

Anexo I - Landing Zone (F).xlsx	CUMPLE / NO CUMPLE	Página del sustento en la oferta.
6. CARACTERÍSTICAS Y CONDICIONES DE LOS BIENES A ADQUIRIR		
A continuación, se detalla como mínimo los siguientes componentes y características:		
6.2.1. Componente Data Lake		
Nodos de Gestión del Clúster de data lake		
Producción:		
Cantidad: 96 cores físicos		
Capacidad: 768 GB Memoria RAM		
Desarrollo:		
Cantidad: 32 cores físicos		
Capacidad: 256 GB Memoria RAM		
Calidad:		
Cantidad: 32 cores físicos		
Capacidad: 256 GB Memoria RAM		
Nodos de procesamiento de data lake		
Producción:		
Cantidad: 448 cores físicos		
Capacidad: 3.5 TB Memoria RAM		
Desarrollo:		
Cantidad: 32 cores físicos		
Capacidad: 256 GB Memoria RAM		
Calidad:		
Cantidad: 64 cores físicos		
Capacidad: 512 GB Memoria RAM		
Nodos de Utilitarios o Multipropósito**		
Producción:		
Cantidad: 64 cores físicos		
Capacidad: 512 GB Memoria RAM		
Desarrollo:		
Cantidad: 16 cores físicos		
Capacidad: 128 GB Memoria RAM		
Calidad:		
Cantidad: 16 Cores físicos		
Capacidad: 128 GB Memoria RAM		
Almacenamiento - Data Lake		
Producción:		
Capacidad: 420 TB Efectivos		
Desarrollo:		
Capacidad: 10 TB Efectivos		
Calidad:		
Capacidad: 30 TB Efectivos		
6.2.1.1. Nodos de Gestión de Clúster de data lake:		
Los nodos deberán ser rackeables de altura máximo 2 UR (unidades de rack)		
Las fuentes de alimentación deberán ser redundantes, de intercambio en caliente. HotSwap.		
Los ventiladores deberán ser redundantes N+N o N+1, de manera que la falla de un ventilador no deba impedir el normal funcionamiento del nodo ofertado.		
A. Producción:		
Almacenamiento de RAW de 14 TB en total por el conjunto de nodos		
Como mínimo los discos de 12 Gbps, de tipo SSD y reemplazables en caliente.		
Por cada nodo deberá incluir como mínimo:		
Procesador: Dos (02) procesadores de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 16 Cores como mínimo y de al menos 3.0 GHz por procesador.		
Memoria RAM: 256 GB de tecnología DDR4 o superior.		
Almacenamiento para sistema operativo:		
Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1.		
Los discos ofertados deben ser reemplazables en caliente.		
Conectividad:		
Un (01) puerto de 01Gbps Cobre (para gstión).		
04 puertos de 25 Gbps (fibra).		
Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.		
B. Calidad:		
Almacenamiento de RAW de 9 TB en total por el conjunto de nodos		
Como mínimo los discos de 12 Gbps, de tipo SSD y reemplazables en caliente.		
Por cada nodo deberá incluir como mínimo:		
Procesador: Un (01) procesador de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 16 Cores como mínimo y de al menos 3.0 GHz		
Memoria RAM: 128 GB de tecnología DDR4 o superior.		
Almacenamiento para sistema operativo:		
Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1.		
Los discos ofertados deben ser reemplazables en caliente.		
Conectividad:		
Un (01) puerto de 01Gbps Cobre (para gstión).		
04 puertos de 25 Gbps (fibra).		
Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.		
C. Desarrollo:		
Almacenamiento de RAW de 9 TB en total por el conjunto de nodos		
Como mínimo los discos de 12 Gbps, de tipo SSD y reemplazables en caliente.		
Por cada nodo deberá incluir como mínimo:		
Procesador: Un (01) procesador de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 16 Cores como mínimo y de al menos 3.0 GHz		
Memoria RAM: 128 GB de tecnología DDR4 o superior.		

Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.		
Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.		
6.2.1.2. Nodos de Utilitarios o Multipropósito		
Los nodos deberán ser rackeables de altura máximo 2 UR (unidades de rack)		
Las fuentes de alimentación deberán ser redundantes, de intercambio en caliente. HotSwap.		
Los ventiladores deberán ser redundantes N+N o N+1, de manera que la falla de un ventilador no deba impedir el normal funcionamiento del nodo ofertado.		
A. Producción:		
Almacenamiento de RAW de 14 TB en total por el conjunto de nodos		
Como mínimo los discos de 12 Gbps, de tipo SSD y reemplazables en caliente.		
Por cada nodo deberá incluir como mínimo:		
Procesador: Dos (02) procesadores de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 16 Cores como mínimo y de al menos 3.0 GHz por procesador		
Memoria RAM: 256 GB de tecnología DDR4 o superior.		
Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.		
Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.		
B. Calidad:		
Almacenamiento de RAW de 3 TB en total por el conjunto de nodos		
Como mínimo los discos de 12 Gbps, de tipo SSD y reemplazables en caliente.		
Por cada nodo deberá incluir como mínimo:		
Procesador: Un (01) procesador de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 16 Cores como mínimo y de al menos 3.0 GHz		
Memoria RAM: 128 GB de tecnología DDR4 o superior.		
Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.		
Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.		
C. Desarrollo:		
Almacenamiento de RAW de 3 TB en total por el conjunto de nodos		
Como mínimo los discos de 12 Gbps, de tipo SSD y reemplazables en caliente.		
Por cada nodo deberá incluir como mínimo:		
Procesador: Un (01) procesador de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 16 Cores como mínimo y de al menos 3.0 GHz		
Memoria RAM: 128 GB de tecnología DDR4 o superior.		
Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.		
Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.		
