

| 12. Operaciones con Reconocimiento del Tenant

Cuando se crea un entorno multi-tenant, gran parte del esfuerzo se concentra en construir una experiencia unificada que permita gestionar, operar y desplegar el entorno desde un panel único de control.

En muchos sentidos, esta visión operacional de una solución multi-tenant es la que mejor refleja si se ha construido un sistema que logra la agilidad, la innovación y la eficiencia que promete el modelo SaaS.

| La mentalidad operacional de SaaS

En este modo de trabajo, pensamos en cada paso del recorrido del cliente y monitoreamos, medimos y analizamos continuamente la calidad de la experiencia de servicio del cliente en cada punto de contacto con el sistema.

Para entender mejor este modelo operacional más amplio, veamos cómo SaaS influye en la mentalidad de distintas partes de la organización. El punto de partida más natural es la visión "clásica" de las operaciones, donde un equipo técnico está en primera línea monitoreando y midiendo la actividad, la escalabilidad y el estado de la aplicación SaaS.

Este equipo enfrenta nuevos desafíos en un entorno SaaS, donde cualquier interrupción o degradación del rendimiento puede tener un efecto en cascada sobre todos los tenants del sistema.

En esencia, los equipos necesitan vistas mucho más orientadas al tenant que les permitan identificar, reaccionar y responder eficazmente a los eventos operacionales.

Lo que resulta aún más interesante es cuando consideramos aspectos operacionales fuera de la visión tradicional. Aquí nos alejamos de la urgencia de mantener un entorno sin interrupciones y nos enfocamos más en la experiencia del cliente.

El onboarding de tenants, por ejemplo, representa un momento operacional significativo para un proveedor SaaS. Queremos que los clientes avancen por el proceso de onboarding con la menor fricción posible, pasando de la incorporación a la generación de valor real en el menor tiempo posible.

Esta misma mentalidad puede extenderse a otras áreas de la organización. El equipo de Customer Success, por ejemplo, debería tener acceso a información que le permita monitorear la actividad continua del cliente: qué funcionalidades utiliza, dónde podría estar encontrando obstáculos, etc.

Los equipos de product management también están vinculados a esta historia operacional. Pueden necesitar, por ejemplo, acceso a datos sobre tendencias de consumo de tenants que moldeen la estrategia de niveles y precios del entorno. Pueden querer emplear canary releases para probar funcionalidades en poblaciones específicas de tenants o analizar patrones de interacción para detectar puntos de fricción en la experiencia de usuario.

La conclusión clave es que las operaciones SaaS deben verse como una experiencia más holística que abarca múltiples roles en una organización. Esto, por supuesto, requiere un cambio cultural significativo en algunas organizaciones.

Para algunos, asumir esta perspectiva operacional ampliada puede no ser algo natural. Aquí es donde el liderazgo debe marcar el tono operacional adecuado para la organización.

En mi opinión, si realmente nos enfocamos en construir una experiencia de servicio SaaS rica, el negocio debería priorizar la construcción de los cimientos operacionales y la cultura que impulsen el crecimiento y el éxito del negocio SaaS. Los equipos y la organización deben apostar por las capacidades operacionales del servicio —incluso a expensas de funcionalidades y características.

A medida que evolucionan los requisitos de los tenants, la arquitectura, el mercado y los equipos, debemos reevaluar continuamente las herramientas, mecanismos y métricas utilizados para gestionar y analizar el estado operacional del negocio.

| Métricas operacionales multi-tenant

Los equipos SaaS suelen tener hambre de datos e información que abarquen todo el espectro de perspectivas técnicas y de negocio. Los product owners, arquitectos, desarrolladores SaaS, marketing, directores generales —todos deberían tener un interés genuino en operar con base en métricas, utilizando datos para evaluar continuamente el rendimiento del negocio y si está satisfaciendo las necesidades de los tenants.

| Métricas de actividad de tenant

En un entorno SaaS, el equipo necesitará información sobre las actividades específicas de los tenants individuales. Estos datos ayudarán a construir una visión más completa de cómo los tenants interactúan con los distintos elementos del entorno y, en algunos casos, a correlacionar esa actividad con otras métricas que puedan revelar patrones y tendencias interesantes.

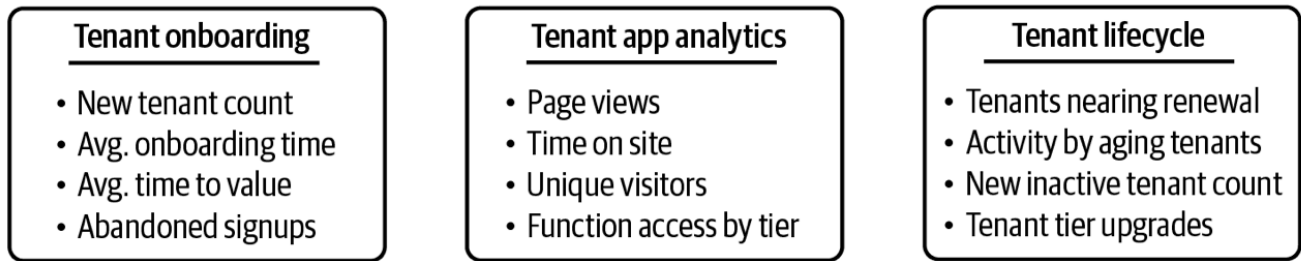


Figure 12-1. Examples of tenant activity metrics

Medir la repetibilidad, estabilidad y escalabilidad del onboarding es fundamental para evaluar el estado del negocio SaaS. Además, representa la primera impresión que se le da al nuevo tenant. **Las métricas de onboarding tienen especial importancia en cualquier flujo de onboarding de autoservicio.**

La siguiente categoría de métricas de actividad de tenant, el análisis de la aplicación por tenant, se muestra en el centro del diagrama. Representa las métricas clásicas utilizadas para registrar la interacción de un tenant con la aplicación en sí (similar a las analíticas web). **Esta es un área bien conocida, pero la idea de capturarla por nivel o por tenant agrega una nueva capa de consideraciones.**

