

Sasank Varanasi

+1 (617) 906 1433 - varanasidbzt@gmail.com - [LinkedIn Profile](#) - [github-college](#) - [github-personal](#)

EDUCATION

University of Massachusetts Boston

MS Computer Science

Boston, MA, USA

Sep 2022 - May 2024

National Institute of Technology, Arunachal Pradesh

B.Tech Electrical and Electronics Engineering

Yupia, Arunachal Pradesh, India

Aug 2012 - Jun 2016

TECHNICAL SKILLS

Programming Languages: Python, SQL, Java, JavaScript, ABL (Progress 4GL), C

Data Engineering Tools: Git, Docker, ETL workflows, Tableau, Confluence, JIRA

Database Management: Relational Databases (MySQL, PostgreSQL, SQL Server), NoSQL (MongoDB, Oracle)

Frameworks: Django, Flask, REST API, FastAPI

Libraries & Utilities: Pandas, Numpy, Sklearn, OpenCV

Big Data & Data Quality Tools: Internal EDQ tools, data rule validation systems

Version Control: GitHub, BitBucket

Visualization Tools: Tableau, Kibana, Excel, Google Sheets

Operating Systems: Linux, Windows 10/8/7, Windows Server 2012 R

WORK EXPERIENCE

Data Engineer

Jun 2023 - Jan 2024

University of Massachusetts Boston, Boston, MA, US

- Transformed Python-wrapped SQL scripts for operational workflows and ETL pipelines.
- Analyzed datasets for rule validation, converting rules into Python scripts to process data.
- Validated data quality using internal tools (EDQ) to streamline data rule testing and file generation.
- Built workflows to execute scripts on sample data and evaluate results against benchmarks.
- Supported iterative development of data processing pipelines, focusing on scalability.
- Maintained GitHub repositories for collaborative code review and version control.
- Tools: Python, SQL, Git, Google Sheets, MongoDB, ReactJS.

Data Engineer

Nov 2021 - Jul 2022

Datycs, Hyderabad, India

- Designed and implemented ETL pipelines to extract, transform, and load structured and unstructured datasets into FHIR-databases for seamless data integration.
- Developed a Python-based framework to validate data quality, ensuring consistency and compliance with predefined business rules.
- Conducted advanced data cleaning and preprocessing techniques to prepare datasets for analytics and machine learning pipelines.
- Automated unit testing frameworks for data extraction, transformation, and rule-based validation using Pytest and Python.
- Migrated legacy codebases to PEP-8 compliant standards to improve maintainability and performance of data pipelines.
- Collaborated with stakeholders to understand business requirements and translate them into actionable data models and transformation logic.
- Maintained and version-controlled data engineering workflows using Git and BitBucket, ensuring seamless team collaboration.
- Optimized data workflows to improve data processing speeds and accuracy, enabling timely delivery of analytics-ready datasets.
- Tools: Python, SQL, FHIR-database, Git, Pytest, BitBucket, JIRA.

- 4+ years of experience in software development with strong expertise in Python/Django web application development and Progress 4GL (ABL).
- Managed and processed large datasets in Excel, specializing in data migration, updates, and upgrades for JIRA and HP ALM applications.
- Acted as an administrator and application owner for the Fuel tracker application, streamlining fuel management processes at depots using VINs and PINs.
- Proficient in handling RFSs, CRQs, solution planning, Change Approval Board (CAB) calls, ETL workflows, and decommissioning of servers.
- Coordinated with field personnel to map terrain and implemented RESTful services using the Django framework.
- Demonstrated expertise in managing source code with version control systems, handling large databases, and establishing Python-based connections.
- Knowledgeable in OOP, XML parsing, web scraping, CI/CD pipelines for code deployment, and migrating business objects between versions.
- Experienced in all phases of the software development lifecycle, including requirement gathering, technical design, development, and post go-live activities.
- Tools: Python, ABL (Progress 4GL), HTML, CSS, ElasticSearch, Git, JIRA, HPQC.

PROJECTS

CAMS

Nov 2021 - Jul 2022

University of Massachusetts, Boston, US

- Spearheaded the design and development of a data engineering pipeline to optimize student allocation to projects using the Min-Cost Flow algorithm.
- Engineered Python scripts to process, clean, and validate input datasets, ensuring alignment with business rules.
- Integrated Python-based backend logic with ReactJS front-end for seamless data visualization and functionality.
- Deployed the complete solution to Render, ensuring scalability and operational efficiency.
- Conducted extensive testing using sample datasets to validate results and ensure data quality.
- Managed data-related operations, including extraction, transformation, and load (ETL) processes for student project data.
- Utilized Git for version control, enabling effective collaboration and ensuring code integrity.
- Tools & Tech: Python, SQL, MongoDB, ReactJS, Git.