6.2.1.3. Nodos de procesamiento de data lake		
Los nodos deberán ser rackeables de altura máximo 2 UR (unidades de rack)		
Las fuentes de alimentación deberán ser redundantes, de intercambio en caliente. HotSwap.		
Los ventiladores deberán ser redundantes N+N o N+1, de manera que la falla de un ventilador no deba impedir el normal funcionamiento del nodo ofertado.		
A. Producción:		
Almacenamiento de RAW de 32.5 TB en total por el conjunto de nodos		
Como mínimo los discos de 12 Gbps, de tipo SSD y reemplazables en caliente.		
Por cada nodo deberá incluir como mínimo:		
Procesador: Dos (02) procesadores de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 32 Cores como mínimo y de al menos 2.7 GHz por procesador		
Memoria RAM: 512 GB de tecnología DDR4 o superior.		
Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. 2. Los discos ofertados deben ser reemplazables en caliente.		
Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.		
B. Calidad:		
Almacenamiento de RAW de 15 TB en total por el conjunto de nodos		
Como mínimo los discos de 12 Gbps, de tipo SSD y reemplazables en caliente.		
Por cada nodo deberá incluir como mínimo:		
Procesador: Dos (02) procesadores de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 16 Cores como mínimo y de al menos 3.0 GHz		
Memoria RAM: 256 GB de tecnología DDR4 o superior.		
Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.		

<p>Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.</p>		
<p>C. Desarrollo:</p>		
Almacenamiento de RAW de 15 TB en total por el conjunto de nodos		
Como mínimo los discos de 12 Gbps, de tipo SSD y reemplazables en caliente.		
Por cada nodo deberá incluir como mínimo:		
<p>Procesador: Un (01) procesador de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 16 Cores como mínimo y de al menos 3.0 GHz</p>		
<p>Memoria RAM: 128 GB de tecnología DDR4 o superior.</p>		
<p>Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.</p>		
<p>Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.</p>		
<p>6.2.1.4. Almacenamiento (Data Lake):</p>		
Se permitirán dos tipos de arquitecturas tecnológicas: Almacenamiento distribuido o Almacenamiento centralizado.		
Con una capacidad efectiva de 420TB para Producción, 20TB para Desarrollo y 30TB para Calidad.		
Este almacenamiento en su totalidad (data Lake) deberá soportar data estructurada, semiestructurada y no estructurada.		
<p>6.2.1.4.1. Arquitectura con Almacenamiento Distribuido:</p>		
La solución deberá incluir la implementación del sistema de almacenamiento distribuido, que permita un crecimiento a futuro, tanto en volumen como en rendimiento. Los discos de este almacenamiento podrán formar parte de los nodos de procesamiento o podrán ser nuevos nodos dedicados al almacenamiento distribuido.		
<p>Almacenamiento (Data): Discos de 7.68 TB SSD de 12 Gbps o NVMe. Deberán ser discos de 2.5” Deberá soportar discos de tipo SSD. Los discos ofertados deben ser reemplazables en caliente.</p>		
<p>6.2.1.4.2. Arquitectura con Almacenamiento centralizado:</p>		
Deberá incluir un sistema de almacenamiento del tipo Scale-Out de alta performance como repositorio de la información (Data Lake) que permita un crecimiento a futuro, tanto en volumen como en rendimiento.		
El sistema de almacenamiento deberá cumplir con las siguientes características mínimas:		
<p>Capacidad efectiva: La capacidad total usable deberá ser de al menos 470 TB (se distribuirá según la tabla 1). Con tolerancia a falla de un nodo y con tolerancia a falla simultánea de dos discos, sin pérdida de datos ni de servicio. La capacidad usable no toma en cuenta ratios de deduplicación ni compresión.</p>		
<p>Protección de Datos: Deberá incluir la capacidad de protección N+1 (algoritmo Reed Salomon o equivalente). Se aceptará también protección basada en “mirroring”.</p>		
<p>File System: La arquitectura de la solución deberá ser un único filesystem, o la inclusión e implementación de un único Global Namespace. Deberá incluir la abstracción de todas las informaciones sobre archivos, espacio para almacenamiento, “shares” integración con servicios de directorio, definiciones de permisos de acceso a los datos: en un “Global Namespace” con identidad única.</p>		
<p>Escalabilidad: Deberá tener la capacidad de crecer en tamaño a 20 PB (Petabyte) en un solo Filesystem y/o Name space con la agregación de nodos en modo no disruptivo y sin requerir cambios de configuración en los clientes del sistema de almacenamiento. Deberá contar con todo lo necesario para soportar el crecimiento mínimo requerido y garantizar la no interrupción del servicio durante la adición de componentes a la misma.</p>		
<p>Discos soportados: Deberá soportar discos SSD o NVMe</p>		
<p>Redistribución de datos: Ante el incremento en la cantidad de discos por la adición de componentes, la solución deberá incluir la capacidad de redistribuir la información existente de manera transparente para el usuario (sin afectar la continuidad operativa del servicio).</p>		
<p>Procesamiento: Deberá estar basado en arquitectura x86 con al menos 192 cores distribuidos entre todos los nodos.</p>		
<p>Memoria cache: Deberá incluir una memoria cache no menor a 3 TB de memoria cache nivel L1 (RAM) distribuida entre todos los nodos. La memoria cache deberá ser de lectura y escritura.</p>		
<p>Conectividad front-end: Deberá contar como mínimo con dos (02) puertos de 25 Gbps (fibra) y al menos 1 puerto de 1GbE RJ45 (gestión) por nodo o controladora del sistema de almacenamiento. Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.</p>		
<p>Protocolos soportados e incluido: Deberá soportar e incluir como mínimo el protocolo de almacenamiento de red: NFS, CIFS, HTTP, FTP, SNMP, LDAP, ADS y HDFS (Hadoop Distributed File System). Los protocolos HDFS o HTTP o FTP podrán ser soportados de manera nativa o no nativa.</p>		
<p>Cantidad y tipo de discos configurados: Deberá incluir discos no mayores de 16TB de capacidad en tecnología SSD.</p>		
