Convocatoria FORDECYT 2019-06 CONACyT

Etapa 2 - Reporte Técnico

Proyecto **41756**

Responsable Técnico:
Dr. José Luis González Compeán
joseluis.gonzalez@cinvestav.mx
Profesor-Investigador
Cinvestav Tamaulipas

Muyal-Ilal:

Plataforma tecnológica para la gestión, aseguramiento, intercambio y preservación de grandes volúmenes de datos en salud y construcción de un repositorio nacional de servicios de análisis de datos de salud

Fecha: 25/01/2023	*Publicación:*	*Versión: 1.0*
Seguimiento	**Nombre completo**	**Fecha**
Elaboró:	M. C. Diana Elizabeth Carrizales Espinoza	10/11/2021
Revisó:	Dr. Julio Noe Hernandez Torres	25/11/2021
Autorizó:	Dr. José Luis González Compean	

Control de cambios

Versión	Fecha	Bitácora
1.0	25/01/2023	Documento Inicial

Resumen

En este proyecto multidisciplinario se desarrolla Muyal-Ilal, una plataforma de servicios en la nube para la gestión, aseguramiento, intercambio, procesamiento, análisis y preservación de grandes volúmenes de datos médicos. Muyal-Nez crea sistemas eficientes de e-salud para el procesamiento y manejo de datos médicos. Muyal-Xelhua crea sistemas de analítica (big data) basados en inteligencia artificial para convertir datos (históricos estadísticos, notas clínicas, datos de dispositivos médicos, imagenología, etc.) en información. Nez y Xelhua guían al personal médico en la creación de sistemas portables de e-salud de toma de decisiones o asistencia a diagnósticos. Estos sistemas pueden incluir y/o acoplarse con sistemas existentes y se pueden transferir fácilmente a otras instituciones, lo cual reduce costos de desarrollo y resuelve la dependencia tecnológica entre instituciones y proveedores de software y servicios. Muyal-Chimalli asegura y verifica que los datos manejados por los sistemas creados en Muyal cumplan, en forma automática y transparente, las normas nacionales (NOM-024-SSA3-2010 y NOM-004-SSA3-2012) e internacionales (ISO-270001-13, COBIT5, NIST) referentes a tolerancia a fallos de servicios/servidores, privacidad, confidencialidad, integridad, disponibilidad y trazabilidad de los datos médicos. Chimalli crea reportes sobre cumplimiento de cada norma incluyendo guías para el cumplimiento de las mismas. Muyal-Painal construye sistemas de almacenamiento y distribución de datos para que las instituciones soporten escenarios de intercambio ininterrumpido de catálogos de bases de datos, resultados/información y/o sistemas e-Salud a través de intra/internet. Muyal-Alwa crea servicios de repositorios (estandarizados y FAIR) para facilitar el acceso a catálogos publicados por instituciones de salud. En Muyal se construyen sistemas e-salud de inteligencia artificial para diagnóstico asistido de cáncer de hueso largo y pulmones, estudios espaciotemporales de enfermedades de alta prevalencia con georreferenciación, calculadoras de medición de riesgo de enfermedades cardiovasculares que producirán bases de datos y repositorios para la comunidad científica y las instituciones de salud.

Índice

Índice de Figuras

1. Introducción

Los sistemas de expediente clínico electrónico (*SECE*) han sido herramientas clave para mejorar los procesos de atención de pacientes. Sin embargo, existen aún áreas de mejora en dos vertientes:

a) La primera tiene que ver con el intercambio de contenidos en información médica entre múltiples roles de profesionales de la salud, múltiples niveles de atención e incluso múltiples instituciones de salud o gubernamentales donde se debe dar cumplimiento a requisitos de interoperabilidad, eficiencia, seguridad, resistencia a fallas y trazabilidad de transacciones establecidas por normas nacionales (*NOM-024-SSA3-2010* y *NOM-004-SSA3-2012*) e internacionales (ISO-270001-13, COBIT5, NIST) para el manejo de datos sensibles en el área de la salud. Tanto la transferencia de datos y contenidos como el cumplimiento de los requerimientos no son provistos, en su conjunto, por los *SECES* actualmente disponibles en México.

b) La segunda tiene que ver con la posibilidad de que sistemas computarizados de analítica de datos puedan convertir tanto las fuentes de datos de los *SECEs* (datos históricos) como información producida por la práctica médica, sensores y dispositivos médicos en información útil que soporten procesos de toma de decisiones.

