

II - Servicios de nube

El presente lineamiento representa una guía de referencia previa para la elaboración de un requerimiento técnico abierto destinado a contratar servicios de nube.

La idea principal de las siguientes recomendaciones y prácticas es facilitar la elaboración del requerimiento, promoviendo una adecuada provisión de servicios y tendiendo a disminuir eventuales problemas al momento de la evaluación y/o adjudicación de las propuestas de los oferentes.

Servicios de nube

El servicio de computación en Nube, también denominado "*Cloud Computing*", refiere a un modelo de gestión de TI que permite el acceso, bajo demanda y a través de Internet, a un conjunto de recursos compartidos y configurables (tales como redes, servidores, capacidad de almacenamiento, aplicaciones y servicios) que pueden ser rápidamente asignados y liberados con una mínima gestión por parte del proveedor del servicio.

El avance tecnológico y la evolución de aspectos tales como la capacidad de procesamiento, la potencia de la virtualización, el ahorro de energía, la velocidad de las conexiones, la capacidad de almacenamiento, entre otras, permitieron construir infraestructuras que puedan brindar servicios de nube. Esto último fomenta un cambio en el consumo de la tecnología trasladando el foco desde los bienes hacia los servicios.

El Instituto Nacional de Estándares y Tecnología (NIST por sus siglas en inglés) de los Estados Unidos en la publicación especial SP 800-145, determinó una serie de características esenciales que caracterizan a los servicios de Nube:

1. Servicio a demanda, el usuario puede acceder a capacidades de forma automática a medida que las va requiriendo sin necesidad de una interacción con su proveedor.
2. Amplio acceso a la red, los recursos son accesibles por medio de la red y de mecanismos estándar que son utilizados por una amplia variedad de dispositivos.
3. Puesta en común de recursos, los recursos (almacenamiento, memoria, ancho de banda, capacidad de procesamiento, máquinas virtuales, etc.) de los proveedores son compartidos por múltiples usuarios, a los que se van asignando capacidades de forma dinámica según sus peticiones.
4. Elasticidad, los recursos se asignan y liberan rápidamente, muchas veces de forma automática, lo que da al usuario la posibilidad de contar con recursos ilimitados y disponibles en todo momento.
5. Servicio medido, el proveedor es capaz de medir el servicio efectivamente entregado lo que posibilita el pago por el uso real de los servicios.

Existen, asimismo, tres tipos principales de entornos en la Nube, también conocidos como modelos de implementación en la Nube, según sus requisitos específicos:

1. Nube Pública - Un entorno de Nube pública es propiedad de un proveedor de Nube subcontratado y es accesible para muchas organizaciones a través de Internet en un modelo de pago por uso. Este modelo de implementación proporciona servicios e infraestructura a empresas que desean ahorrar dinero en costos operativos de TI de una forma rápida y sencilla.
2. Nube Privada - Este modelo de implementación en la Nube es una infraestructura a medida y propiedad de una sola empresa. Ofrece un entorno más controlado y el acceso a los recursos es centralizado dentro de la organización. Este modelo se puede alojar externamente o se puede administrar internamente. Aunque el alojamiento es más costoso por su autonomía, ofrece un mayor nivel de seguridad.
3. Nube Híbrida - Es la combinación de los anteriores modelos (Pública y Privada), un modelo de Nube híbrida proporciona una solución de TI más personalizada que cumple con los diferentes requisitos específicos (acceso, seguridad, políticos, comerciales, etc.).

A su vez, existen tres modelos de servicios que se montan en la Nube: el uso del término *como servicio (as a -cloud- service)* es un sufijo que describe la capacidad de cómputo de dar cumplimiento a las cinco características de la computación en la nube.