<p>Conectividad back-end: La solución deberá incluir los elementos de red necesarios (puertos, switches, cables, entre otros) para garantizar que el movimiento de datos entre distintas tecnologías de discos, o redistribución entre misma tecnología de discos, se realice mediante una red de alta velocidad separada de la red de front-end, sin generar competencia del tráfico en la red y con un ancho de banda no menor a 100 GbE (fibra) por puerto de manera redundante.</p>		

<p>Balanceo de carga : La solución deberá incluir de un mecanismo propio o externo de balanceo de acceso de los clientes entre sus interfaces front-end de red, permitiendo la distribución de las cargas de I/O entre las mismas y las controladoras de la solución provista. La solución deberá incluir la capacidad de balancear en forma automática todos sus componentes, garantizando que dicho balanceo ocurra ante datos nuevos o cuando se agreguen componentes (controladoras, discos) al sistema, y que dicho balanceo se realice mediante una red independiente de la red de servicio (Front-end).</p>		
<p>Administración y monitoreo : Software de administración y monitoreo a través de una consola gráfica basada en web para administración local y remota. Deberá contar con la capacidad de crear niveles de acceso de usuario. Deberá disponer de una herramienta de monitoreo de utilización de sus componentes con almacenamiento de datos históricos.</p>		
<p>Funcionalidades: La solución deberá disponer de la capacidad de realizar Snapshots para los datos almacenados. Deberá incluir la capacidad de expansión del área útil de almacenamiento en caliente. La solución deberá soportar deduplicación. La solución deberá permitir la asignación de cuotas del File System.</p>		
<p>Seguridad de los datos : Deberá soportar la configuración de reglas WORM (write once, read many), de forma que se garantice por el sistema preservar el dato de ser borrado y/o alterado.</p>		
<p>Fuentes de Poder: Fuentes de alimentación redundantes, de intercambio en caliente a nivel del sistema de almacenamiento.</p>		
<p>Ventiladores: ventiladores redundantes de intercambio en caliente a nivel del sistema de almacenamiento, de manera que la falla de un ventilador no deba impedir el normal funcionamiento.</p>		
<p>6.2.2. Componente Machine Learning</p>		
<p>Se permitirán dos tipos de arquitecturas tecnológicas: Contenerizada o Clusterizada.</p>		
<p>6.2.1.1. Arquitectura Contenerizada</p>		
<p>Nodos de gestión servicios Contenerizados: 96 cores físicos / 768 GB de Memoria RAM</p>		
<p>Nodos de procesamiento de Machine Learning: 896 cores físicos / 7 TB de Memoria RAM</p>		
<p>6.2.2.1.1. Nodos de gestión servicios Contenerizados:</p>		
<p>Los nodos deberán ser rackeables de altura máximo 2 UR (unidades de rack).</p>		
<p>Las fuentes de alimentación deberán ser redundantes, de intercambio en caliente. HotSwap.</p>		
<p>Los ventiladores deberán ser redundantes N+N o N+1, de manera que la falla de un ventilador no deba impedir el normal funcionamiento del nodo ofertado.</p>		
<p>El Proveedor deberá incluir como mínimo un almacenamiento RAW de 14 TB en total por el conjunto de nodos, esta capacidad deberá ser distribuida como parte de la opción de almacenamiento que proponga el postor fuera de los nodos o de lo contrario deberá distribuirlo en forma equitativa en las bahías internas de los nodos. Como mínimo los discos ofertados deben ser de 12 Gbps, de tipo SSD y reemplazables en caliente. No se aceptará configuraciones que propongan cumplir lo solicitado con solo un par de discos por nodo.</p>		
<p>Por cada nodo deberá incluir como mínimo:</p>		
<p>Procesador: Dos (02) Procesadores de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 16 cores como mínimo y de al menos 3.0 GHz por procesador.</p>		
<p>Memoria RAM: 256 GB de tecnología DDR4 o superior.</p>		
<p>Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.</p>		
<p>Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.</p>		
<p>6.2.2.1.2. Nodos de procesamiento de Machine Learning</p>		
<p>Los nodos deberán ser rackeables de altura máximo 2 UR (unidades de rack).</p>		
<p>Las fuentes de alimentación deberán ser redundantes, de intercambio en caliente. HotSwap.</p>		
<p>Los ventiladores deberán ser redundantes N+N o N+1, de manera que la falla de un ventilador no deba impedir el normal funcionamiento del nodo ofertado.</p>		
<p>El Proveedor deberá incluir como mínimo un almacenamiento RAW de 43.5 TB en total por el conjunto de nodos, esta capacidad deberá ser distribuida como parte de la opción de almacenamiento que proponga el postor fuera de los nodos o de lo contrario deberá distribuirlo en forma equitativa en las bahías internas de los nodos. Como mínimo los discos ofertados deben ser de 12 Gbps, de tipo SSD y reemplazables en caliente. No se aceptará configuraciones que propongan cumplir lo solicitado con solo un par de discos por nodo.</p>		
<p>Por cada nodo deberá incluir como mínimo:</p>		
<p>Procesador: Dos (02) Procesadores de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 64 cores como mínimo y de al menos 2.0 GHz por procesador.</p>		
<p>Memoria RAM: 1 TB de tecnología DDR4 o superior.</p>		
<p>Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.</p>		
<p>Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.</p>		
<p>El postor podrá ofertar el componente de Machine Learning que forme parte del componente equivalente de data lake (nodo de base de datos y/o storage y/o servidor de gestión). Para este tipo de oferta, el postor deberá incrementar las capacidades de procesamiento solicitadas en la tabla 3. Para este caso, el almacenamiento de procesamiento de data local deberá ofertarse como mínimo 57.5 TB SSD.</p>		
<p>6.2.1.2. Arquitectura Clusterizada</p>		
<p>Nodos de Gestión: Produccion Cantidad / Capacidad: 2</p>		
<p>Nodos de Inferencia: Produccion Cantidad / Capacidad: 2</p>		
<p>Nodo de procesamiento de Machine Learning: Produccion Cantidad / Capacidad: 4</p>		
<p>A. Nodos de Gestión:</p>		

Se deberán cumplir con las siguientes características:		
<ul style="list-style-type: none"> Los nodos deberán ser rackeables de altura máximo 2 UR (unidades de rack). 		
