

# Piyush Jain

## Data Engineer

✉ piyushjain1999kota@gmail.com

☎ +91 9079893819

🌐 [LinkedInProfile](#)

### Skills

- Programming Language: Python
- Frameworks & Libraries: PySpark, Pandas, Tableau Hyper API, Elasticsearch, Hadoop, cx\_Oracle, REST API, SFTP
- Database Management: Oracle SQL
- Tools & Software: Pycharm, Tectia (UNIX), Oracle SQL developer, Jenkins, Autosys. GIT, MS Office

### Experience

**LTIMindtree Ltd.**

**Jul 2021 - Present - 2.1 years**

#### Project\_1: Elastic Search Indexer (CITI Bank Project - Banking & Finance Domain)

- Designed and implemented a mission-critical Elasticsearch integration project to introduce an intuitive auto-complete search feature for a robust reporting engine.
- Created a data pipeline, enabling lightning-fast indexing of 3-5 million documents per batch.
- Leveraged advanced PySpark techniques to streamline data processing, resulting in a significant reduction in storage overhead and improved data ingestion performance.
- Implemented Elasticsearch Spark Connector with precision, introducing bulk indexing to the pipeline, which decreased indexing time by 79% and alleviated resource contention.
- Expertly fine-tuned resource allocation and refresh intervals, resulting in a 25% boost in indexing efficiency, showcasing exceptional optimization skills.

#### Project\_2: Cloud Reference Data Consumption (CITI Bank Project - Banking & Finance Domain)

- Designed and implemented a dynamic data pipeline that transformed unstructured data into actionable insights, elevating reporting processes.
- Devised a comprehensive Python and PySpark solution, orchestrating seamless data integration and processing for diverse sources, including MongoDB.
- Architected efficient data retrieval from MongoDB database, through REST API and SFTP services, optimizing JSON content extraction while handling large requests.
- Processed extensive datasets (up to 20 million records) utilizing Python/PySpark, with an 83% improvement in data processing time and efficiency, expediting data ingestion (cx\_Oracle for small datasets) with a consistent flow of accurate information.
- Amplified reporting capabilities by delivering refined table formats to multiple reporting touchpoints, enhancing data accessibility and utility enabling stakeholders to make informed decisions efficiently.

#### Project\_3: Tableau Hyper File Project (CITI Bank Project - Banking & Finance Domain)

- Spearheaded an end-to-end Python automation project focused on seamless data extraction, conversion, and visualization for an agile analytics workflow.
- Designed and executed a streamlined process to extract source data in JSON format and seamlessly transform it into CSV files.
- Leveraged Python's `tableauhyperapi` library to create Tableau hyperfiles and automated publication to server, with robust data validation, ensuring data accuracy.
- Implemented data retention strategies, archiving old data as new datasets were published, and optimizing data management.
- Significantly enhanced data accessibility and business insights by providing immediate visibility of dynamic datasets on dashboards without any manual intervention.

### Education

Bachelor of Technology in Mechanical Engineering  
*Indian Institute of Technology (ISM) Dhanbad, Jharkhand, India*

Jul 2017 - May 2021