Finalmente, en el extremo derecho del diagrama, se observan las métricas o eventos del ciclo de vida del tenant. Aquí el sistema captura datos sobre tenants que pueden estar acercándose o atravesando diferentes transiciones de estado. **Esto, combinado con el hecho de que se acercan a su renovación, ayudaría al equipo a identificar tenants que están considerando abandonar el sistema.**

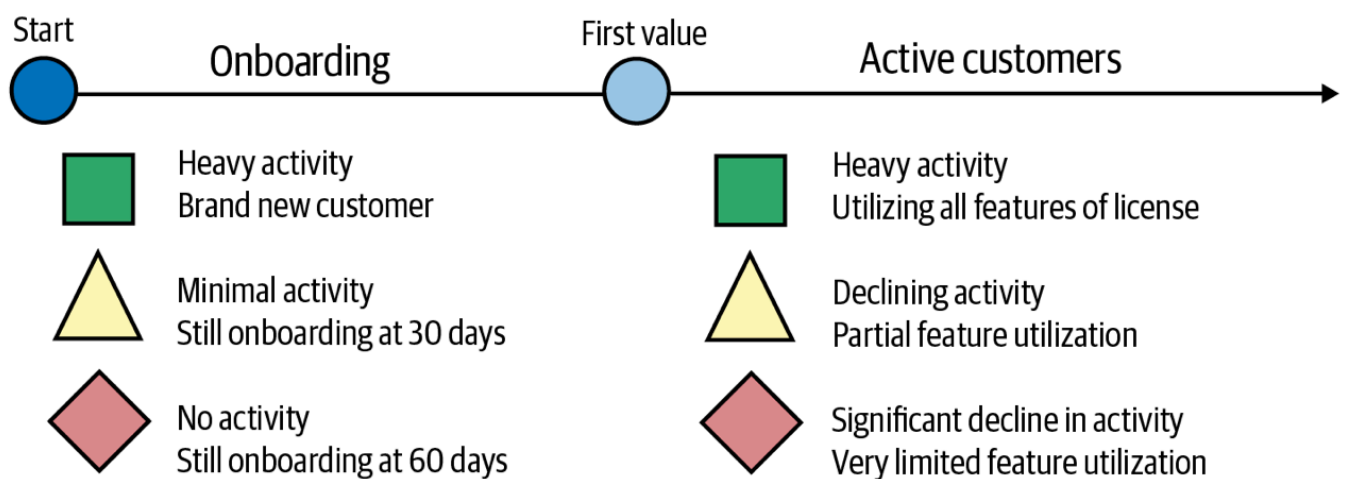


Figure 12-2. Correlating tenant activity to business events

Í Métricas de agilidad

La agilidad se utiliza para plantear cualquier cantidad de preguntas operacionales.

- ¿Están preparados para absorber una ráfaga de nuevos tenants?
- ¿Con qué eficacia se están incorporando nuevas funcionalidades y capacidades?
- ¿Con qué rapidez y proactividad puede el equipo responder a problemas de rendimiento, escalabilidad o funcionalidad?

Estas son todas las áreas donde los mecanismos y herramientas de la experiencia operacional deben demostrar su valor.

Í Disponibilidad

Medir la agilidad operacional comienza con la métrica más fundamental: la disponibilidad. Los problemas con el tiempo de actividad socavarán cada uno de los demás aspectos de la historia de agilidad.

Los equipos con dificultades de disponibilidad o estabilidad tienden a limitar los lanzamientos por temor a que introducir nuevas funcionalidades también signifique más interrupciones.

Rastrear estos datos y medir la respuesta del sistema permitirá evaluar la capacidad del sistema para detectar y responder a los desafíos antes de que afecten a los tenants.

Í Frecuencia de despliegue/lanzamiento

En un entorno multi-tenant, la construcción y el despliegue suelen presentar nuevos desafíos. La huella de recursos siloed y pooled de los tenants implica que las herramientas de despliegue deben considerar cómo aplicar actualizaciones de manera contextual, basándose en el perfil de infraestructura único del entorno de cada tenant.

Esto incluye considerar cómo aplicar cambios de configuración o de esquema a los tenants en un entorno sin tiempo de inactividad.

Aquí es donde se verá una superposición con las métricas de DevOps Research and Assessment (DORA), a las que muchos recurren para medir la eficacia de su huella de recursos DevOps.

Despliegues fallidos

Esta métrica ofrece una evaluación más concreta de la estabilidad de la automatización del despliegue, destacando potencialmente problemas que pueden estar afectando o afectarán la disponibilidad general del entorno.

Tiempo de ciclo

El tiempo de ciclo es la medida clave de esta dinámica: mide el tiempo transcurrido entre tener una idea para una nueva funcionalidad y el momento en que esa funcionalidad llega a manos de los clientes.

Tiempo medio de detección/recuperación

Si algún problema se introduce en el sistema, las herramientas y mecanismos deben detectarlo lo más rápido posible, empleando elementos que puedan reparar el entorno con rapidez.

Esto puede consistir en un rollback o en el lanzamiento de un parche. La clave es: ¿con qué rapidez pueden las herramientas y la automatización abordar eficazmente el problema y devolver el sistema a un estado saludable?

Tasa de escape de defectos

La naturaleza de todo o nada de un modelo multi-tenant puede requerir un mayor nivel de inversión en la cobertura general de pruebas del entorno.

Al medir la tasa de escape de defectos, los equipos tendrán una visión mucho más clara de qué tan efectivamente sus estrategias de prueba están capturando e identificando problemas antes de que lleguen al entorno de producción.

Métricas de consumo

En los entornos multi-tenant, los equipos operacionales deben tener visibilidad sobre cómo los tenants consumen los recursos que forman parte de su entorno.

Estas métricas serán esenciales para analizar políticas de escalado y perfilar la eficiencia del consumo de infraestructura. También es probable que influyan en los modelos de niveles y throttling.

