

# Jon Roosevelt

✉ rooseveltadvisors@gmail.com    ☎ (917) 891-9082    📍 New York, NY    🔗 jonroosevelt.com  
🌐 linkedin.com/in/roosevelt-advisors

## Professional Summary

---

Solution Architect and AI/Data Engineering leader with 18 years of experience designing and deploying enterprise-scale data, ML, and GenAI solutions. Deep expertise in large language models (LLMs), distributed data engineering, and cloud-native architectures for healthcare, media, and AdTech. Proven record of delivering measurable business impact and regulatory-compliant solutions for Fortune 500 clients and startups. Specialized in Python, Apache Spark, Databricks, AWS, Azure, and healthcare data interoperability.

## Professional Experience

---

<b>Cofounder &amp; AI Engineer, Arcs Health</b>	01/2023 – Present New York, NY
<ul style="list-style-type: none"><li>• Led the creation of Scribe, a real-time AI system for transcribing clinical conversations using OpenAI Whisper and Llama 3.2 70B, reducing clinician note-taking time by 60% and ensuring HIPAA compliance.</li><li>• Built scalable, multi-format healthcare data pipelines (CCDA, HL7 v2/v3, EDI) with Azure Data Factory, Databricks, and Fabric, standardizing to FHIR for analytics and interoperability.</li><li>• Designed and implemented a RAG-enabled patient engagement chatbot (Twilio, Llama 3.2 70B, LangChain Agentic), automating appointment scheduling and triage for 5,000+ patients/month.</li><li>• Developed digitized intake and document upload workflows, streamlining patient registration and insurance verification.</li><li>• Architected secure, regulatory-compliant data workflows, integrating Azure AD authentication and API Management for privacy and access control.</li></ul>	
<b>Sr. Data Engineer, NBC News</b>	09/2024 – 01/2025 Englewood Cliffs, NJ
<ul style="list-style-type: none"><li>• Automated CCPA/GDPR Data Subject Request (DSR) processing using Airflow and Databricks Delta Lake, reducing manual workload by 85% and ensuring regulatory compliance.</li><li>• Optimized Spark streaming jobs for MSNBC/CNBC, resolving data skew and DAG lineage issues, improving real-time processing efficiency by 35%.</li><li>• Developed real-time BI dashboards (Looker, AWS Kinesis, Kafka, EMR) for subscription and viewership analytics, reducing latency to 5 minutes.</li><li>• Led Terraform-based CI/CD deployments and orchestrated complex data workflows using Airflow.</li></ul>	
<b>Staff Software Engineer, Preveta</b>	03/2023 – 02/2024 Los Angeles, CA
<ul style="list-style-type: none"><li>• Architected HIPAA-compliant data pipelines for major EHR/EMR integrations using Python, Databricks, and Azure Synapse.</li><li>• Designed secure ingestion platforms for bi-directional data flows, supporting PowerBI analytics and regulatory reporting.</li><li>• Developed robust Python libraries for HL7, C-CDA, and EDI standardization, enabling rapid onboarding of healthcare partners.</li></ul>	

**Senior Solution Architect, *The Trade Desk***

11/2019 – 10/2022

New York, NY

- Designed and implemented custom ML models and CI/CD pipelines (Python, Spark) for ad campaign optimization, delivering up to 35% ROI improvement for clients such as McDonald's and Bayer.
- Led multi-million-dollar data science projects, including custom bidding algorithms and supply path optimization, reducing cost per viewable impression by 12%.
- Built real-time analytics solutions integrating DSP, ad server, and audience data for programmatic marketing automation.

**Founding Director, *Intellinum Analytics Inc***

02/2017 – 01/2023

New York, NY

- Led AI/data engineering projects for AdTech, healthcare, and retail clients, from ideation to delivery.
- Optimized Spark jobs on Kubernetes, reducing operational costs by 70–80% and accelerating development cycles.
- Designed and deployed ensemble ML models for campaign optimization, achieving up to 112% test campaign performance uplift.

**Research Staff, *IBM Research***

2015 – 2018

Yorktown Heights, NY

- Developed city analytics and customer segmentation solutions using Python ETL, Spark, and IBM Big Insights.
- Built predictive models for healthcare and media, delivering actionable insights from large-scale, multi-source data.
- Led real-time analytics platform development for telecommunications, supporting campaigns for 35M+ users.

**Solutions Architect, *EMC***

2014 – 2015

New York, NY

- Provided backend engineering and infrastructure consulting for clients including Monsanto, specializing in VCE converged infrastructure and Python-based automation.