<ul style="list-style-type: none"> Las fuentes de alimentación deberán ser redundantes, de intercambio en caliente. HotSwap. 		
<ul style="list-style-type: none"> Los ventiladores deberán ser redundantes N+N o N+1, de manera que la falla de un ventilador no deba impedir el normal funcionamiento del nodo ofertado. 		
<ul style="list-style-type: none"> El Proveedor deberá incluir como mínimo un almacenamiento RAW de 12 TB en total por el conjunto de nodos, esta capacidad deberá ser distribuida como parte de la opción de almacenamiento que proponga el postor fuera de los nodos o de lo contrario deberá distribuirlo en forma equitativa en las bahías internas de los nodos. Como mínimo los discos ofertados deben ser de 12 Gbps, de tipo SSD y reemplazables en caliente. No se aceptará configuraciones que propongan cumplir lo solicitado con solo un par de discos por nodo. 		
<ul style="list-style-type: none"> Por cada nodo deberá incluir como mínimo: 		
Procesador: Dos (02) Procesadores de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 64 cores como mínimo y de al menos 2.45 GHz por procesador.		
Memoria RAM: 512 GB de tecnología DDR4 o superior.		
Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.		
Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.		
B. Nodos de Inferencia:		
Se deberán cumplir con las siguientes características:		
<ul style="list-style-type: none"> Los nodos deberán ser rackeables de altura máximo 2 UR (unidades de rack). 		
<ul style="list-style-type: none"> Las fuentes de alimentación deberán ser redundantes, de intercambio en caliente. HotSwap. 		
<ul style="list-style-type: none"> Los ventiladores deberán ser redundantes N+N o N+1, de manera que la falla de un ventilador no deba impedir el normal funcionamiento del nodo ofertado. 		
<ul style="list-style-type: none"> El Proveedor deberá incluir como mínimo un almacenamiento RAW de 12 TB en total por el conjunto de nodos, esta capacidad deberá ser distribuida como parte de la opción de almacenamiento que proponga el postor fuera de los nodos o de lo contrario deberá distribuirlo en forma equitativa en las bahías internas de los nodos. Como mínimo los discos ofertados deben ser de 12 Gbps, de tipo SSD y reemplazables en caliente. No se aceptará configuraciones que propongan cumplir lo solicitado con solo un par de discos por nodo. 		
<ul style="list-style-type: none"> Por cada nodo deberá incluir como mínimo: 		
Procesador: Dos (02) Procesadores de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 64 cores como mínimo y de al menos 2.45 GHz por procesador.		
Memoria RAM: 512 GB de tecnología DDR4 o superior.		
Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.		
Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.		
C. Nodos de procesamiento de Machine Learning		
Se deberán cumplir con las siguientes características:		
<ul style="list-style-type: none"> Los nodos deberán ser rackeables de altura máximo 2 UR (unidades de rack). 		
<ul style="list-style-type: none"> Las fuentes de alimentación deberán ser redundantes, de intercambio en caliente. HotSwap. 		
<ul style="list-style-type: none"> Los ventiladores deberán ser redundantes N+N o N+1, de manera que la falla de un ventilador no deba impedir el normal funcionamiento del nodo ofertado. 		
<ul style="list-style-type: none"> El Proveedor deberá incluir como mínimo un almacenamiento RAW de 25 TB en total por el conjunto de nodos, esta capacidad deberá ser distribuida como parte de la opción de almacenamiento que proponga el postor fuera de los nodos o de lo contrario deberá distribuirlo en forma equitativa en las bahías internas de los nodos. Como mínimo los discos ofertados deben ser de 12 Gbps, de tipo SSD y reemplazables en caliente. No se aceptará configuraciones que propongan cumplir lo solicitado con solo un par de discos por nodo. 		
<ul style="list-style-type: none"> Por cada nodo deberá incluir como mínimo: 		
Procesador: Dos (02) Procesadores de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 64 cores como mínimo y de al menos 2.45 GHz por procesador.		
Memoria RAM: 512 GB de tecnología DDR4 o superior.		
Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.		
Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.		
GPU: Deberá contar con un GPU de al menos 48GB de memoria con un rendimiento TF32 Tensor Core no menor a 164 TFlops; o Dos GPUs de al menos 24 GB de memoria cada uno, interconectados con links de al menos 200GB/s, con un rendimiento TF32 no menor a 82 TFlops por cada GPU..		
6.2.3. Componente de Ingesta de Datos		
Nodos de ingesta de datos: Produccion Cantidad: 128 Cores Fisicos Capacidad: 512 GB RAM		
Nodos de ingesta de datos: Desarrollo Cantidad: 32 cores fisico Capacidad: 128 GB RAM		
Nodos de ingesta de datos: Calidad Cantidad: 64 cores fisicos Capacidad: 256 GB RAM		
6.2.3.1. Nodos de Ingesta de Datos		
<ul style="list-style-type: none"> Los nodos deberán ser rackeables de altura máximo 2 UR (unidades de rack). 		

