

PROCEDIMIENTO PARA DISEÑAR LOS NODOS DE CÓMPUTO DE UNA NUBE PRIVADA/CENTRO DE DATOS VIRTUALIZADO

El procedimiento para diseñar los recursos de cómputo de una Nube Privada (NP)/Centro de Datos Virtualizado (CDV) a (re)diseñar es mostrado en la [Figura 1](#). A continuación, se describen detalladamente el conjunto de procesos y actividades a desarrollar.

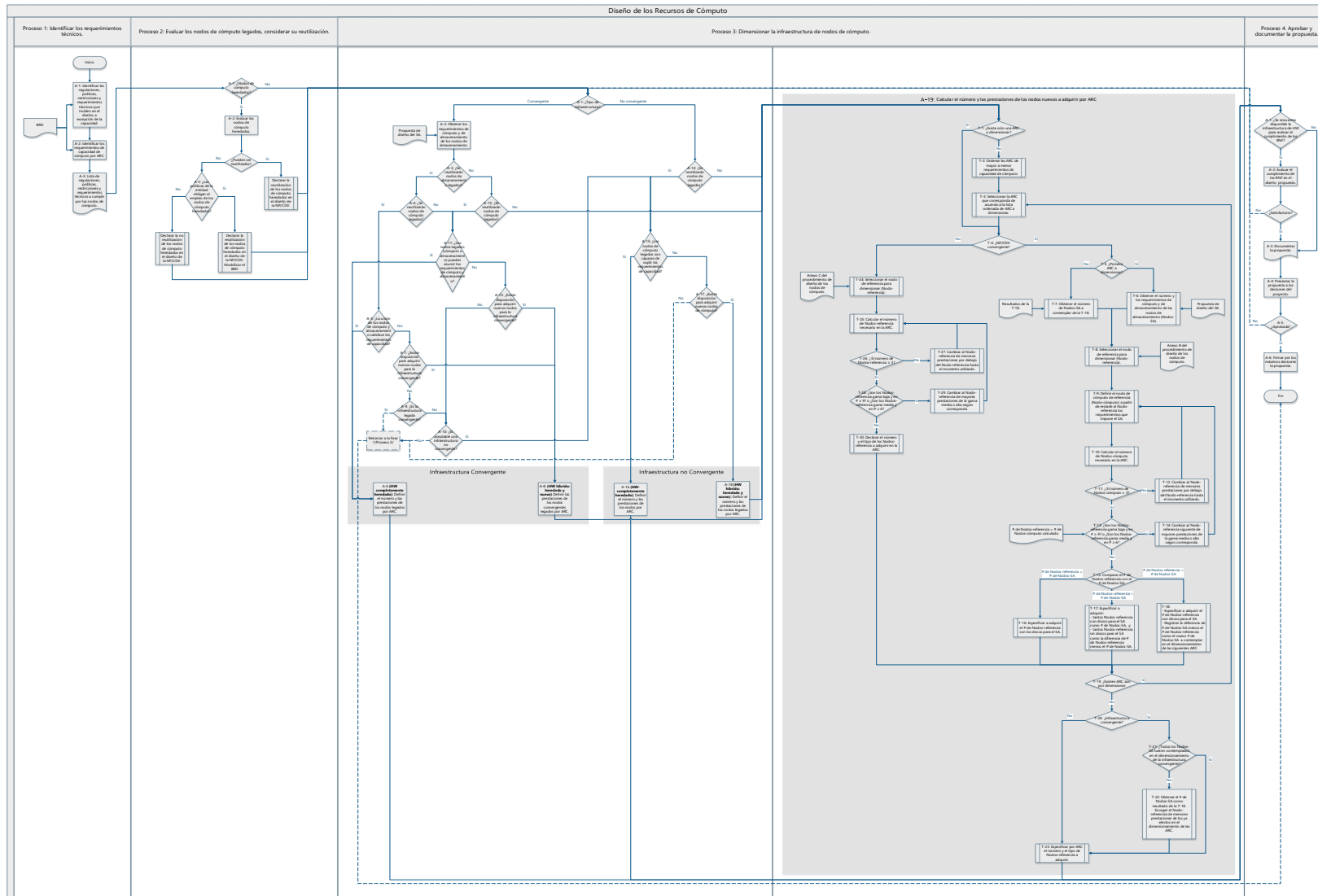


Figura 1. Procedimiento para diseñar los nodos de cómputo de una NP/CDV

Proceso 1. Identificar los requerimientos técnicos

En esta actividad se identifican las regulaciones, políticas, restricciones y requerimientos técnicos que influyen directamente en el diseño físico de los recursos de cómputo. Se divide en tres actividades fundamentales descritas a continuación.

Actividad 1 (A-1). Identificar las regulaciones, políticas, restricciones y requerimientos técnicos que inciden en el diseño, a excepción de la capacidad

Tarea 1 (T-1): Extraer del [Documento de Requerimientos del Negocio](#) (BRD¹) o del documento [“Instrumento para identificar políticas, regulaciones, estándares, recomendaciones, restricciones y requerimientos a cumplir en el diseño de la Nube Privada”](#):

- Presupuesto para comprar nodos de cómputo: _____.
- Planificación del consumo de energía eléctrica para los tres años posteriores a la puesta a punto de la NP/CDV: 1^{er.} año _____, 2^{do.} año _____, y 3^{er.} año _____.
- Presupuesto para el mantenimiento del equipamiento de las Tecnologías de la Información y las Comunicaciones (TIC) y soporte: _____.
- Presupuesto para la adición de equipamiento de las TIC ante el crecimiento esperado en los tres años posteriores a la puesta a punto de la NP/CDV: 1^{er.} año _____, 2^{do.} año _____, y 3^{er.} año _____.

¹ Siglas correspondientes al término en inglés: [Business Requirement Document](#).

Tarea 2 (T-2): Extraer del [BRD](#) o del documento “[Instrumento para identificar políticas, regulaciones, estándares, recomendaciones, restricciones y requerimientos a cumplir en el diseño de la Nube Privada](#)” las regulaciones/resoluciones a cumplir por los nodos de cómputo.

Tarea 3 (T-3): Extraer del [BRD](#) o del documento “[Instrumento para identificar políticas, regulaciones, estándares, recomendaciones, restricciones y requerimientos a cumplir en el diseño de la Nube Privada](#)” los estándares y recomendaciones que deben ser soportados por los nodos de cómputo.

Tarea 4 (T-4): Extraer [BRD](#) o del documento “[Instrumento para identificar políticas, regulaciones, estándares, recomendaciones, restricciones y requerimientos a cumplir en el diseño de la Nube Privada](#)” las restricciones y/o preferencias en el empleo de tecnologías:

- Política ante el despliegue del tipo de infraestructura: convergente o no.
- Política ante la reutilización del Hardware (HW) heredado.
- Las prestaciones que deben soportar de forma obligatoria u opcional los nodos de cómputo.

Tarea 5 (T-5): Extraer del [BRD](#) o de la [Fase 1, Proceso 3, Actividad 2](#), y de las capacidades de la Plataforma de Gestión de Nube (CMP²) seleccionada, el número de clústeres y nodos que deben soportar la plataforma de virtualización y el CMP respectivamente.

Tarea 6 (T-6): Identificar la disponibilidad de espacio para los nodos de cómputo inminentes y a largo plazo de acuerdo a: las dimensiones del local en donde residirá

² Siglas correspondientes al término e inglés: [Cloud Management Platform](#).

la infraestructura de la NP/CDV, los estándares³: TIA 942-B 2017 y la ISO/IEC 11801-5:2017, el espacio ya destinado al Sistema de Almacenamiento (SA) de forma inminente y a largo plazo, y los que pueden demandar el resto de los bloques de la nube como la red y los recursos facilitadores.

Actividad 2 (A-2). Identificar los requerimientos de capacidad por Agrupación de Recursos de Cómputo

Los requerimientos totales de los recursos de cómputo se encuentran compuestos por: la capacidad de los servicios de usuario y soporte identificados en la [Sub-fase 2.2 / Proceso 4 / Actividad 11](#); los requerimientos de la plataforma de virtualización; y los requerimientos del CMP de la NP/CDV, que es un servicio de soporte, pero como constituye un subsistema de la NP/CDV que se diseña posterior a la estimación de la demanda, se realiza la distinción. Por tanto, para poder identificar los requerimientos de capacidad por Agrupación de Recursos de Cómputo (ARC) se deben ejecutar las siguientes **seis Tareas**:

Tarea 1 (T-1): De la [Sub-fase 2.2 / Proceso 4 / Actividad 11](#), deben extraerse los requerimientos de capacidad de cada una de las ARC identificadas: requerimientos de capacidad inminente para el soporte de los servicios durante los horarios típicos de la jornada laboral, y los requerimientos de capacidad inminente para el soporte de los servicios durante los horarios picos, que a su vez se corresponderán con los nodos de cómputos necesarios para garantizar una rápida recuperación ante fallos; y los requisitos mínimos que demanda la plataforma de virtualización en cada nodo:

³ Debe ser comprobada la actualidad de los estándares cuando se requiera su empleo.

Unidad Central de Procesamiento (CPU⁴), Memoria de Acceso Aleatorio (RAM⁵) y almacenamiento, junto al overhead que causa debido a la gestión de las Instancias Virtuales (IV). De corresponderse con una ARC-Bare Metal (BM), este último dato no es necesario. La Tabla 1 muestra los datos a obtener.

Tabla 1. Métricas de capacidad de recursos de cómputo a especificar por ARC_{hipervisor/ (BM)}

CPU (GHz)	RAM (GB)	Almacenamiento			Red, Ancho de Banda (AB) (Mbps)	
		Capacidad (GB) ⁶	<u>Throughput</u>		Transmisión (TX)	Recepción (RX)
			Operaciones de Entrada/Salida por Segundo (IOPS ⁷)	<u>Throughput</u> (Mbps)		

Tarea 2 (T-2): Identificar los requerimientos de capacidad del CMP de la NP/CDV seleccionado en la [Fase 2, Sub-fase 2.1](#). Si el gestor de la NP/CDV se encuentra compuesto por diferentes aplicaciones y servicios, los requerimientos de capacidad deben ser ubicados por ARC, es decir, si corresponde a una ARC de tipo BM, o a una con una plataforma de virtualización específica, así como sus requerimientos inminentes y futuros.

Tarea 3 (T-3): Sumar las capacidades de las **Tareas 1 y 2**, para obtener la demanda de capacidad por ARC.

Tarea 4 (T-4): Identificar si las plataformas de virtualización pueden coexistir en el mismo nodo. En esto influye el principio de funcionamiento de los hipervisores y del

⁴ Siglas correspondientes al término en inglés: Central Processing Unit.

⁵ Siglas correspondientes al término en inglés: Random Access Memory.

⁶ En este caso no se corresponde con la capacidad estimada en la Fase 1, ya que esa será soportada por un SA distribuido. En este caso se propone por la autora de la presente investigación sean dos discos pequeños. Se especifican dos, en post de lograr alta disponibilidad.

⁷ Siglas correspondientes al término en inglés: Input/Output Operations Per Second.

CMP. Proxmox por ejemplo permite que tanto Kernel-based Virtual Machine (KVM) como Contenedores Linux (LXC⁸) coexistan en el mismo nodo.

Tarea 5 (T-5): Agrupar en una sola ARC aquellas plataformas de virtualización que pueden coexistir en un nodo. Por ejemplo, en el caso de que el gestor elegido fuese Proxmox, y que se fuesen a emplear LXC y KVM, ambos quedarían en una misma ARC, ARC_{LXC, KVM}.

Tarea 6 (T-6): Identificar finalmente el número de ARC y sus requerimientos de capacidades de cómputo, contando de existir los clústeres de tipo BM.

Actividad 3 (A-3). Obtener la lista de regulaciones, políticas, restricciones y requerimientos técnicos a cumplir por los nodos de cómputo

La Actividad 3 tiene como objetivo documentar las políticas, regulaciones, resoluciones, recomendaciones, estándares, restricciones y requerimientos técnicos a cumplir por los nodos de cómputo de la NP/CDV. Este documento, o [lista de chequeo](#), constituye el primer acápite del [informe final de la propuesta de los nodos de cómputo](#).

