






Rahul Singh — Data Analyst | AI/ML Engineer | Data Warehouse Specialist

 India |  Remote Available
9+ Years Experience |  Data Analytics |  AI/ML |  Automation

Professional Summary

Results-driven **Data Analyst & AI/ML Engineer** with over 9 years of experience in **data analytics, data warehousing, and business intelligence**. Proven expertise in transforming complex datasets into actionable insights, building scalable data pipelines, and implementing AI-driven solutions.

Specialized in:

- Web analytics & usage data analysis
- Data warehouse optimization (performance + cost)
- AI-powered automation and predictive analytics
- Dashboarding & business reporting

I help businesses **turn raw data into revenue-driving insights**.

Core Skills

Data & Analytics

- SQL (Advanced), MySQL, PostgreSQL
- Data Modeling (Star Schema, Fact-Dimension Design)
- Data Warehousing (Hive, EMR, Lakehouse Architecture)
- ETL Pipelines, Data Cleaning, Aggregation

AI / Machine Learning

- Python (Pandas, NumPy, Scikit-learn, TensorFlow)
- NLP, OCR (LaTeX OCR Model with CTC Loss)
- Predictive Modeling
- LLM Integration (AI automation, chatbots)

Visualization & Tools

- Power BI / Tableau / Plotly Dash
- Advanced Reporting Dashboards

- Excel (Advanced Analytics)

Tech Stack

- Python, Django, Flask
 - Spark (PySpark)
 - AWS (S3, EMR)
 - REST APIs, JWT Auth
 - Git, CI/CD
-

Professional Experience

Manager – MPS Insight & Analytics

Current Role

- Managing **web metrics analytics for global publishers**
 - Designing and optimizing **large-scale data warehouse systems**
 - Building **aggregated reporting tables** for high-performance queries
 - Delivering actionable insights to improve **content usage and engagement**
 - Handling **multi-source data ingestion pipelines (logs, APIs, analytics)**
-

Senior Data Analyst / Data Engineer

Previous Experience

- Built scalable ETL pipelines using **Spark & SQL**
 - Optimized reporting queries reducing execution time by **40–60%**
 - Designed dashboards for business stakeholders
 - Worked on **data migration (Hive → modern lakehouse architecture)**
-

Key Projects

1. AI-Powered LaTeX OCR System

- Built an end-to-end OCR model using **CNN + LSTM + CTC Loss**
- Converts mathematical images into LaTeX expressions
- Implemented:
 - Custom tokenizer
 - Beam search decoding
 - CER (Character Error Rate) evaluation
- Optimized for GPU (Apple Metal backend)

👉 Tech: TensorFlow, Python, Image Processing

◆ 2. Data Warehouse Optimization & Migration

- Migrated legacy Hive/EMR system to **modern lakehouse architecture**
- Designed optimized **fact & dimension tables**
- Reduced storage cost and improved query performance

👉 Tech: Hive, Spark, AWS S3

◆ 3. Advanced Analytics Dashboard (Plotly Dash + Django)

- Built interactive dashboards with:
 - Bar charts, scatter plots, radar charts
 - Real-time data fetching via Django APIs
- Embedded dashboards directly into web apps (no iframe)

👉 Tech: Python, Dash, Django

◆ 4. Web Usage Analytics Platform

- Developed reporting system for publisher analytics
- Created aggregated tables for:
 - Platform usage
 - Monthly reporting metrics
- Improved reporting speed and accuracy

👉 Tech: SQL, Data Modeling

🏆 Certifications

- Google Associate Data Analyst Certification
 - Data Warehouse Management Certification
-

Services I Offer on Upwork

- ✓ Data Analysis & Reporting
 - ✓ Data Warehouse Design & Optimization
 - ✓ AI/ML Model Development
 - ✓ Dashboard Development (Power BI / Dash / Tableau)
 - ✓ ETL Pipeline Development
 - ✓ API Development (Django / Flask)
 - ✓ Business Intelligence Solutions
 - ✓ AI Automation (Chatbots, NLP, OCR)
-

Why Hire Me?

- 9+ years real-world experience
 - Strong combination of **data + AI + engineering**
 - Focus on **business impact, not just code**
 - Fast delivery + clear communication
 - Scalable and production-ready solutions
-

Let's Work Together

If you're looking to:

- Turn your data into insights
- Build AI-powered solutions
- Optimize your data systems

👉 Let's connect and discuss your project!