KUBECON + CLOUDNATIVECON INDIA 2026

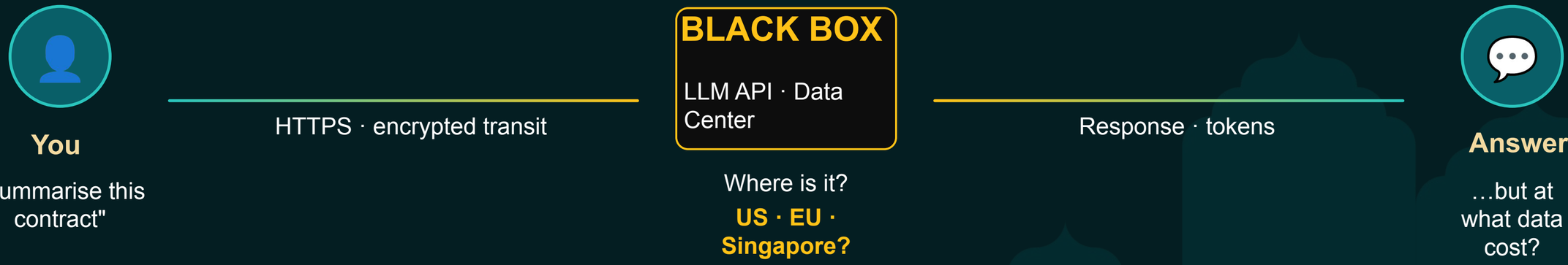
YOUR PROMPT IS A CROSS-BORDER DATA TRANSFER

What really happens after you press Enter?

Sudhanshu Prajapati · Sr. Developer Advocate, Improving

THE SETUP

WHAT HAPPENS WHEN YOU PRESS ENTER?



📄 Contract text & PII

🏢 Company data

🌍 Cross-border transfer

🔍 Zero observability

ABOUT THE SPEAKER

Sudhanshu Prajapati

Sr. Developer Advocate · Improving (formerly InfraCloud)

- Data & backend engineer turned developer advocate - 5+ years experience
- Milvus Ambassador · Co-organizer CNCG Lucknow · Active in CNCF Initiatives
- GenAI · AI Agents · DevOps · Open Source · Developer Relations
- Contributor to CNAI, open source projects & CNCF initiatives

Scan to connect on LinkedIn

THE CRITICAL QUESTION

DO YOU KNOW WHERE THE DATA IS GOING?

…and does it even matter?

02

WHY SOVEREIGNTY? WHY NOW?

Why is everyone suddenly talking about AI sovereignty?

FROM INDIA AI SUMMIT

THE CASE FOR SOVEREIGN AI

HPCL

Oil & Gas · National Critical Infrastructure

Moved to **on-premise AI** not just for security, but for **cost predictability** (avoiding per-token cloud costs) and to handle sensitive National Critical Information Infrastructure.

Speaker: Ritwik Rath, Executive Director - HPCL

NABARD

Banking / Agriculture Finance

Chose on-premise to maintain **Strategic Independence** (not relying on one hyperscaler) and comply with strict **financial data localization laws** - DPDP Act.

Speaker: Balasubramanian V, Chief General Manager - NABARD

Source: India AI Summit — Sovereign and Trusted AI: Building India's Intelligence Infrastructure for the Next Decade

THE DILEMMA

VISIBILITY VS. VELOCITY

Build Everything from Scratch

Own models, own infra, full control. Maximum security - **maximum time.**

- Full IP Ownership
- No External Dependency

OR

Use External Providers

Fast to market. Easy integration. Zero visibility - **zero trace.**

- Speed to Market
- Low Ops Overhead

Should we build everything from scratch to stay secure?

THE RISK

WHY WE CAN'T RELY ON EXTERNAL PROVIDERS

1 **Model Deprecation**
OpenAI historically phases out models, forcing migrations or unexpected price hikes.

2 **Unpredictable SLAs**
Shared environments mean you share the downtime. No guaranteed Quality of Service.

3 **Explainability & QoS**
In-house models let you track time-to-first-token, guarantee QoS, and debug model weights directly.

