The Data Canary

How Netflix Validates Catalog Metadata

Celina Amados

Site Reliability Engineer, **Netflix**

What We'll Cover

01 **Context**
The incident, and why validating data is harder than it looks

02 **Architecture**
How the Data Canary actually works, under the hood

03 **Experimentation**
How we proved it, by trying to break it

04 **Takeaways**
What your team can borrow from this

Why Catalog Metadata Matters

Playability

Defines what titles exist, whether they play, and where they're available.

Continuously Transformed

Transformed and redistributed across Netflix's infrastructure on a continuous, tight cycle.

High Stakes

Accurate data creates a moment of joy. Corrupted data may break streaming instantly, everywhere.

What Went Wrong

The Trigger
A production incident exposed a critical gap in our resilience strategy.

The Mystery
No code was deployed. No configuration changed.

The Cause
A mitigation step from an earlier incident quietly corrupted a data feed.

The Impact
Calls failed, blocking manifest generation and breaking playback for members.

Every deployment signal said healthy.

The only thing that moved was **Starts Per Second**: real people failing to hit play.

What if we canaried the data, too?

Same discipline we already apply to code, just **harder**, **faster**, and on **live customers**.

Why This Is Hard

Time Constraints

Traditional canary analysis needs 30–60 minutes of statistical confidence. Our data cycle gives us a fraction of that to detect, decide, and block.

Production Traffic Is Essential

Shadow traffic can't replicate the full playback lifecycle across services. Catching real impact requires real traffic.

Emergent Issues

Every upstream source is validated independently, but some problems only appear once everything is transformed together.

Limited Blast Radius

Testing with production traffic means any regression has to be caught and contained immediately, not reviewed after the fact.

Why This Is Hard

Time Constraints

Traditional canary analysis needs 30–60 minutes of statistical confidence. Our data cycle gives us a fraction of that to detect, decide, and block.

Production Traffic Is Essential

Shadow traffic can't replicate the full playback lifecycle across services. Catching real impact requires real traffic.

Emergent Issues

Every upstream source is validated independently, but some problems only appear once everything is transformed together.

Limited Blast Radius

Testing with production traffic means any regression has to be caught and contained immediately, not reviewed after the fact.

Why This Is Hard

Time Constraints

Traditional canary analysis needs 30–60 minutes of statistical confidence. Our data cycle gives us a fraction of that to detect, decide, and block.

Production Traffic Is Essential

Shadow traffic can't replicate the full playback lifecycle across services. Catching real impact requires real traffic.

Emergent Issues

Every upstream source is validated independently, but some problems only appear once everything is transformed together.

Limited Blast Radius

Testing with production traffic means any regression has to be caught and contained immediately, not reviewed after the fact.

Enter: The Data Canary

Bringing the same discipline we apply to **code** over to **data**.

How Catalog Data Moves

How Catalog Data Moves

Innovation 1: The Orchestrator Pattern

Orchestration

A dedicated orchestrator coordinates the entire validation flow, including verifying cluster versions, and triggering the canary test.

Dual Clusters

Baseline on known-good catalog and canary running continuously.

Generic Interface

Results report back via generic interface for other pipelines to reuse.

Innovation 1: The Orchestrator Pattern

Orchestration

A dedicated orchestrator coordinates the entire validation flow, including verifying cluster versions, and triggering the canary test.

