



DEFINICIÓN DE LOS CUATRO ARQUETIPOS DEL EDGE Y SUS REQUISITOS TECNOLÓGICOS

Introducción

Durante los últimos años, el “Edge Computing” se ha convertido en una de las tendencias informáticas que más ha dado que hablar, y con razón. Grand Valley Research prevé una [tasa de crecimiento anual compuesto del 41 % para el Edge Computing](#) entre 2018 y 2025. Casi todos los sectores de la industria reconocen las limitaciones de sustentar a los usuarios y las tecnologías emergentes a través de infraestructuras informáticas centralizadas, de modo que están acercando el almacenamiento y los recursos informáticos a los usuarios y los dispositivos.

Este cambio es necesario debido a la mayor conectividad de los dispositivos y las personas, así como a los grandes volúmenes de datos que generan y consumen. De acuerdo con el [Cisco Visual Networking Index](#) el tráfico IP global fue de 1,2 zettabytes en 2016. Esta cifra casi se triplicará para 2021, llegando a alcanzar los 3,3 zettabytes. También para 2021, Cisco estima que el número de dispositivos conectados a redes IP equivaldrá al triple de la población mundial. Esto representa más de 23.000 millones de dispositivos conectados en tan solo tres años. [Otras empresas están realizando predicciones similares](#): para 2020, Gartner calcula 20.800 millones de dispositivos conectados, según IDC serán 28.100 millones y, de acuerdo con IHS Markit, se alcanzarán los 30.700 millones.

Una parte considerable de estos datos de IoT serán datos de sensores móviles que se deben transmitir mediante redes inalámbricas o móviles, no a través de conexiones de Internet cableadas, lo cual supone una gran carga para la infraestructura de redes móviles. [Se calcula que el tráfico IP móvil se multiplicará por siete para 2021](#), es decir, el doble de la velocidad de crecimiento en el tráfico IP fijo.

Los cambios en la infraestructura de cálculo y almacenamiento necesarios para sustentar el futuro inteligente y conectado, en especial localmente, serán profundos.

No obstante, si observa la información disponible en la actualidad sobre el Edge Computing, descubrirá que hay pocos recursos, en caso de existir, que proporcionen una visión exhaustiva del ecosistema del edge. Un análisis detallado del mercado revela una amplia gama de ejemplos de uso actuales y emergentes y, aunque comparten algunas similitudes gracias a la definición general del Edge Computing, también difieren de forma significativa.

Vertiv ha analizado los ejemplos de uso que conforman el ecosistema del edge a fin de comprender mejor estas diferencias y sus implicaciones para la infraestructura subyacente. Como resultado de este análisis, hemos identificado cuatro arquetipos principales para las aplicaciones del edge:

- Aplicaciones con uso intensivo de datos
- Aplicaciones sensibles a la latencia humana
- Aplicaciones sensibles a la latencia máquina-máquina
- Aplicaciones cruciales para la vida

Este artículo ofrece una descripción de cada arquetipo con los ejemplos de uso más destacados, además de una descripción general de los requisitos de conectividad con concentradores locales, urbanos y regionales, que representan la capa de transmisión y el núcleo del edge, en ocasiones diferenciados como Edge Computing (informática perimetral), Fog Computing (informática en la niebla) y Cloud Computing (informática en la nube).

Ejemplos de uso del Edge Computing

Para identificar los cuatro arquetipos, ha sido necesario comprender antes los ejemplos de uso de la tecnología del edge. El equipo de investigación de Vertiv ha identificado y revisado más de 100 ejemplos de uso de la tecnología del edge y ha reducido la lista inicial a los 24 que tienen un mayor impacto en la infraestructura informática para realizar un análisis más detallado.

Este análisis examina los requisitos de rendimiento de cada ejemplo de uso en términos de latencia, disponibilidad y crecimiento previsto, así como los requisitos de seguridad tales como la necesidad de cifrado, autenticación y cumplimiento normativo. También se ha evaluado la necesidad de integración con las aplicaciones existentes o heredadas y con otras fuentes de datos, así como el número de ubicaciones potenciales necesarias para sustentar el ejemplo de uso.

