

# YATIN CHADHA

E-mail: [yatinchadha95@gmail.com](mailto:yatinchadha95@gmail.com) | Mobile: 9205448520 | [linkedin.com/in/yc95/](https://www.linkedin.com/in/yc95/)

---

## SUMMARY

- Result-driven Data Engineer with over 5 years of experience specializing in the design, development, implementation, and testing of various projects in Data Engineering, Data Warehousing, and Data Analysis, utilizing Agile Methodologies.
- Demonstrated proficiency in Big Data technologies such as PySpark, SQL, Hive, and others, coupled with expertise in ETL and reporting tools used in Business Intelligence. Proficient in Python programming.
- Acquired hands-on experience in creating comprehensive end-to-end data pipelines in Microsoft Azure, enhancing data flow and efficiency.
- Established proficiency in working with Apache Airflow, Spark Streaming and Kafka.
- Experienced in scripting and automating applications using Shell Scripts, contributing to efficient data pipeline management.
- Quick learner and excellent team player

## WORK EXPERIENCE

### Associate Staff Engineer, Nagarro

*(Aug. 2023 - Present)*

- Engaged in Big Data Analytics Projects for a leading Retail company, driving data-driven decision-making processes.
- Optimized existing data pipelines, resulting in a significant reduction in execution time and enhancing overall efficiency.
- Implemented real-time data streaming solutions by leveraging technologies such as Apache Kafka and Spark Streaming.
- Gained hands-on experience with Apache Airflow, contributing to the management and scheduling of complex workflows.
- Migrated the existing repositories from Jenkins to LOCI, our proprietary Continuous Integration/Continuous Deployment (CI/CD) system, ensuring seamless project transitions.

### Data Engineer, Optum

*(Oct. 2021 – July 2023)*

- Worked on Big Data Analytics Projects for one of the largest Healthcare services providers to process and develop end-to-end data pipelines.
- Collaborated with cross-functional teams to identify and analyze data sources to support strategic business decisions.
- Extracted data from various sources including APIs, websites, databases, and unstructured data.
- Developed data warehousing solutions using UNIX Shell, Hive, and PySpark and worked with Bot processing teams to extract data from the main operations database of Optum for its Specialty medicines and uploaded to the database using Hive.
- Designed and developed data pipelines using ADF, Databricks, Azure SQL, Delta Lake, Key Vault, SHIR, Unity Catalog, etc.
- Worked with Data Science team to develop NER and predictive models.
- Developed custom algorithms and scripts to extract, clean and structure unstructured data and automate using PySpark, pandas, Hive, Sqoop and Shell scripts.
- Mentored junior team members on data engineering best practices and techniques.

## Senior Systems Engineer, Infosys Ltd.

(July 2018 – Oct. 2021)

- Worked in Prepaid Account for leading Financial Services Company.
- Extracted and analyzed client's customer data to identify key metrics; developed business intelligence reports and dashboards in Tableau, driving effective business decisions involving millions of users.
- Developed Tableau data visualization using cross tab, bar chart, pie chart and scatter plots.
- Worked extensively on creating calculated fields, sets, hierarchies, and parameters using Tableau Desktop.
- Also, worked in Prepaid Account DB Migration from DB2 to YellowBrick database.
- Responsible for migration of DB components like complex stored procedures, views, and reporting queries along with UNIX Shell scripts to call the YellowBrick DB components from the execution framework.
- Developed Informatica Job connection/override queries and design changes due to incompatibility/limitations of YellowBrick and PostgreSQL connectors.
- Parallel run/data validation for maintaining data integrity.

## SKILLS

Programming Languages	:	Python
Operating Systems	:	Windows, Linux
Database	:	MySQL, PostgreSQL, SQL Server, Oracle, MongoDB
Orchestration Tool	:	Airflow
Big Data	:	Spark, Hive, Kafka, HDFS, YARN
Azure Services	:	Databricks, Data Factory, ADLS, Blob Storage
CI/CD	:	Git, Bitbucket
Python Libraries	:	BeautifulSoup, Numpy, Pandas, Scikit-Learn, Scipy, NLTK, Spacy
Data Visualization Tool	:	Tableau
IDE Tools	:	Anaconda Navigator, Google Colab, PyCharm, VSCode
Agile Methodologies	:	Jira Scrum Boards, Confluence

## PROJECTS UNDERTAKEN

### Loyalty Wallet Events Pipeline

- The objective of this project was to design and implement an end-to-end ETL pipeline using Kafka, HDFS, Hive, PySpark, Python libraries and orchestrate/monitor the pipeline using Airflow.
- Ingested raw data from Kafka topics, performed data compaction for Spark efficiency, and segregated the JSON data based on different values in various "resources" key. Also, implemented the logic to handle corrupt records from the Kafka topic.
- Created Silver Tables (snapshot and history) from the raw JSON data to create structured tables. These tables will be used for intermediate data storage. Also validated this silver tables data against the source tables (upstream tables) based on product team metrics.
- Developed the final set of Gold Tables for dashboard creation by the Analysts and Product team in Superset.

### Rebate UMCS Data Digitization Pipeline

- Developed end-to-end data pipeline to extract data from UMCS Word documents by implementing extensive data cleaning and transformations using python libraries.
- Used Azure Databricks to create PySpark code to fetch Delta Lake tables data and store it in ADLS.
- Integrated ADLS files in the main code to use those as lookup while fetching KPIs from Word documents.
- Scheduled notebook inside Azure Databricks for monthly data load refresh in ADLS.
- Collaborated with Data Science team to develop predictive models and integrate them in data pipeline code.
- Stored cleaned data in JSON files and shared with the UMCS team.

## Order Processing Automations

- Led the project to enable Bot processing on selected orders satisfying eligibility criteria.
- Extracted orders from the main operations database of Optum for Specialty medicines and uploaded them to the HIVE tables by designing SQL queries from scratch.
- Carried out attributes manipulation to cater to data requirements for Bot processing.
- Created hive external tables with partitioned columns.
- Developed shell scripts to create or update intermediate and final hive tables for report generation.
- Also, developed end-to-end data pipelines using Azure services like ADF, Databricks, SHIR, ADLS, Key Vault, etc.
- Managed multiple sub-projects following the same methodology, including Refill-Too-Soon, Day-Supply, RX Data Check validation automations.

## EDUCATIONAL QUALIFICATIONS

**B. Tech in Computer Science** - GGSIPU, Delhi (Percentage: 73%)

*(August 2014 - June 2018)*

Coursework – Operating Systems, Discrete Mathematics, Data Structures, Object-Oriented Programming, Algorithms, Databases, Computer Architecture

## CERTIFICATIONS

- Python for Data Science - *IBM*
- Distributed Computing with Spark SQL - *Coursera*
- Tableau for Data Science & Business Analytics - *Udemy*
- DP-900 Azure Data Fundamentals - *Microsoft*