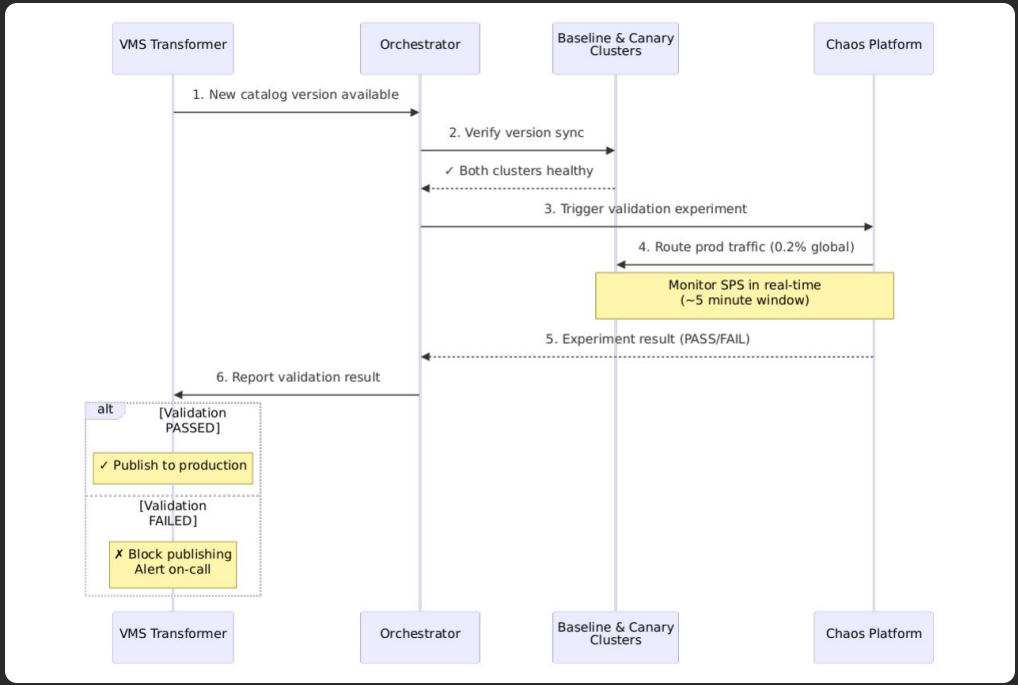
Dual Clusters

Baseline on known-good catalog and canary running continuously.

Generic Interface

Results report back via generic interface for other pipelines to reuse.

Data Canary Workflow



Trigger & Sync

New catalog version triggers version sync check across baseline & canary clusters.

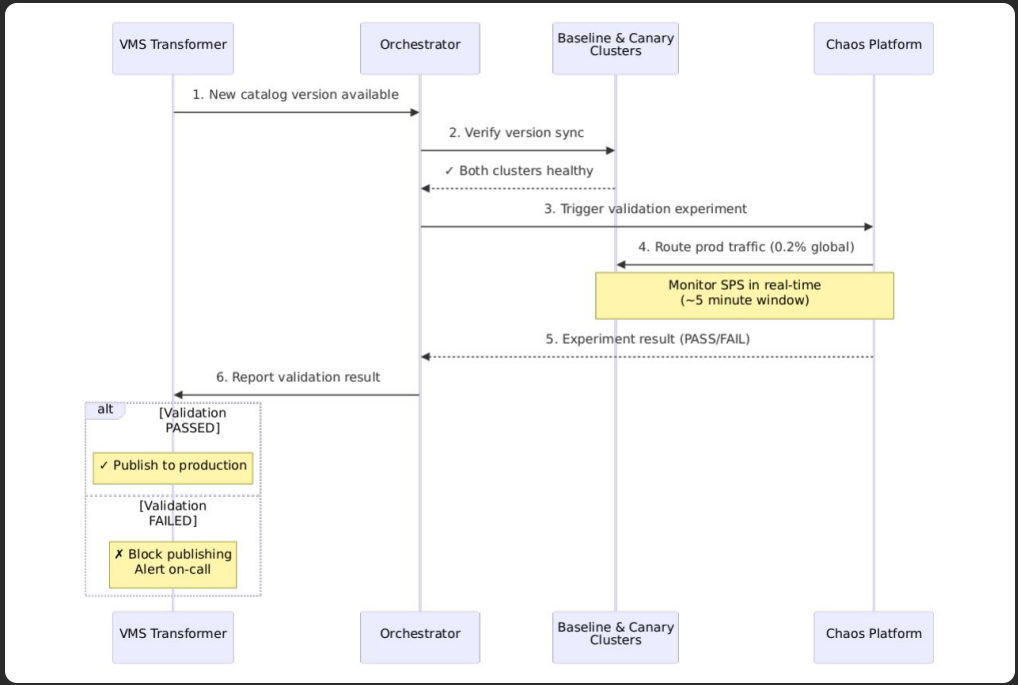
Real-Time Validation

ChAP routes 0.2% prod traffic to clusters; monitors SPS during ~5 min window.