DATYCS Data Framework

Nov 2021 - Jul 2022

Datycs, Hyderabad, India

- Built ETL workflows to handle large-scale structured and unstructured data, ensuring seamless ingestion into FHIR-compliant databases.
- Created a Python-based data validation framework to verify data quality, detect anomalies, and enforce business logic rules.
- Automated data preprocessing pipelines, enabling efficient transformation and loading of datasets.
- Enhanced data engineering workflows by migrating legacy scripts to PEP-8 compliant standards.
- Developed and tested unit-testing modules for data transformation and rule validation using Pytest.
- Collaborated with cross-functional teams to translate business requirements into actionable data engineering solutions.
- Optimized SQL queries for faster data retrieval and reporting, improving pipeline performance.
- Tools & Tech: Python, SQL, FHIR-database, Pytest, BitBucket, Git.

MediaWorks Application (Data Engineering)

Tech Mahindra, Hyderabad, India

Oct 2019 - Jun 2021

- Developed a scalable Django-based application to process and manage user data, supporting RESTful APIs for seamless data exchange.
- Designed and integrated job schedulers to automate data collection and transformation workflows.
- Implemented Elasticsearch for advanced search capabilities, enabling rapid data retrieval and analytics.
- Built interactive dashboards using Kibana to visualize critical data metrics and enhance decision-making.
- Established CI/CD pipelines for automated code deployment, ensuring operational continuity.
- Integrated Gmail and Facebook APIs for user authentication and streamlined data onboarding processes.
- Managed and maintained version control for data engineering workflows using Git.
- Tools & Tech: Python, Django, Elasticsearch, Kibana, Git, SQL.

OGE Fuel Tracker System (Data Engineering)

Tech Mahindra, Hyderabad, India

Jan 2018 - Oct 2019

- Designed and implemented ETL processes to migrate and synchronize data between JIRA and HP ALM platforms using JIRA Bridge.
- Automated data extraction and reporting workflows using MySQL and Python scripting, improving reporting accuracy and timeliness.
- Integrated Fuel Tracker with SAP MM modules, enabling seamless data flow and improving operational efficiency.
- Developed a proprietary geospatial mapping system to track fleet movements and tank farm locations using geocoordinates.
- Decommissioned legacy servers in coordination with database teams, streamlining data storage processes.
- Created and distributed automated Excel reports from PDF data using Python, reducing manual effort by 70%.
- Supported real-time data visualization on Tableau dashboards for offshore and onshore teams.
- Tools & Tech: Python, SQL, SAP, Tableau, JIRA, MySQL.

Thomson Reuters (Data Engineering)

Tech Mahindra, Hyderabad, India

Jan 2017 - Jan 2018

- Utilized Progress 4GL (ABL) to implement data processing logic and patch existing code based on client requirements.
- Performed unit testing to validate data pipelines and ensure code accuracy and reliability.
- Analyzed and optimized data workflows to improve system performance and reduce latency.
- Participated in Agile development processes, including daily stand-up meetings and sprint reviews, ensuring timely delivery of data-related tasks.
- Collaborated with cross-functional teams to gather requirements and develop data-driven solutions tailored to business needs.
- Maintained detailed documentation for data engineering processes and shared daily progress reports with stakeholders.
- Integrated quality checks within data pipelines to ensure compliance with client specifications.
- Tools & Tech: Progress 4GL (ABL), SQL, Python, Agile methodology.

TRAININGS AND CERTIFICATIONS

- The Ultimate MySQL Bootcamp: Go from SQL Beginner to Expert Udemy, 2021
- Python Basic HackerRank, 2020
- Crash Course on Python Coursera.org, 2020
- Python Project: pillow, tesseract, and opencv Coursera.org, 2020
- Capstone: Retrieving, Processing, and Visualizing Data with Python Coursera.org, 2020
- Python for Everybody Coursera.org, 2020
- Python Classes and Inheritance Coursera.org, 2020
- Data Collection and Processing with Python Coursera.org, 2020
- Python Functions, Files, and Dictionaries Coursera.org, 2019
- Using Python to Access Web Data Coursera.org, 2020
- Programming for Everybody (Getting Started with Python) Coursera.org, 2019