THE PROBLEM REMAINS

WE HAVE THE INFRASTRUCTURE. WE LACK THE TRACE.

- ⚡ Even with local LLMs, agents use MCP to call external tools - APIs, web searches.
- ⚡ Data can **still cross borders** through those tool calls.
- ⚡ We need **visibility at the request level.**

THE SOLUTION

USING OPENTELEMETRY FOR GENAI OBSERVABILITY

What It Is

The **open-source standard** for observability - traces, metrics, and logs across distributed systems.

GenAI Semantics

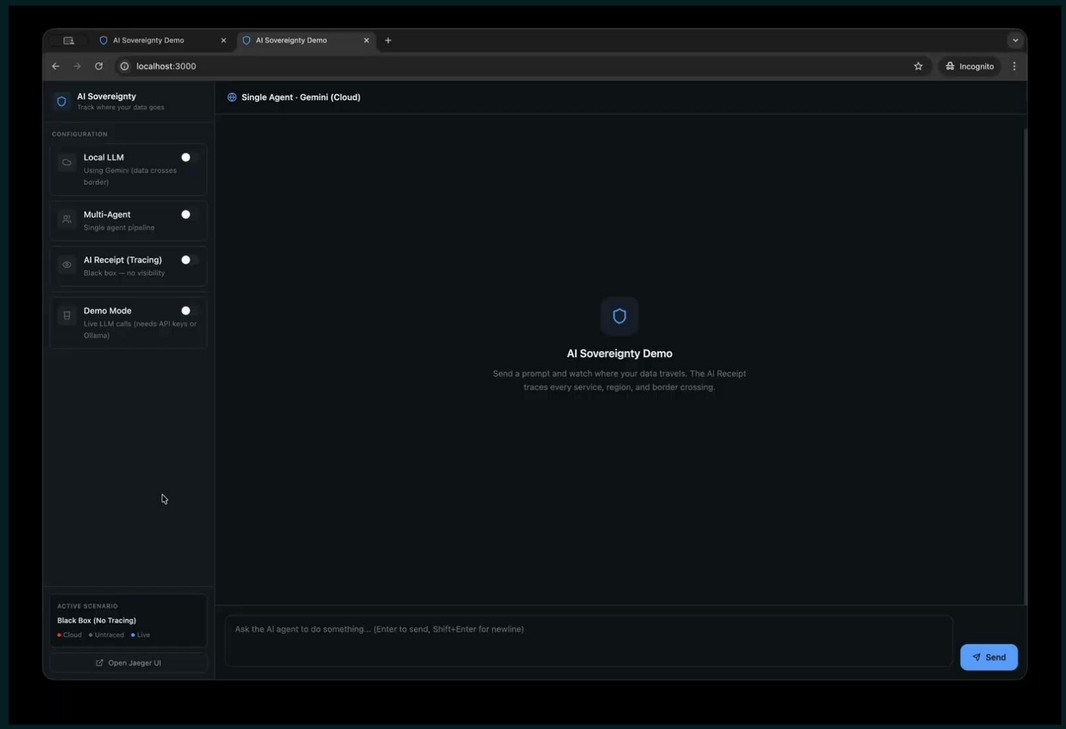
Ongoing OTel developments to **standardly track** token usage, model names, and MCP tool calls.

Why It Matters

It turns **black-box agent workflows** into verifiable, debuggable waterfalls you can audit.

THE "AI RECEIPT"

- A **cryptographically sound, span-by-span audit** of the data's journey.
- **Flags cross-border transfers instantly** - before it's too late.
- Built on OTel - open, vendor-neutral, **community-driven.**