Gated Execution

Automated gate publishes to production on success, or blocks release & alerts on-call.

Data Canary Workflow



Trigger & Sync

New catalog version triggers version sync check across baseline & canary clusters.

Real-Time Validation

ChAP routes 0.2% prod traffic to clusters; monitors SPS during ~5 min window.

Gated Execution

Automated gate publishes to production on success, or blocks release & alerts on-call.

Innovation 2: Extending Our Chaos Platform

Custom Tuning

Standard chaos experiment thresholds were too conservative for a sub-10-minute window.

Sticky Canaries

Session affinity keeps a user's traffic on one cluster for the whole experiment, preventing cross-contamination.

Behavioral Metrics

Starts Per Second (real playback attempts) proved far more reliable than latency or error rates for catching catalog corruption.

Immediate Abort

We stream metrics live and stop the moment we see a problem, trading some statistical confidence for speed.

Innovation 2: Extending Our Chaos Platform

Custom Tuning

Standard chaos experiment thresholds were too conservative for a sub-10-minute window.

Sticky Canaries

Session affinity keeps a user's traffic on one cluster for the whole experiment, preventing cross-contamination.

Behavioral Metrics

Starts Per Second (real playback attempts) proved far more reliable than latency or error rates for catching catalog corruption.

Immediate Abort

We stream metrics live and stop the moment we see a problem, trading some statistical confidence for speed.

Innovation 3: Production-Hardened Engineering

In-Flight Experiments

The orchestrator must resume validation mid-cycle during redeploys, ensuring active experiments are never abandoned.

Leader Election

While multiple orchestrator instances can run during a deploy, leader election guarantees only one experiment fires per version.

Fail-Safe by Default

If the validation system itself encounters a failure, the engine defaults to allowing publication rather than blocking good data.

Validating the Validator

Because you don't trust a safety system until you've tried to defeat it!

We Broke Things on Purpose

Controlled Corruption

Ran controlled experiments deliberately corrupting catalog data to simulate realistic scenarios: denylisting high-profile titles and setting `goLive` flags to false.

Proactive Incidents

Coordinated as proactive incidents, during business hours, with operations teams on standby.

0.2% Traffic Routing

Routed roughly 0.2% of global traffic through the validation flow: small enough to limit impact, large enough for a clean signal.

We Broke Things on Purpose

Controlled Corruption

Ran controlled experiments deliberately corrupting catalog data to simulate realistic scenarios: denylisting high-profile titles and setting `goLive` flags to false.

Proactive Incidents

Coordinated as proactive incidents, during business hours, with operations teams on standby.

0.2% Traffic Routing

Routed roughly 0.2% of global traffic through the validation flow: small enough to limit impact, large enough for a clean signal.

We Broke Things on Purpose

Controlled Corruption

Ran controlled experiments deliberately corrupting catalog data to simulate realistic scenarios: denylisting high-profile titles and setting `goLive` flags to false.

Proactive Incidents

Coordinated as proactive incidents, during business hours, with operations teams on standby.

0.2% Traffic Routing

Routed roughly 0.2% of global traffic through the validation flow: small enough to limit impact, large enough for a clean signal.

Key Results

2.5–4

MINUTES DETECTION

Fast detection speed, varying slightly depending on the specific client type.

10x

ERROR DIFFERENTIAL

A clear, unmistakable signal between the canary and the baseline traffic.