<ul style="list-style-type: none"> Las fuentes de alimentación deberán ser redundantes, de intercambio en caliente. HotSwap. Los ventiladores deberán ser redundantes N+N o N+1, de manera que la falla de un ventilador no deba impedir el normal funcionamiento del nodo ofertado. 		
A. Producción		
El Proveedor deberá incluir como mínimo un almacenamiento RAW de 12 TB en total por el conjunto de nodos, esta capacidad deberá ser distribuida como parte de la opción de almacenamiento que proponga el postor fuera de los nodos o de lo contrario deberá distribuirlo en forma equitativa en las bahías internas de los nodos. Como mínimo los discos ofertados deben ser de 12 Gbps, de tipo SSD y reemplazables en caliente. No se aceptará configuraciones que propongan cumplir lo solicitado con solo un par de discos por nodo.		
Por cada nodo deberá incluir como mínimo:		
Procesador: Dos (02) Procesadores de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 32 cores como mínimo y de al menos 2.9 GHz por procesador.		
Memoria RAM: 512 GB de tecnología DDR4 o superior.		
Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.		
Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.		
B. Calidad		
El Proveedor deberá incluir como mínimo un almacenamiento RAW de 3 TB en total por el conjunto de nodos, esta capacidad deberá ser distribuida como parte de la opción de almacenamiento que proponga el postor fuera de los nodos o de lo contrario deberá distribuirlo en forma equitativa en las bahías internas de los nodos. Como mínimo los discos ofertados deben ser de 12 Gbps, de tipo SSD y reemplazables en caliente. No se aceptará configuraciones que propongan cumplir lo solicitado con solo un par de discos por nodo.		
Por cada nodo deberá incluir como mínimo:		
Procesador: Dos (02) Procesadores de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 32 cores como mínimo y de al menos 2.9 GHz por procesador.		
Memoria RAM: 512 GB de tecnología DDR4 o superior.		
Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.		
Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.		
C. Desarrollo		
El Proveedor deberá incluir como mínimo un almacenamiento RAW de 3 TB en total por el conjunto de nodos, esta capacidad deberá ser distribuida como parte de la opción de almacenamiento que proponga el postor fuera de los nodos o de lo contrario deberá distribuirlo en forma equitativa en las bahías internas de los nodos. Como mínimo los discos ofertados deben ser de 12 Gbps, de tipo SSD y reemplazables en caliente. No se aceptará configuraciones que propongan cumplir lo solicitado con solo un par de discos por nodo.		
Por cada nodo deberá incluir como mínimo:		
Procesador: Dos (02) Procesadores de última generación lanzada al mercado para el modelo de servidor o nodo ofertado, 32 cores como mínimo y de al menos 2.7 GHz por procesador.		
Memoria RAM: 256 GB de tecnología DDR4 o superior.		
Almacenamiento para sistema operativo: Dos (02) discos de 480 GB SSD o dos discos 480 GB SSD M.2, configurados en RAID 1. Los discos ofertados deben ser reemplazables en caliente.		
Conectividad: Un (01) puerto de 01Gbps Cobre (para gestión). 04 puertos de 25 Gbps (fibra). Debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución.		
6.2.4. Gabinetes		
Las características técnicas mínimas son:		
Altura: 42U estándar de la industria.		
Puertas: Frontal y posterior. Ambas con cerradura Con ventilación necesaria para el adecuado flujo de refrigeración de los componentes.		
Power Device Units (PDU) : Se debe incluir los PDU necesarios de acuerdo con su diseño de manera que soporte el suministro eléctrico redundante de todos los componentes ofertados.		
Energía eléctrica: 200-240 V y 50-60 Hz. Tener en cuenta que SUNAT utiliza tomacorrientes del tipo IEC60309		
El Proveedor deberá incluir todos los accesorios necesarios para su implementación.		
En el Data Center de la SUNAT el Proveedor deberá proveer los cables vulcanizados libres de halógenos (2 conductores más línea a tierra) tomacorrientes del tipo IEC60309 (macho) conforme a las capacidades de consumo de corriente, debidamente etiquetado de acuerdo con la numeración de los circuitos eléctricos provisto.		
La Entidad tiene un servicio de housing, en ese sentido, para el cableado se dará todas las facilidades para el Proveedor, cuyas coordinaciones con este se realizará por intermedio de la SUNAT. Es responsabilidad del Proveedor realizar el cableado necesario para el funcionamiento de la solución ofertada.		
6.2.5. Switch LAN		
Los switches ofertados deberán considerar mecanismos de alta disponibilidad o redundancia entre ellos (en la misma sede). No se aceptará configuración de stacking.		
Para la conectividad a los switches core de la Entidad deberá realizarse mediante puertos 10 Gbps (fibra). Para la conectividad, el Proveedor deberá proporcionar cuatro transceivers (2 por Switch core de la Entidad) para los switches core del modelo HPE 12908E.		
