



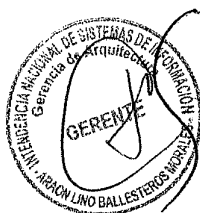
**INFORME TECNICO PREVIO DE EVALUACION
DE SOFTWARE N° 002-2018 / 1U4100**

- 1. NOMBRE DEL ÁREA**
DIVISION DE ARQUITETURA DE INFORMACION Y APLICACIONES
- 2. RESPONSABLES DE LA EVALUACIÓN**
CESAR FORTUNATO ROLDAN BALUIS
- 3. CARGOS**
CONTRATO ADMINISTRATIVO DE SERVICIOS
- 4. FECHA**
01/08/2018
- 5. JUSTIFICACIÓN**

SUNAT está con la necesidad de incrementar los niveles de recaudación fiscal, requiere la implementación de un repositorio para datos no estructurados y semi-estructurados, esto con el fin de incluir información valiosa de fuentes de datos internas y externas que permita realizar análisis de contribuyentes desde todos los puntos de datos disponibles, esto permitirá a SUNAT a generar modelos probabilísticos donde con todos los datos disponible permita identificar problemas de elusión de tributos.

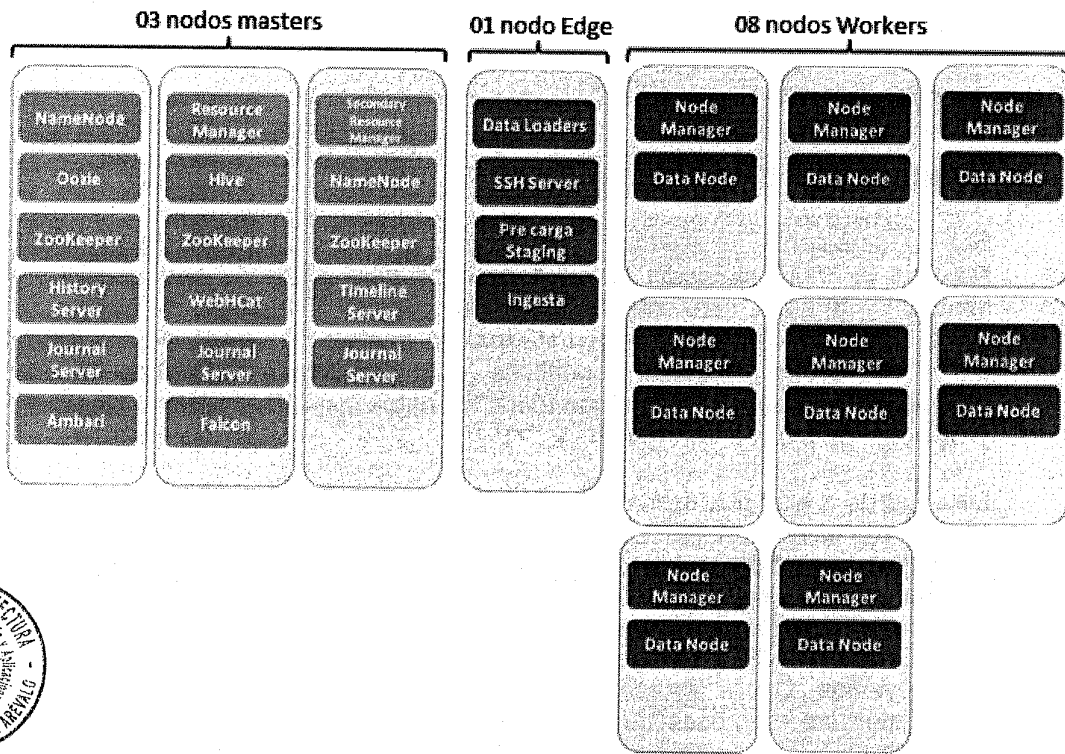
De acuerdo a los tipos de datos, semi-estructurados y no estructurados, se ha identificado diversas soluciones tecnológicas que permitan almacenar, procesar y analizar este tipo de datos. Principalmente las tecnologías utilizadas se centran en soluciones Hadoop, teniendo HDFS como sistema de archivos y permitiendo procesar los datos con soluciones incluida en las plataformas como SQL, Lenguaje Python, Lenguaje estadístico R, entre otros, todos apalancados en proyectos Open Source que posteriormente han sido soportados y distribuidos las empresas de tecnologías tales como Microsoft, IBM, Oracle, Teradata.

Sobre la plataforma Hadoop se ingestara (Integración en formato crudo o RAW) todas las fuentes de datos disponible, posteriormente se procesará con las capacidades SQL (ANSI) y procesamientos en Python o R, para luego ser procesada por los científicos de datos de SUNAT para la generación de modelos predictivos para identificación de elusión fiscal.



La configuración será realizada por clúster de servidores virtualizados sobre VMWare, permitiendo de esta manera generar eficiencias económicas al no ser necesario la compra de nuevos activos de procesamiento (servidores adicionales). A nivel de Hadoop el clúster estará compuesto por 10 máquinas virtuales, de las cuales 1 será definida como Nodo de Enlace, en donde se tendrán las herramientas de integración con los otros sistemas de SUNAT, 3 nodos de administración donde se distribuirán los servicios propios de Hadoop con alta disponibilidad para evitar punto de fallas que permite el crecimiento del clúster para soportar más capacidades de almacenamiento y procesamiento. y para terminar con 6 nodos de datos donde se almacenará los datos en HDFS.

Adjunto una figura donde muestra la arquitectura propuesta en base a las buenas prácticas definidas para Hadoop.



