



**Ravi Roshan**  
Principal Data Engineer

Noida, Uttar Pradesh, India

+91 8527212992, 8527499482

[raviroshan494@gmail.com](mailto:raviroshan494@gmail.com)

[linkedin.com/in/ravi-roshan/](https://www.linkedin.com/in/ravi-roshan/)

#### Technical Skills:

##### Programming:

Python, VBA, JavaScript, HTML

##### Backend & APIs:

REST APIs, OAuth 2.0, Webhooks

##### Frameworks:

FastAPI, SQLAlchemy, Pandas, BeautifulSoup, Selenium, Requests & more

##### Databases:

PostgreSQL, SQLite, Snowflake, BigQuery

##### Cloud (AWS):

EC2, S3, Athena, RDS, Lambda, CloudWatch, IAM, ECR

##### AI/LLMs:

ChatGPT, Gemini, Perplexity, AI Agents, MCP Server

##### Other Tools:

Excel, Airbyte, Zyte Proxy, Github, ZeroBounce, Jira

## SUMMARY

Principal Data Engineer with 10+ years of experience in large-scale data engineering, backend/API development, automation, and cloud-based data systems. Deep expertise in web scraping, high-volume data pipelines, data validation frameworks, and cloud architecture. Skilled in building scalable platforms, designing intelligent automation systems, and leading engineering teams. Passionate about solving complex data challenges using modern engineering, AI agents, and automation to deliver high-impact data products.

## EXPERIENCE

### June 2023 – feb 2026

#### Principal Data Engineer @ [iCustomer.ai](#) , USA (Boston) (Remote)

- Lead end-to-end planning, execution, and delivery for large-scale data engineering and automation initiatives.
- Architect high-volume web scraping and data extraction systems for complex, distributed, and dynamic data sources.
- Build robust automation frameworks that ensure continuous, reliable ingestion, validation, and transformation of data.
- Design and implement ETL & Reverse ETL pipelines for both structured and unstructured datasets across multiple product lines.
- Develop internal rule engines, QA automation layers, and data standardization APIs to ensure accuracy, consistency, and governance.
- Build, deploy, and maintain scalable REST APIs for data enrichment, identity resolution, taxonomy mapping, and operational workflows.
- Deliver multiple integration pipelines with platforms such as Salesforce, HubSpot, Mailchimp, Amplitude, and Google Analytics.
- Architect and enhance the Visitor Intelligence Engine, combining behavioral signals, event tracking, and enrichment data.
- Lead development of AI-driven automation tools, including prompt-engineered agents for research, enrichment, and similarity scoring.
- Build custom developer tools using MCP (Model Context Protocol) to streamline engineering workflows and AI-powered automations.
- Own data quality, governance, observability, and system reliability across the company's data ecosystem.

### Aug 2015 – Jun 2023

#### Lead Data Cloud Engineer @ Zylotech inc.(Acquired by Terminus) (Noida)

- Led teams across web scraping, data extraction, transformation, and QA automation.
- Designed advanced scraping pipelines processing millions of global records.
- Built industry-specific prospect databases using social media, directories, and deep web sources.
- Developed Python tools, APIs, and automation scripts improving efficiency and data accuracy.
- Evaluated and validated datasets to ensure accuracy, integrity, and consistency.

## AWARDS

- Winner's Award (2016)
- Spotlight of the Year (2017)
- Innovator of the Year (2019, 2021)

## Languages

- Hindi
- English

## Education

2020 - 2021

IGNOU

Master of Public Administration

2024 - 2026

Kurukshetra University

Master of Computer Application

## Interests

Nature & Travel · Photography ·  
Sci-Fi Movies

- Created backend internal APIs for data quality checks, enrichment, and transformation workflows.
- 

## Oct 2013 – Jan 2015

### Research Analyst @ Denave India Pvt. Ltd. (Noida)

- Managed research and lead-generation projects for Microsoft, Cisco, CtrlS Datacenters, GetIT, and Yash Technologies.
  - Conducted IT infrastructure research, cloud demo lead generation, and customer onboarding support.
  - Provided technical support for Microsoft Lumia smartphone users.
- 

## CORE EXPERTISE

- Leading data engineering teams, defining roadmap, and driving platform reliability.
- Architecting end-to-end data ecosystems (Collection — ingestion → standardization → enrichment → delivery).
- Designing scalable web-scraping frameworks and high-volume extraction pipelines.
- Implementing validation engines, rule-based QA automation & data governance systems.
- Delivering integrations across CRM, marketing, and analytics platforms.
- Developing AI-powered APIs and agents for research, enrichment, and scoring.