Las características mínimas son las siguientes:		
Factor de Forma : Rackeable de altura máxima 2 UR (unidad de rack)		

<p>Puertos de Datos: Puertos o enlaces de 25Gbps (fibra). Debe contar con los puertos activos, con sus respectivos conectores y licenciamiento necesarios para el correcto funcionamiento de la solución ofertada. La arquitectura del switch deberá ser no bloqueante.</p>		
<p>Puertos Uplink Dos (02) puertos de 40 Gbps (fibra) o 100 Gbps (fibra), con sus respectivos conectores y licenciamiento requeridos para implementar una configuración activo – activo en la misma sede.</p>		
<p>Puertos hacia el Switch Core: Dos (02) puertos de 10Gbps (fibra) como mínimo (debe contar con todos los puertos activos con sus respectivos conectores y licenciamiento requeridos para el funcionamiento de la solución).</p>		
<p>MTU: El switch debe soportar MTU de 9000 o más Jumbo Frames.</p>		
<p>Switching capacity (full duplex) Se deberá ofertar la capacidad de switching full duplex necesaria para el correcto funcionamiento de la solución ofertada, de manera que sea como mínimo la suma de las velocidades de todos los puertos ofertados en full duplex y sin sobresuscripción.</p>		
<p>Tecnología: Soporte de Capa 2 y Capa 3 en IPv4 e IPv6</p>		
<p>Latencia: Menor a 1 milisegundo</p>		
<p>Buffer Size: 32MB como mínimo</p>		
<p>Fuentes de Poder: Fuentes de alimentación redundantes (instaladas y operativas), de intercambio en caliente.</p>		
<p>Ventiladores: Ventiladores redundantes N+N o N+1, de manera que la falla de un ventilador no deba impedir el normal funcionamiento.</p>		
<p>6.2.6. Consola KVM, Switch Administración</p>		
<p>El Proveedor deberá incluir una consola y switch de tipo KVM por cada gabinete para el acceso a los servidores o nodos físicos ofertados.</p>		
<p>El Proveedor deberá incluir como mínimo un switch LAN de administración de dispositivos de hardware. Deberá incluir 48 puertos de 1 Gbps (RJ45), SUNAT proveerá dos puertos LAN de 1Gbps (RJ45) en su switch LAN para la conexión del switch LAN de administración de dispositivos de hardware.</p>		
<p>6.2.7. Características de la Plataforma Landing Zone Big Data</p>		
<p>Los roles y su alcance mínimos que deben ser configurados en la solución Landing Zone Big Data en los ambientes de Producción, Calidad y Desarrollo son:</p>		
<p>Roles de usuarios para la solución Landing Zone Big Data (a nivel de software)</p>		
<p>Administrador global Todos los permisos sobre la solución</p>		
<p>Administrador de usuarios: Gestionar las cuentas de usuarios, grupos, roles y permisos.</p>		
<p>Administrador del clúster: Gestión, mantenimiento y monitoreo de los componentes operativos del clúster.</p>		
<p>Administrador de copias de seguridad: Ejecutar y definir operaciones de backup de scripts, metadata y archivos de configuración de la solución.</p>		
<p>Usuarios de solo lectura: Ver la configuración, información de monitoreo, eventos, logs de la solución</p>		
<p>Usuario de ciencia de datos: Usuario con permisos y accesos al ambiente de ciencia de datos: preparación, construcción y despliegue de modelos.</p>		
<p>Usuario Ingeniero de datos: Usuario con permisos y accesos para la ingesta y preparación de datos.</p>		
<p>El PROVEEDOR podrá sugerir la implementación de nuevos roles y su respectivo alcance según las mejores prácticas o recomendaciones de la solución que implementará.</p>		
<p>La solución Landing Zone Big Data deberá cumplir con características mínimas, las cuales se detallan a continuación:</p>		
<p>A. ORQUESTACIÓN DE LA SOLUCIÓN</p>		
<p>La solución Landing Zone Big Data debe tener la capacidad de configurar y administrar los componentes necesarios para implementar una o más cargas de trabajo.</p>		
<p>El componente de Administración debe tener la capacidad de instalar y distribuir paquetes y librerías sin afectar la continuidad operativa de la solución, asimismo; debe permitir administrar y monitorear todos componentes de la solución.</p>		
<p>Las cargas de trabajo deben ser asignadas y/o aprovisionadas sobre las capacidades de hardware con las que cuenta la solución no se aceptará delegar funciones de procesamiento a equipos personales o locales (laptops o PCs).</p>		
<p>La solución debe proporcionar una interfaz gráfica de usuario que permita la gestión y/o administración de flujos de trabajo que vinculan múltiples aplicaciones y componentes.</p>		
<p>La solución debe permitir monitorear las cargas de trabajo a detalle y el estado del sistema para confirmar que se cumplen los requisitos específicos de calidad del servicio para cada carga de trabajo.</p>		
<p>La solución debe permitir asignar y proporcionar de manera elástica recursos físicos o virtuales adicionales para cumplir con los nuevos requisitos de carga de trabajo producidos por los cambios o incrementos de los datos, número de usuarios o transacciones.</p>		
<p>La solución debe permitir administrar diferentes versiones de librerías y lenguajes en entornos separados, aprovisionando para esto los recursos propios de la solución.</p>		
<p>La solución debe garantizar la asignación de recursos para cada proyecto, equipo o usuario.</p>		
<ul style="list-style-type: none"> La solución debe ejecutar sus procesos aprovechando las capacidades de la capa de procesamiento para el data lake y Machine Learning para la analítica según la necesidad que corresponda. Los componentes de seguridad no podrán utilizar los recursos de la capa de procesamiento de data lake, almacenamiento del data lake, gestión del cluster de data lake, nodos de ingesta de datos o del componente machine learning 		
<p>B. GESTIÓN DE DATOS</p>		
<ul style="list-style-type: none"> La solución debe contar con herramientas o funcionalidades para la gestión y administración del ciclo de vida del dato (linaje). La solución Landing Zone Big Data debe permitir que los datos estén disponibles para todos los roles o funcionalidades o componentes de la misma solución. 		

<ul style="list-style-type: none"> La solución debe contar con herramientas para la gestión de diversos tipos de fuentes de datos (como datos en bruto o datos previamente transformados por otro sistema) y ponerlos a disposición a través de diferentes herramientas y protocolos. 		
<ul style="list-style-type: none"> Las fuentes de datos a considerar como mínimo son registros internos o públicos, audio, imágenes, videos, datos de sensores, web logs, registros de auditoría, logs de aplicaciones, cookies HTTP, archivos XML, archivos JSON, bases de datos relacionales, bases de datos NoSQL, bus de eventos. 		
