

CONTACT

- ✉
- ➔ [linkedin.com/in/paddyiyer](https://www.linkedin.com/in/paddyiyer)
- 📧 paddyspeaks.com
- ☰ github.com/paddyiyer
- 📍 San Francisco Bay Area

TECHNICAL SKILLS

- Spark
- Hive
- Hadoop
- Databricks
- Delta Lake
- Snowflake
- Redshift
- dbt
- Airflow
- Datascrum
- AWS
- Azure
- GCP
- Terraform
- Kubernetes
- Python
- SQL
- Scala
- LLM/RAG Pipelines
- Claude Code
- AI Agents
- Prompt Eng.
- LangChain
- GitHub Copilot
- MLflow
- Diff. Privacy
- Unity Catalog
- Data Mesh

EDUCATION

Data Engineering Cert.

UC Santa Cruz, 2013–14

B.S. Electrical Engineering

Sardar Patel College of Eng.

Diploma, Electronics & Comms

Technical Board, Tamil Nadu

CERTIFICATIONS

- VMware SaaS Essentials
- Hadoop Fundamentals
- NoSQL Databases
- NoSQL for SQL Professionals

PUBLICATIONS

- Cloud Computing
- All About Big Data
- Future Trends in BI

AI & THOUGHT LEADERSHIP

- 160+ articles on paddyspeaks.com
- Claude Code skill development
- Vibe coding methodology
- AI agent architecture writing

Paddy Iyer

DATA ENGINEERING LEADER · STRATEGIC DATA ARCHITECT · AI & PRIVACY ENGINEERING · AI INNOVATOR

15+ years driving enterprise data transformation for global tech leaders across cloud/SaaS, fintech, gaming, and consumer platforms. Expert in cloud-native architectures, real-time analytics, privacy-first pipelines, and AI-augmented data operations. Proven builder of high-trust, high-performance data teams at petabyte scale. Published writer and thought leader on data architecture, AI agents, and the intersection of ancient wisdom with modern technology.

- Modern Data Architecture
- Privacy-First Design
- Consent Architecture
- AI Agents & LLM Pipelines
- Prompt Engineering
- Data Governance (C1–C5)
- Petabyte-Scale Systems
- Global Team Leadership
- Data Storytelling

FEATURED EXPERIENCE

Senior Consultant — Privacy & Consent Architecture

Apr 2024 – Present

Meta | Safe Ads Program

Spearheading Meta's transformation from broad data access to consent-first advertising

Privacy Remediation & Compliance Engineering

- Operated within a 20-person cross-functional team (8 data engineers, 12 software engineers) facing a 6-month regulatory deadline to remediate 4,000+ tables/assets — initial manual approach surfaced systemic issues: incorrect classifications, pipeline failures, and unscalable ad-hoc workflows
- Executed the end-to-end 8-stage remediation pipeline (classification → lineage → topology sort → pipeline dev → diff → deploy → compliance → downstream migration) across 700+ personally mitigated assets, helping the team convert a fragile manual process into a repeatable, systematic workflow
- Achieved 100% high-risk asset remediation for AAP and Consent Revocation — directly unblocking Meta's Safe Ads launch timeline and enabling downstream teams to resume development on compliant data, eliminating a critical-path blocker for the broader 4,000+ asset program
- Contributed to AI-accelerated surge-readiness that achieved ~98% overall remediation burndown, delivered ~21 engineering-year-equivalent impact, and pulled delivery forward ~6 months — helping the team meet a regulatory deadline that manual methods alone could not have hit
- Landed 300+ code changes, 1,000+ tasks; established a triage framework that reduced per-asset cycle from ~1.86 → ~0.53 weeks (71.5% reduction) — enabling the team to absorb expanding scope without schedule slippage or additional headcount
- Closed privacy gaps across Dataswarm pipelines involving 3+ producer teams — resolving multi-producer DAG dependencies and preventing compliance audit failures that would have delayed the program

AI-Assisted Remediation & Claude Code Skill

- Championed the team's pivot from manual to AI-assisted remediation after early manual sprints revealed classification errors and pipeline rework that threatened the 6-month deadline — AI adoption increased team productivity and eliminated the bottlenecks that had put the timeline at risk
- Built an AI-assisted remediation "Claude Code Skill" (YAML + Markdown) that standardized how the 20-person team creates, debugs, and tests remediation pipelines — contributing to the 71.5% unit-effort reduction validated across 10–12 representative assets
- Codified best practices (consent-filtered datasets, naming conventions, critical rules, workflows, troubleshooting) that reduced classification errors and rework — directly addressing the quality issues that had plagued the manual approach
- Validated and "dogfooded" an AI remediation assistant, strengthening workflow robustness through schema/dataset hygiene, safer configuration standards, and repeatable lint/schedule/test validation

DataSwarm Pipeline Engineering & Operational Quality

- Strengthened DataSwarm pipeline structure — improved schema/dataset definitions, removed hardcoded paths, streamlined lint/schedule/test workflows — establishing patterns the full team adopted to reduce production incidents
- Documented repeatable validation steps that accelerated onboarding for new engineers joining the surge effort, reducing time-to-first-deploy and minimizing production rollbacks
- Applied SQL/Spark-style ETL best practices and automated checks that improved debuggability and consistency, sustaining the team's 300+ code-change velocity across the engagement

EARLIER CAREER

Lead Data Engineer

Meta

Sep 2015 – Sep 2017

80+ Dataswarm pipelines for petabyte-scale ads; Cross Device Insights, Outcomes Datamart + Norms DB

Data Engineer

GREE International

2014 – 2015

PII masking, Vertica → Redshift, 1,000+ table optimization

Data Architect

Chegg Inc.

