THREE HORIZONS, EL CAMINO RED HAT · PLAYBOOK

# De Azure crudo a una Agentic DevOps Platform en 90 a 180 días.

Three Horizons es el acelerador para empresas que quieren la plataforma con soporte del vendor, certificada y con propiedad conjunta de Red Hat y Microsoft: Red Hat Developer Hub sobre Azure Red Hat OpenShift, OpenShift Pipelines y GitOps, Trusted Application Pipeline, y la fundación de IA de Microsoft por debajo. Este playbook recorre el stack opinado, la secuencia de tres horizontes de la fundación segura a los agentes gobernados, y el framework de decisión que dice cuándo este camino es el correcto.


AUTORA

Paula Silva

CARGO

AI-Native Software Engineer

CONTACTO

in/paulanunes


TIEMPO DE LECTURA

~ 40 minutos, de punta a punta

| 3 | 6 | 22 |
|---|---|---|
| HORIZONTES | CAPÍTULOS | PLANTILLAS DE GOLDEN PATH |

90-180

DÍAS HASTA EL PRIMER VALOR EN PRODUCCIÓN

CAPÍTULOS

Chapter 00

## El argumento

Qué es el acelerador, la propuesta de valor en tres partes (velocidad, coherencia comercial, postura regulatoria), los números de referencia y la escalera de madurez que lo nombra.

El Argumento →

Chapter 01

## El stack opinado

RHDH como IDP, ARO como clúster operado conjuntamente, Pipelines más GitOps como entrega, Trusted Application Pipeline como cadena de suministro, y la capa de asistencia de IA en tres superficies.

Arquitectura →

Chapter 02

## Horizonte 1, la fundación

El runtime Azure seguro en 90 días: red zero-trust, ninguna credencial estática en ningún lugar, registry y datos en private endpoints, recuperación probada y gobernanza de costos desde el primer día.

Horizonte 1 →

Chapter 03

## Horizonte 2, el producto interno

GitOps app-of-apps, Developer Hub como puerta de entrada, el catálogo de 22 plantillas de Golden Path, y el stack de observabilidad con más de 30 alertas en cinco capas.

Horizonte 2 →

Chapter 04

## Horizonte 3, la capa agéntica

Azure AI Foundry como Terraform, siete Golden Paths de IA, la flota de 17 agentes con integraciones MCP, y los gates de evaluación que hacen gobernable la autonomía.

Horizonte 3 →

Chapter 05

## La decisión y el despliegue

El crosswalk de los Cuatro Pilares, la estructura comercial conjunta, los perfiles de cliente que inclinan hacia cada lado, y la secuencia de tres fases anclada en el hito de primer valor.

La Decisión →

CONSEJOS DE LECTURA

> ATAJOS DE TECLADO
>
> Presiona `/` para buscar. Usa `j` y `k` para moverte entre capítulos. Presiona `t` para cambiar tema. Presiona `g` para ir al inicio.

---

**Paula Silva**

AI-NATIVE SOFTWARE ENGINEER

in/paulanunes

v1.0.0 · 2026-07-03

Building the future of software development with AI and Agentic DevOps

00 EL ARGUMENTO

# Un acelerador, un stack opinado: de Azure crudo a una Agentic DevOps Platform.

Three Horizons es un acelerador desplegado por el cliente que materializa una Agentic DevOps Platform sobre el stack Red Hat corriendo en Azure. Existe porque ensamblar esa plataforma desde componentes crudos toma de 9 a 18 meses, y porque una clase de empresas necesita que el resultado tenga soporte del vendor, certificaciones y auditabilidad desde el primer día. Este capítulo define qué es el acelerador, enuncia su propuesta de valor y presenta la escalera de madurez que le da nombre.

## Qué es Three Horizons

Three Horizons es una arquitectura de referencia y un patrón de despliegue por el cliente, no un producto con SKU. Combina productos existentes de Red Hat y Microsoft en una fundación integrada y opinada: Red Hat Developer Hub como Internal Developer Platform, Azure Red Hat OpenShift como clúster, OpenShift Pipelines y OpenShift GitOps como sustrato de entrega, Red Hat Trusted Application Pipeline para seguridad de la cadena de suministro, y la fundación Microsoft (Entra ID, Azure AI Foundry, GitHub Enterprise) por debajo. Los componentes se licencian y soportan de forma independiente; el acelerador es la composición.

Físicamente, el acelerador se entrega como un único repositorio template en GitHub: más de 120 archivos y unas 20.000 líneas de código curado y documentado, incluyendo 16 módulos Terraform que cubren todas las capas de infraestructura Azure, 22 plantillas de Golden Path para scaffolding self-service, 17 agentes especializados de Copilot Chat, 13 integraciones de servidores MCP y más de 30 reglas de alerta de Prometheus. El cliente clona el template y personaliza política y contenido. El cliente no ensambla integraciones.

> DEFINICIÓN
>
> El principio operativo de todo acelerador de este cuerpo de trabajo: el cliente personaliza, el cliente no ensambla. El trabajo de integración es la parte más cara de la construcción de plataforma, y es exactamente la parte que el acelerador absorbe.

## La propuesta de valor en tres partes

El primer componente es velocidad. Una plataforma green-field que tomaría de 9 a 18 meses ensamblar desde componentes crudos se comprime a 90 a 180 días, porque el acelerador entrega los patrones de integración, el catálogo de Golden Paths, el cableado de cadena de suministro, la configuración GitOps y la infraestructura de agentes preintegrados. El segundo componente es coherencia comercial. Red Hat y Microsoft operan los servicios subyacentes como una alianza estratégica conjunta, así que el cliente tiene un despliegue de fundación con dos dueños comerciales de clase mundial y conversaciones unificadas de soporte, licenciamiento y roadmap.