Aquí es donde, para un recurso determinado, podemos entender qué porcentaje fue consumido por tenants individuales. Esto incluye desglosar el consumo de cómputo y almacenamiento de la base de datos relacional.

Existe una amplia variedad de estrategias para capturar estos datos. El enfoque variará según la naturaleza de la solución. Generalmente es más fácil comenzar pensando en las diferentes capas de la arquitectura e identificar dónde conviene introducir la instrumentación para capturar y publicar los datos de métricas de consumo.

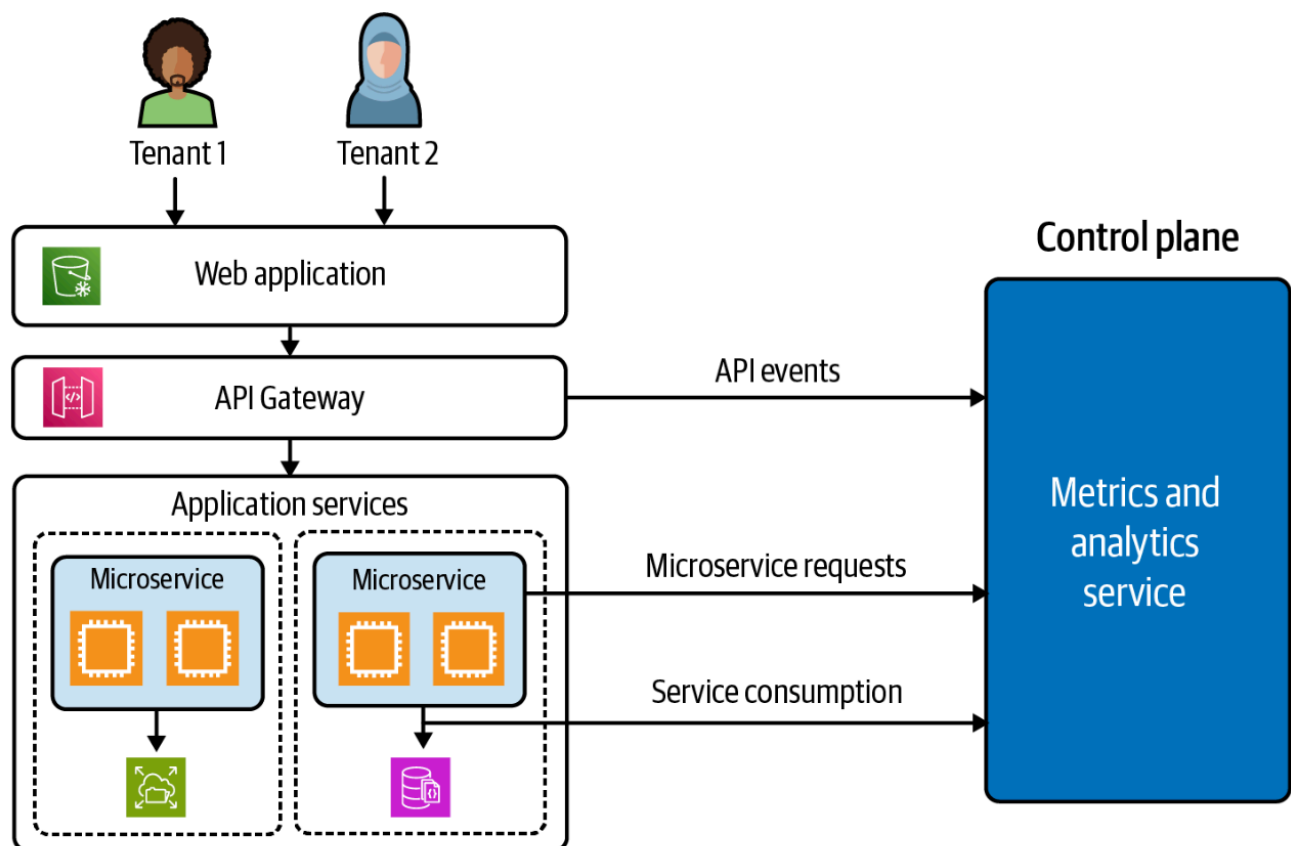


Figure 12-4. A layered approach to gathering consumption metrics

En conjunto, esto puede parecer un esfuerzo considerable. Sin embargo, contar con estos datos es esencial para construir y evolucionar la huella de recursos del entorno SaaS. Tiene un valor e impacto de largo alcance que puede moldear la eficiencia operacional, la eficiencia en costos, el escalado y las decisiones de niveles.

Métricas de costo por tenant

El objetivo es esencialmente asociar un costo con cada tenant y nivel, y utilizar esos datos para entender mejor cómo los costos de infraestructura de los tenants se relacionan con las estrategias de precios y niveles de la arquitectura, ofreciendo una visión más clara de los márgenes reales del entorno SaaS.

Son estos datos los que ayudan al equipo a determinar cómo y dónde las políticas de niveles y precios se alinean con los costos generales de infraestructura del entorno. De manera más amplia, los equipos necesitan acceso a estos datos para perfilar continuamente los costos a medida que se introducen nuevas funcionalidades y capacidades.

Para calcular estos datos de costo por tenant, será necesario incorporar capacidades adicionales al servicio de métricas y analíticas del sistema. Una vez obtenidos los datos de consumo (descritos anteriormente), se necesitará acceso a la información de facturación de la infraestructura.