Proceso 2. Evaluar los nodos de cómputo legados, considerar su reutilización

El objetivo de este proceso es caracterizar los nodos de cómputo heredados, de ser pertinente, e identificar la posibilidad de su reutilización. Consta de tres actividades, las que se desarrollan a continuación:

⁸ Siglas correspondientes al término en inglés: Linux Container.

Actividad 1 (A-1): Identificar la existencia de nodos de cómputo heredados o no. De ser positivo pasar a la **Actividad 2**, de lo contrario proceder al [Proceso 3](#) “Dimensionar la infraestructura de nodos de cómputo”.

Actividad 2 (A-2): Identificar si es posible la reutilización de los nodos de cómputo legados en función de las restricciones y necesidades del cliente identificadas, para lo cual se hace necesario a su vez, la caracterización y evaluación del estado técnico de los nodos de cómputo heredados. Se considera que deben ser reutilizados si al ser evaluadas sus prestaciones, empleando el proyecto de pruebas propuesto en el [Anexo A](#), se obtenga una evaluación igual o superior a “Regular”. En caso positivo se declara la reutilización de los nodos de cómputo y se pasa al [Proceso 3](#); de lo contrario se procede a la **Actividad 3**.

Actividad 3 (A-3): Identificar si las políticas de la entidad obligan al empleo de los nodos de cómputo legados, aun cuando no satisfagan los requerimientos inicialmente declarados. De ser negativa la respuesta, se declara la no reutilización de los nodos de cómputo heredados y se pasa al [Proceso 3](#); de lo contrario se procede a declarar los nodos de cómputo heredados como solución de cómputo de la NP/CDV a diseñar, y a confeccionar una nueva versión del [BRD](#), en donde se registren los nuevos requerimientos técnicos, regulaciones, estándares y/o restricciones que la entidad cliente consideró suficientes, para poder reutilizar el HW legado. A continuación, se debe proceder al [Proceso 3](#).

Proceso 3: Dimensionar la infraestructura de nodos de cómputo

El Proceso 3 posee las siguientes Actividades:

Actividad 1 (A-1) “Identificar el tipo de infraestructura: convergente o no”: debe ser consultada la documentación técnica del CMP, e incluso de la integración CMP-plataformas de virtualización-solución Almacenamiento Definido por Software (SDS⁹), en post de identificar si puede ser desplegada una infraestructura convergente en la NP/CDV, o no. Se aboga por una infraestructura convergente dada las bondades que brinda ante la EH y la eficiencia. De ser posible el despliegue de una infraestructura convergente proceder a la **Actividad 2**, de lo contrario a la [Actividad 14](#).

Infraestructura Convergente

Actividad 2 (A-2) “Obtener los requerimientos de cómputo y de almacenamiento de los nodos de almacenamiento”: Obtener de la **Subfase 2.2** los requerimientos de cómputo y de almacenamiento de los nodos de almacenamiento definidos. Proceder a la **Actividad 3**.

Actividad 3 (A-3) “Identificar si los nodos de almacenamiento a emplear son heredados”: Identificar si los nodos de almacenamiento a emplear son heredados. De ser positivo, se debe proceder a la **Actividad 4**, de lo contrario a la **Actividad 13**.

Actividad 4 (A-4) (Nodos de almacenamiento legados) “Identificar si es necesario reutilizar los nodos de cómputo legados”: identificar si es necesaria la reutilización de nodos de cómputo heredados. De ser positivo se procede a la **Actividad 5**, de lo contrario a la **Actividad 11**.

⁹ Siglas correspondientes al término en inglés: Software-Defined Storage.

Actividad 5 (A-5) (Nodos de almacenamiento y cómputo legados) “Identificar si los nodos de cómputo y almacenamiento legados soportan el despliegue de la infraestructura convergente”: Identificar si mediante la combinación de las prestaciones de los servidores de ambos subsistemas, almacenamiento y cómputo, pueden suplirse los requerimientos de capacidad por ARC calculados en la [Actividad 2 / Proceso 1](#), más los requerimientos de cómputo y red que impone por nodo de almacenamiento el SA y su SDS. Esta Actividad posee **dos Tareas**:

Tarea 1 (T-1): Identificar los requerimientos de capacidad de cada ARC, una vez sumados los requerimientos de cómputo y red que impone por nodo de almacenamiento el SA y su SDS.

Tarea 2 (T-2): Combinar las prestaciones de los nodos de cómputo y almacenamiento heredados en vistas a identificar si los requerimientos de capacidad inminente¹⁰, quedan cubiertos en cada ARC. De ser posible, debe pasarse a la **Actividad 6**, de lo contrario a la **Actividad 7**.

Actividad 6 (A-6) (Nodos de almacenamiento y cómputo legados) “Definir las prestaciones de los nodos legados y su número por ARC”: declarar por ARC el número y las prestaciones de los nodos legados para suplir los requerimientos de capacidad en horarios típicos y picos de la infraestructura convergente. Cubrir el [Excel “Nodos heredados infraestructura convergente”](#) y proceder al [Proceso 4 “Documentar la propuesta de diseño de los nodos de cómputo”](#).

Actividad 7 (A-7) “Identificar la disposición de la entidad cliente para adquirir nuevos nodos para la infraestructura convergente”: identificar si la entidad

¹⁰ Requerimientos de capacidad para horarios típicos y pico.

cliente tiene la disposición de adquirir como parte del proyecto de diseño de la NP/CDV los nuevos nodos necesarios para poder completar la infraestructura convergente. De ser positiva la respuesta debe procederse a la **Actividad 8**, de lo contrario a la **Actividad 9**.

Actividad 8 (A-8) (HW híbrido: legados y nuevos a adquirir) “Definir las prestaciones de los nodos convergentes legados por ARC”: declarar por ARC el número y las prestaciones de los nodos convergentes legados para suplir sus requerimientos de capacidad en horarios típicos y picos a corto plazo. Las **Tareas** a desarrollar son:

Tarea 1 (T-1): Identificar los requerimientos de capacidad de cada ARC, una vez sumados los requerimientos de cómputo y red que impone por nodo de almacenamiento el SA y su SDS.

Tarea 2 (T-2): Cubrir las capacidades de cada ARC con los nodos de cómputo y/o almacenamiento legados. Debe ser completado el [Excel “Nodos de cómputo heredado infraestructura convergente”](#).

Tarea 3 (T-3): Calcular por ARC la capacidad que no es cubierta por los nodos legados. Proceder a la [Actividad 19](#).

Actividad 9 (A-9): Identificar si la infraestructura legada es convergente o no. Si la infraestructura legada no es convergente proceder a la **Actividad 10**; de lo contrario, se debe retornar a la [Fase 1/Proceso 3](#), para poder reajustar los requerimientos de capacidad de cómputo y almacenamiento a soportar, ya que, hasta este punto del procedimiento, solo puede cubrirse con la infraestructura convergente legada los requerimientos del SA, y no se cuenta con nodos legados para una infraestructura no convergente.

Actividad 10 (A-10): Identificar si la entidad cliente acepta el despliegue de una infraestructura no convergente. De ser positivo proceder a la [Actividad 14](#), de lo contrario se debe retornar a la [Fase 1/Proceso 3](#), para poder reajustar los requerimientos de capacidad de cómputo y almacenamiento a soportar, y así poder desplegar una infraestructura convergente con el HW legado.

Actividad 11 (A-11) “Identificar si los nodos de almacenamiento o de cómputo legados soportan solos el despliegue de la infraestructura convergente”:

Identificar si los nodos de almacenamiento o de cómputo legados pueden satisfacer solos los requerimientos de capacidad inminente¹¹: requerimientos de capacidad por ARC calculados en la [Actividad 2 / Proceso 1](#), más los requerimientos de cómputo y red que impone por nodo de almacenamiento el SA y su SDS. Esta Actividad posee **dos Tareas**:

Tarea 1 (T-1): Identificar los requerimientos de capacidad de cada ARC, una vez sumados los requerimientos de cómputo y red que impone por nodo de almacenamiento el SA y su SDS.

Tarea 2 (T-2): Identificar si los nodos de almacenamiento o de cómputo legados pueden asumir solos la infraestructura convergente con los requerimientos de capacidad estimados por ARC. De ser posible, debe pasarse a la **Actividad 6**, de lo contrario a la **Actividad 12**.

Actividad 12 (A-12) “Identificar la disposición de la entidad cliente para adquirir nuevos nodos para la infraestructura convergente”: identificar si la entidad cliente tiene la disposición de adquirir como parte del proyecto de diseño de

¹¹ Requerimientos de capacidad para horarios típicos y pico.

la NP/CDV los nuevos nodos necesarios para poder completar la infraestructura convergente. De ser positiva la respuesta debe procederse a la **Actividad 8**, de lo contrario a la **Actividad 10**.

Actividad 13 (A-13) (Nodos de cómputo legados) “Identificar si es necesario reutilizar los nodos de cómputo legados”: identificar si es necesaria la reutilización de nodos de cómputo heredados. De ser positivo se procede a la **Actividad 11**, de lo contrario a la [Actividad 19](#).

Infraestructura no convergente:

Actividad 14 (A-14) “Identificar si es necesario reutilizar los nodos de cómputo legados”: Identificar si es necesaria la reutilización de nodos de cómputo heredados. De ser positivo se procede a la **Actividad 15**, de lo contrario a la [Actividad 19](#).

Actividad 15 (A-15) (Nodos de cómputo legados) “Identificar si los nodos de cómputo legados soportan solos los requerimientos de capacidad”: Identificar si los nodos de cómputo legados pueden satisfacer los requerimientos de capacidad por ARC calculados en la [Actividad 2 / Proceso 1](#). Esta Actividad posee **dos Tareas**:

Tarea 1 (T-1): Identificar los requerimientos de capacidad de cada ARC.

Tarea 2 (T-2): Identificar si los nodos de cómputo legados pueden asumir los requerimientos de capacidad estimados por ARC. De ser posible, debe pasarse a la **Actividad 16**, de lo contrario a la **Actividad 17**.

Actividad 16 (A-16) (Nodos de cómputo legados) “Definir las prestaciones de los nodos legados y su número por ARC”: declarar por ARC el número y las prestaciones de los nodos de cómputo legados para suplir sus requerimientos de

capacidad en horarios típicos y picos. Cubrir el [Excel “Nodos de cómputo heredado infraestructura no convergente”](#), en donde quede especificado el número de nodos legados y sus prestaciones por ARC. Proceder al [Proceso 4 “Documentar la propuesta de diseño de los nodos de cómputo”](#).

Actividad 17 (A-17) “Identificar la disposición de la entidad cliente para adquirir nuevos nodos de cómputo”: Identificar si la entidad cliente tiene la disposición de adquirir como parte del proyecto de diseño de la NP/CDV los nuevos nodos de cómputo necesarios para poder completar los requerimientos de capacidad por ARC. De ser positiva la respuesta debe procederse a la **Actividad 18**, de lo contrario se debe retornar a la [Fase 1/Proceso 3](#), para poder reajustar los requerimientos de capacidad de cómputo a soportar, y así poder desplegar una infraestructura no convergente con el HW legado.

Actividad 18 (A-18) (HW híbrido: legado y nuevo a adquirir) “Definir las prestaciones de los nodos de cómputo legado y su número por ARC”: declarar por ARC el número y las prestaciones de los nodos de cómputo legados para suplir sus requerimientos de capacidad en horarios típicos y picos a corto plazo. Las Tareas a desarrollar son:

Tarea 1 (T-1): Identificar los requerimientos de capacidad por ARC calculados en la [Actividad 2 / Proceso 1](#).

Tarea 2 (T-2): Cubrir las capacidades de cada ARC con los nodos de cómputo legados. Debe ser completado el [Excel “Nodos de cómputo heredado infraestructura no convergente”](#).

Tarea 3 (T-3): Calcular por ARC la capacidad que no es cubierta por los nodos legados. Proceder a la [Actividad 19](#).