1. Software como servicio (SaaS – *Software as a Service*), la provisión de este servicio permite al usuario utilizar las aplicaciones del proveedor que se encuentran corriendo en infraestructura de Nube; puede accederse a las aplicaciones desde distintos dispositivos a través de una interfaz de cliente o de programa. El usuario no gestiona ni administra la infraestructura de Nube sobre la que se está montando el software ni sobre las redes, servidores, sistemas operativos, almacenamiento o las capacidades individuales de la aplicación; con la potencial excepción de poder configurar los ajustes.
2. Infraestructura como servicio (IaaS – *Infrastructure as a Service*), el servicio contempla la provisión de procesamiento, almacenamiento, redes y otros recursos informáticos fundamentales sobre los que el usuario podrá desplegar y ejecutar su software, que puede incluir sistemas operativos y aplicaciones. El usuario no gestiona ni administra la infraestructura de Nube, pero tiene control sobre los sistemas operativos, el almacenamiento y las aplicaciones desplegadas, y posiblemente acceso limitado a algunos elementos de la red.
3. Plataforma como servicio (PaaS – *Platform as a Service*), se provee al usuario con la capacidad de desplegar sobre la infraestructura de Nube aplicaciones creadas específicamente para el consumidor o adquiridas, creadas usando

lenguajes de programación, librerías, servicios y herramientas cuyo soporte brinda el proveedor. El usuario no gestiona ni administra la infraestructura de Nube incluyendo las redes, servidores, sistemas operativos o almacenamiento, pero tiene control sobre las aplicaciones desplegadas y los ajustes para el entorno de hosting de las aplicaciones.

Elaboración del requerimiento técnico abierto para servicios de Nube

Lo primero que se debe tener en cuenta es que los servicios de nube se contratan como servicios, es decir, el objeto de la contratación siempre será “servicios”. Sobre esto último es importante destacar que cuando se contrata un servicio, lo que se transfiere al adjudicatario es únicamente la ejecución de un conjunto de tareas, prestaciones y demás acciones que conforman el alcance del servicio. En ningún caso, y bajo ninguna circunstancia se transfiere la responsabilidad en la continuidad, disponibilidad y los resultados esperados para dicho servicio.

Estructura del pliego. Es importante remarcar que, al contratar servicios de nube, se deben incluir los tres componentes básicos detallados para servicios en la *Guía para la incorporación de nuevas tecnologías en el Sector Público Nacional*, es decir, alcance, disponibilidad y penalidades específicas.

Alcance. En el alcance se definen todas las prestaciones requeridas para el servicio, para el caso de los servicios de nube se recomienda definir e incluir los siguientes puntos según se ajusten a las necesidades del organismo:

Servicios requeridos. Se pueden requerir cualquiera de los tipos de servicios antes mencionados, SaaS, PaaS, e IaaS.

Para la solicitud de servicios tipo SaaS, se debe recordar que, si bien cambia la modalidad de prestación, dado que en este caso no sólo se incluye la utilización del software, sino que además se provee el derecho de uso de la infraestructura tecnológica necesaria para su correcto funcionamiento, la descripción del software en cuestión se deberá realizar a través de la enumeración de las necesidades técnico funcionales relevadas en los usuarios, con el fin de elaborar un requerimiento técnico abierto.

Para la solicitud de servicios tipo PaaS y/o IaaS, aparece la posibilidad de definir otras variables que van más allá de la utilización de un aplicativo. En consecuencia, los organismos deberán verificar que el servicio que se describa en el requerimiento cuente con la capacidad de aprovisionar la infraestructura necesaria (máquinas virtuales, contenedores, bases de datos relacionales y no relacionales, sistemas de archivos, componentes de seguridad, y demás componentes) mediante distintas herramientas digitales que permitan realizar

las altas, bajas y modificaciones necesarias para definir la infraestructura y/o plataforma requerida, su dimensionamiento, elasticidad y escalabilidad según la demanda.