Y, lo que es más importante, el equipo ha estudiado las características de los datos correspondientes a los casos de uso y ha descubierto que las aplicaciones que actúan como base presentan siempre un conjunto de requisitos de carga de trabajo centrados en los datos además de los requisitos de disponibilidad y seguridad. Entre ellos están el volumen de datos, cómo se accede a ellos, los requisitos de transmisión de datos, la integridad de los datos y el análisis de los datos. Esta propuesta basada en los datos, filtrada según los requisitos de disponibilidad y seguridad, es crucial para entender y categorizar los requisitos de los distintos ejemplos de uso.

La Figura 1 incluye una lista de los 24 ejemplos de uso organizados por arquetipo.

El ecosistema del edge

APLICACIONES CON USO INTENSIVO DE DATOS	APLICACIONES SENSIBLES A LA LATENCIA MÁQUINA-MÁQUINA	APLICACIONES CRUCIALES PARA LA VIDA	APLICACIONES SENSIBLES A LA LATENCIA HUMANA
<ul style="list-style-type: none"> • Conectividad restringida • Smart Cities • Smart Factories • Smart Home/Building • Distribución de contenido HD • Informática de alto rendimiento • Realidad virtual • Digitalización de oil and gas 	<ul style="list-style-type: none"> • Smart Security • Smart Grid • Distribución de contenido de baja latencia • Mercado bursátil • Análisis en tiempo real • Simulaciones militares 	<ul style="list-style-type: none"> • Salud digital • Coches conectados/autónomos • Drones • Smart Transportation • Robots autónomos 	<ul style="list-style-type: none"> • Optimización web • Realidad aumentada • Smart Retail • Procesamiento del lenguaje natural

Figura 1: Arquetipos

Arquetipo 1: Aplicaciones con uso intensivo de datos

Ancho de banda	Latencia	Disponibilidad	Seguridad
Alto	Media	Alta	Media

El arquetipo de aplicaciones con uso intensivo de datos representa ejemplos de uso en los que la cantidad de datos provoca que no resulte práctica la transferencia a través de la red directamente a la nube, o bien desde la nube al punto de uso, debido al volumen de los datos, al coste o a problemas de ancho de banda.

Probablemente, el ejemplo más debatido de una aplicación del edge con un uso intensivo de datos sea el suministro de contenido de alta definición. [En 2016, el vídeo supuso el 73 % de todo el tráfico IP, y se espera que este porcentaje alcance el 82 % para 2021](#) ya que la transmisión de vídeo y la realidad virtual siguen en aumento. Los principales proveedores de contenido, como Amazon y Netflix, se están asociando de forma activa con proveedores de colocation a fin de ampliar sus redes de difusión y de acercar la transmisión de vídeo con un uso intensivo de datos a los usuarios para reducir los costes y la latencia.

Actualmente, [el 35 % del contenido al que accede un usuario de Internet norteamericano se envía desde el término municipal donde se encuentra](#). Se espera que este porcentaje aumente hasta el 51 % para 2021, ya que los proveedores de contenido siguen extendiendo sus redes hacia el edge. Sin embargo, esto solo representa la primera ola del Core-to-Edge Computing (informática del núcleo al edge). A medida que aumente la demanda de vídeo de alta definición, los concentradores locales respaldarán cada vez más los concentradores urbanos actuales a fin de reducir aún más los costes de ancho de banda y los problemas de latencia.

Otro buen ejemplo del arquetipo de aplicaciones con uso intensivo de datos es el uso de redes IoT para crear hogares, edificios, fábricas y ciudades inteligentes. Una encuesta realizada en 2018 por 451 Research y Vertiv determinó que, aunque solo el 33 % de las 700 organizaciones encuestadas había desplegado IoT de forma amplia, el 56 % indicó que al menos el 25 % de su capacidad informática admite IoT actualmente. A pesar de que IoT sigue estando en sus etapas iniciales, las organizaciones ya están intentando gestionar el volumen de datos que se genera.

Aquí, el reto es el contrario que en el caso de la difusión de contenido de alta definición. En lugar de acercar los datos a los usuarios, estas aplicaciones deben trasladar las grandes cantidades de datos generados por los dispositivos y los sistemas desde el origen a una ubicación central para su procesamiento. Esto requerirá la evolución de una arquitectura de red desde el edge al núcleo (edge-to-core).