Ambas áreas de oportunidad representan un enorme desafío para las instituciones de salud debido a la disponibilidad de recursos computacionales, así como la heterogeneidad tanto de los datos como de los sistemas computarizados por dichas instituciones.

En este proyecto, de carácter multidisciplinario, colaborarán investigadores y especialistas en telecomunicaciones, ciencias de datos, informática médica y tecnologías de la información para diseñar y desarrollar una plataforma de gestión, aseguramiento, intercambio y preservación de grandes volúmenes de datos en el área de la salud (*Big Data*) llamada *Muyal-Ilal* (combinación de palabras de lenguas mayas que se traduce como nube de medicina o cuidado). *Muyal-Ilal* incluye una nube de servicios llamado *Muyal-Nez* que provee a las instituciones de salud (específicamente el área de tecnologías de la información) servicios para crear, en minutos, sistemas de e-Salud que permitan el intercambio de datos y contenidos médicos en forma intra-institucional (entre diferentes profesionales de la salud) e inter-institucionales (entre múltiples instituciones de salud y/o gubernamentales) sin que este personal realice tareas complejas de instalación o programe dichos sistemas. *Muyal-Nez* crea automáticamente los sistemas de almacenamiento y distribución de datos para asegurar la entrega de los contenidos a sus destinatarios, los cuales los reciben como un sistema de paquetería tradicional. De esta forma *Muyal-Nez* resuelve la dependencia de las instituciones con los proveedores de software y servicios web o en la nube. Usando *Muyal-Nez* se han creado sistemas de e-Salud para el diagnóstico asistido de cáncer de hueso largo, que crea flujos automáticos desde tomógrafos, pasando por PACS, radiólogos, a un sistema de inteligencia artificial que crea un pre-diagnóstico computarizado, el cual es automáticamente recibido por el profesional de oncología que lo ha solicitado.

Múltiples sistemas de e-Salud construidos por *Muyal-Nez* están actualmente disponibles en *Muyal* (por ejemplo, creación de imágenes 3D y detección de picos QRS en ECGs) y otros están bajo construcción (identificación mediante redes neuronales de nódulos en pulmón, Covid y tuberculosis). *Muyal-Chimalli* verifica que cada sistema de e-Salud cumpla, de forma automática y transparente, las normas nacionales e internacionales garantizando privacidad, confidencialidad, integridad y disponibilidad de los contenidos, así como estableciendo tolerancia a fallas de servicios/servidores y creando registros inmutables en una red privada (blockchain) para dar trazabilidad al manejo de los datos. *Muyal-Chimalli* crea automáticamente redes de cripto-contendores y blockchain para los sistemas *Nez* sin la intervención del personal de salud. *Muyal-Xelhua* es un sistema similar a *Muyal-Nez* pero que permite a las instituciones de salud crear, en minutos, sistemas de analítica de datos (big data) para convertir datos (históricos estadísticos, textos de diagnosis, etc.) en información para procesos de toma de decisiones y soportar posibles intervenciones.

Muyal-Xelhua se conecta con *Muyal-Nez* para crear sistemas de e-Salud de analítica y con *Muyal-Chimalli* para garantizar que tanto los datos como los tomadores de decisiones son aptos para realizar procesos de análisis y/o expuestos. Tanto *Muyal-Nez* como *Muyal-Xelhua* permiten crear conexiones con sensores de dispositivos médicos para que las instituciones de salud puedan acceder a los datos en tiempo real y los puedan usar como insumos de sus sistemas de e-Salud. Con *Muyal-Xelhua* se han creado sistemas de big data para realizar estudios espaciales temporales con mapas de riesgo basados en bases de datos de egresos de algunas enfermedades crónicas. *Muyal-Xelhua* incluye un amplio repositorio de sistemas de preparación de datos (para la curación de bases de datos), preprocesamiento (imputación, interpolación, identificación de variables de incidencia, disminución de dimensionalidad, etc.), procesamiento/análisis automático mediante inteligencia artificial (máquinas de soporte de vectores, redes neuronales, minado de texto, etc.) y esquemas de visualización (georreferencia por sistemas de información geográfica).