El tercer componente es postura regulatoria y de auditoría. El stack Red Hat carga certificaciones FedRAMP, FIPS y Common Criteria que algunos sectores tratan como criterio bloqueante de compras, y el acelerador entrega defaults de política opinados para regímenes latinoamericanos: LGPD en Brasil, BACEN en servicios financieros, ANEEL en energía, ANS en salud. El cliente hereda esa postura en lugar de negociarla. Para bancos, utilities, salud y gobierno, este tercer componente es con frecuencia el decisivo.

## El impacto, en números

Los despliegues de referencia del acelerador reportan cuatro efectos principales. Velocidad de despliegue: releases cerca de 60% más rápidos, porque los pipelines Tekton automatizados y la promoción vía GitOps recortan los ciclos de semanas a horas. Eficiencia operacional: cerca de 40% menos toil, porque la infraestructura como código, la reconciliación y los portales self-service quitan trabajo manual de la semana del equipo de plataforma. Experiencia del desarrollador: las encuestas de adopción reportan mejoras de satisfacción en torno a 3x cuando llegan los Golden Paths y los flujos asistidos por

Copilot. Compliance: la política como código pone cada despliegue bajo los mismos estándares aplicados.

Trata estos números como debe tratarlos un líder de ingeniería: evidencia direccional de despliegues de referencia, no garantías. Los compromisos medibles que importan son los que la propia plataforma emite cuando está corriendo: lead time y tasa de fallo de cambios DORA para los equipos incorporados, atribución de costos por workload en los dashboards de gasto, y el hito de primer valor, la semana en que el primer equipo entrega por un Golden Path más rápido que su propio baseline anterior.

> DEFINICIÓN
>
> Cuatro números de referencia: 60% más velocidad de despliegue, 40% menos toil operacional, 3x de satisfacción del desarrollador y 100% de los despliegues bajo política como código. Verifícalos en tu propio parque con métricas DORA y atribución de costos, ambas entregadas por defecto por la plataforma.

## La escalera de madurez que da nombre al acelerador

Three Horizons toma su nombre del modelo de madurez que operacionaliza. Horizonte 1, Optimizar el Presente, es la fundación creíble: clúster, red, identidad, pipelines y la línea base de seguridad, apuntando a los primeros 90 días. Horizonte 2, Capacidades Emergentes, convierte la infraestructura en producto interno: el portal Developer Hub, Golden Paths a escala, observabilidad y self-service gobernado, en los meses tres a seis. Horizonte 3, Transformar el Futuro, es la capa agéntica: AI Foundry, workloads RAG y multiagente, y operaciones autónomas con revisión humana en las fronteras de promoción, en los meses seis a doce.

La escalera es un modelo de secuenciación: le dice al equipo de plataforma qué construir después y se niega a dejar que la capa agéntica aterrice sobre una fundación inacabada. Es compatible con el CNCF Platform Maturity Model, que es un modelo de medición: el Horizonte 1 corresponde aproximadamente al nivel Operational, el Horizonte 2 a Scalable y el Horizonte 3 a Optimizing o AI-Native en la mayoría de las dimensiones. Los capítulos

02 a 04 recorren cada horizonte en profundidad operacional, y el capítulo 05 cierra con el framework de decisión y la secuencia de despliegue.

# Referencias

Fuentes primarias para la definición del acelerador, el stack y el modelo de madurez.


- Documentación de Red Hat Developer Hub, la distribución enterprise de Backstage que sirve de IDP al acelerador.
- Documentación de Azure Red Hat OpenShift, el servicio OpenShift gestionado y operado conjuntamente sobre el que corre la plataforma.
- CNCF Platforms Whitepaper, la definición de referencia de las capacidades de plataforma que el acelerador materializa.
- CNCF Platform Engineering Maturity Model, el modelo de medición sobre el que se mapea la escalera H1/H2/H3.
- DORA 2025, State of AI-assisted Software Development, la lente de métricas de entrega usada para verificar las afirmaciones de impacto del acelerador.
- Red Hat Trusted Application Pipeline, el stack de seguridad de cadena de suministro cableado por defecto.


---


**Paula Silva**

AI-NATIVE SOFTWARE ENGINEER

in/paulanunes



v1.0.0 · 2026-07-03

Building the future of software development with AI and Agentic DevOps

01 ARQUITECTURA

# Cada componente nombrado, cada componente integrado: el stack de un vistazo.

El stack de Three Horizons es opinado a propósito. Cada capa tiene un componente nombrado, y cada componente llega integrado con los demás al momento del despliegue. Este capítulo recorre las cinco decisiones estructurales: la Internal Developer Platform, el clúster, el sustrato de entrega, el stack de seguridad de cadena de suministro y la capa de asistencia de IA, y explica qué compra cada decisión para el cliente.

## Red Hat Developer Hub como IDP

Red Hat Developer Hub es la distribución enterprise de Backstage soportada por Red Hat, y juega respecto del Backstage upstream el mismo papel que Red Hat Enterprise Linux juega respecto del Linux upstream: mismo núcleo, empaquetado endurecido, soporte enterprise, ciclo de vida más largo. RHDH entrega un conjunto curado de unos 30 a 40 plugins probados, incluyendo plugins nativos de OpenShift que traen información de Pipelines, GitOps, ACM y ACS directo al catálogo, más soporte 24x7 con parches de seguridad y asistencia de upgrade. Para industrias reguladas, ese contrato de soporte es criterio de compras, no conveniencia.

Los Software Templates de RHDH son la superficie de Golden Paths del acelerador, organizados por la escalera H1/H2/H3, y el Software Catalog es el registro de cada servicio, componente, API y recurso de la empresa, poblado por descubrimiento automático y por entradas declarativas en YAML junto al código. Para workloads de agentes, el catálogo es la superficie de descubrimiento: un agente que necesita llamar a un servicio consulta el catálogo, encuentra la entrada, recupera la spec de la API y la usa. El catálogo es la expresión operacional de la capa de contexto de la plataforma para información de servicios.