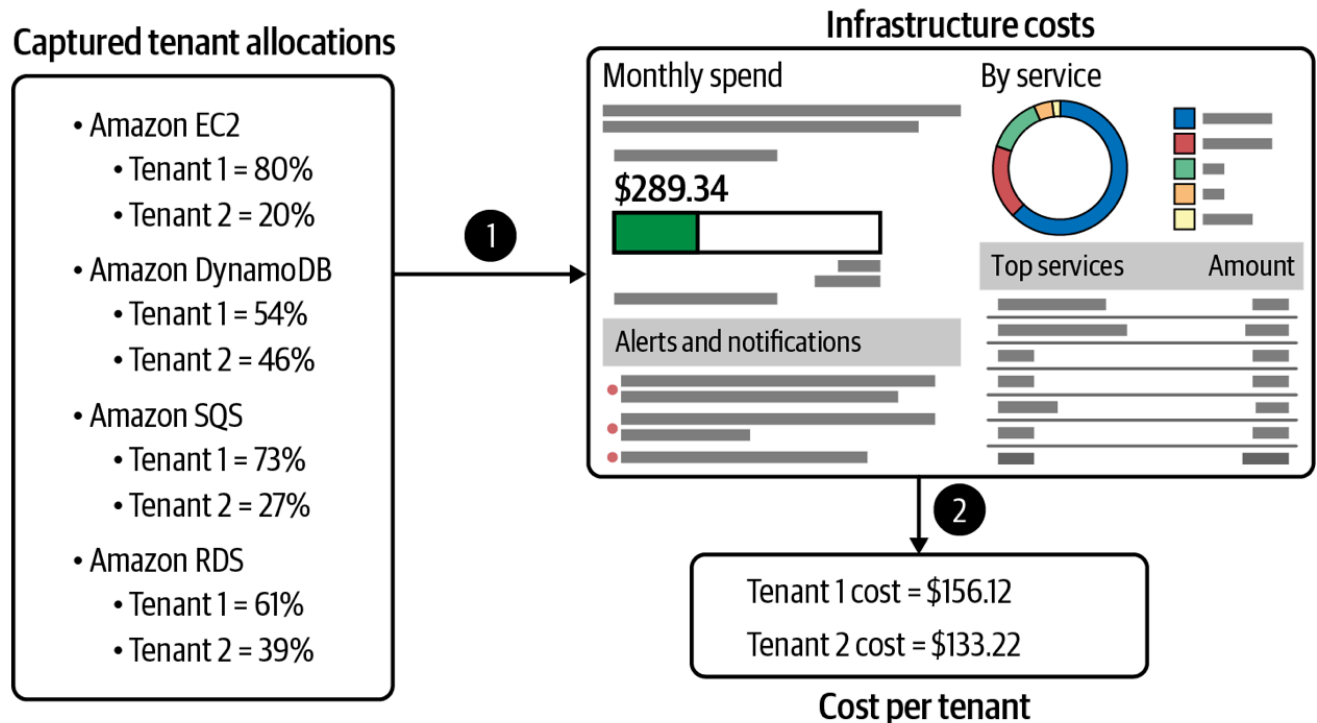


Figure 12-6. Correlating consumption with costs

Es importante señalar que la métrica de costo por tenant está concebida para representar una aproximación de costos. **Esta no es una función contable ni una forma de generar una factura para un cliente** —es una manera de proporcionar una visión operacional de los costos por tenant que puede informar el perfil arquitectónico, de precios y de niveles del entorno.

Para muchos, la estrategia de costo por tenant puede estar directamente influenciada por el perfil de costos general de la infraestructura. Si el cómputo representa el 80% de la factura, por ejemplo, resulta fácil justificar una inversión significativa en capturar los detalles de costo por tenant para esa porción.

Í Métricas de salud del negocio

Estas métricas se centran más en ingresos, marketing e información macro sobre tenants que permiten evaluar tendencias con impacto en la salud general del negocio.

A continuación se presenta una revisión rápida de algunas de las métricas clave que vale la pena destacar:

- **Ingresos recurrentes mensuales (MRR):** Representa la visión más clara de cómo evolucionan los ingresos de la organización.
- **Churn:** En entornos donde los tenants se incorporan y, potencialmente, se van con cierta regularidad, conviene monitorear la tasa de churn para evaluar continuamente la velocidad a la que los tenants rotan dentro del entorno.
- **Costo de adquisición de clientes (CAC):** Es una métrica de negocio clásica que evalúa el costo asociado con la adquisición de nuevos clientes. En un entorno con inversión significativa en marketing, es importante tener alguna noción del costo promedio asociado con cada cliente adquirido.
- **Valor de vida del cliente (CLTV):** Esta métrica mide el ingreso promedio que se obtendrá de un cliente a lo largo de su tiempo usando el sistema.
- **Ratio CLTV/CAC:** Aquí se evalúa la relación entre el costo de adquirir un cliente y el valor total que este aporta al negocio. Un ratio de 1:1, por ejemplo, sugeriría que el dinero invertido en adquirir un cliente equivale exactamente a lo que se ganará de ese cliente. **Obviamente, ese no es el objetivo. Cada organización debe determinar qué ratio tiene sentido para su negocio. Algunos apuntan a 3:1, pero existe debate sobre cuál debería ser ese objetivo.**

Los datos pueden provenir de sistemas contables, herramientas de gestión de relaciones con clientes (CRM) y similares. La forma de agregar y presentar esta información dependerá en gran medida de cómo se obtenga. Existen sistemas especializados en este espacio y, en algunos casos, puede ser necesario construir una solución propia.

Í Métricas compuestas

También es probable que sea necesario desarrollar e introducir métricas propias que se correspondan con los aspectos específicos de la solución o del dominio.

La conclusión clave es que las métricas no siempre se correlacionan directamente con alguna actividad de tenant o consumo de infraestructura. Algunas de las mejores métricas que se pueden introducir provienen de los mecanismos creados para perfilar cargas de trabajo, eventos de negocio lógicos u otras actividades de mayor nivel que forman parte del entorno.

! Métricas de línea base

La infraestructura emitirá naturalmente métricas que ofrecen información fundamental sobre el rendimiento de los componentes de la arquitectura. El cómputo del entorno, por ejemplo, emitiría naturalmente datos sobre actividad de CPU, consumo de memoria, etc.

Estos datos deben seguir considerándose parte del alcance de esta historia de métricas. Conviene incorporarlos y situarlos junto con los demás datos de métricas, utilizándolos para correlacionar patrones de tenants con estas otras métricas (cuando resulte pertinente).

El desafío de estas métricas de línea base es que no siempre pueden vincularse a tenants individuales.

! Instrumentación y agregación de métricas

En primer lugar, será necesario introducir instrumentación en los servicios de la aplicación para que publiquen las métricas en el control plane.