Cuentas de los usuarios. Los organismos deberán verificar que el servicio que se describe en el requerimiento cuente con la capacidad de gestionar las cuentas de los usuarios que acceden a los servicios de nube, mediante las herramientas digitales necesarias que permitan realizar altas, bajas y modificaciones de permisos para organizar y otorgar distintos niveles de acceso a los recursos contratados, según el perfil de usuario que se haya definido.

Estadísticas y reportes de uso. Los organismos deberán verificar que el servicio que se describe en el requerimiento cuente con la capacidad de generar estadísticas de uso de los recursos contratados, mediante herramientas digitales que generen informes, reportes, y estadísticas históricas de uso, con la granularidad de los componentes involucrados en cada uno de los recursos utilizados, con el objetivo de monitorear, según sea el caso, el uso de la infraestructura, plataforma, y/o software contratado, para prever futuras optimizaciones y/o escalamientos.

Usuarios y sistemas que acceden al servicio de nube. Se deberán describir en el requerimiento las funcionalidades que permitan configurar las interfaces para que los distintos usuarios y/o sistemas que accedan al servicio en la nube puedan interoperar sin dificultad para llevar a cabo el intercambio de información demandado por dichos usuarios y/o sistemas.

Redundancia (para obtención de alta disponibilidad). Los organismos deberán verificar que el servicio que se describe en el requerimiento cuente con la funcionalidad que permita habilitar la redundancia para aquellos servicios que sean definidos como críticos. Adicionalmente, todos los servicios que se consideren críticos deberán contar con la capacidad de ser redundados a fin de alcanzar una disponibilidad superior a la ofrecida por los recursos básicos propios del proveedor de nube.

Elasticidad del aprovisionamiento (escalabilidad a demanda). Los organismos deberán verificar que el servicio que se describe en el requerimiento cuente con la capacidad de configurar la elasticidad (escalabilidad a demanda) de los recursos de software, plataforma o infraestructura. En el primer caso a través de una solicitud, y en los últimos dos casos, mediante distintas herramientas digitales que permitan parametrizar y automatizar el aprovisionamiento de distintos recursos de cómputo, almacenamiento, networking, bases de datos, etc., aumentando o reduciendo los recursos disponibles (dentro de ciertos

parámetros configurables por el usuario), con el fin de dar una respuesta adecuada a los cambios imprevistos en la demanda.

Para los servicios de PaaS y de IaaS, se deberá verificar que la elasticidad de los recursos sea configurable para escalar tanto en forma horizontal como vertical, esto es, ya sea que el cambio en la demanda provenga de la capacidad de (cómputo, tráfico de red, almacenamiento, etc.) las aplicaciones, o bien, de la concurrencia de los usuarios.

Infraestructura como código. En caso de requerir dicha funcionalidad, se deberá verificar que el servicio que se describe en el requerimiento permite aprovisionar de manera customizada recursos de infraestructura mediante lenguajes de programación y/o archivos de procesamiento por lotes, con el objetivo de modelar y/o aprovisionar de una manera automatizada, todos los recursos necesarios para las aplicaciones, sin importar la región o la cuenta donde se ejecuten.

Balanceo de carga. En caso de requerir dicha funcionalidad, se deberá verificar que el servicio que se describe en el requerimiento permita el balanceo de carga configurando, mediante herramientas digitales, la distribución automática del tráfico de aplicaciones entrantes a través de varios orígenes. Opcionalmente, y en caso de requerirse, deberán detallarse propiedades de elasticidad, esto es, que el balanceo se pueda configurar, de modo que crezca según la demanda y el consumo de las aplicaciones entrantes.

Cronograma de activación de recursos. En caso de requerir dicha funcionalidad, se deberá verificar que el servicio que se describe en el requerimiento permita la definición de un cronograma de activación de recursos mediante el cual se puedan definir rangos o períodos de actividad, configurando los días y horarios de consumo de uno o más recursos específicos, para que de manera automática se activen únicamente en los días y horarios establecidos, sin perjuicio de que, en los períodos de actividad, se pueda configurar adicionalmente una capacidad de gestión de elasticidad manual o automática.