# Azure Red Hat OpenShift como clúster

ARO es un servicio OpenShift gestionado, operado conjuntamente por Microsoft y Red Hat. El control plane, etcd y el API server son aprovisionados, parchados y monitoreados por la organización conjunta de ingeniería bajo un único SLA, típicamente 99,95%, y una única factura de Microsoft cubre el clúster. Ese modelo operacional conjunto es una de las propiedades diferenciadoras del acelerador: los incidentes que cruzan la frontera Microsoft y Red Hat se resuelven dentro de una sola relación de soporte, en lugar de entre dos vendors culpándose mutuamente.

ARO se integra nativamente con la fundación Microsoft. Entra ID es el proveedor de identidad del clúster con RBAC gobernado por grupos Entra; Workload Identity emite credenciales federadas de corta duración, así que los secretos de servicio de larga vida son innecesarios; Azure Monitor ingiere la telemetría del clúster al plano unificado de observabilidad; Azure Policy aplica gobernanza en la frontera del clúster; y Defender for Cloud provee detección de amenazas en runtime. Para workloads sensibles a residencia de datos, ARO en Brazil South mantiene clúster, workloads y servicios de fundación enteramente en la región, lo que importa operacionalmente bajo LGPD y BACEN.

> DEFINICIÓN
>
> La elección del clúster es una elección de modelo de soporte. ARO significa un SLA, una factura y una ruta de escalamiento cruzando Microsoft y Red Hat. Donde esa relación única es requisito de compras, la decisión del clúster ya está tomada.

# OpenShift Pipelines y OpenShift GitOps

OpenShift Pipelines es la distribución Red Hat de Tekton, y su propiedad estructural es que los pipelines son recursos Kubernetes, no estado de un orquestador externo: un pipeline es un objeto YAML en la API del clúster, observable por el mismo stack que observa los workloads. OpenShift GitOps es la distribución Red Hat de Argo CD, y convierte a Git en la única superficie de escritura del clúster: el estado se declara en Git y se reconcilia continuamente, así que el drift manual se detecta y revierte, y el clúster se autocorrige.

El patrón combinado es un lazo: un desarrollador o agente hace commit, Pipelines construye y valida, Pipelines actualiza el repositorio GitOps con la nueva referencia de imagen, GitOps reconcilia el clúster, y la observabilidad cierra el lazo. Las políticas de sync son por ambiente: dev sincroniza automáticamente, staging exige aprobación explícita, producción exige sync explícito con múltiples aprobadores. Eso es revisión manual aplicada en la frontera de despliegue, y cada cambio es auditable en el historial de Git, lo que satisface la postura de compliance con primitivas que Git ya tiene.

> DEFINICIÓN
>
> Dos propiedades hacen el trabajo pesado: producción no tiene superficie de escritura directa, todo fluye por Git; y cada cambio es auditable en el historial. El patrón de incidente más común en producción, el cambio directo no autorizado, queda prevenido estructuralmente.

## Trusted Application Pipeline y la cadena de suministro

Red Hat Trusted Application Pipeline es el stack de seguridad de cadena de suministro integrado con Pipelines. Genera un SBOM completo en cada paso de build, firma cada artefacto y atesta la procedencia del build en formato compatible con SLSA, escanea dependencias y artefactos producidos en busca de vulnerabilidades, y aplica política que rechaza artefactos reprobados en build, despliegue y runtime vía admission control. El patrón en capas con GitHub es deliberado: GitHub Advanced Security en el commit, TAP en el build, Advanced Cluster Security en el runtime.

Ese cableado responde directamente a una regresión medida: las organizaciones que desplegaron asistentes de codificación con IA antes de madurar su postura de seguridad de plataforma vieron un alza del 38% en vulnerabilidades explotables en los primeros doce meses, según el Global Threat Report 2026 de CrowdStrike. Los cuatro mecanismos de TAP cierran esa brecha estructuralmente: dependencias vulnerables rechazadas en el build, artefactos sin firma rechazados en la admisión, divergencias de procedencia rechazadas en el despliegue, y SBOMs verificados continuamente contra lo que de verdad está corriendo.

## La capa de asistencia de IA en tres superficies

El acelerador integra tres superficies de IA por audiencia. GitHub Copilot Enterprise asiste a los desarrolladores en la capa de código fuente: completion, Copilot Chat y el coding agent de Copilot para trabajo multisesión. Red Hat Lightspeed asiste a los operadores en la capa de plataforma: operación de clúster, generación de playbooks de Ansible, explicación de políticas. Microsoft Foundry Agent Service corre los agentes personalizados que el cliente construye, con Entra Agent ID para identidad y la observabilidad de Foundry para trayectorias. Los desarrolladores encuentran a Copilot en el IDE, los operadores a Lightspeed en la consola, y los agentes de producción corren en Foundry.

La decisión de diseño que vale notar es que la capa de runtime de agentes no varía por acelerador: los agentes personalizados en Three Horizons corren sobre la fundación Microsoft, el mismo runtime que usa la variante Open Horizons, porque Red Hat no entrega un runtime de agentes competidor en 2026. La diferenciación entre los dos aceleradores vive en las capas de IDP y clúster. La capa de primitivas de IA, gateway de modelos, runtime de agentes, evaluación y observabilidad, es fundación compartida.

# Referencias

Documentación de componente para cada elemento nombrado del stack.


- Documentación de Red Hat Developer Hub, el conjunto curado de plugins, software templates y ciclo de soporte.
- Documentación de Azure Red Hat OpenShift, el modelo operacional conjunto, SLA e integraciones Azure.
- Documentación de OpenShift Pipelines, el sustrato de CI/CD nativo de Kubernetes basado en Tekton.
- Documentación de OpenShift GitOps, el modelo de reconciliación basado en Argo CD y las políticas de sync.
- Red Hat Trusted Application Pipeline, SBOMs, firma, atestación y aplicación de política.
- SLSA, Supply-chain Levels for Software Artifacts, el framework de procedencia y atestación que TAP implementa.