<ul style="list-style-type: none"> Los modelos y sus datos deberán poder ser portables a otras plataformas de manera transparente sin grandes modificaciones o adecuaciones. 		
<ul style="list-style-type: none"> La solución debe incluir como mínimo las siguientes actividades que son comunes a la mayoría de los sistemas de gestión de datos: <ol style="list-style-type: none"> Recolección o ingesta de datos; Tableros de control e implementación de métricas para monitorear la recolección de datos en curso; Persistencia de los datos; Proporcionar funciones de transformación para el manejo de datos confidenciales o sensibles; Creación de metadatos que describen las fuentes de datos, políticas de acceso y otros atributos relevantes (linaje); Asegurar el acceso a los datos según los permisos asignados; Hacer que los datos sean accesibles a través de interfaces y se ejecuten múltiples procesos en paralelo y la ejecución de consultas de grandes volúmenes de datos; Publicar la disponibilidad de la información y los medios para acceder a ella. 		
<ul style="list-style-type: none"> La solución debe tener integración como mínimo con las siguientes herramientas: DataStage 11.7 y Kafka. 		
<p>C. GESTIÓN DE APLICACIONES BIG DATA</p>		
<p>La solución debe cumplir como mínimo con las siguientes funcionalidades:</p>		
<ul style="list-style-type: none"> Recolección e ingesta de datos: 		
<p>La recolección de datos, está integrada con la funcionalidad de Gestión de Datos. Se debe considerar como mínimo a las fuentes de recolección como servidor de archivos, servidor web, logs de aplicaciones, Base de datos Oracle, SQL Server, Informix, MongoDB, bus de eventos, archivos planos, archivos XML.</p> <p>Debe tener esquemas de control de registros y versionamiento de los flujos creados.</p> <p>Debe ofrecer trazabilidad de los flujos y componentes creados en conjunto con los componentes del Data Lake y Machine Learning.</p> <p>Debe soportar conexión y creación de flujos con:</p> <ul style="list-style-type: none"> Banco de datos vía ODBC o JDBC Soporte a conexión Kafka. Soporte a consumo de datos por API Rest. Drivers nativos para BD relacionales y no SQL. 		
<ul style="list-style-type: none"> Preparación: 		
<p>Debe permitir integrarse con procesos de un ciclo ETL (Extract, Transform, Load) o ELT (Extract, Load, Transform), apoyada por la funcionalidad de análisis donde también se puede ejecutar tareas de transformación de datos. Entre las tareas que debe permitir esta funcionalidad como mínimo son la validación de datos, limpieza, normalización y disponibilizar los datos finales para el entrenamiento y prueba de los modelos Machine Learning.</p> <p>Debe permitir realizar el procesamiento de datos de manera independiente, que proceso A no compita con el proceso B.</p>		
<ul style="list-style-type: none"> Analítica 		
<ul style="list-style-type: none"> Debe permitir implementar los métodos y técnicas para extraer conocimiento de los datos, debe estar soportada por herramientas o software para el análisis en streaming o por procesamiento por lotes, además de brindar herramientas para la creación, entrenamiento y prueba de modelos. 		
<ul style="list-style-type: none"> Debe soportar como mínimo los lenguajes de programación en su última versión estable: Java, R, Python, Scala además del framework Apache Spark. 		
<ul style="list-style-type: none"> La solución de Machine Learning debe mirar los datos y metadatos directamente del Data Lake (para cualquiera de las dos arquitecturas permitidas: Clusterizada o Contenerizada) sin necesidad de hacer movimiento de datos para el storage de Machine Learning, contemplando como mínimo el procesamiento de datos en Python, Scala y R. 		
<ul style="list-style-type: none"> La solución de Machine Learning debe tener integración de seguridad y gobernanza con el Data Lake o permitir implementar sus propias políticas. Se debe permitir que los usuarios trabajen con los archivos y datos almacenados en el data lake y teniendo la misma política de seguridad o de políticas propias del componente de Machine Learning ofertado, además debe permitir asignar cuotas sobre los archivos y filesystem del Data Lake 		
<ul style="list-style-type: none"> Debe permitir el despliegue de los modelos con imágenes contenerizadas o clusterizadas para facilitar el mantenimiento, portabilidad y escalabilidad. 		
<ul style="list-style-type: none"> Debe permitir realizar prototipos de modelos de Machine Learning de manera automatizada. 		
<ul style="list-style-type: none"> Debe realizar el entrenamiento de los modelos de forma independiente, es decir, el entrenamiento A no debe competir por recursos con el entrenamiento B. 		
<ul style="list-style-type: none"> Debe permitir la instalación de notebooks como Jupyter u otros de características similares. 		
<ul style="list-style-type: none"> Debe integrarse con repositorios de desarrollo de código y control de versiones como GitHub, GitLab o similares. 		
<ul style="list-style-type: none"> Debe desplegar los modelos como web services (API Rest) 		
<ul style="list-style-type: none"> Debe permitir asignar recursos (CPU y RAM) para cada usuario, grupo o proyecto que vaya a crear un modelo de analítica. 		
<ul style="list-style-type: none"> Debe automatizar el monitoreo de performance de los modelos en el tiempo. 		
<ul style="list-style-type: none"> Debe identificar automáticamente cuando los modelos están sufriendo problemas de precisión o sesgos en los modelos desplegados. 		
<ul style="list-style-type: none"> La solución deberá aprovisionar sus propios recursos para el entrenamiento, scoring y publicación de modelos aun cuando este use librerías externas, no se aceptará la ejecución de cualquiera de estas fases de manera local (equipos del usuario). 		