Muyal-Painal es un sistema de publicación y suscripción que permite a las instituciones de salud compartir sus fuentes de datos, resultados (información y contenidos médicos procesados) e inclusos sus sistemas de e-Salud en forma intra-institucional o inter-institucional. *Muyal-Painal*, provee a los usuarios finales herramientas, similares a las usadas en redes sociales, para que puedan intercambiar (mediante publicación y subscripción) catálogos de datos/contenidos intra e inter-institucionalmente. *Muyal-Painal*, en forma transparente, se conecta con *Muyal-Chimalli* para validar y registrar cualquier operación de compartición de datos. *Muyal-Alwa* es un sistema de repositorios (FAIR) basado en sistemas estandarizados de exposición de catálogos, que permite automatizar la publicación y consumo, para uso privado y/o público de datos, componentes o sistemas completos de e-Salud, así como información producida por los usuarios en la plataforma. Cada contenido agregado/descargado en *Muyal-Alwa* posee un pasaporte que es revisado por *Muyal-Painal* y validado por *Muyal-Chimalli* previo a su incorporación a los repositorios de *Muyal-Alwa* o el consumo por parte de los usuarios a través del Internet.

La Figura 1muestra las características principales de la plataforma *Muyal-Ilal*, la cual se encuentra compuesta por los servicios de *Muyal-Nez, Muyal-Chimalli, Muyal-Xelhua, Muyal-Painal y Muyal-Alwa.*