Reserva anticipada de recursos. En caso de requerir dicha funcionalidad, se deberá verificar que el servicio que se describe en el requerimiento cuente con la capacidad de reservar recursos con acceso a distintas facilidades de pago y/o descuentos, para aquellos consumos fijos que se hayan detectado en base a las predicciones resultantes de las estadísticas y reportes de uso, con el fin de lograr la optimización de costos del servicio de nube.

Reserva anticipada de consumo. En caso de requerir dicha funcionalidad, se deberá verificar que el servicio que se describe en el requerimiento incluya la

capacidad de reservar crédito de consumo con acceso a distintas facilidades de pago y/o descuentos, con el fin de lograr la optimización de costos del servicio de nube. Esto se aplicaría, en los casos en que existan consumos fijos que se hayan detectado en base a las predicciones resultantes de las estadísticas y reportes de uso. El crédito disponible debe poder ser transferido a otra cuenta del organismo dentro del proveedor, usado como crédito para reserva de recursos, o transferido como crédito al mes siguiente.

Pruebas de estrés de la arquitectura. En caso de requerir esta funcionalidad, se deberá verificar que se encuentren disponibles herramientas que simulen sobrecargas en el acceso y/o procesamiento de la infraestructura y/o plataforma definida, para establecer distintos niveles de exigencia ante escenarios de alta demanda, con el fin de analizar la capacidad de respuesta del diseño de la arquitectura definida en el servicio de infraestructura o plataforma a contratar.

Este tipo de funcionalidades pueden ser utilizadas tanto para la realización de pruebas de concepto que se ejecuten como mecanismos de comparación de las capacidades de los distintos proveedores de servicios de nube, así como también para que, una vez adjudicado el servicio, se pueda evaluar la capacidad de respuesta del diseño ante ciertas estimaciones de crecimiento en la demanda.

Asistencia técnica. En caso de requerir dicha prestación, se deberá verificar que el servicio que se describe en el requerimiento permita definir y asignar un tipo de plan de soporte técnico acorde a las necesidades que el organismo crea conveniente, en función de los distintos niveles de criticidad que se hayan definido, y con capacidad de exigir cierta disponibilidad. El servicio de asistencia técnica puede ser contratado directamente al proveedor del servicio de nube, o a un socio oficial del proveedor (*partner*). En el primer caso, la propuesta de los distintos niveles de soporte deberá estar definida en el dominio público de modo previo a la adjudicación. En el segundo caso, deberán definirse los requerimientos funcionales estándares que cubran las necesidades de servicio técnico del organismo, sin que éstas restrinjan innecesariamente el espectro de oferentes.

Auditoría de los recursos. Los organismos deberán verificar que el servicio que se describe en el requerimiento, cuente con distintas funcionalidades para la consulta del registro y la gestión de logs, así como distintas herramientas similares para poder realizar la trazabilidad de las acciones e incidentes que hayan ocurrido en todos los recursos de la infraestructura y/o plataforma que se hayan contratado.

Entornos de producción, desarrollo y testing

- **PRODUCCIÓN:** En caso de requerir dicha prestación, se deberá verificar que el servicio que se describe en el requerimiento posea la funcionalidad de gestión de entornos productivos, mediante distintas herramientas digitales, las cuales permitan realizar las altas, bajas y modificaciones necesarias para brindar la continuidad y disponibilidad de una o más aplicaciones operativas que pudieran estar ejecutándose sobre dichos entornos.
- **DESARROLLO:** En caso de requerir esta prestación, se deberá encontrar disponible la capacidad de gestionar el servicio de plataforma para entornos de desarrollo que permita, por ejemplo, escribir código web, de análisis, de extracción de datos de formularios, y realizar *debugging* o *testing* de las aplicaciones, mediante distintas herramientas digitales que permitan el desarrollo colaborativo entre programadores.
- **TESTING:** En caso de requerir esta prestación, se deberá encontrar disponible la capacidad de gestionar entornos de testing, incluyendo distintas herramientas digitales que permitan realizar las pruebas necesarias fuera del entorno de producción, para asegurar los niveles de calidad exigidos, previo a la puesta en producción de las aplicaciones que se encuentren en desarrollo.