Actividad 19 (A-19): Calcular el número y las prestaciones de los nodos nuevos a adquirir por ARC

La Actividad 19 consta de las siguientes **Tareas**:

Tarea 1 (T-1): Identificar si existe solo una ARC a dimensionar. De ser positivo debe procederse a la **Tarea 4**, de lo contrario a la **Tarea 2**.

Tarea 2 (T-2): Ordenar las ARC de mayor a menor requerimientos de capacidad de cómputo. El objetivo de esta tarea es de simplificar el cálculo del número de nodos convergentes en el caso de que la infraestructura sea convergente, al concentrar estos en las menos ARC posibles. Proceder a la **Tarea 3**.

Tarea 3 (T-3): Seleccionar la ARC que corresponda de acuerdo a la lista ordenada de ARC por dimensionar. Proceder a la **Tarea 4**.

Tarea 4 (T-4): Identificar si la infraestructura de la NP/CDV será convergente o no. De ser positivo proceder a la **Tarea 5**, de lo contrario a la **Tarea 24**.

Tarea 5 (T-5): Identificar si es la primera vez que se dimensiona una ARC. De ser positivo proceder a la **Tarea 6**, de lo contrario a la **Tarea 7**.

Tarea 6 (T-6): Identificar el número y los requerimientos de cómputo y de almacenamiento de los nodos de almacenamiento (Nodos_{SA}). Esta información se obtiene del “[Informe de la propuesta de diseño del SA](#)”, [Subfase 2.2/Proceso 5](#). Proceder a la **Tarea 8**.

Tarea 7 (T-7): Obtener el número de nodos de almacenamiento (Nodos_{SA}) a contemplar de la **Tarea 18**. Se mantienen las prestaciones especificadas de los Nodos-SA en el “[Informe de la propuesta de diseño del SA](#)”, [Subfase 2.2/Proceso 5](#). Proceder a la **Tarea 8**.

Tarea 8 (T-8): Seleccionar del [Anexo B](#) el nodo de referencia para dimensionar ($Nodo_{referencia}$). La selección debe regirse en primer lugar por el número y prestaciones de los discos que poseen los nodos de almacenamiento, lo que ubica la selección del nodo de referencia en una de las tres posibles gamas: baja, media y alta. Una vez localizado en una gama el nodo de referencia dados los requerimientos de almacenamiento, debe tomarse el nodo de referencia de mayores prestaciones dentro de la gama, el que debe cubrir los requerimientos de procesamiento de los nodos de almacenamiento. Proceder a la **Tarea 9**.

Tarea 9 (T-9): Definir el nodo de cómputo de referencia ($Nodo_{cómputo}$) a partir de restarle al $Nodo_{referencia}$ los requerimientos que impone el SA. Se propone para obtener el $Nodo_{cómputo}$ restar al $Nodo_{referencia}$ identificado los requerimientos que impone el $Nodo_{SA}$ como indican las Fórmulas 1-4:

$$CPU_{Cómputo} = CPU_{Nodo_{referencia}} - CPU_{SA} \quad (1)$$

En donde:

$CPU_{Cómputo}$: frecuencia del $Nodo_{referencia}$ dedicada al cómputo por parte de los servicios a soportar.

$CPU_{Nodo_referencia}$: frecuencia del $Nodo_{referencia}$ identificado.

CPU_{SA} : frecuencia del $Nodo_{referencia}$ dedicada al SA, es decir, frecuencia del $Nodo_{SA}$.

$$RAM_{Cómputo} = RAM_{Nodo_{referencia}} - RAM_{SA} \quad (2)$$

En donde:

$RAM_{Cómputo}$: capacidad de RAM del $Nodo_{referencia}$ dedicada al cómputo por parte de los servicios a soportar.

$RAM_{Nodo_referencia}$: capacidad de RAM del $Nodo_{referencia}$ identificado.

RAM_{SA}: capacidad de RAM del Nodo_{referencia} dedicada al SA, es decir, RAM del Nodo_{SA}.

$$Disco_{C\acute{o}mputo} = Disco_{Nodo_{referencia}} - Disco_{SA} \quad (3)$$

En donde:

Disco_{C\acute{o}mputo}: capacidad del disco del Sistema Operativo (SO) del Nodo_{referencia} dedicada al c\acute{o}mputo por parte de los servicios a soportar.

Disco_{Nodo_referencia}: capacidad del disco del SO del Nodo_{referencia} identificado.

Disco_{SA}: capacidad del disco del SO del Nodo_{referencia} dedicada al SA, es decir, disco del SO del Nodo_{SA}.

$$Red_{C\acute{o}mputo} = Red_{Nodo_{referencia}} - Red_{SA} \quad (4)$$

En donde:

Red_{C\acute{o}mputo}: capacidad de Red del Nodo_{referencia} dedicada al c\acute{o}mputo por parte de los servicios a soportar.

Red_{Nodo_referencia}: capacidad de Red del Nodo_{referencia} identificado.

Red_{SA}: capacidad de Red del Nodo_{referencia} dedicada al SA, es decir, Red del Nodo_{SA}.

Una vez identificadas las prestaciones del Nodo_{c\acute{o}mputo} proceder a la **Tarea 10**.

Tarea 10 (T-10): Calcular el n\acute{umero de Nodo_{c\acute{o}mputo} necesario en una ARC, para cubrir la capacidad de c\acute{o}mputo inminente¹². El c\acute{alculo consiste en dividir los requerimientos inminentes de CPU, RAM, Disco_{SO}¹³ y Red de la ARC, por las prestaciones del Nodo_{c\acute{o}mputo} calculado en la **Tarea 9**. En el [Anexo C](#) se ejemplifica el c\acute{alculo. Proceder a la **Tarea 11**.

¹² No se incluye en estos momentos la demanda de capacidad para horarios picos.

¹³ Capacidad de almacenamiento dedicado al sistema operativo del nodo.

Tarea 11 (T-11): Identificar si el número de $\text{Nodos}_{\text{c\acute{o}mputo}}$ es mayor o igual a tres. De ser negativo proceder a la **Tarea 12**, de lo contrario a la **Tarea 13**.

Tarea 12 (T-12): Seleccionar el $\text{Nodo}_{\text{referencia}}$ de menores prestaciones por debajo del $\text{Nodo}_{\text{referencia}}$ hasta el momento utilizado sin afectar los requerimientos de cómputo y almacenamiento del Nodo_{SA} . Retornar a la **Tarea 9**.

Tarea 13 (T-13): Asumir que el número de $\text{Nodos}_{\text{referencia}}$ necesarios es igual al número de $\text{Nodos}_{\text{c\acute{o}mputo}}$ calculados, y seguir el dimensionamiento con las prestaciones del $\text{Nodo}_{\text{referencia}}$ hasta el momento definido. Identificar si el número de $\text{Nodos}_{\text{referencia}}$ equivale a un número inferior, pero mayor o igual a tres, de $\text{Nodos}_{\text{referencia}}$ de la gama superior, para poder optimizar el número de nodos a adquirir y así aumentar la relación costo/beneficio, la EH y la EV, manteniendo la alta disponibilidad. Los criterios de diseño a considerar son:

- Tres $\text{Nodos}_{\text{referencia}}$ gama baja = un $\text{Nodo}_{\text{referencia}}$ gama media. Optimizar si el número de $\text{Nodos}_{\text{referencia}}$ gama media resultante es mayor o igual a tres, lo que equivale a como mínimo nueve $\text{Nodos}_{\text{referencia}}$ gama baja.
- Dos $\text{Nodos}_{\text{referencia}}$ gama media = un $\text{Nodo}_{\text{referencia}}$ gama alta. Optimizar si el número de $\text{Nodos}_{\text{referencia}}$ gama alta resultante es mayor o igual a tres, lo que equivale a como mínimo seis $\text{Nodos}_{\text{referencia}}$ gama media.

De resultar positivo algunos de los dos criterios proceder a la **Tarea 14**, de lo contrario a la **Tarea 15**.

Tarea 14 (T-14): Cambiar al $\text{Nodo}_{\text{referencia}}$ de mayores prestaciones de la gama media o alta según corresponda. Retornar a la **Tarea 9**.

Tarea 15 (T-15): Comparar el número de $\text{Nodos}_{\text{referencia}}$ con el número de Nodos_{SA} , si:

- El número de $\text{Nodos}_{\text{referencia}}$ es igual al número de Nodos_{SA} : proceder a la **Tarea 16**.
- El número de $\text{Nodos}_{\text{referencia}}$ es mayor al número de Nodos_{SA} : proceder a la **Tarea 17**.
- El número de $\text{Nodos}_{\text{referencia}}$ es menor que el número de Nodos_{SA} : proceder a la **Tarea 18**.

Tarea 16: Declarar que se deben adquirir el número de $\text{Nodos}_{\text{referencia}}$ hasta el momento calculado, con la capacidad de almacenamiento declarado en los Nodos_{SA} , $\text{Nodos}_{\text{convergentes}}$. Deben ser calculados el número de $\text{Nodos}_{\text{referencia}}$ necesarios para la capacidad inminente en horarios picos, como se explica en el [Anexo C](#). Los nodos para cubrir los horarios picos, deben ser adquiridos solo con los discos para el SO.

Proceder a la **Tarea 19**.

Tarea 17: Declarar que se deben adquirir:

- tantos Nodos-referencia con discos para el SA como el número de Nodos-SA , y
- tantos Nodos-referencia sin discos para el SA como la diferencia del número de Nodos-referencia menos el número de Nodos-SA .

Deben ser calculados el número de $\text{Nodos}_{\text{referencia}}$ necesarios para la capacidad inminente en horarios picos, como se explica en el [Anexo C](#). Los nodos para cubrir los horarios picos, deben ser adquiridos solo con los discos para el SO.

Proceder a la **Tarea 19**.

Tarea 18: Declarar que se deben adquirir el número de $\text{Nodos}_{\text{referencia}}$ con discos para el SA, y se debe registrar la diferencia del número de Nodos_{SA} menos el

número de Nodos-referencia como el número de nodos convergentes aún por dimensionar, es decir, el nuevo número de Nodos_{SA} a incluir en el dimensionamiento de la infraestructura convergente. Además, deben ser calculados el número de Nodos_{referencia} necesarios para la capacidad inminente en horarios picos, como se explica en el [Anexo C](#). Los nodos para cubrir los horarios picos, deben ser adquiridos solo con los discos para el SO.

Proceder a la **Tarea 19**.

Tarea 19 (T-19): Identificar si aún existen ARC por dimensionar. De ser positivo retornar a la **Tarea 3**, de lo contrario proceder a la **Tarea 20**.

Tarea 20 (T-20): Identificar si la infraestructura que está siendo dimensionada es convergente o no. De ser convergente debe procederse a la **Tarea 21**, de lo contrario a la **Tarea 23**.

Tarea 21 (T-21): Identificar si todos los Nodos_{SA} fueron contemplados en el dimensionamiento de la infraestructura convergente. De no existir proceder a la **Tarea 23**, de lo contrario a la **Tarea 22**.

Tarea 22 (T-22): Identificar el número y las prestaciones de los Nodos_{SA} nuevos a adquirir. Para ello se proponen las siguientes acciones:

Acción 1 (A-1): Identificar el número de nodos dedicados al soporte de procesamiento en horarios picos. De no existir nodos dedicados a los horarios picos proceder a la **Acción 3**, de lo contrario a la **Acción 2**.

Acción 2 (A-2): Cubrir con nodos picos el número posible de Nodos_{SA}, especificar que deben ser adquiridos los discos para el SA en estos nodos. Si aún quedasen Nodos_{SA} por dimensionar proceder a la **Acción 3**, de lo contrario a la **Acción 4**.

Acción 3 (A-3): Escoger el Nodo_{referencia} de menores prestaciones de los ya electos en el dimensionamiento de las ARC. Declarar a adquirir el número de nodos de almacenamiento necesarios. Proceder a la **Acción 4**.

Acción 4 (A-4): declarar y actualizar por ARC el número y las prestaciones de los Nodo_{referencia} dedicados a los horarios pico que actuarán como nodos convergentes, y el número de Nodo_{referencia} nuevos a adquirir dedicados solo a almacenamiento. Proceder a la **Tarea 24**.