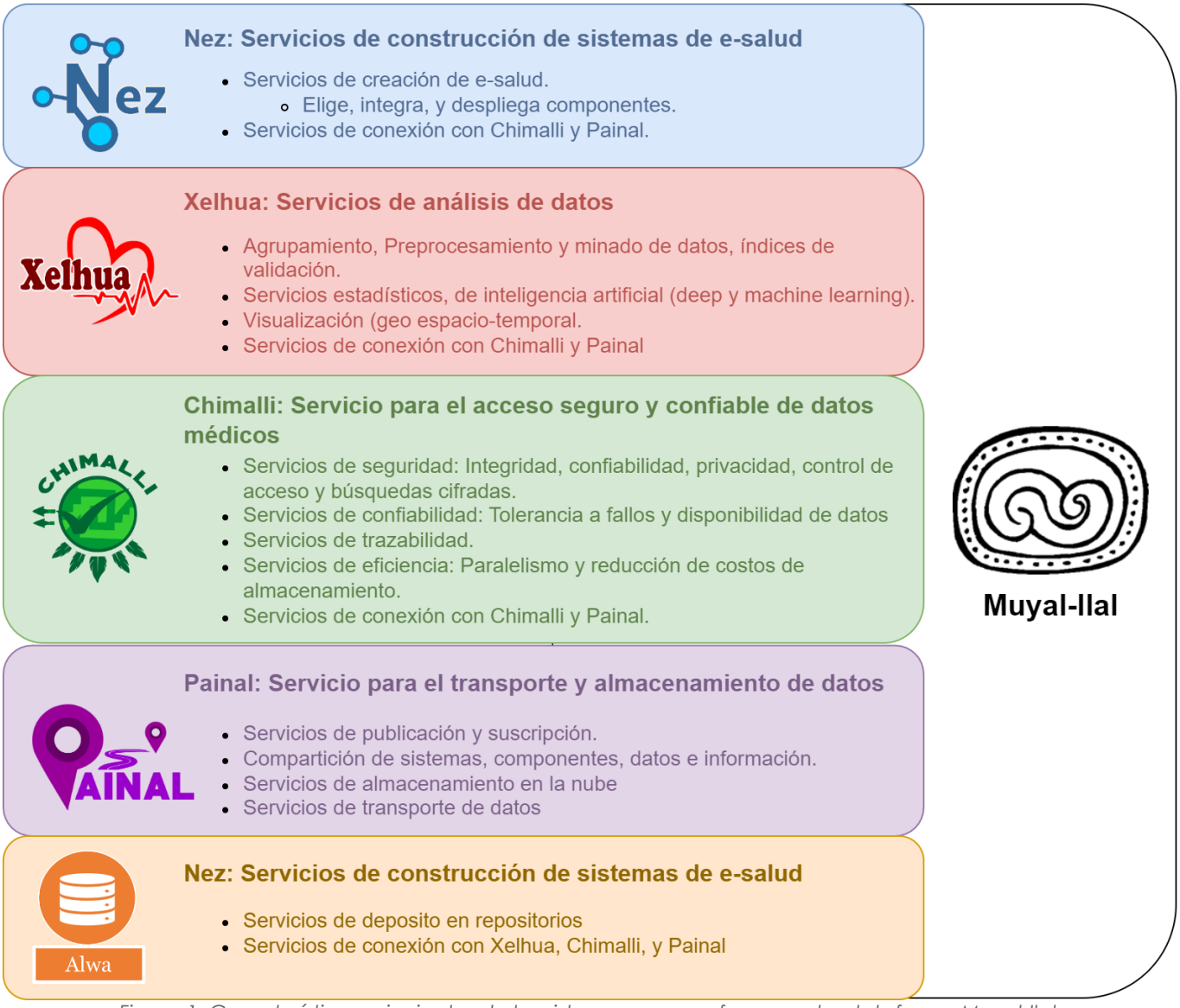

Figura 1. Características principales de los sistemas que conforman a la plataforma Muyal-Ilal.

2. Antecedentes

Los sistemas de expediente clínico electrónico (*SECE*), dispositivos médicos y plataformas para el manejo de datos de salud mejoran los tiempos de respuesta de la atención a los pacientes del sector salud de México. Además, estos componentes posibilitan el fortalecimiento de mecanismos de control de los sistemas de salud y permiten a las instituciones mantenerse en línea con las diversas políticas, normas (por ejemplo, *NOM-024-SSA3-2010* y *NOM-004-SSA3-2012 así como* ISO-270001-13, COBIT5, NIST) y estándares (por ejemplo, *DICOM* y *HL7*). En este contexto, el término de sistema de *e-Salud* se refiere al uso de las diversas tecnologías de la información, telecomunicaciones y manejo de operaciones para integrar los componentes antes mencionados un solo sistema coherente y de fácil manejo cuya operación se base en las normas y protocolos oficiales.

Previamente, en nuestro grupo de trabajo, se han realizado desarrollos tecnológicos con el fin de cumplir con la norma oficial mexicana, en su rubro de conservación y almacenamiento de datos clínicos. Específicamente, el Instituto Nacional de Rehabilitación (*INR*), que forma parte de nuestro grupo de trabajo, implantó un sistema llamado *PACS-INR* que permite manejar, distribuir, almacenar y recuperar imágenes médicas con calidad de diagnóstico en formato *DICOM*.
En 2015, *INR* en colaboración con la empresa INFOTEC desarrollaron una integración de *PACS-INR* con *BABEL*, un sistema distribuido de almacenamiento tolerante a fallos, el cual almacena a la fecha 82 millones de imágenes de diferentes modalidades (*RM, TC, US, MN, y RX*) distribuidas en 73 TB de almacenamiento físico.

Bajo este contexto, el *INR* desarrolló funcionalidades de almacenamiento confiable basado en tecnología nacional eliminando costos de licencias y cambiando hardware de gama alta por equipo de cómputo de gama media. Este sistema, sin embargo, actualmente no considera mecanismos de intercambio de datos en forma segura, la construcción de flujos de trabajo inter/intrainstitucionales o la incorporación de herramientas de mejora funcional de preprocesamiento ni sistemas de diagnóstico asistido por inteligencia artificial.
En este sentido, el manejo de datos sensibles (como lo son las imágenes médicas) implica la utilización de estructuras de datos que permitan acceder tanto a los datos como los metadatos, así como sistemas de colocación y localización que permitan compartir, cargar, descargar, visualizar y/o eliminar datos/metadatos.

Actualmente *Muyal-Nez* permite al INR conectar su PACS-INR con las diferentes estaciones de los profesionales de la salud (por ejemplo, radiología, consulta, y oncología). Se pretende extender a epidemiología y demás departamentos del Instituto, así como con otros institutos u hospitales referentes que envían pacientes al INR. *Muyal-Chimalli* ha permitido alcanzar un 70% de las regulaciones estandarizadas en forma internacional para el manejo seguro de datos sensibles y cubre todas las fases de interconexión establecidas por las normas oficiales. El resumen de seguridad creado por *Chimalli* también ha permitido revelar las tareas de ciberseguridad que no dependen de la plataforma *Muyal*, sino de actividades realizadas por personal de