- Red Hat Lightspeed, la capa de asistencia de IA del lado del operador en el portafolio Red Hat.
- CrowdStrike, 2026 Global Threat Report, la regresión de seguridad medida que el cableado de cadena de suministro atiende.

---

**Paula Silva**

AI-NATIVE SOFTWARE ENGINEER

in/paulanunes

v1.0.0 · 2026-07-03

Building the future of software development with AI and Agentic DevOps

02 HORIZONTE 1

# Optimizar el Presente: una fundación Azure segura y observable en los primeros 90 días.

El Horizonte 1 es la fundación creíble. En los primeros 90 días el acelerador aprovisiona un ambiente Azure de grado de producción vía 16 módulos Terraform: el clúster ARO, red zero-trust, identidad y secretos, registry y bases de datos, disaster recovery y gobernanza de costos, e incorpora a los tres primeros equipos mediante seis plantillas H1 de Golden Path. Todo lo que está encima de esta capa asume que quedó bien hecha, y por eso viene primero y nada en ella es opcional.

## El alcance de los primeros 90 días

El Horizonte 1 entrega la plataforma mínima viable: el clúster ARO aprovisionado con Workload Identity, Key Vault e integración con Azure Monitor; Red Hat Developer Hub desplegado, con marca aplicada y asegurado con GitHub OAuth; OpenShift GitOps con políticas de sync por ambiente; OpenShift Pipelines; y el catálogo inicial de plantillas H1 cargado. La secuencia de despliegue corre el trabajo de fundación en las semanas cero a cuatro y la incorporación de equipos de la semana cuatro en adelante, con mediciones DORA activas para cada equipo incorporado desde el inicio.

El hito que cierra el Horizonte 1 es observable, no ceremonial: el primer equipo de aplicación entrega un workload por un Golden Path, y su lead time de PR a producción, medido en la nueva plataforma, supera el baseline anterior del propio equipo. Eso suele llegar en la semana seis a ocho. Una vez que el primer equipo entrega, los tres siguientes se incorporan con menos fricción, porque los patrones de despliegue están validados y la documentación fue probada en campo.

## Red zero-trust

El módulo de red aprovisiona una red virtual segmentada: subredes dedicadas para los nodos master y worker de ARO, private endpoints y tráfico de gestión, con network security groups restringiendo los flujos de entrada y salida en cada frontera de subred. Todos los servicios PaaS, Key Vault, el registry de contenedores y las bases de datos, se alcanzan exclusivamente por private endpoints resueltos por zonas de Azure Private DNS. Ningún servicio de datos tiene exposición pública en ninguna configuración que el acelerador entrega.

La postura es deliberadamente preventiva en lugar de detectivesca: la superficie de ataque se reduce antes de que lleguen los workloads, no se monitorea hasta tomar forma después. El control de egress, la resolución privada y el aislamiento de subredes se gestionan con Terraform, lo que significa que las propiedades de seguridad de la red son revisables en un pull request y reproducibles entre ambientes, en lugar de vivir como estado de consola que sufre drift.

## Identidad, secretos y protección de runtime

El módulo de seguridad aplica una regla en todas partes: ninguna credencial estática. Cada servicio usa managed identity; GitHub Actions y los workloads de Kubernetes se autentican por Workload Identity Federation sobre OIDC, así que no existe secreto de larga vida que filtrar; y las asignaciones RBAC son de mínimo privilegio, con alcance por resource group. Azure Key Vault, con soft delete y purge protection habilitados, es la fuente única de verdad para los secretos que deben existir.

La protección de runtime viene de Microsoft Defender for Containers, que vigila el clúster ARO en busca de actividad sospechosa de procesos, escalación de privilegios y movimiento lateral, escanea imágenes por vulnerabilidades y evalúa el clúster contra benchmarks CIS. El RBAC completo integra Azure AD con OpenShift OAuth, y las Security Context Constraints se aplican a nivel de clúster. El modelo de identidad es la parte del Horizonte 1 en la que la capa agéntica más se apoyará después: los agentes reciben identidades igual que los servicios.

DEFINICIÓN

Una regla, aplicada por construcción: sin contraseñas, sin secretos de service principal en código, sin credenciales de larga vida en ningún lugar. Managed identity más Workload Identity Federation es el default, y la lista de excepciones está vacía.

## Registry y servicios de datos

Azure Container Registry corre con geo-replicación para latencia de pull multirregión, escaneo de vulnerabilidades integrado en el push, private endpoint en lugar de exposición pública, y autenticación de pull por managed identity desde ARO, lo que elimina las credenciales de registry como clase de fuga. El módulo de bases de datos aprovisiona PostgreSQL Flexible Server como backend del Developer Hub con backups automáticos y alta disponibilidad de producción, Redis para sesiones y caché, y opcionalmente Cosmos DB detrás de una flag de Terraform para acceso distribuido globalmente.

Cada uno de estos servicios se alcanza solo por private endpoints. La consecuencia práctica para el equipo de plataforma es que las conversaciones de compliance de la capa de datos se vuelven cortas: no hay ruta pública que enumerar, y la cadena de verificación de SBOM a runtime del capítulo 01 se extiende hasta las imágenes que el registry admite.

## Resiliencia, costo y nomenclatura

El disaster recovery es explícito en lugar de aspiracional. Velero respalda el estado del clúster y los volúmenes persistentes en Azure Blob Storage con agenda configurable; las storage accounts y las bases replican a una región secundaria con geo-redundancia; y los objetivos de punto y tiempo de recuperación se definen por perfil de dimensionamiento, del mejor esfuerzo en dev a SLAs estrictos de producción. La recuperación es una propiedad probada de la plataforma, no un documento.