Oct 2013 – Apr 2014

Technical Director

Model N

2011 – 2013

Life Sciences BI, cloud migration, ETL modularization

Architect

CallidusCloud

2005 – 2011

Incentive comp analytics, BusinessObjects XI

DW Architect

Hewlett Packard

2001 – 2005

Enterprise DW + 8 datamarts; ETL 18 → 3 hrs; 40–50% sales lift via clickstream

Data Architect / Sr. Engineer

Xoriant · Dept of Electronics, India

1990 – 2001

Lead Data Architect — Partner Data Engineering

Jan 2022 – Jan 2024

VMware

Strategic architect for VMware's Partner Data Platform — unifying fragmented partner data into a single governed analytics layer

Data Architecture & Platform Unification

- Designed and built the Partner Data Platform from the ground up, unifying fragmented partner data from 10+ disparate sources into a single governed analytics layer — eliminating data silos that had blocked cross-functional decision-making for years
- Designed an intuitive data catalog and governance model (Python/Confluence) that drove broad self-service adoption, cutting reliance on tribal knowledge by 60% and reducing ad-hoc data request volume to the data team by ~40%
- Implemented VMware's first structured data governance for partner data — classification, lineage tracking, and policy enforcement — achieving audit-ready status and enabling the team to pass compliance reviews without remediation
- Improved data consistency by 60% through standardized naming conventions, automated quality checks, and cross-team data contracts — directly reducing downstream reporting errors that had caused quarterly restatements

AI & Performance Engineering

- Pioneered an AI copilot for retrieval optimization, coding, and governance best practices — one of VMware's earliest LLM-assisted data engineering deployments, reducing routine query development time by ~30%
- Achieved 40%+ Spark processing time improvements through broadcast joins, caching strategies, and skew management — enabling SLA-critical reports to complete within their delivery windows for the first time
- Established automated monitoring and proactive alerting that replaced reactive firefighting, reducing pipeline incident response time from hours to minutes and cutting weekend on-call escalations significantly

Leadership & Cross-Functional Alignment

- Drove cross-team alignment across 5+ data teams via shared roadmaps, reducing duplicate pipeline development efforts by ~25% and accelerating delivery cycles
- Led workshops and knowledge-sharing sessions that upskilled 30+ partner data consumers, building organizational data literacy and reducing the "data team as bottleneck" pattern
- Engaged director/VP-level stakeholders to define a unified data vision, securing budget and headcount for the platform roadmap — resulting in the team growing from 3 to 8 engineers during the engagement

Senior Consultant — Ads, Commerce & Privacy

Jul 2019 – Dec 2021

Meta

Led high-impact projects across Meta's Ads, Commerce, and Privacy teams, architecting secure, scalable, and compliant data platforms

Commerce Data Architecture

- Architected a centralized data warehouse that unified product and seller data into a single analytics layer — enabling commerce teams to reduce decision-making cycle time from weeks to days and directly informing category investment strategies
- Designed and launched the Category Management Data Warehouse, providing deep cross-vertical analysis that empowered commerce teams to identify underperforming seller segments and reallocate resources, contributing to measurable GMV growth
- Owned the SMB funnel redesign, improving query performance by 3x while maintaining key functionality — enabling the pipeline to scale from thousands to tens of thousands of small business advertisers without infrastructure changes

Privacy Engineering & Compliance

- Led privacy remediation across 1,200+ SMB 2.0, Customer Journey, and BPO tables — resolving critical design flaws that had left the team exposed to compliance risk, and establishing remediation patterns later adopted at scale in the Safe Ads program
- Collaborated with Hive Anonymization engineers to enhance anonymization processes, enabling the ads and commerce data products to pass privacy audits without manual intervention — reducing audit preparation effort by ~60%
- Drove Salesforce ID deprecation and Facebook ID integration across 1,000+ pipelines, eliminating an external vendor dependency that had caused data consistency issues affecting ~15% of cross-system joins

Data Platform Migration

- Spearheaded migration of 1,000+ Hive pipelines to Spark, automating ~70% of conversion tasks — reducing what was projected as a 12-month migration to under 6 months and enabling commerce/payments teams to benefit from Spark's performance gains ahead of schedule
- Built chargeback representation and leakage dashboards (FGF, Dataswarm, Unidash) that identified ~\$XM+ in previously undetected revenue leakage, enabling the finance team to recover funds and close monitoring gaps
- Designed Marketplace App reliability dashboard covering payin/payouts and user stability — reducing MTTR for payment-related incidents by providing real-time visibility that replaced manual log analysis

ADDITIONAL RECENT EXPERIENCE

Data Engineer Consultant

Dec 2017 – Apr 2018

[LinkedIn](#)

- Led GDPR compliance efforts: encrypted sensitive customer data into LinkedIn Dali storage using Hive, Python, and Pig; retired legacy data sources and feeds; migrated to a stable, compliant framework

Lead Data Engineer Consultant

Sep 2015 – Sep 2017

[Meta](#)

- Developed 80+ complex Dataswarm pipelines for petabyte-scale ads data across Sales Operations and Marketing Intelligence: Cross Device Insights, Global Account Pipeline, Gaming Apps Reporting, Outcomes Datamart + Norms DB, Facebook Media & Live Monetization
- Deep expertise in petabyte-scale ads data, partnering with cross-functional teams to optimize revenue generation strategies and deliver scalable analytics