Tarea 23 (T-23): Especificar por ARC el número y las prestaciones de los nodos nuevos a adquirir a corto y largo plazo como indica el [Excel “Nodos nuevos a adquirir por ARC”](#).

Tarea 24 (T-24): Seleccionar del [Anexo B](#) el nodo de referencia para dimensionar (Nodo_{referencia}). Debe seleccionarse el nodo de referencia de mayores prestaciones de la gama más baja. Proceder a la **Tarea 25**.

Tarea 25 (T-25): Calcular el número de Nodos_{referencia} necesario en una ARC, para cubrir la capacidad de cómputo inminente¹⁴. El cálculo consiste en dividir los requerimientos inminentes¹⁵ de CPU, RAM, Discos¹⁶ y Red de la ARC, por las prestaciones del Nodo_{referencia}. En el [Anexo C](#) se ejemplifica el cálculo. Proceder a la **Tarea 26**.

Tarea 26 (T-26): Identificar si el número de Nodos_{referencia} es mayor o igual a tres. De ser negativo proceder a la **Tarea 27**, de lo contrario a la **Tarea 28**.

¹⁴ Requerimientos de capacidad para horarios no picos.

¹⁵ Requerimientos de capacidad para horarios no picos.

¹⁶ Capacidad de almacenamiento dedicado al sistema operativo del nodo.

Tarea 27 (T-27): Seleccionar el Nodo_{referencia} de menores prestaciones por debajo del Nodo_{referencia} hasta el momento utilizado. Retornar a la **Tarea 25**.

Tarea 28 (T-28): Identificar si el número de Nodos_{referencia} equivale a un número inferior, pero mayor o igual a tres, de Nodos_{referencia} de la gama superior, para poder optimizar el número de nodos a adquirir y así aumentar la relación costo/beneficio, la EH y la EV, manteniendo la alta disponibilidad. Los criterios de diseño a considerar son:

- Tres Nodos_{referencia} gama baja = un Nodo_{referencia} gama media. Optimizar si el número de Nodos_{referencia} gama media resultante es mayor o igual a tres, lo que equivale a como mínimo nueve Nodos_{referencia} gama baja.
- Dos Nodos_{referencia} gama media = un Nodo_{referencia} gama alta. Optimizar si el número de Nodos_{referencia} gama alta resultante es mayor o igual a tres, lo que equivale a como mínimo seis Nodos_{referencia} gama media.

De resultar positivo algunos de los dos criterios proceder a la **Tarea 29**, de lo contrario a la **Tarea 30**.

Tarea 29 (T-29): Cambiar al Nodo_{referencia} de mayores prestaciones de la gama media o alta según corresponda. Retornar a la **Tarea 25**.

Tarea 30 (T-30): Declarar el número y el tipo de los Nodos_{referencia} a adquirir en la ARC a corto plazo¹⁷. Deben ser calculados el número de Nodos_{referencia} necesarios para la capacidad inminente en horarios picos, como se explica en el [Anexo C](#). Proceder a la **Tarea 19**.

¹⁷ Requerimientos de capacidad para horarios no picos.

Proceso 4: Aprobar y documentar la propuesta de diseño de los nodos de cómputo

El Proceso 4 posee seis posibles Actividades en vistas a aprobar la propuesta del diseño de la infraestructura de los nodos de cómputo. A continuación, se desarrollan las Actividades:

Actividad 1 (A-1). Identificar si es posible evaluar el cumplimiento de los RNF en la propuesta de diseño:

Debe ser identificado si la entidad cuenta con los recursos de HW necesarios para corroborar el cumplimiento de los RNF a cumplir en el proyecto. Esta posibilidad se enmarcaría en dos posibles contextos:

- El HW a emplear es íntegramente legado.
- El ciclo de aprovisionamiento del nuevo HW a adquirir posee un tiempo lo suficientemente corto como para no detener el proyecto de diseño de la NP/CDV por la ejecución de las pruebas.

De ser posible la ejecución de las pruebas se procede a la **Actividad 2**, de lo contrario a la **Actividad 3**.

Actividad 2 (A-2). Evaluar el cumplimiento de los RNF en el diseño propuesto:

La evaluación debe ser realizada aplicando las pruebas definidas en el [Anexo A](#). De ser satisfactoria la propuesta debe procederse a la documentación de la propuesta, **Actividad 3**; de lo contrario debe regresarse al [Proceso 3](#) para suplir las deficiencias identificadas y/u optimizar la propuesta.

Actividad 3 (A-3): Documentar la propuesta realizada, y su proceso de concepción.

El formato del documento se encuentra en la [Plantilla](#). Proceder a la **Actividad 4**.

Actividad 4 (A-4): Presentar la propuesta a los decisores del proyecto.

Actividad 5 (A-5): Aprobar la propuesta, o no, por parte de los decisores del proyecto. Der ser aprobada proceder a la **Actividad 6**, de lo contrario retornar al [Proceso 3](#).

Actividad 6 (A-6): Firmar por los máximos decisores la propuesta realizada como muestra la [Plantilla](#).

Anexos

Anexo A. Requerimientos no Funcionales de los recursos de cómputo

Las prioridades a otorgar al cumplimiento de los Requerimientos no Funcionales (RNF) de los recursos de cómputo deben ser heredadas de las prioridades asignadas a los RNF de la NP/CDV como un todo y/o especificadas de forma particular para este bloque. Las soluciones a evaluar deben recibir un valor final atendiendo a dos posibles opciones:

- La primera es a través del Indicador de Calidad (α). Esta es la manera más sencilla y rápida de evaluar ya que se otorga igual prioridad a todos los RNF. [1]
- La segunda requiere un trabajo matemático extra ya que propone otorgar prioridades a los RNF como se propone al inicio del presente escrito. Esta posición se materializa a través del Indicador de Calidad Ponderado (β). [1]

Sean x_1, x_2, \dots, x_n los valores normalizados de los n atributos a medir de los recursos de cómputo, se define como α a la media aritmética de estos n valores, como muestra la Fórmula A.1: [1]

$$\alpha = \frac{1}{n} * (\sum_{i=1}^n x_n) = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (A.1)$$

Sean x_1, x_2, \dots, x_n los valores normalizados de los n atributos a medir en los recursos de cómputo, y w_1, w_2, \dots, w_n de los coeficientes ponderados o prioridades de cada uno de los x_i respectivamente, se define como β a la media aritmética ponderada de estos n valores normalizados, como muestra la Fórmula A.2: [1]

$$\beta = \frac{\sum_{i=1}^n x_i * w_i}{\sum_{i=1}^n w_i} = \frac{x_1 * w_1 + x_2 * w_2 + \dots + x_n * w_n}{w_1 + w_2 + \dots + w_n} \quad (\text{A.2})$$

Una vez evaluada la solución, se calculan sus indicadores de calidad. Los valores de α y β dan una medida de la calidad de la solución seleccionada en cuestión. Es deseable $\alpha < \beta$ ya que da una medida de que los RNF que interesa priorizar en el diseño han tenido mejor cumplimiento que los que no interesa priorizar. Mientras más grande sea la diferencia entre los dos coeficientes más eficiente habrá sido la elección. Mientras más cercano sea β a la unidad mejor será la solución, es decir, cumple con los RNF propuestos. [1]

A su vez en función de los valores de α y/o β el subsistema debe ser evaluado de:

- Excelente: $0,80 < \alpha \text{ y/o } \beta \leq 1$
- Muy buena: $0,60 < \alpha \text{ y/o } \beta \leq 0,80$
- Buena: $0,40 < \alpha \text{ y/o } \beta \leq 0,60$
- Regular: $0,20 < \alpha \text{ y/o } \beta \leq 0,40$
- Insatisfactoria: $\alpha \text{ y/o } \beta \leq 0,20$

Dimensión de Adaptabilidad

Indica la capacidad de los recursos de cómputo de escalar horizontalmente ante la demanda de recursos y el nivel de aprovechamiento del 100% de su escalabilidad vertical; así como el grado de personalización que ofrecen para ajustar

tecnológicamente la solución, ante las necesidades presentes y futuras de la entidad cliente.

Categoría de Escalabilidad

Es la capacidad que tiene la infraestructura de la NP/CDV de crecer en lo que a recursos de cómputo se refiere. Posee dos atributos: la Escalabilidad Vertical (EV) y la Escalabilidad Horizontal (EH).

Atributo de Escalabilidad Vertical

Es la cantidad de recursos de cómputo que se permite aumentar en los nodos físicos. Por lo general se tienen en cuenta factores como: el número máximo de núcleos lógicos/físicos y capacidad de RAM presentes en un nodo que soporta/restringe el hipervisor, ya sea por limitaciones tecnológicas o de licencias; el número máximo de sockets de CPU y RAM que tienen las placas de los nodos; el número máximo de NIC; la facilidad de crecimiento de las prestaciones de hardware (HW); la obsolescencia tecnológica; y los costos de crecimiento [2]–[5]. La autora de la presente investigación aboga por dimensionar físicamente con el 100% de la EV.

Métricas y Procedimiento de Evaluación

La métrica que se propone para evaluar el cumplimiento de la EV en el subsistema de los recursos de cómputo es “EV en los RC” definida por la Fórmula A.3:

$$EVRC = (\sum_{k=1}^n EVARCK) * 100\% \quad (A.3)$$

En donde:

EVRC: EV en los recursos de cómputo, se evaluará de Excelente (E), Bien (B) o Mal (M) en función del rango en que se encuentre el resultado:

- Excelente: si $90\% < EV_{RC} \leq 100\%$.
- Regular: si $80\% < EV_{RC} \leq 90\%$.
- Mal: si $EV_{RC} \leq 80\%$.

EVARCk: EV en cada Agrupación de Recursos de Cómputo que posea la infraestructura. Como criterio de diseño se considera que en cada ARC los nodos físicos deben tener las mismas prestaciones. Se define por tanto como indica la Fórmula A.4:

$$EVARC = \sum EV_{RAM} + EV_{CPU} + EV_{NIC} \quad (A.4)$$

En donde:

EVARC: se le asignará a una evaluación de Excelente (E), Bien (B) o Mal (M) en función del rango en que se encuentre el resultado:

- Excelente: si $90\% < EV_{ARC} * 100\% \leq 100\%$.
- Regular: si $80\% < EV_{ARC} * 100\% \leq 90\%$.
- Mal: si $EV_{ARC} * 100\% \leq 80\%$.

EV_{RAM}: EV del subsistema de RAM en un nodo físico, definido por la Fórmula 5:

$$EV_{RAM} = \frac{\# \text{ de ranuras RAM utilizadas}}{\text{Total de ranuras para RAM}} \quad (A.5)$$

Se le asignará a EV_{RAM} una evaluación de Excelente (E), Bien (B) o Mal (M) en función del rango en que se encuentre el resultado:

- Excelente: si $90\% < EV_{RAM} * 100\% \leq 100\%$.
- Regular: si $80\% < EV_{RAM} * 100\% \leq 90\%$.
- Mal: si $EV_{RAM} * 100\% \leq 80\%$.

EV_{CPU} : EV del subsistema de CPU en un nodo físico, definido por la Fórmula A.6:

$$EV_{CPU} = \frac{\# \text{ de sockets de CPU utilizados}}{\text{Total de sockets de CPU}} \quad (A.6)$$

Se le asignará a EV_{CPU} una evaluación de Excelente (E), Bien (B) o Mal (M) en función del rango en que se encuentre el resultado:

- Excelente: si $90\% < EV_{CPU} * 100\% \leq 100\%$.
- Regular: si $80\% < EV_{CPU} * 100\% \leq 90\%$.
- Mal: si $EV_{CPU} * 100\% \leq 80\%$.

EV_{NIC} : EV del subsistema de red en un nodo físico, definido por la Fórmula 7:

$$EV_{nic} = \frac{\# \text{ de ranuras para NIC utilizadas}}{\text{Total de ranuras para NIC}} \quad (A.7)$$

Se le asignará a EV_{NIC} una evaluación de Excelente (E), Bien (B) o Mal (M) en función del rango en que se encuentre el resultado:

- Excelente: si $90\% < EV_{NIC} * 100\% \leq 100\%$.
- Regular: si $80\% < EV_{NIC} * 100\% \leq 90\%$.
- Mal: si $EV_{NIC} * 100\% \leq 80\%$.

