

KARTHIK TEJA KATREDDI

Frisco, Texas 75033 | (317) 712-7510 | [katreddikarthik@gmail.com](mailto:katreddikarthik@gmail.com) | [Linkedin/](#) | [github](#) | [Portfolio](#)

EDUCATION

- Data Science Master’s Degree**  
Indiana University-Purdue University Indianapolis, IN
  - Bachelor of Technology in Computer Science and Engineering**  
Jawaharlal Nehru Technological University Kakinada, India
- GPA: 3.74/4  
May 2020

GPA: 8.65/10  
May 2017

TECHNICAL SKILLS

- Programming Languages** : Python, Java, R programming, HTML, CSS, JavaScript, D3.js, PHP
- Query Language** : MS SQL, PL/SQL, PySpark, MongoDB, Apache Spark, Cassandra
- Machine Learning Libraries** : NumPy, Pandas, Scikit- Learn, TensorFlow, NLTK, Keras, Matplotlib
- Methodologies** : SDLC, Agile and Scrum Methodologies
- Cloud Technologies** : Azure SQL, Azure Blob, Azure Data Lake, Azure Data Factory, AWS
- Tools and Technologies** : ERwin, SQL Server Management Studio, Tableau, Git/GitHub, Jupyter, Visual Studio

PROFESSIONAL EXPERIENCE

Cerner Corporation | Kansas City, Missouri Feb’2021 – Present

- Data Engineer
- Worked on building modern **data pipelines** to ingest, process, transform, and model data into Cerner's Corporate Data Platform.
  - Participate in architecture discussions, design sessions and code reviews to review for functional correctness, architectural maintainability, and performance.
  - Use big data tools to design batch and streaming feature pipelines and pipelines to populate a **data lake**.
  - Develop and document the strategy to maintain and support the platform, pipelines, and tooling.
  - Worked on **Kafka APIs** to ingest the data into azure data lake.
  - Created ETL pipelines from corporate systems using a combination of **Azure Data Factory, Azure Data Lake, Azure Databricks, Snowflake and data analytics**.
  - Created clear, well-constructed strategic technical designs for large or complex scope projects with assistance.
  - Worked on **python** and **spark** for transforming the data in data lake.
  - Worked on **GIT, GitHub** for version control and deployment.

Cerner Corporation | Kansas City, Missouri Aug’2019 – Jan’2021

- Data Engineer
- Writing **python** scripts for automating to fetch and integrate data into **SQL** servers, API from the web.
  - Reviewed, Debugged, and improved the performance of **R, Python, PySpark** scripts, and created clear technical documentation.
  - Created Pipelines in **ADF** using Datasets/Pipeline/ to **Extract, Transform, and load** data from different sources like Azure SQL, Blob storage, Azure SQL Data warehouse.
  - Created different pipelines using **data factory** for migrating on-prem **HDFS** server to **Azure server**.
  - Data taken from the Database Management team is converted into Datasets by writing SAS scripts.
  - Experienced in debugging, manipulating data, cleaning, analyzing, and extraction from web servers with a vast range of python libraries like **NumPy, Pandas, BeautifulSoup, Matplotlib, Seaborn, ggplot**.
  - Developed various dashboards using **Tableau** having data points coming from various resources like flat files, MySQL databases, and Hadoop File systems.
  - Helped in creating a Master Data Management System using Hadoop Ecosystem technologies namely, **PIG, HIVE, and HBase**. Sqoop is used for integrating preexisting SQL databases and **Oozie** is used to create Job schedulers.
  - Developed real-time data pipelines to extract streaming data using **Apache Kafka** and Spark Streaming.

Indiana University | Indianapolis, Indiana Feb’2019 – Jul’2019

- Research Assistant
- Used **PHP** and **XAMPP** framework for application development for cancer patients.
  - Designed and developed the UI of the website using **HTML, AJAX, CSS, and JavaScript**.
  - Worked on CSS Bootstrap to develop web applications.

Tata Consultancy Services | Hyderabad, India Jun 2017 – Jan 2019

- Data Engineer
- Involved in daily **Scrum** meetings, Sprint planning and estimation of the tasks for the user stories, participated in retrospective and presenting Demo at end of the sprint. Communication with external teams, verbally and written daily.
  - Worked as a Data Engineer for making automation scripts for data extraction and data loading to the **SQL** servers using python and **SSIS** Packages which are later used for **performing analytics**.
  - Participated in the full product development life cycle including analysis, design, development, deployment, and operations support.
  - Various types of data and big data which are in HDFS are connected to Tableau and created interactive Dashboards.
  - Based on the Analytics provided by Tableau, KPIs are found. The **SparkML** library is then used in the **PySpark** to predict the performance based on the KPIs given.
  - Developed complex jobs to perform CRUD operations and analysis of data in Cassandra distributed database using Spark Data Frame API.
  - Created Oozie workflows that collect data from the live **SQL** servers and load them into **HDFS** using **Sqoop**.

PROJECTS

- Development of Classification Models and Predictive Analysis using Loan data**

  - Managed and analyzed a large data of 3 GB using Spark RDD’s transformation, Spark SQL. Performed feature selection to figure out optimal feature set using Chi-Squared test in Python. And building the pipeline for vectorizing the features.
- Decoding American Sign Language with Deep Learning**

  - Pre-processed the 1 GB of image data using one hot encoding. Used CNN models and deeper networks like Alex net, VGG16 to predict the images.
  - Compared the accuracy of test images and visualized the predicted images using confusion matrix.