La gobernanza de costos llega en el Horizonte 1 porque retrofitearla después nunca sucede. Los presupuestos de Azure llevan alertas de umbral enrutadas a correo y Teams, las políticas de tagging son obligatorias en cada recurso, y los dashboards de análisis de gasto aterrizan en Grafana junto a las métricas de plataforma. Todos los recursos siguen

las convenciones de nombres del Cloud Adoption Framework, manteniendo el parque buscable y auditable entre ambientes y suscripciones.

> DEFINICIÓN
>
> Seis plantillas H1 de Golden Path incorporan a los primeros equipos: basic-cicd, documentation-site, infrastructure-provisioning, new-microservice, security-baseline y web-application. Cada una trae template.yaml, código esqueleto y documentación.

# Referencias

Documentación de componente para la fundación del Horizonte 1.


- Documentación de Azure Red Hat OpenShift, arquitectura de clúster, clústeres privados y autoscaling por MachineSet.
- Documentación de Azure Key Vault, soft delete, purge protection y acceso a secretos con alcance RBAC.
- Documentación de Microsoft Entra Workload ID, federación de identidad de workload sobre OIDC para pipelines y workloads.
- Microsoft Defender for Containers, detección de amenazas en runtime y evaluación contra benchmarks CIS.
- Documentación de Velero, backup y restore del estado del clúster y volúmenes persistentes.
- Convenciones de nombres del Cloud Adoption Framework, el estándar de nombres aplicado por el módulo CAF de naming.


---


**Paula Silva**

AI-NATIVE SOFTWARE ENGINEER

in/paulanunes



v1.0.0 · 2026-07-03

Building the future of software development with AI and Agentic DevOps

03 HORIZONTE 2

# Capacidades Emergentes: la infraestructura se convierte en producto interno.

El Horizonte 2 es donde la infraestructura cruda se convierte en algo que los desarrolladores de verdad quieren usar. En los meses tres a seis la plataforma gana su columna GitOps a escala, su puerta de entrada en Developer Hub, su catálogo completo de 22 plantillas de Golden Path, y el stack de observabilidad que hace la operación basada en evidencia. El criterio de salida es adopción: al menos la mitad de los workloads elegibles corriendo por la plataforma, catálogo poblado y TechDocs en uso en toda la organización.

## GitOps a escala: app-of-apps y políticas de sync

El acelerador configura Argo CD con el patrón app-of-apps: una aplicación raíz gestiona todas las aplicaciones hijas, así que el stack completo de la plataforma se despliega en orden declarado mediante sync waves. Las políticas de ambiente codifican el modelo de revisión en configuración, no en documentos de proceso. Dev sincroniza automáticamente, los cambios entran de inmediato; staging es semiautomático, sincroniza con aprobación; producción es manual, exige aprobación humana explícita. El control de acceso corre por SSO con Azure AD vía Dex, con roles de platform-admin y developer aplicados por RBAC.

A la escala del Horizonte 2, la disciplina que importa es la estructura de repositorios. Los parques multi-tenant producen repositorios GitOps con cientos o miles de aplicaciones, y la relación aplicación-repositorio se vuelve la superficie operacional real. El acelerador entrega convenciones opinadas de repositorio exactamente por eso: el patrón sobrevive a la escala solo cuando la estructura se decide antes de que llegue la aplicación número cincuenta, no después.

# Developer Hub como puerta de entrada

Developer Hub es el punto de entrada unificado de cada desarrollador: un solo lugar para descubrir servicios, crear aplicaciones por scaffolding, navegar documentación y revisar la salud del sistema. El Software Catalog es el registro centralizado de cada servicio, API y componente; TechDocs renderiza documentación versionada junto a cada entidad del catálogo; y la integración con GitHub provee autenticación OAuth y acceso vía GitHub App a repositorios y workflows.

El Horizonte 2 fija metas explícitas de adopción para el portal porque un portal sin adopción es mobiliario. Se espera que el Service Catalog alcance al menos 80% de cobertura de los servicios reales, y que TechDocs esté en uso en toda la organización. Ambos son medibles desde el propio catálogo, lo que mantiene la conversación honesta: o el parque es descubrible por el portal, o no lo es.

> DEFINICIÓN
>
> La prueba del portal es cobertura, no existencia. Un catálogo con 80% de cobertura cambia cómo agentes y humanos descubren servicios. Un catálogo con 20% le enseña a todos a esquivar el portal, y la plataforma pierde su puerta de entrada.

# El catálogo de 22 plantillas de Golden Path

El Horizonte 2 libera el catálogo completo de Golden Paths: 22 plantillas a través de los tres horizontes, cada una con template.yaml, código esqueleto y documentación. Las nueve plantillas H2 cubren el centro del parque: api-gateway con rate limiting y políticas de autenticación, api-microservice con OpenAPI y validación, batch-job como CronJob con reintentos y alertas, data-pipeline conectado a Event Hub, event-driven-microservice con integración a Service Bus o Event Hubs, gitops-deployment para incorporación rápida, reusable-workflows como biblioteca compartida de Actions, ado-to-github-migration para migraciones guiadas, y una plantilla de microservice de uso general.

La función del catálogo es estandarización por conveniencia, no por decreto. Cuando el camino sancionado es también el más rápido, los equipos lo toman voluntariamente, y cada servicio creado por plantilla llega con CI/CD, línea base de seguridad, observabilidad y registro en el catálogo ya cableados. La meta de salida del Horizonte 2 es adopción de

la plataforma en 50% o más de los workloads elegibles, medida, como todo aquí, con los datos de la propia plataforma.

## El stack de observabilidad

El stack es Prometheus, Grafana, Alertmanager y Jaeger, desplegados y configurados por el acelerador. Prometheus corre dos réplicas con retención de 15 días y 50GB de almacenamiento; Grafana autentica por SSO de Azure AD y entrega tres dashboards curados: Platform Overview como centro de mando diario del equipo de plataforma, Cost Management con atribución por equipo y namespace, y Golden Path Application con las cuatro señales doradas de cada servicio creado por plantilla. Alertmanager enruta a PagerDuty y Microsoft Teams con reglas de silencio e inhibición, y Jaeger rastrea solicitudes entre servicios.