Atributo de Escalabilidad Horizontal

El atributo de EH en los nodos de cómputo es generalmente definido como la cantidad de nodos de cómputo que permite aumentar la capacidad sin afectar el

funcionamiento del CD. Se tiene en cuenta factores como: la facilidad de crecimiento en número de nodos, las Inversiones de Capital (CAPEX¹⁸), los Gastos de Operaciones (OPEX¹⁹), el consumo de energía eléctrica, el número máximo de nodos y clústeres que soporta el gestor/hipervisor de la nube, y el número de unidades de rack disponibles que deja la propuesta del proveedor [2]–[5].

La autora de la presente investigación se encuentra en concordancia con la definición mencionada, pero considera que la evaluación de la EH de este subsistema no debe ser de forma aislada, sino integrado al resto de los subsistemas y parámetros que restringen la adición de nodos a la infraestructura como se encuentra plasmado en la definición de la EH de una NP/CDV: los estándares de la TIA 942-B 2017, la ISO/IEC 11801-5:2017 y el ANSI/ASHRAE Standard 90.4-2019, el CAPEX y el OPEX anuales dedicados a TI en la entidad, el consumo de energía eléctrica y los requerimientos de Calidad de Servicio (QoS²⁰) que imponen los diferentes bloques de la nube. Estos factores abarcan a todos los bloques funcionales de la nube como el SA, la red y los recursos facilitadores, razón por la cual debe ser evaluada la EH de la nube como un todo integrado, y no como islas, en donde no se verían los compromisos e interacciones entre los subsistemas.

Categoría de Personalización

La categoría de Personalización indica las capacidades que brindan los recursos de cómputo para adaptar tecnológicamente la infraestructura de la NP/CDV, así como

¹⁸ Siglas correspondientes al término en inglés: Capital Expenditure.

¹⁹ Siglas correspondientes al término en inglés: Operational Expenditures.

²⁰ Siglas correspondientes al término en inglés: Quality of Service.

sus servicios de soporte y de usuarios, ante las necesidades presentes y futuras de la entidad cliente.

Atributo de Flexibilidad

Indica la capacidad de personalizar los componentes de HW en función de los requerimientos y necesidades [6] [7] [8]. La autora de la presente investigación considera que este atributo se cumple siempre que se diseñe con HW tipo Cots of the Shell (COTS), valor 1, de lo contrario 0.

Atributo de Compatibilidad

Soporte de diferentes tecnologías y estándares en vistas a satisfacer los requerimientos del diseño.

Métricas y Procedimiento de Evaluación

Se propone la métrica “Nivel de Compatibilidad (NC)”. Su valor será cuantitativo y se encontrará en función de cuántos RF, de los especificados en el Anexo A.1²¹, cumplen los nodos físicos de cada ARC, como muestra la Fórmula A.8.

$$NC = \min \{NC_ARCn\} \quad (A.8)$$

En donde:

NC_ARCn: es el nivel de compatibilidad de cada ARC. Tomará un valor de:

- Muy Alto: si los nodos cumplen con el 90% de los RNF.
- Alto: si los nodos cumplen con el 80% de los RNF.
- Normal: si los nodos cumplen con el 70% de los RNF.
- Bajo: si los nodos cumplen con el 60% de los RNF.

²¹ Pueden ser agregados a la Tabla C.1 nuevos Requerimientos Funcionales (RF).

- Muy Bajo: si los nodos cumplen con menos del 60% de los RNF.

Dimensión de Calidad de Servicio

Indica los requerimientos de desempeño y disponibilidad que cumplen los nodos de cómputo de la infraestructura de la NP/CDV.

Categoría de desempeño

La categoría de desempeño indica la capacidad de los nodos de cómputo de lograr los requerimientos relacionados con la eficiencia energética, la capacidad y el poder de cómputo.

Atributo de Capacidad

Indica la capacidad total de los nodos de cómputo de la infraestructura de la NP/CDV en términos de:

- CPU: total de GHz.
- RAM: capacidad total de RAM (GB).
- SA: capacidad de acceso al SA de L/E: IOPS y throughput (GHz); y capacidad del(os) disco(s) (GB).
- Red: capacidad de E/S a la red (GHz).

Constituye el resultado de la sumatoria de las capacidades de cada ARC, las que a su vez son el resultado de la sumatoria de las capacidades de los nodos que la conforman.

La métrica que se propone para evaluar el atributo de capacidad en un diseño de NP es el “Factor de Precisión (FPRC)”. FPRC indica cuánto se ajusta la capacidad lograda²², a los recursos de cómputos estimados. Su valor constituye el peor valor de precisión de las ARC, como muestra la Fórmula A.9.

$$FPRC = valor_máx\{FP_{ARC1}; \dots FP_{ARCn}\} \quad (A.9)$$

En donde:

FP_{ARCn} : Factor de precisión de la ARC número n. El factor de precisión de una ARC viene dado por la Fórmula A.10.

$$FP_{ARCn} = valor_máx\{FP_{CPU}; FP_{RAM}; FP_{ALMcap}; FP_{ALMthr}; FP_{RED}\} \quad (A.10)$$

En donde FP_{CPU} , FP_{RAM} , FP_{ALMcap} , FP_{ALMthr} y FP_{RED} , son los valores de precisión de cada uno de los subsistemas de capacidad de una ARC. El FP de cada subsistema es el resultado de aplicar la Fórmula A.11.

$$FPRC = Capacidad\ lograda - Capacidad\ estimada \quad (A.11)$$

Mientras más cercano a cero (valor=1) el resultado, más preciso fue el diseño logrado. De ser negativo el resultado, el diseño debe ser considerado insuficiente, dado un subdimensionamiento, y ha de realizarse un rediseño. De ser positivo, y la diferencia ser menor o igual a:

- el 25% de la capacidad estimada, la evaluación de la precisión debe ser evaluada de Muy buena; (Valor 0,25)
- el 50% de la capacidad estimada, la evaluación de la precisión debe ser evaluada de Aceptable. (Valor 0,5)

²² O existente, si se fuese a caracterizar un sistema inicial. De ser el caso debe ser evaluado si el HW heredado es suficiente o no en cuanto a la capacidad para el soporte de los servicios a desplegar.

- Un valor superior arroja una evaluación de mala precisión, y debe considerarse un rediseño por sobredimensionamiento. (Valor 0)

Atributo de Utilización

Indica los índices de utilización de cada uno de los subsistemas fundamentales que componen a un nodo de cómputo: CPU, RAM, almacenamiento y red. Los índices de utilización deben ser monitorizados a nivel de clúster o ARC, así como en cada uno de los nodos.

Métricas y Procedimiento de Evaluación

Las métricas que se consideran mínimas necesarias a monitorizar, para poder dimensionar dinámicamente y/o estimar las necesidades de cómputo en función de la demanda se muestran en la Tabla A1. Debe conocerse el número de nodos totales en el clúster además de las características y ubicación de cada nodo como muestra el [Excel Prestaciones nodos de cómputo](#).

Tabla A.1. Métricas primarias a monitorizar a nivel de clústeres y nodos.

Métrica	Unidades de medida	Significado	Herramientas para su medición
CPU_total_clúster	MHz	Total de ciclos de máquina que presenta el clúster-	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
%utilización_CPU_clúster	%mínimo %máximo %promedio P95	% de utilización de CPU del clúster.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
RAM_total_clúster	GB	Total de capacidad de RAM que presenta el clúster.	Gestor, herramientas de configuración y desempeño propias

			de cada subsistema o generales como Zabbix.
%utilización_RAM_clúster	%mínimo %máximo %promedio P95	% de utilización de la capacidad de la RAM del clúster dedicada a soportar servicios de usuarios.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
Alma_cap_total_local_clúster	GB	Sumatoria del almacenamiento local de los nodos.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
Alma_IOPS_total_local_clúster	IOPS	Sumatoria de los IOPS de los nodos del clúster.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
Alma_throughput_total_local_clúster	Mbps	Sumatoria del throughput de L/E de los nodos del clúster.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
%utilización_cap_Alma_clúster	GB	% de utilización de la capacidad de almacenamiento local del clúster.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
%utilización_iops_Alma_clúster	%mínimo %máximo %promedio P95	% de utilización de la capacidad de IOPS del clúster.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
%utilización_throughput_Alma_clúster	%mínimo %máximo %promedio P95	% de utilización de la capacidad de L/E de los nodos del clúster al SA distribuido.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
Red_cap_total_clúster	Mbps	Capacidad de acceso a la red del clúster, TX/RX.	Gestor, herramientas de configuración y desempeño propias

			de cada subsistema o generales como Zabbix.
%utilización_red_clúster	%mínimo %máximo %promedio P95	% de utilización de la capacidad de acceso a la red del clúster, TX/RX.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix, lperf.

A su vez en cada nodo del clúster deben ser monitorizadas las métricas que muestra la Tabla A.2.

Tabla A.2. Métricas obligatorias a monitorizar en los nodos de cómputo.

Métrica	Unidades de medida	Significado	Herramientas para su medición
CPU_total_nodo	MHz	Total de ciclos de máquina que presenta el nodo.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
%utilización_CPU_nodo	%mínimo %máximo %promedio P95	% de utilización de CPU del nodo.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
RAM_total_nodo	GB	Total de capacidad de RAM que presenta el nodo.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
%utilización_RAM_nodo	%mínimo %máximo %promedio P95	% de utilización de la capacidad de la RAM del nodo.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
Alma_cap_local_nodo	GB	Capacidad de almacenamiento local en el nodo.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.

Alma_IOPS_local_nodo	IOPS	Capacidad de IOPS de un nodo.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
Alma_BW_local_nodo	Mbps	Capacidad de L/E al almacenamiento distribuido de un nodo.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
%utilización_cap_Alma_nodo	GB	% de utilización de la capacidad de almacenamiento local del nodo.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
%utilización_iops_Alma_nodo	%mínimo %máximo %promedio P95	% de utilización de la capacidad de IOPS del nodo.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
%utilización_throughput_Alma_clúster	%mínimo %máximo %promedio P95	% de utilización de la capacidad de L/E de los nodos del clúster al SA distribuido.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
Red_cap_total_nodo	Mbps	Capacidad de acceso a la red del clúster, TX/RX.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix.
%utilización_red_nodo	%mínimo %máximo %promedio P95	% de utilización de la capacidad de acceso a la red del nodo, TX/RX.	Gestor, herramientas de configuración y desempeño propias de cada subsistema o generales como Zabbix, lperf.

Atributo de Throughput

Posee la misma definición, métricas y pruebas que el Throughput definido en “[RNE pruebas NP/CDV](#)”.

Atributo de Tiempo de Respuesta y Demoras

Posee la misma definición, métricas y pruebas que el Throughput definido en “[RNE pruebas NP/CDV](#)”.

Atributo de eficiencia energética

Posee la misma definición, métricas y pruebas que el Throughput definido en “[RNE pruebas NP/CDV](#)”.

Categoría de disponibilidad

Indica la confiabilidad, tolerancia a fallos y facilidades ante la recuperación ante fallos de los nodos de cómputo de la infraestructura de la NP/CDV. Posee los atributos: % de servicio activo, confiabilidad, tolerancia ante fallos y recuperación ante fallos. Para integrar el atributo al sistema general que evalúa a los recursos de cómputo se debe emplear la métrica general Disponibilidad_{RC}, a través de la Fórmula A.12, que permite una escala de 0 a 1.

$Disponibilidad_{SA} =$

$$\frac{\% \text{ Servicio Activo} + \text{Confiabilidad} + \text{Tolerancia a Fallos} + \text{Recuperación ante Fallos}}{4} \quad (\text{A.12})$$

Atributo de “Porcentaje de servicio activo”

Se adoptó la propuesta de plantilla de métricas para el porcentaje de servicio activo propuesto por el Instituto Nacional de Estándares y Tecnología (NIST²³) en [9]. La Figura A.1 muestra un diagrama de la plantilla de la métrica propuesta por el NIST en [9]. Debe ser evaluado de satisfactorio (1) o no (0).