La otra mitad de la historia de métricas es la ingestión y agregación de los datos de métricas. Aquí se identifican las herramientas y tecnologías que se utilizarán para procesar y almacenar estos datos.

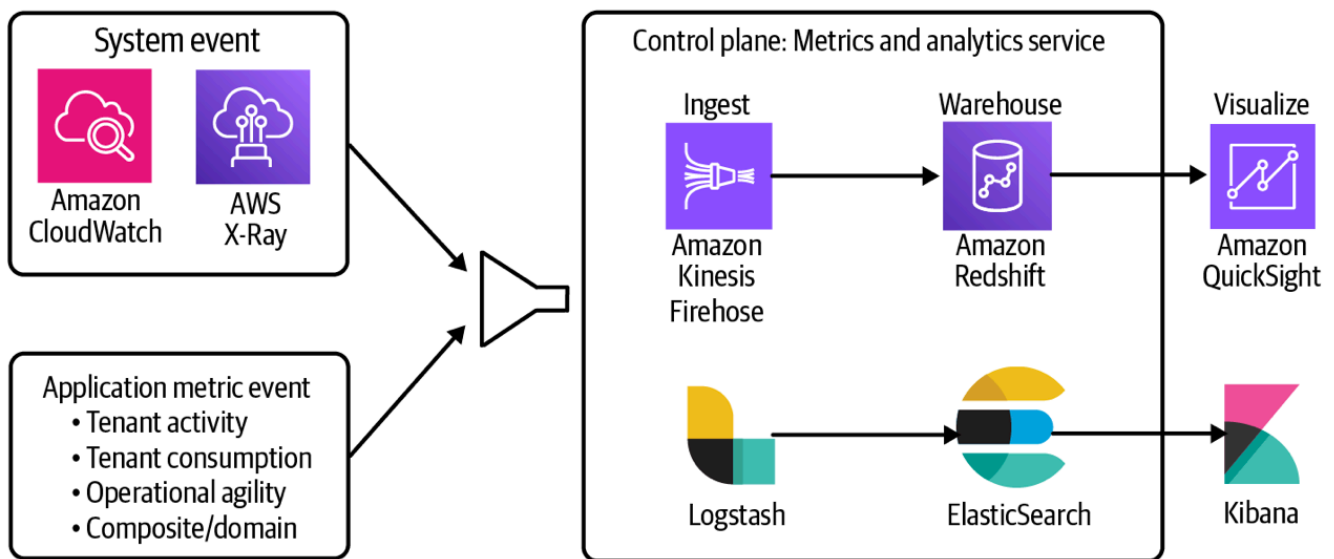


Figure 12-7. Metric event ingestion and aggregation

En la parte izquierda de este diagrama se muestran las diversas fuentes de datos de métricas. En la parte inferior izquierda se encuentran las diferentes categorías de datos de métricas que se instrumentarían en la aplicación SaaS.

También se representan los eventos "de sistema" para incorporar los otros eventos integrados que generarán los servicios de infraestructura. Aquí se capturan y publican los conceptos de métricas de línea base comunes (CPU, memoria, etc.) junto con las demás métricas.

Estos datos se publicarán en el servicio de métricas y analíticas del control plane, que incluirá herramientas y servicios para ingerir y agregar esta información (mostrados en el centro del diagrama).

En la parte inferior se muestra cómo Logstash podría utilizarse para ingerir los datos y publicarlos en Elasticsearch, un motor de búsqueda que puede emplearse para analizar los datos de métricas. Este conjunto de herramientas se combina con Kibana, que se utilizaría para construir los distintos paneles de control que permiten analizar los datos (en el extremo derecho).

Es importante señalar que los datos agregados aquí podrían utilizarse en múltiples contextos y por múltiples roles de la organización. **La clave es que estos datos de métricas no deben considerarse como propiedad exclusiva de los equipos técnicos.**

Como parte de este modelo, también habrá que decidir por cuánto tiempo deben conservarse estos datos. La vida útil de los datos estará determinada por las necesidades específicas del negocio.

! Construcción de una consola de operaciones con reconocimiento de tenant

Quiero explorar cómo llevar estas métricas y datos a la superficie, creando las herramientas que permitirán a los equipos construir una experiencia de gestión que responda a las exigencias particulares de una solución multi-tenant.

Existen innumerables herramientas listas para usar que permiten a los equipos operacionales ver logs y obtener información sobre las métricas principales de su sistema. Si bien estas herramientas pueden ofrecer valor en un entorno SaaS, por lo general no incluyen ninguna noción de contexto de tenant o de niveles como parte de su solución.

Imaginemos un escenario en el que la visión global del estado del sistema muestra "verde". En apariencia, todos los indicadores clave de salud sugieren que el sistema no presenta condiciones de rendimiento, escalabilidad o fallas que requieran atención. **AI**

mismo tiempo, llega la notificación de que un tenant en particular está reportando problemas de rendimiento.

Para poder atender este y otros escenarios operacionales con reconocimiento de tenant de manera eficaz, se necesitan herramientas y mecanismos que permitan interactuar con los datos operacionales a través del prisma de tenants y niveles individuales.

