# YASH RAYTHATHA

yashbraythatha@gmail.com | (973)380-6309 | LinkedIn | Connecticut

## PROFESSIONAL SUMMARY

Experienced Data Engineer with expertise in designing efficient ETL pipelines, automating data workflows, and optimizing query performance to drive business insights. Proficient in Python, SQL, and AWS technologies, with a strong focus on building scalable, robust data solutions. Skilled in documenting data infrastructures to enhance collaboration and knowledge sharing. Driven to apply technical expertise in innovative projects, delivering actionable, data-driven solutions that empower forward-thinking organizations to achieve their goals.

## SKILLS

**Languages**: Python, C/C++, Unix, HTML, CSS.
**Data Pipelines and Cloud Application**: Pyspark, Spark SQL, AWS
**Data Visualization:** Tableau, Matplotlib, Seaborn, Power Bi, Quicksight
**Database:** SQL, Snowflake, MongoDB, Oracle
**Certifications:** Azure Administrator (AZ-104), Power BI Analyst Associate (PL-300), Oracle SQL Database Associate (1ZO-071)

## PROFESSIONAL EXPERIENCE

**Data Engineer** Aug 2019 - Mar 2023
HCL Technologies, Nagpur, India

- Led the development of scalable ETL pipelines using Apache Spark, increasing data processing efficiency by 20% and reducing latency in downstream analytics.
- Optimized Spark SQL queries, reducing execution time from 2 hours to just 10 minutes, saving over 100 compute hours per month and cutting infrastructure costs by 30%.
- Automated data ingestion and transformation tasks using Python and AWS Lambda, reducing manual intervention by 80% and improving workflow efficiency.
- Processed and maintained datasets exceeding 10TB, ensuring high data integrity and availability for analytics and reporting teams.
- Authored and maintained technical documentation for 100% of data sources and ETL pipelines, improving team collaboration and onboarding efficiency.

**Application Support Engineer** Nov 2018 - Aug 2019
Savantis Solutions, Hyderabad, India

- Improved application performance by 30% by identifying and resolving bottlenecks in SQL queries and backend processes.
- Reduced incident resolution time by 40%, minimizing downtime and improving user experience by implementing a proactive monitoring system.
- Led the implementation of a new incident management process, reducing ticket backlog by 60% and enhancing response times.
- Conducted root cause analysis for 50+ critical incidents, implementing preventative solutions that reduced recurrence by 70%.

## ACADEMIC PROJECTS

**Twitter Data Analysis Using Apache Airflow**

- Developed and managed a robust Apache Airflow ETL pipeline on AWS EC2 to automate the extraction, transformation, and loading of Twitter data from S3.
- Enhanced data integrity and streamlined processing by optimizing transformation and storage operations, resulting in improved efficiency.
- Leveraged Amazon Quick Sight to create dynamic visualizations of the transformed data, facilitating advanced analytics and informed decision-making.

**Biomedical Named Entity Recognition (NER) using Transformer Models**

- Conducted a comparative analysis of transformer-based models (Bio BERT, BERT) combined with BiLSTM and CRF for Biomedical NER on the BC5CDR dataset.
- Implemented advanced pre-processing and fine-tuning strategies to achieve the highest F1-score of 90% using Bio BERT with CRF.
- Evaluated models on key metrics such as precision, recall, and F1-score, demonstrating robust entity recognition for diseases and chemicals.
- Highlighted the importance of domain-specific pretraining and structured sequence modeling for handling complex biomedical texts.

**Campus Image Semantic Segmentation Using DeepLabV3**

- Implemented a semantic segmentation model using DeepLabV3 with ResNet-50 backbone to classify campus images into doors, stairs, and background.
- Created and annotated a custom dataset of 200 images with pixel-wise precision using Roboflow.
- Achieved 63% mIoU and 83% pixel-wise accuracy through transfer learning, data augmentation, and hyperparameter tuning.
- Visualized segmentation results in validating model performance.

## EDUCATION

**Bachelor of Engineering: Computer Technology** May 2017
*Rajiv Gandhi College of Engineering Research & Technology - Chandrapur, Maharashtra, India*

**Master of Science: Data Science** May 2025
*University of New Haven, West Haven, CT*