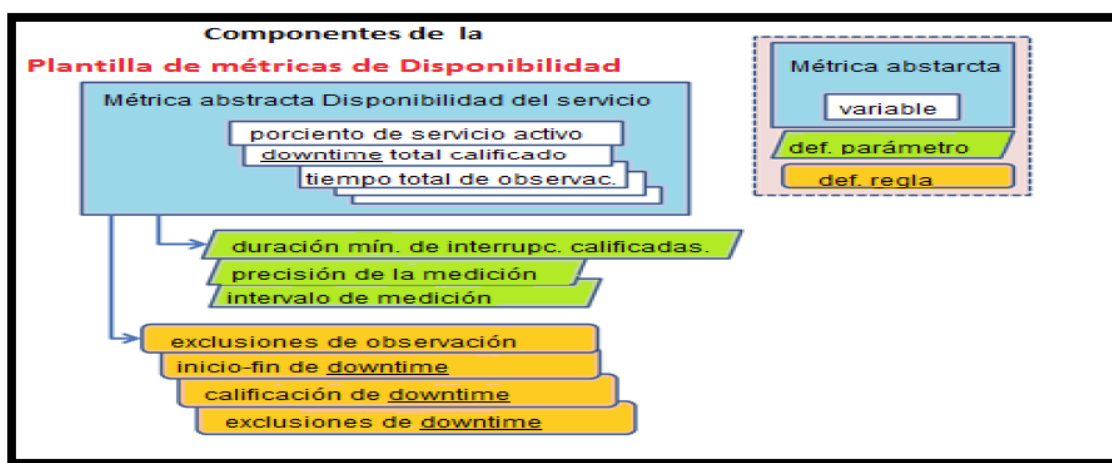


Figura A.1. Plantilla de la métrica de Disponibilidad del NIST [9].

La Tabla A.3 muestra una definición precisa de la plantilla de la métrica.

Tabla A.3. Definición de variables, parámetros y reglas de la plantilla en [9].

Definición de variables primarias		
nombre	tipo	expresión
porcentaje de servicio activo	porcentaje	$= (\text{tiempo total de observación} - \text{downtime total calificado}) * 100 / \text{tiempo total de observación}$
Definición de variables secundarias		
tiempo total de observación	duración	= sum (intervalos de observación)
downtime total calificado	duración	= sum (intervalos <u>downtime</u> calificados)
intervalos de observación	intervalos de tiempo	= (aplicar (reglas de exclusiones de observación) a intervalo de medición)
intervalos de <u>downtime</u> calificados	intervalos de tiempo	= intervalos de observación intersecto (aplicar (reglas de

²³ Siglas correspondientes al término en inglés: National Institute of Standards and Technology.

		exclusiones de <u>downtime</u>) a intervalos de <u>downtime</u> reales)
intervalos de <u>downtime</u> reales	intervalos de tiempo	= aplicar (reglas de calificación de <u>downtime</u> , reglas de inicio-fin de <u>downtime</u>) a intervalo de medición
Definición de parámetros		
intervalo de medición	intervalos de tiempo	Define el(los) período(s) real(es) de tiempo usado(s) para calcular el porcentaje de disponibilidad. Parámetro <i>dinámico</i> (se instancia en el momento de evaluación de la métrica)
precisión de la medición	duración	Cantidad mínima de tiempo usada en todas las mediciones y que expresa la precisión o granularidad de las medidas (expresada en segundos). Parámetro <i>estático</i> (se instancia en el momento de definición de la métrica)
duración mínima de las interrupciones calificadas	duración	Define la duración mínima, bajo la cual un <u>downtime</u> no será calificado. Parámetro <i>estático</i> (se instancia en el momento de definición de la métrica)
Definición de reglas		
exclusiones de observación	intervalos de observación	Reglas de este tipo determinan qué períodos de tiempo no cuentan como parte del periodo de observación (ej.: mantenimientos programados)
inicio-fin de <u>downtime</u>	intervalos de <u>downtime</u> reales	Reglas de este tipo determinan como identificar el principio y fin de un <u>downtime</u> .
calificación de <u>downtime</u>	intervalos de <u>downtime</u> reales	Reglas de este tipo definen la naturaleza precisa de un <u>downtime</u> .
exclusiones de <u>downtime</u>	intervalos de <u>downtime</u> calificados	Estas reglas determinan problemas en los servicios que no califican como <u>downtimes</u> (ej.: problemas causados por un evento externo incontrolable como un terremoto, o pérdidas del servicio por parte del cliente). Una de estas reglas a menudo establece una duración mínima para el <u>downtime</u> , por debajo de la cual los <u>downtime</u> no se cuentan (ej.: un <u>downtime</u> debe tardar al menos 10 min. Este tiempo mínimo es contemplado por un parámetro: duración mínima de las interrupciones calificadas).

De la Tabla A.3 se debe señalar la diferencia entre las variables y los parámetros. Una variable está asociada con una expresión que calcula su valor durante la aplicación de la métrica, mientras que un parámetro es un rango para un valor que debe ser determinado antes de que la métrica sea realmente usada. Puede ser establecido en el momento en que la métrica es creada de la plantilla en caso de que el valor sea el mismo para todas las aplicaciones. Ese tipo de parámetros son conocidos como parámetros estáticos. De otra manera puede ser establecido en cada aplicación de la métrica si este cambia de una aplicación a otra. Este tipo de parámetros se conoce como parámetros dinámicos. Otro aspecto a señalar es el uso de operadores genéricos, como el operador **sum** de la suma, el operador **aplicar** para la aplicación de reglas y el operador **interseccion** correspondiente a la operación matemática entre conjuntos intersección. [9]

La Figura A.2 muestra el conjunto de definiciones brindadas. En ella se aprecia como los intervalos de downtime calificados son el resultado de la intersección de los intervalos de observación con los intervalos de downtime reales una vez que les han sido aplicadas a estos últimos las reglas de exclusiones de downtime.

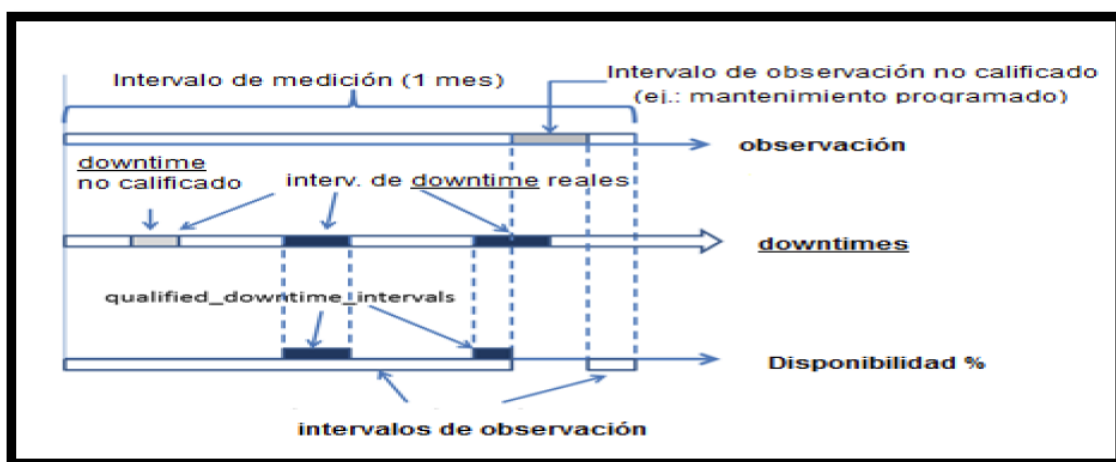


Figura A.2. Intervalo de medición con parámetros y reglas aplicadas.

Para cada nodo de cómputo debe ser identificado su “porcentaje de servicio activo”, para posteriormente calcular el promedio de todos los nodos.

Atributo de “Confiabilidad”

Indica la capacidad del nodo de cómputo de operar sin fallas bajo condiciones dadas durante un período de tiempo determinado [10] [6] [11] [1] [12]–[14]. Se propone como métrica para su evaluación el Tiempo Medio entre Fallos (MTBF²⁴). Para cada nodo de cómputo debe ser identificado su “MTBF”, para posteriormente calcular el promedio de todos los nodos. Debe ser evaluado de satisfactorio (1) o no (0).

Atributo de “Tolerancia ante fallos”

Indica la capacidad de un nodo de cómputo para continuar operando adecuadamente ante la ocurrencia de fallos planificados o no, en uno o más componentes. Se propone que en cada nodo sea evaluado el Nivel de Redundancia de sus Componentes (NRC), en donde son considerados esenciales la redundancia en: las Unidades de Fuentes de Alimentación (PSU²⁵), elementos de enfriamiento como extractores, y discos de almacenamiento locales. La evaluación del NRC de un nodo será de:

- Excelente, valor numérico 1: si cumple con la redundancia en todos los componentes especificados.
- Bueno, valor numérico 0,75: si cumple con la redundancia en dos de los componentes especificados.

²⁴ Siglas correspondientes al término en inglés: Mean time Between Failures.

²⁵ Siglas correspondienteados s al término en Inglés: Power Supply Units.

- Regular, valor numérico 0,50: si cumple con la redundancia en uno de los componentes especificados.
- Malo, valor numérico 0: no cumple con la redundancia en ninguno de los componentes especificados

La evaluación final de la Tolerancia a Fallos de los nodos de cómputo será el promedio de los valores de NRC de los nodos de cómputo de la infraestructura.

Atributo de “Recuperación ante Fallos”

Indica cuán rápido la capacidad de los nodos de cómputo de continuar operando adecuadamente ante la ocurrencia de fallos planificados o no, en uno o más de sus componentes. Se propone sea evaluado a través de la métrica “Tiempo Promedio para la Reparación” de fallos (MTTR), que viene dada por la Fórmula A.13.

$$MTTR = \frac{\sum_{i=1}^n TTRn}{n} \quad (A.13)$$

En donde:

TTR, Tiempo para la Restauración del fallo, como indica la Fórmula A.14: Diferencia en tiempo entre el momento de restauración y ocurrencia del fallo:

$$TTR = \text{Tiempo de Reacción} + \text{Tiempo de Recuperación} \quad (A.14)$$

En donde:

Tiempo de reacción: Demora entre la ocurrencia de un fallo y la primera reacción del sistema de gestión de la disponibilidad.

Tiempo de recuperación: tiempo desde la primera reacción hasta que el componente dañado vuelva a estar activo nuevamente.

Este atributo debe ser evaluado de satisfactorio (1) o no (0).

Dimensión de Factibilidad

Indica el grado de factibilidad de los nodos de cómputo desde las perspectivas de: costos, facilidades de gestión y robustez de la solución.

Categoría de Facilidades de Gestión

Indica el soporte de herramientas de gestión SLCA en los nodos de cómputo que facilitan la administración a los administradores de TI.

Atributo de facilidad de gestión con herramientas SLCA

Se propone la métrica “Nivel de Facilidades de Gestión con SLCA (NFG_{SLCA})”. Su valor será cuantitativo y se encontrará en función del soporte de las siguientes características y RF²⁶ por parte del sistema de gestión de los nodos de cómputo:

- Libre
- Abierto
- Out of Band (OOB)
- Dentro de banda
- Soporte para la gestión remota
- Con capacidades de integración con herramientas de gestión SLCA.
- Con capacidades para conectarse al gestor de tipo de SLCA.

La Fórmula A.15 indica la evaluación final del NFG_{SLCA} tomando en cuenta las características de los nodos pertenecientes a cada ARC.

$$NFG_{SLCA} = \min \{NFG_{SLCA_{ARCn}}\} \quad (A.15)$$

En donde:

²⁶ Pueden ser agregados nuevos requerimientos en función de los intereses y necesidades del cliente.

