the data to Client specific cloud storage location (**S3**). House, data validation tool is integrated to take care of data quality checks. Setting up notification service through **Amazon SNS (Simple Notification Service),** which in turn triggers **AWS Lambda** function to pick the data from one **S3** to a Centralized S3 location.
- The data load framework is being used to load the data from **S3 location** to the Cloud **Snowflake Data warehouse**. Ensure appropriate services and effective design in a cost-effective method.
- Intake Business managers/product owner request their strategies and future vision and turn those business requirements into **data visualization** products using the **Tableau** Business Intelligence Data **Visualization** tool.

- Created DataStage jobs using different stages like Transformer, **Aggregator, Sort, Join, Merge, Lookup, Data Set, Funnel, Remove Duplicates, Copy, Modify, and Filter. Change Data Capture, Change Apply, Sample, Surrogate Key, Column Generator, Row Generator Etc.,**
- Expertise in Creating, Debugging, Scheduling and Monitoring jobs using **Airflow** for ETL batch processing to load into **Snowflake** for analytical processes.
- Worked in building **ETL pipeline** for data ingestion, data transformation, and data validation on cloud service AWS, workingalong with data steward under data compliance.
- Worked on scheduling all jobs using **Airflow** scripts using python and added different tasks to **DAG** and **LAMBDA**.
- Experience in Developing **Spark applications using Spark - SQL** in **Databricks** or data extraction, **transformation**, and **aggregation** from multiple file formats for analyzing & transforming the data to uncover insights into the customer usagepatterns.
- Performed data storage on **AWS S3**, and defined **MapReduce** jobs in multi-node Hadoop clusters on AWS **EMR** and **EC2**
- Developed ETL pipelines to ingest data to **AWS RDS** for data processing, data analysis using AWS Glue and **AWS Athena**, and **data warehousing** and visualization using **AWS Redshift and AWS Quick Sight**
- Upkeep AWS Kinesis pipelines to process near real-time server logs, analyze data with **AWS Elasticsearch** and populate data in **AWS DynamoDB** using **AWS Lambda**
- Developed and implemented ETL pipelines using Python, SQL, Spark and PySpark to ingest data and updates to relevant databases

University of Central Missouri   May. 2021 – Jun 2022
Power BI Developer   *Reality Studio, UCMO*

- Extensive data cleansing and analysis, using pivot tables, formulas (v-lookup and others), **data validation**, conditional formatting, and **graph** and **chart** manipulation. Facilitate program and project meetings. Prepare and facilitate weekly
- PMO meetings; captured meeting decisions and content into Excel-based tracking logs.
- Involved in creating **MDX**, **DAX**, Cubes and Tabular models using **SQL Analysis Services** (SSAS). Used different types of **visualization** (**Stacked bar Chart, Clustered bar Chart, Scatter Chart, Pie Chart, Donut Chart, Line & Clustered Column Chart, Map, Slicer** etc.) in Power BI according to the requirements.
- Conduct quality assurance activities regarding data migration and folder permissions. Adhere to and communicateRecords Management procedures to customers. Perform SharePoint administration for the team.
- As Dashboard Developer, Primary activities included designing a technology roadmap for Power BI; product installation and implementation development of insightful dashboards; serving as **SME** point of contact for Power BI and delivery ofsolutions adhering to BI industry best practices
- Implemented metrics, filters, reports, and dashboards and created advanced chart types, using calculations to manipulate thedata and enable user-driven parameters for visualizations and drilled down the level of detail using sets

Yashitha Software Inc   Dec 2019 – Feb 2021
Data Engineer   *Hyderabad, India*

- Collaborate with ML Engineers and Data Scientist to build data and model pipelines and help in running machine learning tests and experiments
- Experience with Statistical Modeling, Data Extraction, Data cleaning, Data screening, Data Exploration and Data Visualization of structured and unstructured datasets

- Ability to implement large scale Deep Learning and Machine Learning algorithms to deliver resourceful insights and inferences
- Using machine learning techniques supervised, unsupervised, reinforcement learning as applied to the travel, transport, or logistics industry
- Collaborates with business and team members to understand company needs, employs predictive modeling, text mining, and forecasting techniques, Keeps up-to-date
- with analytics technology and
- Expertise in designing and deploying **Hadoop clusters** and different Big Data analytic tools, including **Pig, Hive, HBase, Oozie, Sqoop, Flume, Spark, and Impala.**
- Effectively use metrics and data visualization to report progress and support data-driven decision-making at the team or organization level. Perform structural modifications using **Map - Reduce, HIVE** and analyze data using visualization/ reporting tools.
- Effectively use metrics and data visualization to report progress and support data-driven decision-making at the team ororganization level. Perform structural modifications using **Map - Reduce**, HIVE and analyze data using visualization/ reporting tools.
- Worked on importing data from **MySQL** DB to **HDFS** and vice-versa using **Sqoop** to configure **Hive** metastore with**MySQL**, which stores the metadata for Hive tables.
- Collaborated with team members using version control systems such as Git to organize modifications and assign tasks.
- Imported data from sources like different logs, **API's**, and databases like **MySQL**, **Neo4**), **Cassandra**, and **DynamoDB** to understand the customer behavior and make the necessary reports as per client requirement
- Worked on **Python / Scala APIs** to get data from **DynamoDB**, **MYSQL**, and **Cassandra** and loaded it into **Databricks** for further analysis
- Loaded python scripts into **AWS S3**, **DynamoDB** and **Snowflake** to read CSV, JSON and parquet files from S3 buckets
- Extracting data wif **Apache Kafka** and using Scala for converting the distributed collection of data organized into named columns, developing predictive analytics using **Apache Spark Scala APIs**
- Importing data from S3 into **AWS Glue** to perform the ETL operations and store the data into **Redshift**
- Involved in converting **Hive / SQL queries** into Spark transformations using **Spark RDDs** and **Scala**
- Developed Scala scripts, UDF's using both data frames / SQL and RDD in Spark for data aggregation, queries and writing back into S3 bucket

CERTIFICATIONS

- Advanced Excel Skills for Business.
- Machine Learning, by Andrew Ng.
- Deep Learning, by Andrew Ng.
- IBM Data Engineering.
- Data Analysis Using Pyspark
- Python for Data Science, AI & Development

Extracurricular

International Student Organization (UCMO)   **Jun 21 – Sep 22**
Vice president   *University of Central Missouri*
- Work with the zone chairperson to ensure every club in your zone operates efficiently and follows the Association'sConstitution & Bylaws.

- Discuss ways of helping clubs that need assistance with membership growth or leadership development.
- Promote various club functions and events, such as inter-club meetings, installation of club officers, induction of newmembers, or ceremonies honoring Key Award recipients.

DOT Club (IBS-HYD)   **Jul 17 – Mar 19**
Tech Head   *University of Central Missouri*

- Determining project requirements and developing work schedules for the team.
- Delegating tasks and achieving daily, weekly, and monthly goals
- Keeping up to date with industry trends and developments.
- Updating work schedules and performing troubleshooting as required.
- Motivating staff and creating a space where they can ask questions and voice their concerns.