Más de 30 reglas de alerta vienen por defecto, cubriendo cinco capas: alertas de infraestructura para presión de nodos, disco, red y salud de componentes de ARO; alertas de aplicación para SLOs de tasa de error, umbrales de latencia, crash loops y despliegues pendientes; alertas de workloads de IA para cuota de tokens, latencia de inferencia, disponibilidad de endpoints de modelos y fallos de evaluación; alertas de GitOps para fallos de sync y aplicaciones fuera de sincronía; y alertas de seguridad para hallazgos de Defender, violaciones de política, accesos inusuales y expiración de certificados. El grupo de alertas de IA existe en el Horizonte 2 a propósito: la plataforma queda instrumentada para agentes antes de que lleguen los agentes.

## Secretos, runners e ingress

Tres servicios operacionales completan el horizonte. El External Secrets Operator sincroniza continuamente secretos desde Azure Key Vault hacia Kubernetes Secrets mediante un ClusterSecretStore, con rotación automática y cero secretos guardados en Git o variables de entorno. Los runners self-hosted de GitHub en ARO dan a los workflows de CI/CD acceso a recursos privados, ACR, Key Vault, el propio clúster, sin exposición pública; los runners escalan según la profundidad de la cola, autentican por managed identity y son efímeros, así que cada job parte de estado limpio.

Ingress y TLS quedan con el NGINX Ingress Controller con ruteo por camino y host, rate limiting y capacidad de WAF; cert-manager aprovisiona y renueva certificados de Let's Encrypt sin gestión manual; y External DNS crea y actualiza registros en Azure DNS conforme los recursos de ingress aparecen y cambian. Ninguno de estos tres servicios es glamoroso, y los tres son la diferencia entre una plataforma que se opera sola y una que despierta al equipo con el pager.

> DEFINICIÓN
>
> El flujo de secretos en una línea: Azure Key Vault es la fuente de verdad, el External Secrets Operator sincroniza y rota, los pods consumen Kubernetes Secrets en runtime. Nada sensible vive en Git, en variables de entorno ni en definiciones de pipeline.

## Referencias

Documentación de componente para los servicios de plataforma del Horizonte 2.


- Documentación de Argo CD, el patrón app-of-apps, sync waves y políticas de sync.
- Documentación de Backstage, las fundaciones de catálogo, TechDocs y scaffolder bajo Developer Hub.
- Documentación de Prometheus, recolección de métricas y el modelo de reglas de alerta.
- Documentación de Grafana, dashboards y single sign-on con Azure AD.
- Documentación del External Secrets Operator, el patrón ClusterSecretStore y la rotación de secretos.
- Documentación de cert-manager, aprovisionamiento y renovación automatizados de certificados TLS.
- Documentación de Jaeger, tracing distribuido entre los servicios de la plataforma.



**Paula Silva**

AI-NATIVE SOFTWARE ENGINEER

in/paulanunes

v1.0.0 · 2026-07-03

Building the future of software development with AI and Agentic DevOps

04 HORIZONTE 3

# Transformar el Futuro: la capa agéntica aterriza en una plataforma terminada.

El Horizonte 3 es la razón de que existan los otros dos. En los meses seis a doce la plataforma gana su infraestructura de IA como código, sus siete Golden Paths de IA, su flota de 17 agentes especializados de Copilot Chat, y la maquinaria de evaluación y observabilidad que hace gobernable la autonomía. El criterio de salida es un proceso de negocio real corriendo de punta a punta por un agente, con revisión humana en las fronteras de promoción y cada trayectoria, token y costo contabilizados.

## Azure AI Foundry, aprovisionado como código

El módulo de AI Foundry aprovisiona la infraestructura completa de IA como Terraform: el hub de Foundry, proyectos, despliegues de modelos y Azure AI Search, reproducibles y versionados como cualquier otra capa de la plataforma. Los despliegues de referencia incluyen GPT-4o para trabajo multimodal, o3 para razonamiento más difícil y previews de próxima generación conforme llegan a la suscripción, todos expuestos por managed online endpoints para inferencia y por el gateway de Foundry para ruteo, medición y política.

La propiedad estructural que vale internalizar es que la infraestructura de IA deja de ser especial. Los despliegues de modelo viven en el mismo flujo de pull request que las reglas de red; el vector store se aprovisiona por el mismo pipeline que la base de datos; y los rollbacks funcionan igual en todas partes. Los equipos piden capacidad de IA por Golden Paths, no por tickets, y la plataforma mide lo que consumen por workload desde el primer día.

## Los siete Golden Paths de IA

El Horizonte 3 entrega siete plantillas enfocadas en IA. rag-application crea por scaffolding una aplicación completa de generación aumentada por recuperación: ingestión de

documentos, embeddings vectoriales en Azure AI Search y una interfaz de chat, con observabilidad incluida. multi-agent-system provee un patrón de orquestador con subagentes especialistas, integraciones de herramientas MCP, protocolos de handoff y estado compartido. mlops-pipeline cubre el ciclo completo de ML de entrenamiento a evaluación, registro, despliegue y monitoreo de drift. foundry-agent crea aplicaciones de agente único en Foundry Agent Service con identidad y evaluación cableadas.

Las tres restantes cierran el lazo de calidad y operación: ai-evaluation-pipeline convierte la evaluación de modelos en un pipeline reutilizable, copilot-extension crea extensiones personalizadas de GitHub Copilot apoyadas en servicios de IA de Azure, y sre-agent-integration conecta el triaje de incidentes asistido por IA, la ejecución de runbooks y el análisis de causa raíz al flujo operacional de la plataforma. Juntas, las siete plantillas significan que un equipo de aplicación puede levantar un workload de IA gobernado la misma tarde en que decide construirlo, y cada uno de esos workloads llega observable y evaluable por construcción.