Activity type

- Number of requests
- Latency of calls
- Storage consumption
- SLA threshold violations
- Most active services
- Least active services

Tenant filters

- All tenants
- Tier:
- Tenant:

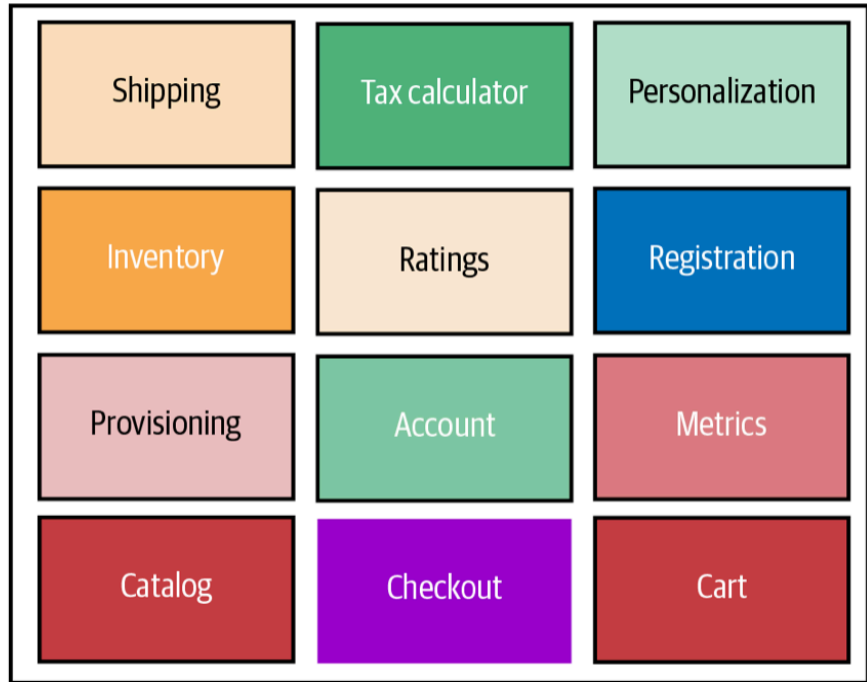


Figure 12-8. Adding tenant context to your operations console

Este ejemplo ofrece una visión centrada en microservicios de la experiencia operacional, utilizando un mapa de calor en el lado derecho del diagrama para ilustrar el estado de salud de los servicios individuales.

Sin esta capacidad de ver los datos operacionales con contexto de tenant, resultaría muy difícil identificar con precisión problemas específicos de un tenant. En un entorno operacional multi-tenant, se necesitan herramientas que permitan detectar y resolver estos problemas con rapidez, anticipándose a desafíos que podrían tener un efecto en cascada sobre múltiples tenants.

La Figura 12-9 ofrece una pequeña muestra de cómo la dimensión de tenancy podría reflejarse en los paneles de operaciones multi-tenant.

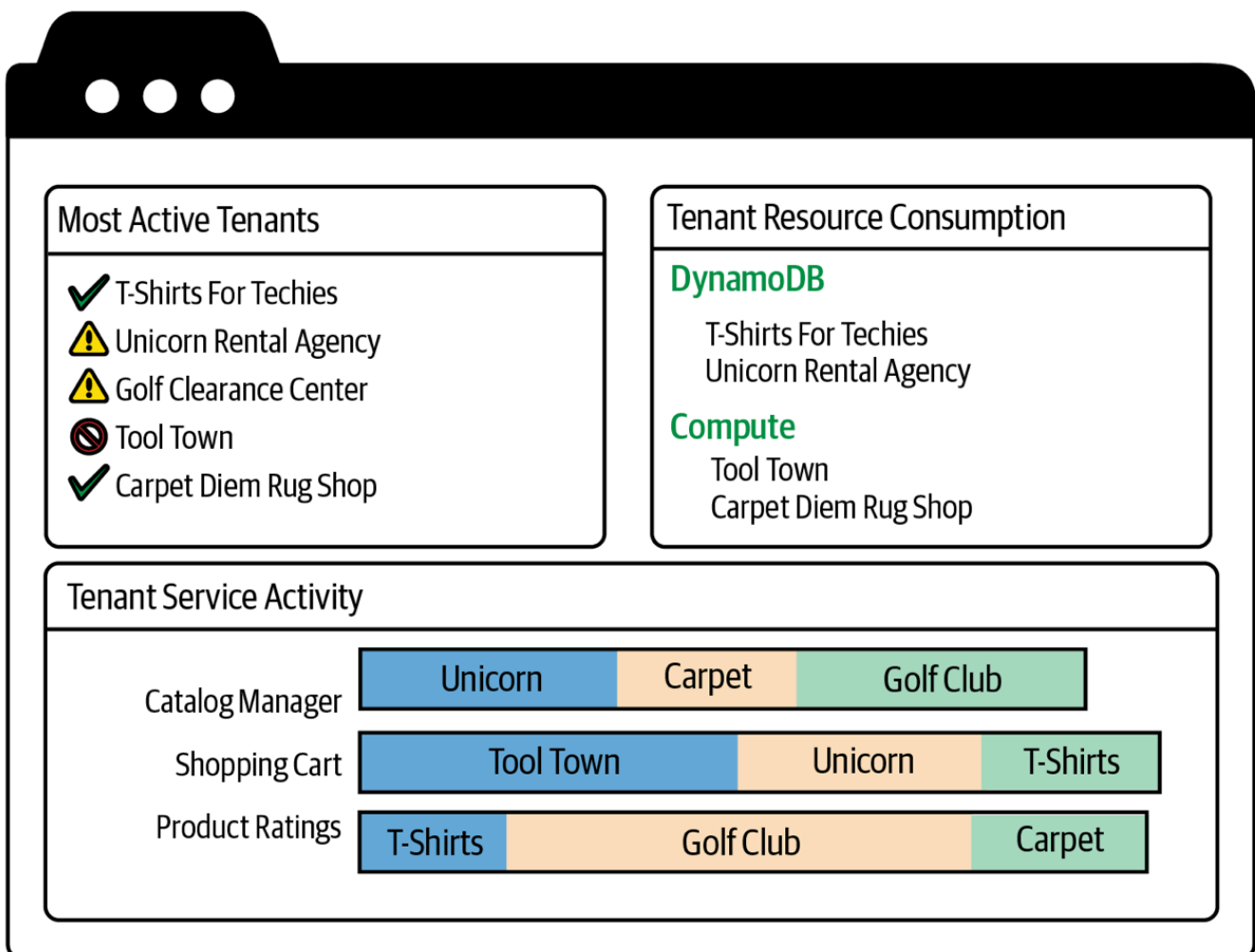


Figure 12-9. Surfacing tenant-aware operational insights

El objetivo fue destacar la necesidad de crear herramientas operacionales con reconocimiento de tenant que respondan a las necesidades particulares de los entornos multi-tenant. Intentar armar perspectivas operacionales revisando manualmente grandes volúmenes de datos simplemente no es viable en un entorno SaaS.