NFG_SLCA_{ARCn} : es el NFG_{SLCA} de cada ARC. Tomará un valor de:

- Muy alto: si se cumple con el 90% de los requerimientos.
- Alto: si se cumple con el 80% de los requerimientos.
- Normal: si se cumple con el 70% de los requerimientos.
- Bajo: si se cumple con el 60% de los requerimientos.
- Muy bajo: si se cumple con menos del 60% de los requerimientos.

Categoría de Robustez de la Solución

Indica el grado de calidad y soporte que tienen los nodos de cómputo de la infraestructura de la NP/CDV. [1]

Atributo de “Consolidación de los Proveedores”

Indica el grado de prestigio y posicionamiento en el mercado de los proveedores, y por consiguiente la calidad de sus productos. Se propone sea identificado el lugar que ocupan los proveedores de los diferentes nodos de cómputo de la infraestructura de la NP/CDV en el mercado, en consultoras reconocidas internacionalmente como Gartner, IDC y Forbes; y las posiciones que han ocupado tres años atrás. De acuerdo a los resultados debe emitirse por ARC una evaluación del prestigio y posicionamiento en el mercado de sus proveedores de: Alto (0,90), Medio (0,70) y Bajo (0,50); para posteriormente emitir la evaluación a nivel de infraestructura.

Atributo de “Soporte”

Indica el tipo y tiempo de garantía, así como el tipo y tiempo de soporte de los nodos de cómputo de la infraestructura. En cada ARC debe evaluarse el tipo de garantía

y soporte de los diferentes modelos de nodos físicos, y emitirse una evaluación de Bueno (0,90), Medio (0,70) y Bajo (0,50), para posteriormente emitir un criterio general del soporte de la infraestructura de cómputo.

Respecto a la garantía deben ser tomados en cuenta los criterios:

- Tipo de garantía: reposición de piezas, equipo y/o envío de especialistas para reparar el desperfecto?
- Tiempo de garantía.

Respecto al soporte deben ser tomados en cuenta los criterios:

- Documentación oficial disponible.
- Tipos de soporte comercial y sus costos.
- Tiempo de soporte aún disponible por parte del proveedor al equipamiento en cuestión. Se recomienda que el equipamiento se encuentre en sus primeros tres años de soporte.

Categoría de Factibilidad Económica

Define los costos totales de inversión, directos e indirectos a evaluar ante la selección de una solución. Se consideran indispensables a evaluar:

- Costos totales de inversión: compra del equipamiento.
- Costos de producción total anual:
 - Costos de soporte.
 - Costos de mantenimiento.
 - Costos de Operación. En el caso de los recursos de cómputo se mide en consumo de energía.
 - Costos de entrenamiento y capacitación.

Debe serle asignado un valor de 1 a este atributo, si se cumple con el presupuesto asignado, de lo contrario será de 0.

Anexo A.1. Requerimientos funcionales obligatorios y opcionales a soportar por los nodos de cómputo

La Tabla A.1.1 muestra los requerimientos obligatorios y opcionales²⁷ que deben cumplir los nodos físicos.

Tabla A.1.1. Requerimientos de los nodos de cómputo.

Requerimiento Funcional	Especificidades	Clasificación	
		Obligatorio	Opcional
CPU:	Soporte de tecnología <u>hyperthreading</u> .	*	
	<u>Turbo Boost</u> .	*	
	Funcionalidades específicas para la virtualización:		
	- Virtualización Asistida por Hardware (HAV ²⁸) ²⁹	*	
	<i>Mecanismos para el soporte de la virtualización anidada:</i>		
	- VMCS <u>shadowing</u> ³⁰		*
RAM:	NUMA		*
	Funcionalidades específicas para la virtualización:		
	- RVI/EPT	*	
	<i>Mecanismos para el soporte de la virtualización anidada:</i>		
	- EPT anidada.		*
Red y NIC:	Gestión Fuera de Banda (OOB ³¹):		*
	- Compatible con estándares como:	*32	

²⁷ Pero deseables.

²⁸ Siglas correspondientes al término en inglés: Hardware Assisted Virtualization.

²⁹ Ya sea: VT-x para chips de Intel o AMD-v para chips AMD.

³⁰ Funcionalidad que permite a un hipervisor anidado acceder a las extensiones de virtualización del procesador directamente, lo cual mejora el desempeño de las MV anidadas. [15]

³¹ Siglas correspondientes al término en inglés: Out of Band.

³² De ser considerada la OOB en el diseño.

	IPMI, DCMI y SMASH-DTMF.		
	- Soporte para la integración con herramientas de gestión de redes.		*
	- Soporte para la integración con CMP.		*
	Funcionalidades a soportar:		
	- Tramas Jumbo		*
	- <u>IPsec Task Offload</u> ³³		*
	- <u>Stateless offload</u>		*
	- <u>TCP Offload Engine (TOE)</u> ³⁴		*
	Funcionalidades específicas para la virtualización:		
	- <u>Remote Direct Memory Access (RDMA)</u> :		*
	o RoCE		*
	o InfiniBand		*
	o iWARP		*
	- SR-IOV		*
	- <u>Receive-Side Scaling (RSS)</u> ³⁵		*
	- <u>Transmit-Side Scaling (TSS)</u>		*
Almacenamiento:	Tipo de discos: ³⁶	*	
	- HDD	*	
	o Sector de 512B	*	
	o Sector de 4kB ³⁷		*
	- SSD		*
	- Discos híbridos.		*
	Velocidad Rotacional del Disco ³⁸ (RPM) ³⁹ :		
	o 10k RPM	*	
	o 15k RPM		*
	Interfaces de E/S:	*	
	- SDI:		*
	o Infiniband		*

³³ IPsec protege la red mediante la autenticación y encriptación de todos o determinados paquetes. IPsec Task Offload utiliza las capacidades de HW de las NIC de los servidores para descargar el procesamiento generado por el IPsec. Esto reduce el overhead del CPU generado por la encriptación/descriptación del Ipsec. [16]

³⁴ TCP Chimney descarga el procesamiento de la transferencia de datos del protocolo TCP a las NIC. [16]

³⁵ Receive-Side Scaling (RSS) distribuye las interrupciones sobre los diferentes procesadores, por lo que un solo procesador no tiene que manejar todas las interrupciones de E/S. [16]

³⁶ Tiene que ser especificado el tipo de disco. Los considerados obligatorios son los básicos, pueden ser sustituidos por los opcionales, pero son más costosos.

³⁷ Mejor capacidad de corrección de errores y por tanto mejor razón de señal/ruido.

³⁸ Siglas correspondientes al término en inglés: Rotational Speed of the Drive.

³⁹ Tiene que ser especificado el RPM. Los considerados obligatorios son los básicos, pueden ser sustituidos por los opcionales, pero son más costosos.

	○ Ethernet	*40	
	- Convergentes:	*	
	○ iSCSI	*	
	○ FCoE		*
	- No convergentes:		
	○ FC		*
	Arquitectura de controladores de discos:		
	- SATA III HBA	*	
	- SAS HBA	*	
	- PCIe/SAS HBA		*
	- PCIe RAID/Clúster RAID		*
	- FC HBA		*
	Arquitectura de la interfaz:		
	- SAS	*	
	- SATA	*	

Anexo B. Servidores de referencia en las gamas: baja, media y alta

A continuación, se proponen los nodos de referencia para las tres gamas: baja, media y alta⁴¹:

Gama baja:

- Procesamiento:
 - 1 socket.
 - CPU con 6 núcleos y una frecuencia de 3,8 GHz, o con como promedio un CPU con 22,4 GHz total.
 - Modelo: Intel® Xeon® E Processors, siempre el procesador de mayor actualidad y soporte. En el momento en que se escribe el documento se proponen los modelos: Intel® Xeon® E-2186G e Intel® Xeon® E- 2174G [17].
- RAM:

⁴⁰ De existir una SDI, preferiblemente la E/S al SA debe ser con tecnología Ethernet, Infiniband es muy costosa.

⁴¹ Esta propuesta es realizada en el momento en que se realiza la propuesta del procedimiento y debe tener una vida útil de tres años, en la que obligatoriamente debe ser actualizada. No obstante, de ser posible, puede ser actualizada anualmente.

- Capacidad de 64 GB.

- Red:

La autora considera que el número de NIC y sus capacidades deben encontrarse en función de la demanda, y el número de nodos resultantes por concepto de RAM y CPU; por lo que se propone sea calculado empleando la Fórmula B.1:

$$\text{Número de NIC por servidor} = \frac{BW \text{ identificado como demanda}}{\text{Número de nodos por concepto de CPU y RAM}} \quad (B.1)$$

Los valores deben ser siempre redondeados al número entero superior. Se propone en post de alcanzar un buen nivel de disponibilidad y evitar la competencia entre los distintos tipos de tráfico:

- Como máximo emplear 6 NIC x 1Gbps. Deben ser dedicadas dos NIC a cada uno de los tráficos: el de las instancias virtuales, el de gestión y el de almacenamiento. Las dos NIC dedicadas a cada tipo de tráfico deben ser agrupadas en un mismo enlace lógico.
 - En caso de requerir un número menor de NIC a 1Gbps se pueden emplear tecnologías de agregación de enlaces y VLAN para aislar los diferentes tipos de tráfico.
 - Como mínimo emplear 2 NIC x 1Gbps agrupadas en un mismo enlace lógico, y VLAN para el aislamiento de tráfico.
 - Emplear NIC a 10 Gbps si se requiere por nodo más de 6 NIC x 1Gbps.
- Almacenamiento:

Se propone en este subsistema:

- Para almacenar el SO, y en configuración RAID 1 en post de la disponibilidad:
2 x Enterprise Class SAS HDD 2.5" de 300 GB 10K rpm.

- Para el almacenamiento de datos en una infraestructura homogénea: 16 TB = 4 x Midline Class SAS HDD 3,5" de 4 TB 7,2K rpm. Esta configuración se propone en post de garantizar capacidad, confiabilidad, rápida recuperación ante fallos y desempeño para las aplicaciones/servicios críticos, manteniendo una adecuada relación beneficio/costo.
- Para el almacenamiento de datos basado en Tiers, en el caso que sea factible, se proponen las configuraciones que se muestran en la Tabla B.1.

Tabla B.1. Configuraciones para almacenamiento basado en Tiers, gama baja

Tier 0	Tier 1	Tier 2	Tier 3
SAS SSD 2,5":	Enterprise SAS HDD 10K rpm 2,5":	Midline SAS HDD 7,2K rpm 3,5":	Midline SATA HDD 7,2K rpm 3,5":
3,2TB 8 x 400 GB	19,2 TB 8 x 2,4 TB	112 TB 8 x 14 TB	112 TB 8 x 14 TB
32TB 8 x 4 TB	14,4 TB 6 x 2,4 TB	56 TB 4 x 14 TB	56 TB 4 x 14 TB
32TB 16 x 2 TB		32 TB 8 x 4 TB	32 TB 4 x 8 TB
91,8TB 6 x 15,3 TB		24 TB 2 x 12 TB	32 TB 8 x 4 TB
o		16 TB 4 x 4 TB	24 TB 2 x 12 TB
Discos SSD NVMe			

Gama Media:

- Procesamiento:
 - 1 socket.
 - CPU con como promedio 18 núcleos a 2,15 GHz, o con como promedio un CPU con 38,7 GHz.
 - Modelo: Intel® Xeon® Scalable, siempre el procesador de mayor actualidad y soporte. En el momento en que se escribe el documento y en función de lo propuesto en el punto anterior se proponen los modelos: Intel® Xeon® Gold 6138T Processor, Intel® Xeon® Gold 6138P Processor, Intel® Xeon® Gold 6138F Processor o Intel® Xeon® Gold 6138 Processor [18].