IoT y el Internet de las cosas industriales (IIoT) representan una red de sensores que generan grandes volúmenes de datos por hora. Estos datos admiten un bucle de “detección-inferencia-reacción” que permite la visibilidad y el control de cualquier cosa, desde electrodomésticos a equipos industriales. Solo se transmite un subconjunto de estos datos a un centro de datos local, regional o en la nube para continuar con su procesamiento, lo que implica que en el extremo (edge) se requerirán cantidades ingentes de cálculos a fin de permitir que los dispositivos y los sistemas tomen decisiones y actúen sobre los datos proporcionados por los sensores.

La aplicación más simple de estas, el hogar inteligente, debe dar cabida a varios dispositivos y sistemas con un uso intensivo de datos, incluidos el entretenimiento, la seguridad y los sistemas de aire acondicionado y ventilación.

Aplicaciones con uso intensivo de datos

De acuerdo con IHS Markit, [el mercado mundial de los dispositivos domésticos conectados aumentará de más de 100 millones de unidades en 2017 a aproximadamente 600 millones de unidades en 2021](#).

Las ciudades y las fábricas inteligentes presentan los mismos retos en cuanto a los datos que resultan inherentes a los hogares inteligentes, pero en mayor medida. Muchas ciudades ya están realizando pruebas piloto o evaluando tecnología para ciudades inteligentes a fin de mejorar el flujo del tráfico, apoyar a los servicios de emergencia y reducir costes.

Las fábricas inteligentes, que sacan partido de la convergencia de IoT, los sistemas ciberfísicos y la informática en la nube para permitir a los fabricantes emplear datos en tiempo real a fin de aumentar el rendimiento, reducir costes y adaptarse a los cambios en la demanda, se están promocionando como la próxima revolución industrial. De acuerdo con McKinsey, las fábricas y otros entornos de producción tienen el potencial de provocar el mayor impacto económico desde la aplicación de IoT. Esta empresa ha pronosticado que IIoT generará un [valor económico de entre 1,2 y 3,7 billones de USD](#) para 2025. Este valor tendrá su origen en los nuevos rendimientos energéticos, la productividad laboral, la optimización de inventarios y una mayor seguridad de los trabajadores. Sin embargo, llevar esto a la práctica requerirá una infraestructura local potente.

En el sector del oil & gas, la digitalización ya ha permitido un amplio aumento en el rendimiento de los procesos de prospección y extracción, pero también ha supuesto grandes retos en cuanto a la gestión de los datos. Una única plataforma de perforación puede generar terabytes de datos cada día.

Otros ejemplos de uso que corresponden al arquetipo de las aplicaciones con uso intensivo de datos incluyen la realidad virtual, la informática de alto rendimiento y los entornos con conectividad restringida, tales como las áreas donde se producen operaciones de recuperación después de un desastre natural o un ciberataque.

Lo que tienen en común todos estos ejemplos de uso es la necesidad de trasladar grandes volúmenes de datos a los usuarios donde puedan consumirlos, o bien desde los dispositivos y sistemas donde se generan a un repositorio central.

Arquetipo 2: Aplicaciones sensibles a la latencia humana

Ancho de banda	Latencia	Disponibilidad	Seguridad
Medio	Alta	Media	Media

El arquetipo de las aplicaciones sensibles a la latencia humana abarca ejemplos de uso en los que los servicios se optimizan para el consumo humano. Como su propio nombre indica, la velocidad es la característica que define este arquetipo.

El desafío que supone la latencia humana se puede comprobar en la optimización de la experiencia del cliente. En aplicaciones como el comercio electrónico, la velocidad tiene un impacto directo sobre la experiencia del usuario; los sitios web optimizados en cuanto a velocidad mediante la infraestructura local presentan un aumento tanto en visitas como en ventas.

Aplicaciones sensibles a la latencia humana

Google ha descubierto que un retardo adicional de 500 milisegundos en el tiempo de respuesta de las páginas tuvo como resultado un descenso del 20 % en el tráfico, mientras que Yahoo observó que un retardo de 400 milisegundos provocó una reducción del tráfico de entre el 5 % y el 9 %.

Este efecto se extiende también al procesamiento de los pagos. Amazon detectó que un retardo de 10 milisegundos en el procesamiento de los pagos provocó un descenso del 1 % en los ingresos retenidos. La aprobación centralizada mediante contraseña tardó de media siete segundos. El cambio al procesamiento local redujo este tiempo a 600 milisegundos, lo que supone una mejora de 6.400 milisegundos con cada 100 milisegundos y resulta potencialmente en un 1 % adicional de ingresos retenidos.