Se necesitan paneles de control y herramientas de analíticas que pongan el contexto de tenant y nivel en primer plano, proporcionando vistas cuidadosamente diseñadas que permitan evaluar y navegar los datos operacionales con reconocimiento de tenant de manera ágil.

Combinación de métricas de experiencia y técnicas

En un entorno multi-tenant, existe un espectro bastante amplio de datos que podrían usarse para enriquecer la experiencia operacional.

También hay métricas de negocio que se pueden estar midiendo (agilidad, time to value (tiempo para generar valor), etc.).

Veo valor en incorporar selectivamente algunas de estas métricas a una visión operacional más amplia del entorno. Llevar estos datos a la consola puede permitir establecer correlaciones naturales entre eventos de negocio y técnicos.

Poder ver estas métricas SaaS más amplias junto con las métricas clásicas de salud y actividad podría permitir identificar patrones que de otro modo pasarían desapercibidos. **Una nueva funcionalidad, por ejemplo, podría introducir un problema de rendimiento que esté afectando los tiempos de onboarding.**

Logs con reconocimiento de tenant

Son una parte esencial del modelo general de herramientas operacionales, y representan otra área donde resulta necesario acceder fácilmente a los logs con contexto de tenant y nivel.

Si los logs incluyen contexto de tenant, el siguiente desafío consiste en determinar cómo se desea soportar el acceso a estos logs. Algunos equipos construirán vistas de logs directamente en su consola, proporcionando a los usuarios la capacidad de filtrar y ver fácilmente la actividad de logs según criterios específicos de tenant o nivel.

Otros recurrirán a herramientas existentes listas para usar, diseñadas específicamente para analíticas de logs.

Estrategias proactivas

En un entorno multi-tenant donde se toman todas las medidas posibles para evitar interrupciones, los equipos suelen poner mayor énfasis en implementar estrategias operacionales proactivas que intenten detectar y resolver problemas mediante automatización y políticas.

Esto puede presentarse en forma de alertas que atraigan atención inmediata a condiciones que requieren intervención humana, o en forma de detección proactiva de degradación del rendimiento que ajuste de manera anticipada el escalado del entorno para limitar el impacto en los tenants.

Paneles de control por perfil de usuario

En la realidad, puede haber múltiples vistas de estos datos para los distintos roles que forman parte del negocio.

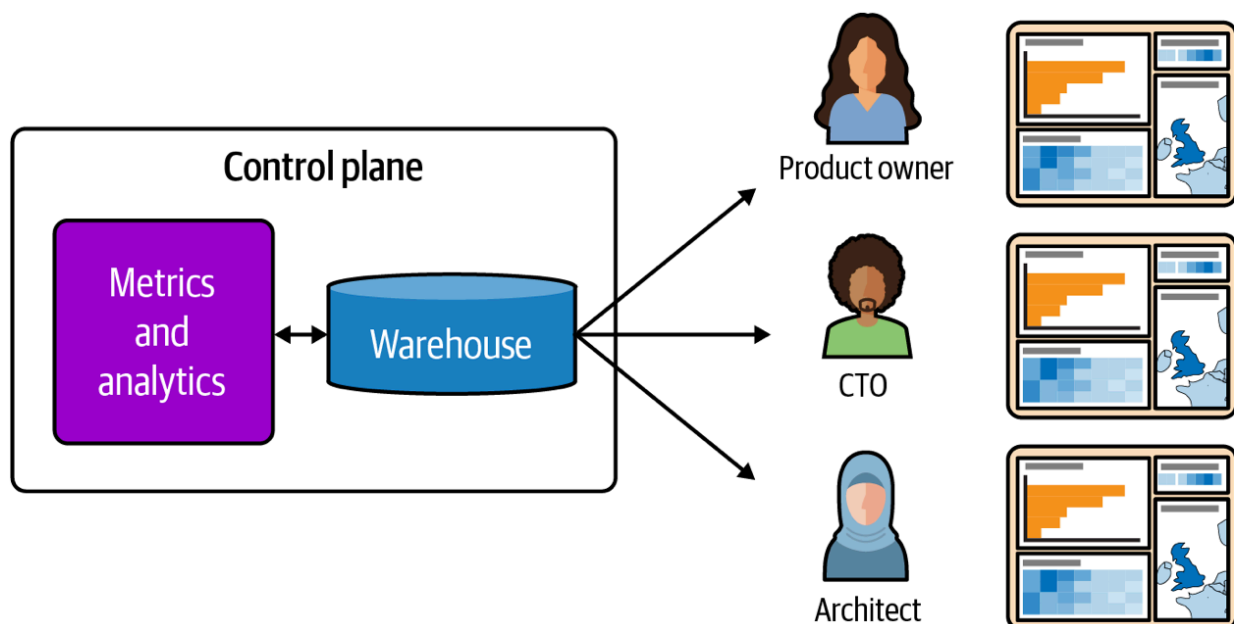


Figure 12-10. Metrics for multiple personas

En un mundo ideal, se vería a los distintos roles del negocio aprovechando activamente los datos de métricas del sistema, desarrollando sus propios paneles con sus propias vistas sobre las tendencias y patrones que aportan valor estratégico e información relevante para su área específica.

Esto juega un papel importante en la promoción de las responsabilidades operacionales compartidas de los distintos roles en una organización SaaS.

| Automatización del despliegue multi-tenant

La naturaleza, el alcance y el rol de la automatización de infraestructura en un entorno multi-tenant a menudo requieren que los equipos evolucionen su forma de pensar para soportar la combinación única de modelos de configuración y despliegue de SaaS.

La multi-tenancy puede requerir descomponer la automatización de manera diferente, separando partes de ella para alinearlas con los flujos y patrones que conforman la experiencia DevOps global de SaaS.

| Alcance de los despliegues

En los entornos SaaS, los equipos suelen buscar formas creativas de controlar cómo se despliegan funcionalidades y capacidades hacia los tenants. Aquí es donde entran en juego conceptos como los feature flags y los canary releases.

