

CJ Fravel

Senior Software Engineer · Data Platforms, Pipelines & Analytics Engineering

Greater Seattle Area · cj@cjfravel.dev · cjfravel.dev · github.com/cjfravel-dev · [LinkedIn](#)

Summary

Senior software engineer who designs the data platforms and processing other engineers build on. At Microsoft, proposed and own a declarative data-contract system that is the platform's core data-processing stage — it replaced the entire legacy U-SQL processing layer in the migration to Spark on Synapse — turning raw data into canonical, governed datasets with schemas, validation, governance, and delivery SLAs, self-served across Partner, Seller, and Royalties orgs without touching platform code. I also build the ADX/Kusto dashboards that report SLA, performance, and cost.

AI-native practitioner with a verification-first workflow: agents do real work on my projects every day, and I've built rigorous testing around them — clear scoping, diff-by-diff review, heavy automated testing, and validating assumptions against real production numbers to catch when an agent is confidently wrong. Strong product mindset; comfortable owning ambiguous problem spaces and pushing them to a shippable bar.

Relevant Skills

Languages	SQL & KQL, Scala, C# / .NET, JavaScript, Ruby, Bash
Data engineering	Contract-driven data processing on Spark, canonical data modeling, ELT, data quality & SLAs, self-serve analytics
Reporting	ADX (Kusto) & Power BI dashboards for SLA, performance, and COGS
Orchestration	Azure Data Factory, Synapse pipelines, Azure Functions cross-region failover
Developer experience	Self-service onboarding, local-dev workflows, CI/CD, observability, SLA / quarantine / alerting
Engineering	Distributed systems, backend services, cross-team API/contract design, performance & cost optimization, debugging production at scale
Cloud	Azure: Synapse, Databricks, Data Factory, ADLS, ADX/Kusto, Functions
Agentic systems	Coding agents (Copilot CLI & others); agent loops, tool calling, prompt iteration, evaluation frameworks, human-in-the-loop review, guardrails
Open source	ariadne (Spark indexing) & nomos (JSON templating) in Scala, ChromaMark (agent-to-human markup) in JS/Python · cjfravel-dev

Experience

Senior Software Engineer · Microsoft	Aug 2021 - Present
• Platform ownership: proposed and own a declarative data-contract system that is the platform's core data-processing stage — it replaced the entire legacy U-SQL processing layer in the migration to Spark on Synapse — and teams use it to produce canonical, governed datasets (schemas, validation, governance, delivery SLAs) self-served across Partner, Seller, and Royalties orgs without modifying platform code. • Data-integrity quality gates: replaced manual contract-review checks with automated validation built into the pipeline, removing humans from the critical path while raising the quality bar across Partner, Seller, and Royalties orgs. • Reporting & metrics: build the ADX (Kusto) dashboards that track SLA, performance, and COGS — the primary way platform delivery health and cost are reported. • Unified cross-team reporting: built Spark observers that aggregate cross-team metrics into a unified reporting layer, and lead cross-team coordination on the next-generation platform for the incentives space.	

- **Resilient orchestration:** Azure Data Factory and Synapse pipelines, plus custom Azure Functions orchestration for cross-region redundancy — automatic fail-over to a backup region when the primary is down, resuming on restore.
- **Developer experience:** drive faster local Spark workflows, self-service onboarding for new entities and teams, and reliability tooling (SLA / quarantine / alerting) so daily work is predictable.
- **Production at scale:** the platform processes 2+ TB/month representing \$110B in revenue; the Spark/Synapse rebuild delivered ~8× end-to-end speedup at 1/5 the cost; also helped move a big-data platform from commercial cloud to a secure federal environment.

Software Engineer · Applied Information Sciences

May 2019 - Aug 2021

- Lead developer on a Microsoft-funded POC for the Air Force Test Center, demonstrating how Azure Storage, Data Lake, Data Factory, and Jupyter notebooks streamline analytics across traditionally siloed test-wing data.
- Technical team lead for a Data Integration Platform: secure cloud ingest to automated pipelines to Power BI dashboards delivering self-serve metrics to stakeholders across previously siloed data.

Developer · Pixel LLC

May 2017 - May 2019

- **Employee #1** — founding-team experience at a small startup: helped build the team and culture and did the late-night 0→1 work, alongside customer-facing web and mobile consulting shipped directly to end users.

Open Source

[cjfravel-dev / ariadne](#) · Scala

Spark utility for generating indexes in a data lake to minimize data read when joining across massive datasets. Designed for situations where data is poorly partitioned (or not at all) and cannot be moved, or where duplication is too costly.

[cjfravel-dev / ChromaMark](#) · JavaScript · Python

Lean rich-markup format for agent-to-human communication: a Markdown superset that adds colored blocks, status pills, and collapsible sections at a fraction of HTML's token cost. Ships a JavaScript renderer (npm/CDN), a byte-identical Python port with Jupyter support, a CLI, and a VS Code extension.

[cjfravel-dev / nomos](#) · Scala

JSON templating engine for Scala: define a template once and generate matching case classes that validate, serialize, and deserialize JSON — one source of truth for the model and its validation, with a small, dependency-free runtime.

Working with AI

AI-native engineer with a verification-first workflow around coding agents. Direction, architecture, and judgment calls are mine; agents execute the well-scoped pieces.

- **Verification-first & safety-minded:** heavy automated testing and validation against real production numbers, so what ships behaves as claimed — catching confidently-wrong output before it reaches production.
- **Evaluation-driven workflow:** scoped task decomposition, prompt iteration, sandbox validation, diff-based human-in-the-loop review, and production guardrails on every change agents land.
- **Autonomous workflows:** design automation and scaffolding that let agents run hands-off on appropriate work, lowering human touchpoints while protecting reliability and developer productivity.
- **Guardrails & reliability:** sandbox before prod, keep secrets out of prompts, stay fluent enough to catch when the agent is confidently wrong.

Education & Certifications

Bachelor's, Computer Science · Wright State University

Dayton, OH

Certifications

- CompTIA Security+ · Jun 2019 - Jun 2022

- Microsoft Certified: Azure Data Engineer Associate · Mar 2021 - Mar 2023