**Education**

---

**M.S. in Computer Science, *New York University***

2012 – 2014

NY

**B.S. in Computer Science, *Donghua University***

2005 – 2009

China

## Technical Skills

---

### Languages

Python  
Scala  
Java  
SQL  
JavaScript  
Go  
Rust  
R  
C#

### Cloud

AWS (EMR, Lambda, S3, SageMaker, EC2)  
Azure (Data Factory, Synapse, OpenAI, Kubernetes, APIM)  
GCP (BigQuery, Dataflow, Vertex AI)

### Healthcare Data

FHIR  
HL7  
C-CDA  
EDI

### Security

Azure AD  
Azure API Management  
OAuth2

### Frameworks

PyTorch  
Spark  
Scikit-learn  
Transformers  
TensorFlow  
Semantic Kernel  
FastAPI

### Data

Databricks  
Delta Lake  
Hive  
Redshift  
Snowflake  
Kafka  
Airflow  
dbt

### DevOps

Docker  
Kubernetes  
Terraform  
Linux  
CI/CD

## Recent Projects

---

### Enterprise Agentic Automations Platform, 2025

- Architected an enterprise-ready AI research agent by porting an open-source solution to Microsoft's Semantic Kernel, enabling modular LLM agent orchestration and plugin-based skill ecosystem.
- Integrated Azure AD authentication and custom API Management (APIM) for secure, role-based access and rate-limiting, supporting 10,000+ research queries per day.
- Implemented the MCP Server/Client using Python MCP SDK to support calling various tools.
- Implemented memory/context management for LLMs, optimizing research workflows and context window usage.
- Containerized the platform for flexible deployment across Azure Kubernetes Service (AKS) and hybrid environments.
- Developed specialized reporting modules for insurance, safety, and financial research, supporting custom compliance workflows.
- Lessons learned: MCP via SSE, LLM context management, authentication flow, and streaming architecture patterns.

**AI Evaluation & Governance Framework, 2025**

- Built a custom evaluation system for LLM agents using Azure AI Evaluation SDK, tailored for enterprise deployments behind Azure API Management (APIM) with custom authentication and header requirements.
- Developed APIM-aware evaluators for metrics including groundedness, factual accuracy, relevance, contextual precision, faithfulness, and fluency, supporting JSON-based scoring and reasoning.
- Implemented a manual evaluation loop for robust aggregation and reporting, overcoming SDK limitations with APIM endpoints.
- Integrated caching and error handling for stability and cost efficiency; enabled detailed pass/fail analytics and compliance reporting.
- Provided actionable insights to improve prompt engineering, retrieval strategies, and system reliability.

**GPU-Accelerated RAG Pipelines, 2025**

- Designed and delivered a GPU-accelerated Retrieval-Augmented Generation (RAG) system for enterprise document intelligence, achieving 5x speedup over CPU-based processing.
- Implemented multi-GPU orchestration for parallel document parsing, embedding generation, and chunking, scaling to 50+ pages/sec.
- Developed memory-aware batching and dynamic chunking strategies to optimize GPU utilization and prevent out-of-memory errors.
- Built robust checkpointing and verification systems to ensure data integrity and enable seamless failover to CPU as needed.
- Containerized the pipeline for reproducible deployment; integrated monitoring for GPU utilization and performance metrics.
- Realized dramatic reductions in processing time and cost per document, enabling real-time analytics on petabyte-scale document collections.

**Publications**

---

**ICCCS 2018**

12/01/2017

Ding X., Yan C., Zhao Y., Yang Z. (2018). Efficient Processing of TopK Dominating Queries on Incomplete Data Using MapReduce. \*ICCCS 2018\*, Cloud Computing and Security, pp. 78–89.

**Certifications**

---

**Databricks Certified Machine Learning Professional**

<https://credentials.databricks.com/aa12012c-d1ae-195-a99c-2b95d99ffa2#acc.apZlwUGe>

**Databricks Certified Data Engineer Professional**

<https://credentials.databricks.com/6abe7e2-163a-3ad-ab2f-bee8999a90f#acc.sgXrZzbq>

**AWS Certified Machine Learning – Specialty**

[https://www.credly.com/badges/33bd7b0-5301-7b-b91d-68b5275e627/public\\_url](https://www.credly.com/badges/33bd7b0-5301-7b-b91d-68b5275e627/public_url)

**AWS Certified AI Practitioner**

[https://www.credly.com/badges/1351a19d-0020-3f3-8fa0-16d8583bceb0/public\\_url](https://www.credly.com/badges/1351a19d-0020-3f3-8fa0-16d8583bceb0/public_url)

**Microsoft Certified: Azure AI Fundamentals**

<https://learn.microsoft.com/en-us/users/jonroosevelt/transcript/dlozriqzx8g9wm>

## Core Competencies

---

### GenAI & LLMs

RAG, Llama Index, LangChain/LangGraph/LangSmith, LLM fine-tuning, evaluation/governance, agentic frameworks, MCP, prompt engineering, vector databases, agentic memory

### Data Engineering

Spark, Databricks, Delta Lake, Airflow, Kafka, Hive, Redshift, Snowflake, AWS EMR, Azure Data Factory, dbt, medallion architecture, structured streaming

### Cloud Platforms

AWS (EMR, Lambda, S3, SageMaker, EC2)

Azure (Data Factory, Synapse, OpenAI, Kubernetes, APIM)

GCP (BigQuery, Dataflow, Vertex AI)

### Healthcare Data

FHIR, HL7 (v2/v3), C-CDA, EDI, HIPAA/GDPR/CCPA compliance, EHR/EMR integration, data normalization, clinical NLP

### Machine Learning

PyTorch, Scikit-learn, Transformers, TensorFlow, ML pipelines, CI/CD, model evaluation, experiment tracking, MLflow

### DevOps & Security

Docker, Kubernetes, Terraform, Linux, CI/CD automation, Azure AD, API Management, secure API design, identity integration