salud para que las instituciones creen un plan para implementarlas. *Xelhua* ha permitido extender el análisis de contenidos (imágenes del PACS) a datos (históricos y bases de datos del sistema de expediente) y extender el servicio no solo a traumatología y oncología, pero también a epidemiología y departamentos asociados a la toma de decisiones y/o intervención de salud pública. *Painal* ahora les permitirá compartir, mediante transferencia tecnológica, sus sistemas con otros hospitales, los cuales podrán ser consumidos en la nube o descargados en su infraestructura. Esto reduce el tiempo de obtención de sistemas de e-Salud a minutos u horas dependiendo de la modalidad de transferencia elegida, lo cual reduce la dependencia con los proveedores de software e infraestructura, así como el pago constante de licencias.

El transporte de los datos adquiridos es uno de los grandes problemas que se presentan cuando se manejan datos sensibles. Lo anterior se debe a que, al ser datos sensibles, es necesario garantizar servicios de seguridad tales como integridad, privacidad, confidencialidad y estrictos controles de acceso. De igual manera, al transportar grandes volúmenes de datos de un sitio a otro, es necesario contar con distintas técnicas que permitan eficientizar dicho proceso. Para hacer frente a este tipo de problema se deben utilizar herramientas que permitan realizar operaciones de cifrado/descifrado para proveer seguridad a los datos, así como operaciones concurrentes/paralelas para mejorar la utilización de los recursos computacionales.

Esquemas criptográficos experimentales para compartir, en forma segura, contenidos en ambientes organizacionales a través de la nube, que hasta la fecha solo se habían probado conceptualmente y mediante modelos matemáticos, mostraron la factibilidad de proteger datos sensibles en escenarios reales en términos de integridad (mediante firmas digitales), confidencialidad y control de acceso (mediante cifrado por atributos), así como privacidad (mediante el uso de sobres digitales). En nuestro grupo de trabajo también se desarrollaron esquemas para el transporte confiable, seguro, flexible y eficiente de contenidos, los cuales fueron probados en escenarios reales para que agencias compartan datos espaciales e imágenes satelitales, y posteriormente probados con éxito en el transporte de imágenes de tomografía.

Estas soluciones incluyen mecanismos para reducir el consumo de almacenamiento producido por procesos de tolerancia a fallos, mecanismos de paralelismo basados en tuberías de procesamiento que mejoran la eficiencia en el procesamiento de datos e imágenes, herramientas de transporte de grandes volúmenes de información a través de flujos de trabajo creados por diferentes usuarios y agencias, y mecanismos que garantizan la integridad de los datos transportados extremo a extremo. En este contexto, las soluciones extremo a extremo permiten a los usuarios proteger sus datos antes de que estos sean enviados a la nube, así como agregar confiabilidad a los mismos (mediante cifrado por atributos), y soportar fallas (mediante técnicas de dispersión de información) que pueden ser provocadas por diversos problemas, tales como interrupciones de los servicios e incidentes de bloqueo por parte de los proveedores.

Debido a ello, en nuestro grupo se han desarrollado esquemas basados en emparejamientos criptográficos los cuales conforman un servicio de almacenamiento de extremo a extremo para modelos de nube híbrida. Estos servicios permiten realizar el intercambio de archivos en escenarios donde los datos son enviados a la nube, y grupos selectos y autorizados de usuarios pueden acceder a los mismos. Los escenarios reales donde nuestras propuestas fueron probadas en el pasado, guardan similitud con los escenarios de movilidad y flujos de trabajo donde se comparten datos clínicos realizados por profesionales de la salud al interior/exterior de instituciones y exhiben marcadas similitudes con los requerimientos de manejo, almacenamiento y transporte de imágenes médicas (seguridad, confiabilidad, eficiencia y confidencialidad).

Actualmente, los sistemas que conforma *Muyal-Ilal* superan a los sistemas actuales experimentales. *Nez* actualmente mejora la eficiencia de los procesos de manejo de imagenología en 4.9 veces al sistema actual del INR y reduce el consumo de almacenamiento en 66%. *Muyal-Chimalli* reduce los tiempos de aseguramiento de datos en hasta 80% con respecto a sistemas criptográficos similares y provee cualidades de integridad, trazabilidad, control de acceso, privacidad, confidencialidad y tolerancia a fallos en un solo sistema sin que los usuarios configuren, programen o instalen nada (esto produce un cumplimiento inmediato de hasta 70 % de los requerimientos de los estándares internacionales).