Disponibilidad. La disponibilidad es un aspecto crítico en los servicios de nube dado que, si no se asegura un porcentaje mínimo mensual, uno o más servicios que el organismo brinde a sus destinatarios podrían ser afectados.

Los organismos deberán encontrar disponibles (en dominio público), los porcentajes de disponibilidad mensual del servicio general contratado y de cada uno de los servicios y recursos contratados, al igual que el porcentaje de acreditación a aplicar como penalidad ante eventuales incumplimientos de la disponibilidad que oportunamente se seleccione.

Los rangos de disponibilidad recomendados son los siguientes:

- **CUALQUIER SERVICIO:** En ningún caso el porcentaje de disponibilidad mensual (para cualquier tipo de servicio) debería ser menor al 95%.
- **SERVICIOS CRÍTICOS:** Para aquellos servicios que se consideren críticos la disponibilidad mensual no debería ser menor al 99.95%.
- En los casos en que el requerimiento se aparte de estos rangos recomendados, se deberá justificar dicha solicitud, en la descripción técnica del proyecto elevado a la ONTI.

- Se deberán incluir cláusulas que establezcan que, una vez adjudicado el servicio de nube, ninguno de los valores de disponibilidad podrá ser modificado de manera unilateral por parte del proveedor.

Tiempo de respuesta de soporte técnico de asistencia. Se deberán encontrar disponibles en dominio público previo a la adjudicación, las prestaciones de cada uno de los tipos/niveles de asistencia técnica (típicamente divididas en tres niveles, por ejemplo, Básico, Standard y Enterprise), incluyendo los tiempos de respuesta propios de cada uno.

Dado que el organismo puede elegir contratar el servicio de soporte técnico propio del proveedor de nube, así como también un servicio técnico de un socio oficial del proveedor, se deberá tener en cuenta que:

- **SERVICIO TÉCNICO DEL PROVEEDOR DE NUBE:** En este caso los niveles de servicio se encuentran definidos y publicados de modo previo a la adjudicación, en consecuencia, el organismo deberá indicar en el pliego, a modo de referencia, por cuál de los tres niveles opta, incluyendo requerimientos que resulten abiertos para el nivel de soporte elegido, y evaluarlos en las propuestas técnicas de los oferentes.
- **SERVICIO TÉCNICO DEL PARTNER OFICIAL:** En este caso el organismo deberá definir los requerimientos funcionales estándares que cubran sus necesidades de servicio técnico sin que restrinjan innecesariamente el espectro de este tipo de oferentes. Es importante destacar que, si bien las tareas poseen un alcance basado en el apoyo y asistencia técnica del socio oficial del proveedor, éstas pueden ser más amplias y flexibles que las de dicho proveedor. No obstante, se advierte que esto último no implica necesariamente el aumento de la disponibilidad de recursos y/o servicios de nube ofrecidos por el proveedor. Asimismo, se deberán incluir en el pliego, cláusulas mediante las cuales se exija a los oferentes adjuntar a la oferta las certificaciones oficiales por parte del proveedor de nube, que acrediten su condición de socio o canal oficial.

Tiempo de respuesta en recursos o servicios elásticos. Los organismos deberán verificar que se encuentren disponibles al momento de efectuar la contratación del servicio, los tiempos de respuesta que ofrecen los proveedores de nube para la creación automática de recursos en aquellos servicios que admiten elasticidad. Esto significa que el organismo deberá solicitar en la cotización el tiempo máximo de respuesta de estas operaciones, y establecer los parámetros de evaluación correspondientes, incluyendo, por ejemplo, cláusulas del tipo *“Eleva en 50 la cantidad de máquinas virtuales en ejecución, no podrá superar en promedio los XX minutos”*, las cuales deberán considerarse a la hora de evaluar las ofertas.