- RAM:
 - Capacidad con como promedio 406 GB, no obstante, las configuraciones estándares más próximas a este valor son $6 \times 64 \text{ GB} = 384 \text{ GB}$ y $16 \times 32 \text{ GB} = 512 \text{ GB}$, siendo esta última capacidad de RAM la que propone la autora de la presente investigación como referencia al dimensionar. Se recalca de igual forma que las configuraciones con capacidad máxima identificadas en esta gama fueron: 192, 256 y 768.
- Red: se propone que se emplee el mismo mecanismo de dimensionamiento propuesto en la Gama Baja.
- Almacenamiento:

Se propone en este subsistema:

- Para almacenar el SO, y en configuración RAID 1 en post de la disponibilidad: 2 x Enterprise Class SAS HDD 2.5" de 300 GB 10K rpm.
- Para el almacenamiento de datos en una infraestructura homogénea se propone como promedio: $48 \text{ TB} = 8 \times \text{Midline Class SAS HDD } 3,5" \text{ de } 6 \text{ TB } 7,2\text{K rpm}$. Esta configuración se propone en post de garantizar capacidad, confiabilidad, rápida recuperación ante fallos y desempeño para las aplicaciones/servicios críticos, manteniendo una adecuada relación beneficio/costo.
- Para el almacenamiento de datos basado en Tiers, en el caso que sea factible, se proponen las configuraciones que se muestran en la Tabla B.2.

Tabla B.2. Configuraciones para almacenamiento basado en Tiers, gama media

Tier 0	Tier 1	Tier 2	Tier 3
--------	--------	--------	--------

SAS SSD 2,5”:	Enterprise SAS HDD 10K rpm 2,5”:	Midline SAS HDD 7,2K rpm 3,5”:	Midline SATA HDD 7,2K rpm 3,5”:
15 TB 16 x 960 GB	32 TB 16 x 2 TB	40 TB 4 x 10 TB	40 TB 4 x 10 TB
61,44 TB 16 x 3,4 TB	38,4 TB 16 x 2,4 TB	80 TB 8 x 10 TB	80 TB 8 x 10 TB
		96 TB 8 x 12 TB	96 TB 8 x 12 TB
o			
Discos SSD NVMe			

Gama Alta:

- Procesamiento:
 - o 1 socket.
 - o CPU con como promedio 20 núcleos a 3,0 GHz, o CPU con un total como promedio de 50,4 GHz.
 - o Modelo: Intel® Xeon® Scalable, siempre el procesador de mayor actualidad y soporte. En el momento en que se escribe el documento y en función de lo propuesto en el punto anterior se proponen los modelos: Intel® Xeon® Platinum 8160T Processor, Intel® Xeon® Platinum 8160F Processor o Intel® Xeon® Platinum 8160 Processor [18].
- RAM:
 - o Capacidad 1 TB.
- Red: se propone que se emplee el mismo mecanismo de dimensionamiento propuesto en la Gama Baja.
- Almacenamiento:

Se propone en este subsistema:

 - o Para almacenar el SO, y en configuración RAID 1 en post de la disponibilidad: 2 x Enterprise Class SAS HDD 2.5” de 300 GB 10K rpm.

- Para el almacenamiento de datos en una infraestructura homogénea se propone como promedio: 84 TB = 14 Midline Class SAS HDD 3,5" de 6 TB 7,2K rpm, con una capacidad que permita ganar en capacidad, confiabilidad, rápida recuperación ante fallos y desempeño para las aplicaciones/servicios críticos, manteniendo una adecuada relación beneficio/costo.
- Para el almacenamiento de datos basado en Tiers, en el caso que sea factible, se proponen las configuraciones que se muestran en la Tabla B.3.

Tabla B.3. Configuraciones para almacenamiento basado en Tiers, gama alta

Tier 0	Tier 1	Tier 2	Tier 3
SAS SSD 2,5":	Enterprise SAS HDD 10K rpm 2,5":	Midline SAS HDD 7,2K rpm 3,5":	Midline SATA HDD 7,2K rpm 3,5":
24 discos	24 discos	24 discos	24 discos
18 discos	18 discos	14 discos	14 discos
12 discos	12 discos		
0			
Discos SSD NVMe			

Anexo C. Ejemplo del cálculo para obtener el número de nodos en una ARC

Ejemplo del cálculo para obtener el número de nodos necesarios en una ARC:

Los requerimientos de una ARC en particular son:

Capacidad.ARC: CPU=500GHz, RAM=1980GB

Capacidad_{disp.picos}.ARC: CPU=10GHz, RAM=40GB

Capacidad a largo plazo: CPU=540GHz, RAM=2138,4GB

Las especificaciones de HW del servidor tomado como ejemplo se muestran en la Tabla C.1.

Tabla C.1. Especificaciones de HW del servidor ejemplo.

Atributo de HW	Especificación
Fabricante	HPE

Formato	1 RU
# de <u>sockets</u>	1
	4
# de núcleos por CPU (Intel)	(Intel® Xeon® E3-1200 v6 product family Intel®, o Core™ i3 Intel® Pentium® Intel®, o Xeon® E3-1200 v5 product family)
Caché	8MB L3 o 3MB L3
MHz por núcleo	4,0GHz
Total de GHz en el nodo	16GHz
Capacidad total de RAM en el nodo	64GB
# de adaptadores de red	4 x 1GHz

El cálculo del número de servidores para la Capacidad.ARC lo define la Fórmula C.1, redondeando siempre a un número entero superior.

$$Nodos_Capacidad_ARC = \max \left\{ \frac{Capacidad_ARC_CPU}{Total\ de\ GHz\ en\ el\ nodo} \mid \frac{Capacidad_ARC_RAM}{RAM\ total\ en\ el\ nodo} \right\} \quad (C.1)$$

En este ejemplo:

$$Nodos_{Capacidad_ARC} = \max \left\{ \frac{500GHz}{16GHz} \mid \frac{1980}{64} \right\} = \max \{31,25 | 30,94\} = 31,25 \approx 32$$

El cálculo del número de servidores para la Capacidad_{disp.picos}.ARC lo define igual Fórmula 5. En este ejemplo:

$$Nodos_{Capacidad_{disp.picos}.ARC} = \max \left\{ \frac{10GHz}{16GHz} \mid \frac{40}{64} \right\} = \max \{0,63 | 0,63\} = 0,63 \approx 1$$

El cálculo del número de servidores para la Capacidad_{LargoPlazo}.ARC lo define igual Fórmula 5. En este ejemplo:

$$Nodos_{Capacidad_{LargoPlazo}.ARC} = \max \left\{ \frac{540GHz}{16GHz} \mid \frac{2138,4GB}{64GB} \right\} = \max \{33,75 | 33,41\} = 33,75 \approx 34$$

En donde entonces se adquirirá tan solo un servidor más a largo plazo, $34 - 33 = 1$.

En relación a las capacidades de red, dígame, número de Tarjetas de Interfaces de Red (NIC) y su capacidad, se propone que se realice una vez se tenga el número de servidores necesarios como indica la Fórmula C.2. Se debería entonces sumar los requerimientos de E/S de red de los servicios a soportar, incluyendo el CMP y la

plataforma de virtualización, y sus requerimientos de E/S al subsistema de SA (Gbps), “BW identificado como demanda”.

$$\text{Número de NIC por servidor} = \frac{\text{BW identificado como demanda}}{\text{Número de nodos}} \quad (\text{C.2})$$

Los valores deben ser siempre redondeados al número entero superior. Se propone en post de alcanzar un buen nivel de disponibilidad y evitar la competencia entre los distintos tipos de tráfico:

- Como máximo emplear 6 NIC x 1Gbps. Deben ser dedicadas dos NIC a cada uno de los tráficos: el de las IV, el de gestión y el de almacenamiento. Las dos NIC dedicadas a cada tipo de tráfico deben ser agrupadas en un mismo enlace lógico.
- En caso de requerir un número menor de NIC a 1Gbps se pueden emplear tecnologías de agregación de enlaces y Red de Área Local Virtual (VLAN⁴²) para aislar los diferentes tipos de tráfico.
- Como mínimo emplear 2 NIC x 1Gbps agrupadas en un mismo enlace lógico, y VLAN para el aislamiento de tráfico.
- Emplear NIC a 10 Gbps si se requiere por nodo más de 6 NIC x 1Gbps

Referencias

- [1] F. Tarrau Prendes, L. R. García Perellada, y Garófalo Hernández, «PROPUESTA DE REQUERIMIENTOS NO FUNCIONALES PARA EL DISEÑO Y EVALUACIÓN DE NUBES PRIVADAS CON SOPORTE PARA INFRAESTRUCTURA COMO SERVICIO», en *18 Convención Científica de Ingeniería y Arquitectura.*, Palacio de Convenciones de La Habana, 25/11 2016, pp. 881-895.
- [2] Frank Ernesto Tarrau Prendes, Lilia Rosa García Perellada, y Alain Abel Garófalo Hernández, «PROPUESTA DE REQUERIMIENTOS NO FUNCIONALES PARA EL DISEÑO Y EVALUACIÓN DE NUBES PRIVADAS CON SOPORTE PARA INFRAESTRUCTURA COMO SERVICIO», Palacio de Convenciones de La Habana, 2016, pp. 881-895.

⁴² Siglas correspondientes al término en inglés: Virtual Local Area Network.

- [3] Martin Hosken, «Architecting a VMware vSphere® Compute Platform for VMware Cloud Providers™». 2018.
- [4] «OpenStack Docs: OpenStack Architecture Design Guide», 2018.
<https://docs.openstack.org/arch-design/> (accedido jul. 20, 2018).
- [5] K. Hwang, X. Bai, Y. Shi, M. Li, W. G. Chen, y Y. Wu, «Cloud Performance Modeling with Benchmark Evaluation of Elastic Scaling Strategies», *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, n.º 1, pp. 130-143, ene. 2016, doi: 10.1109/TPDS.2015.2398438.
- [6] «Microsoft Word - SMI_Overview_140701.docx - SMI_Overview_TwoPointOne1.pdf». Accedido: sep. 27, 2014. [En línea]. Disponible en: http://csmic.org/wp-content/uploads/2014/07/SMI_Overview_TwoPointOne1.pdf.
- [7] «Reference Architecture Red Hat Enterprise Linux OpenStack Platform». Lenovo PRESS, 2015.
- [8] «Challenges of Enterprise Cloud Services1 - Springer», abr. 04, 2013.
http://link.springer.com/chapter/10.1007%2F978-90-481-9846-7_4 (accedido abr. 04, 2013).
- [9] J. Durand, T. Rutt, y F. de Vault, «Cloud Computing Service Metric Templates Primer», National Institute of Standards and Technology, Special Publication 500-xxx, dic. 2014.
- [10] «CSMIC | Defining globally accepted measures for cloud services», nov. 20, 2014.
<http://csmic.org/> (accedido nov. 20, 2014).
- [11] J. Siegel y J. Perdue, «Cloud Services Measures for Global Use: The Service Measurement Index (SMI)», en *2012 Annual SRII Global Conference*, jul. 2012, pp. 411-415, doi: 10.1109/SRII.2012.51.
- [12] NIST, «NIST-Cloud Computing Synopsis and Recommendations». may 2012, Accedido: abr. 06, 2016. [En línea]. Disponible en: <http://csrc.nist.gov/publications/drafts/800-146/Draft-NIST-SP800-146.pdf>.
- [13] L. Badger, T. Grance, R. Patt-Corner, y J. Voas, «DRAFT Cloud Computing Synopsis and Recommendations», National Institute of Standards and Technology, U.S. Department of Commerce, Special Publication 800-146, may 2011.
- [14] «Microsoft Word - CloudFrameworkIR.docx - CloudFrameworkIR.PDF». Accedido: sep. 27, 2014. [En línea]. Disponible en: <http://www.nist.gov/itl/cloud/upload/CloudFrameworkIR.PDF>.
- [15] K. Razavi *et al.*, «Kangaroo: A Tenant-Centric Software-Defined Cloud Infrastructure», en *2015 IEEE*, 2015, p. 10, doi: 10.1109/IC2E.2015.19.
- [16] Microsoft, «Infrastructure-as-a-Service Product Line Architecture». 2016.
- [17] «Procesador Intel® Xeon® E», *Intel*.
<https://www.intel.com/content/www/xl/es/products/processors/xeon/e-processors.html> (accedido ago. 28, 2019).
- [18] «Intel® Xeon® Scalable Processors Product Specifications». <https://ark.intel.com/content/www/us/en/ark/products/series/125191/intel-xeon-scalable-processors.html> (accedido ago. 14, 2019).