Otro ejemplo emergente de una aplicación sensible a la latencia humana es el procesamiento del lenguaje natural. Probablemente la voz será la principal forma de interacción con las aplicaciones informáticas cotidianas en el futuro. El

procesamiento del lenguaje natural para Alexa y Siri se realiza actualmente en la nube. No obstante, a medida que aumente el volumen de usuarios, aplicaciones e idiomas admitidos, será necesario migrar estas capacidades más cerca de los usuarios.

Otros ejemplos de uso sensibles a la latencia humana identificados son la venta minorista inteligente, como las tiendas sin caja Amazon Go, y las tecnologías de inmersión tales como la realidad aumentada, donde pequeños retardos de latencia pueden significar la diferencia entre la diversión y la náusea. En todos los casos, los retardos en el suministro de los datos tienen un efecto directo en la experiencia tecnológica del usuario, como ocurre con el procesamiento del lenguaje y la realidad aumentada, o bien en las ventas y los beneficios de un vendedor minorista, como en el caso de la optimización de sitios web y la venta minorista inteligente. A medida que aumenten los ejemplos de uso, aumentará también la necesidad de concentradores locales para el procesamiento de datos.

Arquetipo 3: Aplicaciones sensibles a la latencia máquina-máquina

Ancho de banda	Latencia	Disponibilidad	Seguridad
Medio	Alta	Alta	Alta

El arquetipo de las aplicaciones sensibles a la latencia máquina-máquina abarca ejemplos de uso en los que los servicios se optimizan para el consumo entre máquinas. Como las máquinas pueden procesar los datos mucho más rápido que los humanos, la velocidad es la característica que define este arquetipo. Las consecuencias de no suministrar los datos a las velocidades necesarias pueden ser incluso mayores en este caso que en el del arquetipo de las aplicaciones sensibles a la latencia humana.

Por ejemplo, los sistemas empleados en las transacciones financieras automatizadas, como en el caso de los productos básicos y el comercio bursátil, son sensibles a la latencia. En estos casos, los precios pueden cambiar en cuestión de milisegundos y los sistemas que no tienen los datos más recientes cuando son necesarios no pueden optimizar las transacciones, convirtiendo así los beneficios potenciales en pérdidas.

Aplicaciones sensibles a la latencia máquina-máquina

De acuerdo con un estudio de Tabb Group, un corredor podría perder **hasta cuatro millones de USD en ingresos por milisegundo** si su plataforma de comercio electrónico estuviera cinco milisegundos por detrás de la competencia.

La tecnología de red eléctrica inteligente también corresponde a este arquetipo. Esta tecnología se está implantando en la red de distribución eléctrica para equilibrar automáticamente la oferta y la demanda, además de gestionar el uso de la electricidad de una manera sostenible, fiable y económica. Permite la autorreparación de las redes de distribución, la optimización de costes y la gestión de las fuentes de energía intermitentes, siempre que los datos correctos estén disponibles en el momento preciso.

Otras aplicaciones sensibles a la latencia máquina-máquina son los sistemas de seguridad inteligentes que dependen del reconocimiento de imágenes, las simulaciones militares y el análisis en tiempo real.

Arquetipo 4: Aplicaciones cruciales para la vida

Ancho de banda	Latencia	Disponibilidad	Seguridad
Medio	Alta	Alta	Alta

El arquetipo de las aplicaciones cruciales para la vida abarca ejemplos de uso que afectan directamente a la salud y la seguridad humanas. En estos casos, la velocidad y la fiabilidad son fundamentales.

Probablemente los mejores ejemplos del arquetipo de las aplicaciones cruciales para la vida sean los vehículos autónomos y los drones, que proporcionan grandes beneficios cuando funcionan según su diseño; no obstante, si toman malas decisiones, pueden poner en peligro la vida humana.

Los vehículos autónomos han avanzado más rápido de lo que muchos esperaban, y varias empresas tecnológicas y de automoción ya están realizando pruebas de manera activa con estos vehículos en la actualidad. La mayoría de ellos cuentan con una persona en el asiento del conductor preparada para anular los controles automáticos si se producen problemas a fin de minimizar el riesgo para la salud humana. Pero, en un futuro cercano, los vehículos de entrega y los sistemas de transporte sin conductor estarán en circulación. Si estos sistemas no cuentan con los datos requeridos cuando los necesitan, las consecuencias podrían ser desastrosas.