3. Motivación y justificación

La atención médica en México es crucial para mejorar el bienestar de los ciudadanos. Esta práctica profesional produce escenarios de grandes volúmenes de datos (big data) producidos por diversas fuentes heterogéneas (sensores, dispositivos médicos, etc.) que deben ser procesados rápidamente (velocidad) por un conjunto heterogéneo de sistemas de expedientes clínicos electrónicos o *SECE* (variedad) que entregan información útil a diferentes repositorios de datos (veracidad-valor). El *Plan Nacional de desarrollo 2019-2024* describe las dimensiones de este escenario: "a finales de 2018, el IMSS contaba con 68.5 millones de derechohabientes, el ISSSTE con más de 13 millones, IMSS-Secretaría de Bienestar con un total de 13 millones, así como 2 millones de Sedena, Semar y Pemex".

Los *SECE* desplegados por las entidades anteriores resultan clave para enfrentar este escenario y su uso se ha consolidado en sectores público y privado. El IMSS, por ejemplo, desplegó su *SECE* en 99% de unidades de atención primaria con adopción programada para hospitales de segundo y tercer nivel. Las normas oficiales *NOM-024-SSA3-2010*, *NOM-004-SSA3-2012* y ley federal de protección de datos personales establecen requisitos funcionales (manejo de datos y metadatos clínicos) que los *SECE* deben cubrir mediante el cumplimiento de estándares internacionales, mientras que requerimientos no-funcionales (seguridad, confiabilidad, confidencialidad, etc.) deben ser garantizados por las instituciones que instalan y operan un *SECE*, lo cual resulta en un gran desafío para las instituciones de salud mexicanas.

Las normas además establecen que los *SECE* deben permitir el intercambio de información e imágenes entre niveles de atención, llevar control administrativo de movimientos de pacientes, y enviar/recibir información, metadatos, imágenes y resultados de laboratorio con garantía de interoperabilidad y seguridad. La dispersión y separación geográfica de las fuentes de datos, los *SECE*, los pacientes, y de los profesionales de la salud que intervienen en el proceso de atención médica aumenta la complejidad del escenario descrito. Por ejemplo, en México existen 25 Centros Estatales de Cáncer (*CEC*), las Unidades de Oncología del Hospital Juárez, el Hospital General de México y el Instituto Nacional de Cancerología, los cuales constituyen la red que proporciona atención médica a pacientes con cáncer. El Instituto Nacional de Rehabilitación Luis Guillermo Ibarra Ibarra, atiende a esta red y a pacientes con cáncer del sistema músculo esquelético. En los estados de Querétaro, Hidalgo, Tlaxcala, Morelos, Quintana Roo y Baja California no existen centros estatales de atención y la población que requiere tratamiento especializado en cáncer, debe trasladarse al centro de referencia más cercano. Se estima que a nivel nacional existen actualmente 735 cirujanos oncólogos, 50 ginecólogos oncólogos, 269 oncólogos médicos, 151 oncólogos pediatras y 180 radio oncólogos para atender a una población estimada de 45 millones de personas que no cuentan con servicios médicos.

Por tanto, y tomando en cuenta el gran volumen de pacientes con respecto al pequeño segmento de profesionales de salud especializados, no solo se requiere que cualquier direccionamiento de pacientes a estos profesionales de la salud sea realizado con la mayor seguridad, eficiencia y prontitud posible, sino que también se les provean de herramientas para mejorar y/o eficientizar los flujos de trabajo e intercambio de datos asociados a las tareas de prognosis, diagnosis y tratamiento así como herramientas de toma de decisiones tales como sistemas de diagnóstico asistido por inteligencia artificial y sistemas de analítica para convertir cúmulos de datos en información útil para tomar decisiones.