> DEFINICIÓN
>
> Siete plantillas, una propiedad: un workload de IA creado por Golden Path llega con identidad, observabilidad, evaluación y medición de costos ya acopladas. El equipo escribe la lógica de dominio. La plataforma provee todo lo que lo hace de grado de producción.

## La flota de 17 agentes y el MCP

El acelerador entrega 17 agentes especializados de Copilot Chat que operan directo en VS Code, cubriendo flujos de desarrollo, despliegue, operación y gestión de incidentes, apoyados en 16 skills operacionales y 13 integraciones de servidores MCP que exponen capacidades de la plataforma como herramientas. Cada agente tiene rol definido, acceso explícito a herramientas y fronteras conductuales de tres niveles: acciones que SIEMPRE ejecuta, acciones donde debe PREGUNTAR PRIMERO, y acciones que NUNCA ejecuta. El modelo de fronteras es la primitiva de gobernanza: la autonomía se concede por clase de acción, no por agente.

La flota es también cómo el equipo de plataforma come de su propia comida. Los primeros agentes adoptados en un despliegue Three Horizons son típicamente internos a la plataforma: triaje de incidentes, autoría de runbooks, revisión de políticas, remediación de drift. Esa adopción interna, al menos una tarea recurrente de la plataforma absorbida por un agente, es la mitad de la prueba de madurez que este cuerpo de trabajo llama el Mandato Dual, y se espera que pase alrededor de la semana seis del despliegue.

## Gates de evaluación y observabilidad de agentes

La evaluación es un gate de despliegue, no un informe. El pipeline de evaluación puntúa las salidas de los modelos por relevancia, coherencia y fundamentación contra datasets de referencia, filtra contenido dañino, sesgado o violatorio de políticas, y alimenta los resultados al CI/CD como gates: una evaluación reprobada bloquea el despliegue automáticamente, igual que lo hace un test reprobado. La promoción de un agente a un radio de acción mayor sigue las mismas fronteras de revisión manual que cualquier cambio de producción.

La observabilidad se extiende a todo lo que hacen los agentes: trazas de trayectorias, costo por solicitud y por workload, muestreo de salidas para revisión de calidad y resultados de evaluación en el tiempo, junto a las alertas de workloads de IA del Horizonte 2. El dashboard de AI Usage se vuelve superficie primaria de KPI de la plataforma. La salida del Horizonte 3 es concreta: al menos un proceso de negocio real corre de punta a punta por un agente, con revisión humana en las fronteras de promoción, a 90 a 180 días del inicio del despliegue.

> DEFINICIÓN
>
> La postura de gobernanza en una frase: los agentes son workloads de primera clase con identidad, presupuesto, gates de evaluación y auditoría, y la autonomía se expande solo conforme la evidencia se acumula. Eso es lo que hace operable el futuro de 100 agentes por humano.

# Referencias

Documentación de componente para la capa agéntica del Horizonte 3.

- Documentación de Azure AI Foundry, el hub, proyectos, despliegues de modelos y gateway.
- Documentación de Azure AI Search, la capa de recuperación vectorial bajo el Golden Path de RAG.
- Microsoft Agent Framework 1.0, el framework sobre el que se construyen los agentes personalizados.
- Model Context Protocol, el estándar de integración de herramientas detrás de los 13 servidores MCP.
- Documentación de GitHub Copilot, Copilot Chat, agentes personalizados y extensiones.
- Convenciones semánticas GenAI de OpenTelemetry, el estándar de telemetría para trayectorias y costos de agentes.

---

**Paula Silva**

AI-NATIVE SOFTWARE ENGINEER

in/paulanunes

v1.0.0 · 2026-07-03

Building the future of software development with AI and Agentic DevOps

05 LA DECISIÓN

# Elígelo por las razones correctas, despliégalo en el orden correcto.

Three Horizons no es la respuesta correcta para todo cliente, y fingir lo contrario desperdicia el año de todos. Este capítulo de cierre le da estructura a la decisión: qué materializa el acelerador en términos de gobernanza, qué cuesta comercialmente y qué compra el soporte conjunto, qué perfiles de cliente inclinan a favor y en contra, y la secuencia de tres fases que lleva una decisión firmada a un hito de primer valor medido.

## Qué compra el crosswalk de pilares

En términos de gobernanza, el acelerador materializa los cuatro pilares CNCF con componentes nombrados. Los Golden Paths son los Software Templates de RHDH con CI/CD Tekton generado, despliegue vía Argo CD e infraestructura Terraform. Los Guardrails son las políticas de Advanced Cluster Security, admission controllers, Azure Policy en las fronteras de nube y clúster, Workload Identity en lugar de credenciales de larga vida, y firma TAP en la admisión. Los Safety Nets son la reconciliación GitOps, el plano unificado de observabilidad, backup con Velero, entrega progresiva vía service mesh y detección de runtime de ACS. La Revisión Manual es el sync con gates por ambiente más la promoción gobernada de agentes.

La propiedad diferenciadora es que estos pilares llegan preintegrados en el despliegue. El cliente no parte de un clúster en blanco a integrar cada pilar; parte de una fundación integrada y personaliza política y contenido. Ese trabajo de integración absorbido es la explicación estructural de la compresión de 9 a 18 meses a 90 a 180 días: la parte más lenta de la construcción de plataforma es exactamente la parte que el acelerador entrega terminada.

## La estructura comercial y el soporte conjunto

Tres líneas comerciales componen el acelerador. Las suscripciones Red Hat cubren RHDH, ARO, Pipelines, GitOps, Service Mesh, ACM, ACS, TAP y Lightspeed donde esté licenciado. El consumo Azure cubre cómputo y almacenamiento del clúster más los servicios de fundación: Key Vault, AI Foundry, Monitor, Entra. GitHub Enterprise cubre licenciamiento por seat con Advanced Security y Copilot Enterprise. Comparando con la variante open-source de esta arquitectura, la línea de suscripciones Red Hat es el principal diferenciador de costo, y el contrato de soporte es exactamente lo que esa línea compra.