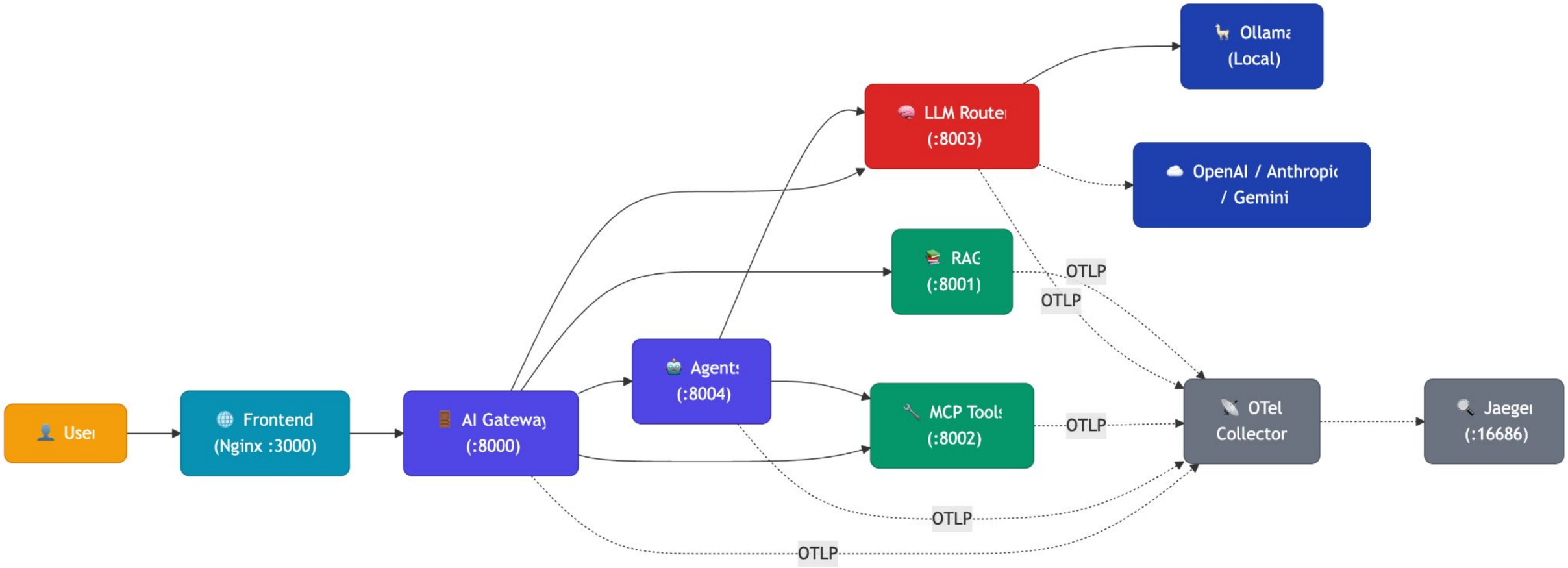


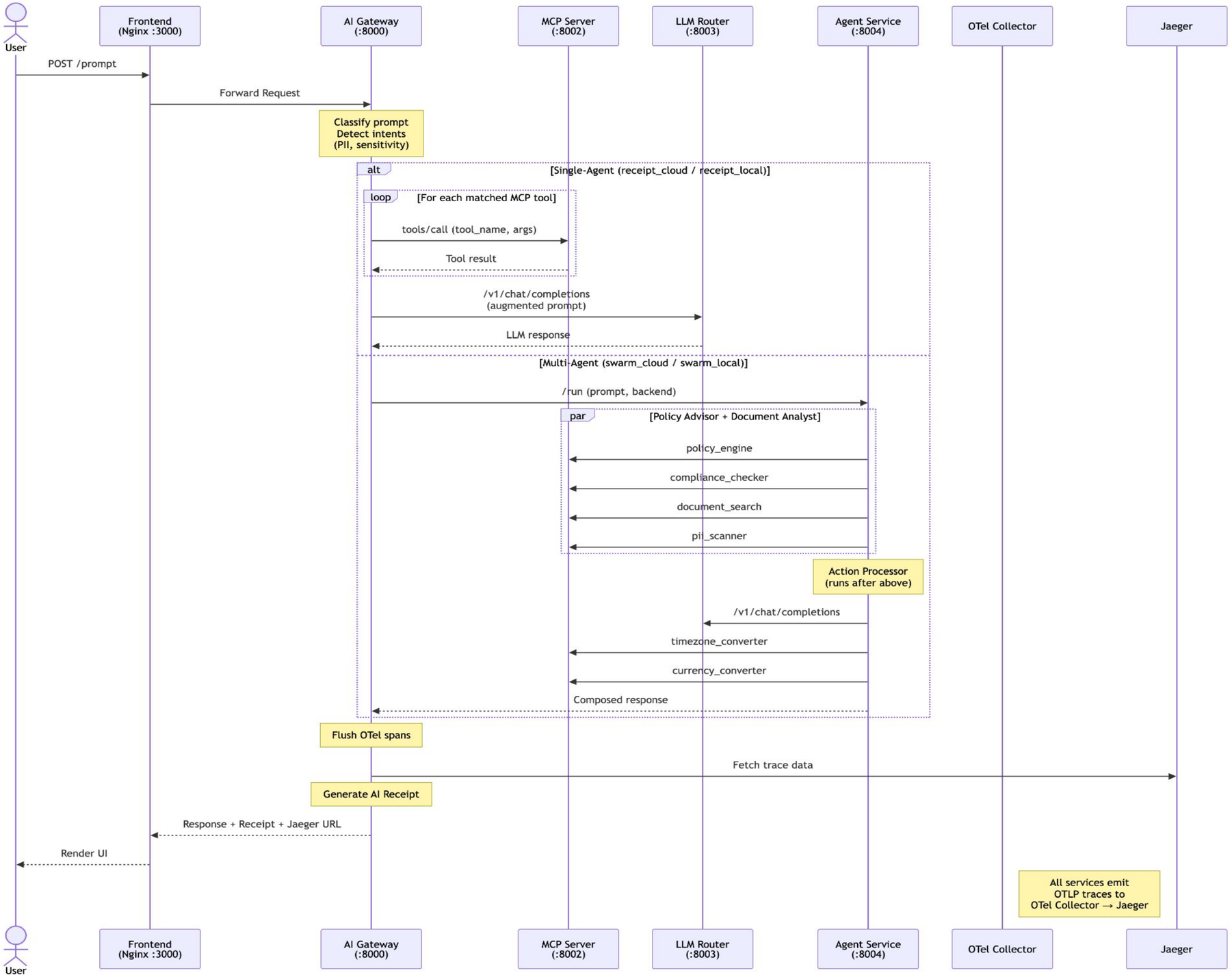
Cautious Adventurism

Adopt AI fast - but trace every step. Don't just trust your AI.

https://github.com/sudhanshu456/ai-sovereignty-demo

Demo Architecture




User
Frontend (Nginx :3000)
AI Gateway (:8000)
MCP Server (:8002)
LLM Router (:8003)
Agent Service (:8004)
OTel Collector
Jaeger
POST /prompt
Forward Request
Classify prompt
Detect intents
(PII, sensitivity)
alt
[Single-Agent (receipt_cloud / receipt_local)]
loop
[For each matched MCP tool]
tools/call (tool_name, args)
Tool result
/v1/chat/completions
(augmented prompt)
LLM response
[Multi-Agent (swarm_cloud / swarm_local)]
/run (prompt, backend)
par
[Policy Advisor + Document Analyst]
policy_engine
compliance_checker
document_search
pii_scanner
Action Processor
(runs after above)
/v1/chat/completions
timezone_converter
currency_converter
Composed response
Flush OTel spans
Fetch trace data
Generate AI Receipt
Response + Receipt + Jaeger URL
Render UI
All services emit
OTLP traces to
OTel Collector → Jaeger

THE TAKEAWAY

WHERE DOES AI RECEIPT FIT IN?

🔐

Span-by-Span Audit

A cryptographically sound record of every step of the data's journey - from prompt to response.

🚩

Instant Flagging

Flags cross-border transfers the moment they happen - real-time compliance signal.

⚖️

Cautious Adventurism

Don't slow down AI adoption - just trace it. Fast *and* accountable.

THANK YOU!

"Don't just trust your AI. Trace it."

Sudhanshu Prajapati
Sr. Developer Advocate · Improving
tinkeringbits.com

SLIDES & RESOURCES
kubecon-india-2026.talk

COMMUNITY
CNCG Lucknow · Milvus Community

CHECKOUT THE PUBLISHED BLOG ON IMPROVING.COM FOR MORE DETAILS

Your Prompt Is a Cross-Border Data Transfer