Los feature flags permiten habilitar o deshabilitar funcionalidades individuales dentro de un lanzamiento, y los canary releases permiten desplegar versiones de un servicio a un conjunto selecto de tenants.

Es fácil imaginar el valor que estos conceptos pueden ofrecer en un entorno SaaS. Con feature flags, un equipo puede activar selectivamente funcionalidades para niveles específicos de tenants, por ejemplo.

Los feature flags pueden representar un punto de tensión para algunas organizaciones. Algunos equipos los verán como una forma de crear funcionalidades a medida para tenants individuales. Esto puede convertirse en un camino resbaladizo.

Si pensamos en los objetivos más amplios del SaaS multi-tenant, estamos intentando deliberadamente alejarnos de cualquier noción de personalización por tenant. Si se termina con 100 tenants distintos, cada uno con sus propios feature flags personalizados, esto podría socavar los objetivos de agilidad y eficiencia de la organización.

Hay que evitar la tentación de ver los feature flags como una forma de eludir los principios del SaaS; el objetivo debe ser tener un sistema único disponible para todos los usuarios que no requiera soporte individualizado para gestionar las personalizaciones de cada tenant.

En el caso ideal, los feature flags se asignarían a nivel de tier para distinguir experiencias diferentes, y cualquier flag por tenant individual sería la excepción en lugar de la norma.

| Lanzamientos dirigidos

Por defecto, al desplegar nuevos lanzamientos en un entorno SaaS, las actualizaciones se envían a todos los tenants de una sola vez.

Para algunas organizaciones, esto puede generar cierta aprensión. Si se ha modificado una funcionalidad o añadido un nuevo flujo al sistema, toda la comunidad de tenants podría verse afectada negativamente si los cambios introducidos no son bien recibidos.

Aquí es donde los equipos buscan maneras de desplegar lanzamientos dirigidos, donde las actualizaciones se liberan a un subconjunto de tenants para evaluar la respuesta y el impacto de un nuevo lanzamiento.

En el mundo DevOps, esto se ha logrado a través de canary releases, donde se identifica un grupo selecto de tenants y se les despliega la actualización como forma de recolectar retroalimentación o evaluar impactos en el sistema sin afectar a todos los tenants.

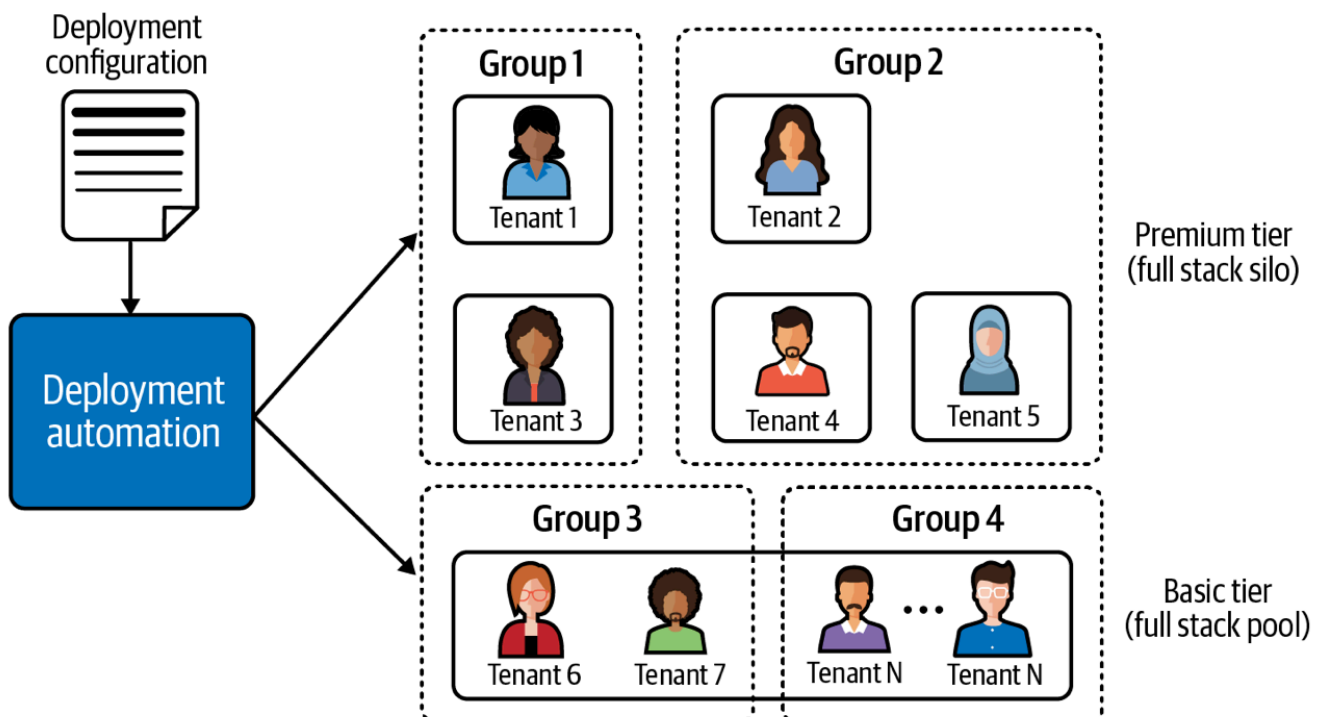


Figure 12-12. Targeted deployment in multi-tenant environments

A la izquierda se encuentran los marcadores de posición para la automatización del despliegue, donde podemos encontrar cualquier número de herramientas y mecanismos DevOps distintos que despliegan actualizaciones a los tenants.

Encaja perfectamente con la idea de desplegar actualizaciones de manera selectiva a un subconjunto de tenants. Estas agrupaciones, por ejemplo, se utilizarían para implementar una estrategia de canary release o de despliegue por oleadas, distribuyendo gradualmente el lanzamiento de cambios y limitando el alcance de posibles impactos operacionales.