CRB
SV34

6. ALTERNATIVAS

No se han considerado otras alternativas de software en la evaluación del presente informe, debido a que la plataforma de Hardware y Software (Aplicaciones desarrolladas) de SUNAT, se encuentran diseñadas para el uso de licencias de software Hortonworks Data Platform.

7. ANÁLISIS COMPARATIVO TÉCNICO

El presente análisis se realizó aplicando la alternativa 3 de la Guía de Evaluación de software Hortonworks Data Platform

a. Propósito de Evaluación

- Determinar las mejores características y atributos del software Hortonworks Data Platform para su adquisición.

b. Identificación el Tipo Producto

Hortonworks Data Platform

c. Especificación Del Modelo de Calidad

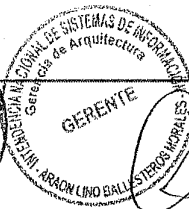
De acuerdo a lo establecido en la Guía de Evaluación del Software para la Administración Pública aprobado mediante la Resolución Ministerial N° 139-2004-PCM, se aplicara el modelo de calidad de software establecido en el mismo.

d. Selección de Métricas

Las métricas fueron seleccionadas en base al análisis de las características técnicas del software seleccionado en el punto 6 (alternativas de software), también a través de internet y plantillas de evaluación.

Las métricas consideradas del presente informe son las siguientes:

ATRIBUTOS	DESCRIPCIÓN
ATRIBUTOS INTERNOS	
Sistemas Operativos / Servidores Virtuales	Linux x86/x64, Linux /zSeries, Linux/Power, AIX/Power, VMWare
Ingesta de Información	Permite llevar datos al clúster de Hadoop mediante técnicas estructuradas como ETL (utilizando IBM DataStage), o utilizando las herramientas de integración que ya posee la plataforma Hadoop, tales como FTP, SFTP, SQL, SSH, Apache Sqoop, Apache Flume.
Procesamiento de datos	Procesamiento de datos dentro del cluster Hadoop utilizando herramientas incluidas en la distribución de Hadoop, como HIVE, HBASE, PIG, HDFS.
Alta Disponibilidad	Por definición Hadoop posee una triple replicación de datos en sus nodos de datos, la implementación de 3 nodos de administración permite contar con alta disponibilidad en administración y en las herramientas de análisis



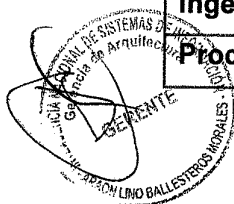
	de información.
Acceso SQL	Acceso SQL ANSI, donde no se deban modificar queries realizadas en SQL ANSI, sino puedan ejecutarse directamente sin necesidad de reescritura.
Migrar entre distribuciones	Apego a las versiones open source del mercado, no corresponde a una distribución cerrada lo que entrega capacidades de migrar a otras versiones.
Idioma	Inglés
ATRIBUTOS EXTERNOS	
Administración	<ul style="list-style-type: none"> • Administración de servidores a través de Ambari. • Prioridades en recursos de sistemas a través de YARN • Registro de eventos en logs de los servicios. • Administrar puede permitir que los usuarios accedan a los datos a los cuales que tiene permisos de acceso, procesamiento y lectura. • Cuenta con una consola de administración y configuración que sea accesible de manera local remota a los servidores.
Alertas	<ul style="list-style-type: none"> • Monitoreo de Recursos • Definición de alertas con umbrales para monitoreo.
ATRIBUTOS DE USO	
Soporte en local	Contar con soporte local que permita la correcta administración de la plataforma.
Facilidad de Uso	Utilizar herramientas a través de consolas Web.
Soporte Técnico	Soporte 24x7
Servicio de mantenimiento	Parches, soporte técnico telefónico, compromisos de respuesta del servicio, asistencia vía web desde el sitio oficial.



e. Niveles, Escalas para las Métricas y Comparación de los Criterios

Se muestran las escalas consideradas para cada una de las métricas señaladas

ATRIBUTOS DE EVALUACION	Puntaje	Horizontes
ATRIBUTOS INTERNOS	40	40
Sistemas Operativos / Servidores Virtuales	4	4
Ingesta de Información	5	5
Procesamiento de datos	5	5



CRB
Ar34

Alta Disponibilidad	4	4
Acceso SQL	4	4
Migrar entre distribuciones	3	3
Idioma	4	4
ATRIBUTOS EXTERNOS	20	20
Administración	6	6
Alertas	5	5
ATRIBUTOS DE USO	40	40
Soporte en local	8	8
Facilidad de Uso	8	8
Soporte Técnico	14	14
Servicio de mantenimiento	10	10
Puntaje Total	100	100
Puntaje Mínimo	80	

8. ANALISIS COMPARATIVO DE COSTO-BENEFICIO:

Los costos asociados al producto incluyen:

- Soporte: El costo de suscripción de Soporte y Mantenimiento por tres (03) años de tres (03) SKU¹ Hortonworks es S/. 650 000.00
- Hardware: SUNAT dispone de la plataforma necesaria para la instalación de los servidores de aplicación no siendo necesario realizar inversiones adicionales en hardware.
- Los beneficios se dan al garantizar el adecuado funcionamiento y la actualización tecnológica de los servicios y aplicativos de la Web de la institución.

9. Conclusiones :

- La legislación vigente, dispone que las entidades públicas legalicen el uso del software que emplean en sus computadores, y establece responsabilidad administrativa, penal y civil para las autoridades de la entidad por el incumplimiento de la ley.
- Del análisis realizado se demuestra que el software Hortonworks cumple con los requerimientos técnicos, siendo necesaria la adquisición de nuevas licencias para garantizar el funcionamiento de los servicios y aplicaciones de negocio de la Institución.

¹ 01 SKU = 04 licencias de Hortonworks Data Platform