La complejidad de este escenario aumenta cuando se deben preservar los derechos de los pacientes y/o profesionales de la salud a la privacidad, disponibilidad, integridad, confidencialidad y auditoria de sus datos. La heterogeneidad tanto de las aplicaciones como de las infraestructuras, la dispersión geográfica de los participantes, así como el volumen de los datos producidos constantemente aumentan considerablemente la complejidad del problema que resulta proveer a tanto a pacientes como profesionales de la salud de un ambiente seguro, confiable, controlado y eficiente para mejorar la atención de los pacientes. Un aspecto menor pero que impacta a las instituciones prestadoras de servicios de salud son las dependencias que se presentan cuando se intenta crear estos sistemas de intercambio seguro de datos. Estas dependencias se presentan con los desarrolladores de servicios, quienes tienen tiempos elevados de construcción de sistemas e imponen el pago constante de licencias de uso, lo cual crea una relación funcional estrecha entre proveedor e institución contratante. Esta dependencia también podría presentarse con proveedores de servicio externos (servicios en la nube) a quienes las instituciones podrían delegar procesos de almacenamiento, distribución de datos, procesamiento y análisis (servicios en línea de inteligencia artificial), lo cual implica que las instituciones

también les están delegando control sobre los contenidos y aplicaciones que les delegan. La pérdida de control sobre datos sensibles y/o sistemas críticos se puede llegar a convertir en accesos no controlados, violaciones de integridad, confidencialidad, privacidad o extravío temporal o permanente de los datos. Eliminar dependencias y mantener control sobre datos de carácter sensible resulta crítico para las instituciones de salud.

4. Grupo de trabajo y Colaboraciones Interinstitucionales

A continuación, se listan las instituciones participantes en el proyecto, así como el equipo de trabajo intrainstitucional e interinstitucional.

4.1. Instituciones participantes

Tabla 1. Instituciones participantes en el proyecto

Institución	Tipo de entidad	Área
Instituto Nacional de Rehabilitación Luis Guillermo Ibarra Ibarra (INRII)	Entidad *receptora*	Servicio de Tomografía Computada y ultrasonido, Departamento de Desarrollo Tecnológico
CINVESTAV Tamaulipas	Entidades *entregantes*	Grupo de Gestión de Datos y Redes de Computadoras
Universidad Carlos III de Madrid (UC3M)		Grupo de Arquitectura de computadoras y tecnologías (ARCOS)
Universidad Autónoma Metropolitana (UAM)		Departamento de Ing. Eléctrica, áreas de Redes y Telecomunicaciones/ Procesamiento Digital de Señales e Imágenes Biomédicas
Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE)		Departamento de Electrónica y Telecomunicaciones de la división de física aplicada de Ensenada y Unidad Monterrey.
University of Naples "Parthenope",		Departamento de Ciencia y Tecnología

4.2. Grupo de trabajo intrainstitucional

Tabla 2- Equipo de trabajo intrainstitucional

Institución	Participante	Líneas de investigación	Nivel SNI
	Dr. José Luis González Compeán	Arquitectura de computadores, Seguridad informática, Cómputo en la nube, Sistemas distribuidos, Sistemas de almacenamiento y tolerancia a fallos.	**1**
	Dr. Miguel Morales Sandoval	Seguridad informática, Criptografía no convencional, Sistemas distribuidos	**2**
	Dr. Iván López Arévalo	Minería de datos, minería de texto, representación y manejo de conocimiento, inteligencia artificial.	**1**

Cinvestav Tamaulipas	Dr. Gregorio Toscano Pulido	Optimización evolutiva multi-objetivo	**1**
	Dr. Edwyn Javier Aldana Bobadilla	Machine Learning, Optimización, Procesos Estocásticos, Data Mining, Ingeniería de Software.	**Candidato SNI**
	Dr. Wilfrido Gómez Flores	Análisis de imágenes, reconocimiento de patrones, y aprendizaje de máquinas.	**1**
	MC. Mariana Magdalena Hinojosa Tijerina	Gestión de datos y redes	**N**
	MC. Hugo German Reyes Anastasio	Gestión de gran volumen de datos, y arquitectura e ingeniería de software	**N**
	MC. Dante Domizzi Sánchez Gallegos	Gestión de gran volumen de datos, Workflows para la gestión de datos medioambientales y climáticos.	**N**
	MC Juan Armando Barrón Lugo	Fusión de datos, gestión de gran volumen de datos.	**N**
	MC. Diana Elizabeth Carrizales Espinoza	Gestión de gran volumen de datos, almacenamiento en la nube y en contenedores, sistemas distribuidos.	**N**
	Ing. Catherine Alessandra Torres Charles	Sistemas de preparación de datos médicos, intercambio seguro de datos en infraestructuras heterogéneas.	**N**
	Ing. José Carlos Morín García	Diseño y desarrollo de sistemas de orquestación y fusión de datos (Big data analytics) basados en patrones espaciales-temporales.	**N**
	Ing. Genaro Juan Sánchez Gallegos	Esquemas de patrones de paralelismo.	**N**
	Ing. Jesús Ignacio Castillo Barrio	Esquemas de alta disponibilidad para mejorar la eficiencia de los sistemas de almacenamiento	**N**