El modelo de soporte conjunto es la razón comercial más fuerte por la que los clientes eligen este camino. Microsoft y Red Hat operan ARO conjuntamente, con rutas de escalamiento entre sí, así que un incidente que cruza la frontera se acciona con una parte y se resuelve internamente, en lugar de estancarse mientras dos vendors se culpan. Para algunas organizaciones de compras esa relación única es un requisito no negociable, y donde lo es, la decisión está efectivamente tomada antes de que empiece cualquier comparación técnica.

> DEFINICIÓN
>
> Lee la estructura de costos como intercambio, no como penalidad: la línea de suscripciones Red Hat compra el contrato de soporte conjunto, la postura de certificación y el ciclo de vida enterprise. Los clientes que no necesitan esas tres cosas no deberían pagarlas, y la lógica del capítulo lo dice explícitamente.

## Perfiles de cliente: a favor, en contra e indecisos

El perfil que inclina hacia Three Horizons tiene al menos tres de estas propiedades: inversión Red Hat previa a escala material, RHEL, OpenShift o Ansible; requisitos regulatorios o de auditoría donde las certificaciones FedRAMP, FIPS o Common Criteria bloquean compras; una estrategia de vendors que prefiere una relación conjunta fuerte a varias relaciones flojas; y concentración en servicios financieros, energía, gobierno o salud. Los despliegues de referencia detrás de este playbook están exactamente en esos

sectores, y la biblioteca de políticas viene opinada para LGPD, BACEN, ANEEL y ANS por eso.

El perfil que inclina en contra es igual de claro: ninguna huella Red Hat previa, estrategia de minimizar la cantidad de vendors, fuerte expertise interna en Kubernetes upstream, o una cultura de ingeniería OSS-first que valora transparencia y ritmo por encima del soporte del vendor. Esos clientes quedan mejor servidos por la variante open-source Open Horizons sobre AKS y Backstage upstream. Cuando el liderazgo está dividido o ambas culturas coexisten, corre un descubrimiento comparativo de un día en lugar de discutir: puntúa el parque contra estas propiedades de perfil y deja que la evidencia recomiende.

## La secuencia de despliegue en tres fases

Fase uno, semanas cero a cuatro, es fundación: el clúster ARO aprovisionado con Workload Identity, Key Vault e integración con Monitor; RHDH desplegado, con marca y asegurado con GitHub OAuth; GitOps con sync por ambiente; Pipelines; y el catálogo de plantillas H1 cargado. Fase dos, semanas cuatro a doce, es incorporación: los tres primeros equipos de aplicación llegan por las plantillas H1, el Service Catalog se puebla, TechDocs se adopta, las políticas ACS se aplican, TAP se integra con Pipelines, y la medición DORA queda activa para cada equipo incorporado.

Fase tres, meses tres a doce, es escala: las plantillas H2 se liberan y adoptan, el dashboard de costos entra en producción, el primer agente de Copilot Chat se despliega, la cobertura del catálogo cruza el 80%, el multiclúster se expande vía ACM donde aplique, y las plantillas H3 llegan a los early adopters. El hito que ancla toda la secuencia llega en la semana seis a ocho: el primer equipo entrega por un Golden Path con un lead time más rápido que su propio baseline anterior. Ese número, no una demo, es lo que debe fundamentar las decisiones de expansión.

> DEFINICIÓN
>
> El hito de primer valor es deliberadamente poco impresionante: un equipo, un workload, un número de lead time mejor que su propio baseline. Todo lo que escala después se construye sobre ese hecho verificado, y por eso la secuencia se niega a saltárselo.

## Mapeo de madurez y la prueba permanente

La escalera H1/H2/H3 es un modelo de secuenciación; el CNCF Platform Maturity Model es un modelo de medición, y los dos se componen. El Horizonte 1 corresponde aproximadamente al nivel Operational de CNCF en la mayoría de las dimensiones, el Horizonte 2 a Scalable, y el Horizonte 3 a Optimizing o AI-Native. Usados juntos, la escalera le dice al equipo de plataforma qué construir después y el modelo CNCF le dice qué tan maduro es de verdad el resultado, lo que mantiene las conversaciones de roadmap ancladas en estado evaluado en lugar de entusiasmo.

La prueba permanente de un despliegue saludable es el Mandato Dual, y en Three Horizons es tratable alrededor de la semana seis: al menos una tarea recurrente del propio equipo de plataforma fue absorbida por un agente, y al menos un equipo de aplicación consume el gateway de modelos por un Golden Path. Cuando ambas mitades se cumplen, la plataforma come de su propia comida y sirve a sus clientes al mismo tiempo, y la capa agéntica que sigue aterriza sobre capacidad demostrada en lugar de una diapositiva.

# Referencias

Fuentes primarias para el framework de decisión, el modelo comercial y el mapeo de madurez.


- Documentación de Azure Red Hat OpenShift, la operación conjunta, el SLA y el modelo de soporte en que se apoya el argumento comercial.
- CNCF Platform Engineering Maturity Model, el modelo de medición de cinco dimensiones sobre el que se mapean los horizontes.
- CNCF Platforms Whitepaper, las definiciones de pilares detrás del crosswalk.
- Red Hat Advanced Cluster Security, el componente de seguridad de runtime de las filas de Guardrails y Safety Nets.
- Red Hat Advanced Cluster Manager, política y gobernanza multiclúster para la fase de escala.
- Programa de investigación DORA 2025, las métricas de entrega que anclan el hito de primer valor y la decisión de expansión.

**Paula Silva**

AI-NATIVE SOFTWARE ENGINEER

in/paulanunes

v1.0.0 · 2026-07-03

Building the future of software development with AI and Agentic DevOps