<ul style="list-style-type: none"> La solución debe permitir la ejecución de librerías externas sobre las capacidades de hardware de la capa de procesamiento o Machine Learning según la necesidad correspondiente. 		
<ul style="list-style-type: none"> La solución debe permitir la migración o extracción del código fuente de los modelos y en el caso del modelo entrenado la exportación a formatos estándares como PMML, ONNX, YAML, JSON, HDF5, Pickle, Joblib, Protobuf. 		
<ul style="list-style-type: none"> La solución deberá permitir habilitar recursos a demanda según las necesidades del ciclo de vida de un modelo como: entrenamiento (train), prueba (test, scoring). 		
<ul style="list-style-type: none"> Visualización 		

<p>Debe permitir preparar los datos procesados y los resultados de la funcionalidad de Analítica para su presentación al usuario correspondiente.</p> <p>Debe implementar herramientas para presentar los datos con un formato que permita expresar de manera óptima su significado y el conocimiento que aporta. Los formatos de presentación como mínimo deben ser informes basados en texto o mostrar los resultados del análisis en forma gráfica.</p> <p>Debe permitir la integración a herramientas de visualización como PowerBI y Kibana.</p>		
<p>• Acceso</p>		
<p>Debe permitir el acceso a los datos que gestiona mediante un servicio genérico o estándar.</p> <p>Debe implementar una interfaz entre las funcionalidades de visualización y analítica para responder a las solicitudes del usuario (científico de datos o usuario de toma de decisiones). La interfaz con el usuario de datos puede ser síncrona o asíncrona y se debe implementar mecanismos que atiendan los requerimientos de datos.</p>		
<p>D. GESTIÓN DE RECURSOS DE INFRAESTRUCTURA BIG DATA</p>		
<p>La solución debe permitir gestionar los recursos de infraestructura que tenga asignada. Estos componentes son los recursos de procesamiento, almacenamiento y redes de datos. Los tres subcomponentes que la solución debe gestionar son: la plataforma de infraestructura, plataforma de datos y plataforma de procesamiento, los que detallamos a continuación:</p>		
<p>• Plataforma de Infraestructura:</p> <p>Se debe tener como mínimo las herramientas para el monitoreo y gestión de los siguientes elementos:</p> <ul style="list-style-type: none"> - Redes - Computación (procesadores y memoria) - Almacenamiento 		
<p>• Plataforma de datos</p> <p>Se debe implementar herramientas que permitan la organización y distribución lógica de los datos en un sistema de archivos distribuido.</p> <p>La organización lógica de los datos debe considerar como mínimo archivos planos, almacenes de datos relacionales o por columnas en un entorno distribuido. Se deben considerar como mínimo los métodos de acceso por API y lenguajes como el SQL (Structured Query Language).</p>		
<p>• Plataforma de procesamiento</p> <p>Se debe brindar el soporte para la manipulación de datos tanto por procesamiento por lotes (batch processing) o por streaming.</p> <p>Permitiendo como mínimo tres fases de procesamiento: la ingesta de datos, el análisis de datos y la difusión de datos. El procesamiento por lotes, por streaming o la combinación de ambos se puede aplicar a las tres fases antes mencionadas.</p>		
<p>E. GESTIÓN DEL CONSUMO DE DATOS Y MODELOS</p>		
<p>La solución Landing Zone Big Data debe permitir la gestión del consumo de datos y modelos. Las funcionalidades mínimas que se deben soportar son las siguientes:</p>		
<p>• Plataforma de Infraestructura:</p> <p>• Para consumo de datos:</p> <ul style="list-style-type: none"> - búsqueda y recuperación, - descarga, - análisis local, - informes y - visualización <p>• Para consumo de modelos:</p> <p>Debe permitir exponer los modelos como API Key o API REST basado en los estándares OAuth2 y JWT.</p>		
<p>La gestión del consumo de datos debe utilizar las interfaces o servicios proporcionados por la funcionalidad de gestión de aplicaciones big data de la solución para obtener acceso a la información requerida. Estas interfaces pueden incluir la presentación de datos, recuperación de datos y la representación de datos.</p>		
<p>F. GESTIÓN DE SEGURIDAD Y PRIVACIDAD</p>		
<p>La solución Big Data debe permitir gestionar las políticas de seguridad mínimas necesarias de esta, implementar herramientas de auditoría, permitir la integración con el Directorio Activo de la SUNAT con el fin de autenticar a los usuarios que harán uso de la plataforma, crear roles o perfiles para el acceso a los datos y herramientas que formen parte de la solución Big Data, definir los niveles de accesos para quienes puedan ver los datos y en qué circunstancias se les permite hacerlo. La solución debe tener la trazabilidad de las acciones realizadas por los usuarios en todos los componentes de la solución (Data Lake, Machine Learning e Ingesta de Datos).</p>		
<p>G. CAPA DE GESTIÓN</p>		
<p>La solución Big Data debe implementar herramientas para la gestión del propio sistema, los datos, la seguridad y las consideraciones de privacidad. Esta capa de gestión debe abarcar dos grupos generales de actividades: gestión de sistemas y gestión del ciclo de vida de Big Data.</p> <p>Entre las principales funcionalidades se debe incluir el aprovisionamiento y gestión de recursos, la configuración de la solución, la gestión de paquetes, gestión de software, gestión de copias de seguridad y la gestión de alertas o notificaciones.</p>		
<p>La gestión de copias de seguridad debe incluir como mínimo el respaldo de scripts (código de modelos, especificación de flujo de datos), metadata y archivos de configuración de la solución. Los cuales permitan la recuperación ante fallas o caídas de cualquier componente de la solución.</p> <p>El licenciamiento de Software debe incluir todas las herramientas que permitan implementar las funcionalidades que forman parte de la solución Landing Zone Big Data.</p>		