Lo mismo ocurre con los drones. No sería descabellado pensar en un futuro en el que cientos de drones de entrega sobrevuelan una ciudad en un momento dado.

Aplicaciones cruciales para la vida

Grandes compañías de comercio electrónico y paquetería, tales como Amazon y DHL, ya están experimentando con drones para la entrega de paquetes.

El aumento del uso de tecnología en el cuidado de la salud también representa el arquetipo de las aplicaciones cruciales para la vida. Los historiales médicos electrónicos, la cibermedicina, la medicina personalizada (genotipado) y los dispositivos de autosupervisión están remodelando la atención sanitaria y generando grandes volúmenes de datos.

Otros ejemplos pueden ser el transporte inteligente y los robots autónomos. Los sectores del transporte y la logística están examinando soluciones centradas en los datos para aumentar la seguridad del conductor y los pasajeros, el rendimiento del combustible y la gestión de activos. La tecnología en este ámbito incluirá sistemas de transporte, gestión de flotas y telemática inteligentes, sistemas de guía y control, entretenimiento de los pasajeros y aplicaciones de comercio, reservas, sistemas de peaje y venta de billetes, y sistemas de seguridad y vigilancia.

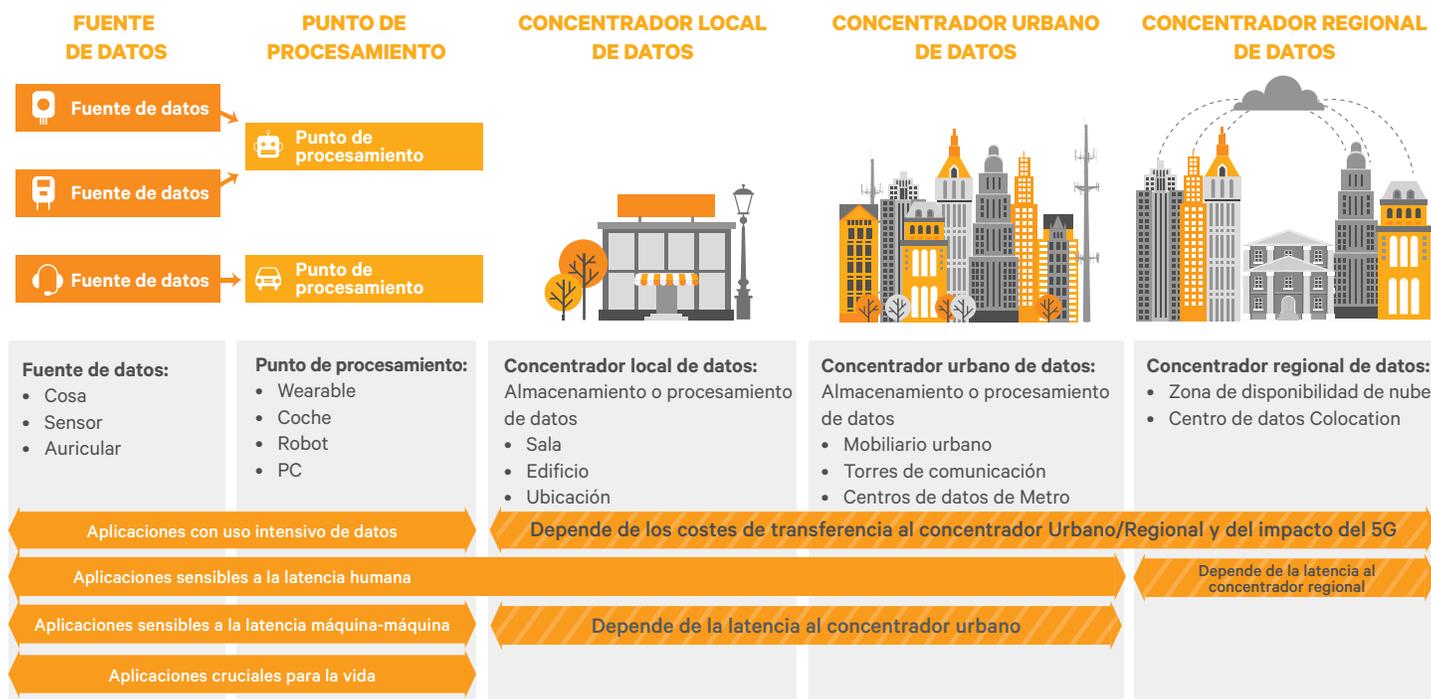
Requisitos tecnológicos de los concentradores locales y regionales

La infraestructura necesaria para respaldar estos ejemplos de uso actuales y establecidos consta de cuatro capas de almacenamiento y cálculo además de la infraestructura de comunicaciones necesaria para mover los datos entre las capas.