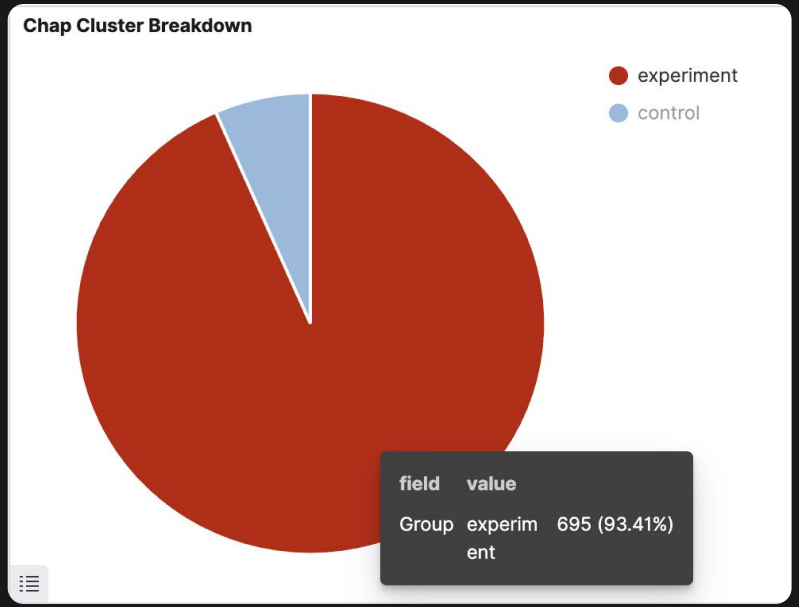
Automatic

PUBLISH BLOCKING

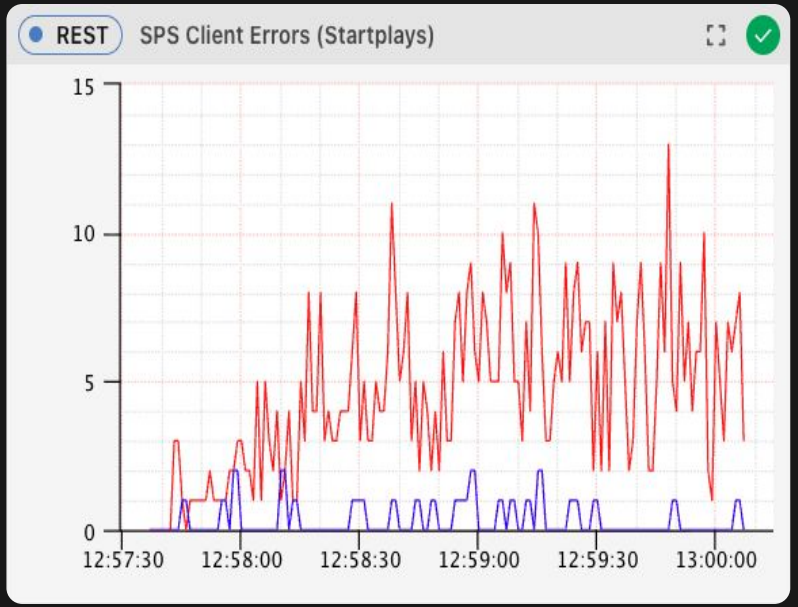
The publishing workflow blocked exactly as designed the moment a regression was detected.

Key Results

Cumulative SPS Error Breakdown



Real-time SPS Error Breakdown



Lessons for Your Team

01 **Data deployments** carry the same production risk as code deployments, and deserve equivalent validation rigor.

02 **Behavioral metrics** can be more reliable than latency or error rates for catching data corruption: find yours.

03 **An orchestrator pattern** with permanent baseline/canary clusters is reusable well beyond catalog metadata.

04 **Production traffic**, with a tightly controlled blast radius, catches what shadow traffic simply can't.

05 **Chaos experiments** can be tuned for tight time windows by prioritizing speed of detection over statistical purity.

Lessons for Your Team

01 **Data deployments** carry the same production risk as code deployments, and deserve equivalent validation rigor.

02 **Behavioral metrics** can be more reliable than latency or error rates for catching data corruption: find yours.

03 **An orchestrator pattern** with permanent baseline/canary clusters is reusable well beyond catalog metadata.

04 **Production traffic**, with a tightly controlled blast radius, catches what shadow traffic simply can't.

05 **Chaos experiments** can be tuned for tight time windows by prioritizing speed of detection over statistical purity.

Data deployments deserve the same rigor as code deployments.

Just because it isn't a binary doesn't mean it can't break production.

Thank You.

Celina Amados
celinaamados@gmail.com

Further Reading

netflixtechblog.com

Special Thanks

Jongyoon Lee, David Su, Zubeen Lalani
(Catalog Foundations & Distribution)

Ales Plsek
(Resilience)