Firmware Update Service – High Level Architecture

Objectives

- Deliver BIOS, BMC, SSD, NIC, GPU, CPLD updates safely at scale.
- Ensure low blast radius with staged rollouts.
- Support rollback and compliance visibility.
- Integrate with existing CI/CD and telemetry pipelines.

Architecture Layers

- Control Plane – orchestration and policy logic.
- Distribution Layer – secure artifact delivery.
- Data Plane – execution on hosts and devices.

Control Plane Components

- Firmware Update API – create campaigns, define targets, trigger waves.
- Metadata & State Store – inventory, version tracking, audit logs.
- Scheduler / Wave Orchestrator – controls concurrency, canary promotion.
- Policy & Compliance Engine – health gates, maintenance windows.
- Artifact Repository – signed, immutable firmware images.
- Telemetry & Audit Service – metrics and dashboards.

Data Plane Components

- Device Agent – polls desired state, validates signatures, applies updates.
- Local Cache/Proxy – reduces WAN load by caching per site.
- Firmware Executors – vendor tools for BIOS/NIC/SSD/etc.
- Health & Recovery Manager – post-update validation and rollback.

Distribution Layer

- Global Object Store (Origin) → Central source for artifacts.
- Regional Caches → Reduce latency and bandwidth consumption.
- Site Proxies / CDN → Local artifact delivery.
- P2P Distribution → Peer-to-peer sharing for large fleets.

Security Model

- Firmware images cryptographically signed (X.509/PGP).
- SBOM and supply-chain scanning on upload.
- mTLS between control and data plane.
- Role-based access (RBAC + ABAC).
- Immutable audit logs for compliance.

Observability and Metrics

- Success rate, rollback count, time-to-update.
- Device reachability and cache hit rate.
- Structured logs with correlation IDs.
- Alerting thresholds for failure spikes.
- Regional compliance dashboards.

Text Architecture Diagram

- Operator UI → Control Plane → Distribution → Data Plane (Host/BMC) → Firmware Executor

Deployment Scale Targets

- Regions: 10+
- Sites per region: 100+
- Hosts per site: up to 10K
- Concurrency: <1% per fault domain
- 99.5% success within maintenance window

Summary

- Central control plane with strong policy and telemetry.
- Secure artifact pipeline with signed binaries and SBOMs.
- Distributed caches and intelligent agents for efficiency.
- Fault-domain aware orchestration with health-based gating.
- Unified visibility, audit, and rollback support.

Architecture Diagram

```
■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■ ■ Control Plane ■ ■ ■■■■■■■■■■■■■■■■
■■■■■■■■■■■■■■■■■■■■■■■■ ■ ■ ■ Update API ■→→ ■ Scheduler / Waves ■ ■ ■ ■■■■■■■■■■■■■■■■
■■■■■■■■■■■■■■■■■■■■■■■■ ■ ■ ■ ■ ■ ■ Metadata DB Artifact Repository ■
■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■ ■ ■ ▼ ▼ ■■■■■■■■■■■■■■■■■■■■■■
■■■■■■■■■■■■■■■■■■■■■■■ ■ Regional Cache ■→→→→→→→→→■ Device Agent (Host) ■ ■■■■■■■■■■■■■■■■■■■■■■
■■■■■■■■■■■■■■■■■■■■■■■ ■ ▼ Firmware Executors / BMC
```