En el origen suele encontrarse el dispositivo que genera o consume los datos y un nodo final de procesamiento. El dispositivo podría ser un sensor que supervisa cualquier cosa, como el estado de encendido de una bombilla, el acceso a una puerta, la temperatura de una habitación o cualquier otra información deseada. El nodo final de procesamiento puede ser tan simple como el PC o la tablet a los que un consumidor está transmitiendo vídeo, o bien podrían ser los microprocesadores incluidos en automóviles, robots o dispositivos móviles. Estos componentes dependen de las aplicaciones y suele diseñarlas el fabricante de equipos o se reacondicionan para los dispositivos existentes.

Todos los arquetipos, excepto el de las aplicaciones cruciales para la vida, se podrían reemplazar en el concentrador de datos local según la aplicación. El concentrador de datos local proporciona almacenamiento y procesamiento muy cerca del origen. En algunos casos, el concentrador local puede ser un centro de datos independiente. Lo más habitual es que se trate de un sistema de tipo "in the rack/row" que proporciona 30-300 kW de capacidad en un armario integrado susceptible de instalación en cualquier entorno.

Estos sistemas de armarios de tipo "in the rack/row" integran comunicación, cálculo y almacenamiento con una protección de la alimentación, controles medioambientales y seguridad



física apropiados. En el caso de los arquetipos que requieren un alto grado de disponibilidad, tales como las aplicaciones sensibles a la latencia máquina-máquina y las aplicaciones cruciales para la vida, el concentrador local debería incluir sistemas de alimentación de reserva redundantes y estar equipado para permitir la gestión y la supervisión remotas. Muchos ejemplos de uso también requerirán el cifrado de los datos y otras funciones de seguridad en el concentrador local.

Todos los arquetipos excepto las aplicaciones cruciales para la vida, el concentrador urbano, el concentrador regional o ambos se podrían usar para respaldar los ejemplos de uso que dependan de los costes de transferencia de datos, el ancho de banda que permite el despliegue de 5G y la latencia en relación con la ubicación física del centro de datos. El concentrador urbano emplea la infraestructura de telecomunicaciones establecida para proporcionar capacidades de cálculo e infraestructura. Se diseñará de acuerdo con los estándares de telecomunicaciones, incluidos la alimentación de CC y el freecooling, y admitirá una gama mucho más amplia de temperaturas y humedad de lo habitual en los centros de datos tradicionales. Es probable que el concentrador regional sea un centro de datos en la nube o de colocation que opere en la misma región que el concentrador urbano local.

En los casos de los concentradores urbanos y regionales, se deben considerar diseños modulares susceptibles de ampliación fácil más allá de las especificaciones de diseño

iniciales a fin de hacer frente a los picos inesperados de demanda. Estas instalaciones deberían diseñarse también para que se puedan ampliar en términos de densidad. Las aplicaciones con un uso intensivo de imágenes, tales como la realidad virtual, así como las aplicaciones con un uso intensivo de procesamiento, tales como el análisis y el aprendizaje automático, probablemente requerirán densidades de rack superiores a la especificación de diseño habitual de 10 kW. Prácticamente en todos los casos, estos concentradores deben proporcionar un nivel de redundancia y seguridad igual o superior al del concentrador local.

El futuro próximo

Al identificar las necesidades de carga de trabajo para los 24 ejemplos de uso analizados, surgieron cuatro arquetipos principales que pueden guiar las decisiones en relación con los requisitos de infraestructura y configuración para los ejemplos de uso examinados, así como los que surgirán en los próximos años. Vertiv ampliará este trabajo inicial con arquetipos a fin de continuar definiendo requisitos tecnológicos y configuraciones específicos para cada arquetipo.