Dado que este tipo de cláusulas podría limitar el espectro de oferentes posibles, los valores especificados deberán justificarse técnicamente (en la descripción técnica del proyecto elevado a la ONTI), acorde a la estimación de la demanda y criticidad del diseño en cuestión.

Es importante destacar que, a diferencia de otros servicios en donde los organismos solicitan unilateralmente el porcentaje de disponibilidad que requieren, en estos casos, dicho porcentaje lo establecen los distintos proveedores del servicio, convirtiéndolo en una característica más a evaluar. Esto último se informa debido a que, si el organismo necesita mayor disponibilidad que la ofrecida por el proveedor para uno o más recursos que componen el diseño involucrado, entonces se verá obligado a contratar servicios redundantes adicionales para lograr los niveles mínimos de disponibilidad requeridos por el proyecto en cuestión, lo que se verá reflejado en mayores costos en la facturación del servicio de nube.

Penalidades. Al igual que sucede con los porcentajes de disponibilidad, en los servicios de nube, los porcentajes de crédito o devolución a asignar como multa o penalización por indisponibilidades de los recursos que ofrecen los proveedores de nube, forman parte de las características del servicio.

Las penalidades específicas aplicables a cada recurso también deberán estar en el dominio público, previo a la adjudicación. Las mismas se encontrarán asociadas a eventuales incumplimientos de los distintos valores de disponibilidad ofertados en la propuesta técnica, dichos valores deberán corresponderse con lo requerido en el apartado de disponibilidad del requerimiento técnico.

Los porcentajes de penalidad a aplicar ante eventuales incumplimientos en la disponibilidad de la propuesta adjudicada, se mantendrán durante el plazo de vigencia de la contratación. Por ello, se deberá incluir en el pliego, una cláusula que establezca que dichos porcentajes no podrán ser modificados de manera unilateral por parte del proveedor, sino mediante coordinación previa con el organismo contratante.

Otras consideraciones importantes respecto de las penalidades. Debido a que distintos proveedores de nube suelen fraccionar los tiempos de indisponibilidad en franjas de tiempo mínimas, es necesario establecer cláusulas que protejan al organismo de situaciones en las que se acumulen varias infracciones sucesivas y periódicas que no superan dicha franja de tiempo mínima por cada falla, lo que llevaría a asumir que el servicio de encuentra disponible cuando esto no es lo que se experimenta.

Las franjas de tiempo utilizadas por algunos proveedores importantes en el mercado son de 5 minutos, 15 minutos y 1 hora. Por lo tanto, hay que tener en cuenta que, bajo estas condiciones, si un proveedor especifica una franja de tiempo de 15 minutos y el servicio no funciona durante 14 minutos, se conserva el 100% de disponibilidad mediante esta métrica.

Por lo tanto, deberán incorporarse cláusulas que contemplen el tiempo mínimo medio entre fallas, para asegurar que las indisponibilidades acumuladas en un día no superen el máximo admisible para el diseño en cuestión.

Por ejemplo, si el organismo pretende asegurar una indisponibilidad no mayor a los 30 minutos por día (24 horas), si el proveedor utiliza una franja mínima de 15 minutos para considerar que el servicio se encuentra indisponible, entonces deberá establecerse un tiempo mínimo medio entre fallas de 12 horas, de modo que se asegure que sólo podrán acumularse dos fallas de menos de 15 minutos en un plazo de 24 horas, no superando así los 30 minutos de indisponibilidad requeridos por el organismo. Es decir, si se producen más de 2 fallas de menos de 15 minutos, al no respetarse el tiempo mínimo medio entre fallas, se penaliza de todos modos.