4.3. Grupo de trabajo interinstitucionales

Tabla 3- Equipo de trabajo interinstitucional

Institución	Participante	Líneas de investigación	Nivel SNI
UC3M	Dr. Jesús Carretero Pérez	Arquitectura y Tecnología de Computadores	**N**
UC3M	Dr. Javier García Blas	Arquitectura y Tecnología de Computadores	**N**
UC3M	MC. Lara Visuña Pérez	Arquitectura y Tecnología de Computadores	**N**
UAM	Dr. Ricardo Marcelín Jiménez	Codificación, tolerancia a fallos, Sistemas distribuidos, sistemas de almacenamiento, redes	**1**
UAM	M. en C. Oscar Yáñez Suárez	Procesamiento digital de señales e imágenes médicas. Inteligencia Artificial	**N**
INRII	M. en C. Marco Antonio Núñez Gaona	Médico especialista en imagenología del sistema músculo esquelético	**N**
INRII	M. en C. Heriberto Aguirre Meneses	Procesamiento digital de señales e imágenes médicas, Tecnologías de la información. Inteligencia Artificial	**N**

INRII	Dr. Garly Daniel González Rosado	Médico especialista en imagenología del sistema músculo esquelético	**N**
INRII	Ing. Laura Aguilar Caballero	Sistemas de información hospitalaria, procesamiento de imágenes médicas	**N**
INRII	M. en C. Nadezhda Aguilar Blas	Desarrollo Web, Redes neuronales artificiales	**N**
CICESE	Dr. Alejandro Galaviz Mosqueda	Sistemas de e-Salud, Internet de las Cosas Médicas - m-IoT, redes de comunicaciones inalámbricas, redes ad-hoc.	**1**
CICESE	Dr. Salvador Villarreal Reyes	Sistemas de e-Salud, Internet de las Cosas Médicas - m-IoT, redes de comunicaciones inalámbricas, redes ad-hoc.	**1**
CICESE	Dr. Andrei Tchernykh	Investigación en redes y en la nube que abordan la optimización de recursos multiobjetivo, seguridad, incertidumbre, programación, heurística y metaheurística, asignación de recursos adaptativos, algoritmos conscientes de la energía e Internet de las cosas.	**2**
University of Naples "Parthenope"	Dr. Raffaele Montella	Computación en red, nube y GPU, predicciones y simulaciones ambientales (tiempo, clima, atmósfera, océano), computación móvil y sistemas integrados, Internet de las cosas, virtualización de GPGPU, middleware para ciencias ambientales computacionales (es decir, flujos de trabajo, acoplamiento de modelos, aprovisionamiento de datos multidimensionales), y ciencia de datos.	**N**
Beca Postdoctorado Conacyt Modalidad 2 asignado al proyecto	Dr. Nelson Emmanuel Ordóñez Guillén	Diseño de sistemas de inteligencia artificial. Desarrollo de algoritmos de aprendizaje máquina y aprendizaje profundo para la medición de riesgo asociado a enfermedades crónica no contagiosas.	**N**

5. Productos conseguidos comprometidos en la Etapa 2[1, 2]

En esta sección se describen los productos comprometidos durante la segunda etapa del proyecto No. **41756.** En este apartado se dan a conocer los aspectos relevantes de cada producto, para los entregables llamados *Muyal-Xelhua, Muyal-Nez, Muyal-Chimalli,* así como para los servicios creados utilizando la plataforma *Muyal-llal.*

5.1. *Muyal-Xelhua*: Servicio de ciencia de datos de alta disponibilidad y tolerante a fallos.

Xelhua es un sistema de big data agnóstica para la construcción, asistida por el diseño, de servicios de ciencia de datos de alta disponibilidad para la toma de decisiones basada en datos. El sistema Xelhua consta de cuatro componentes principales:

[1] Las evidencias del proyecto están disponibles a través del siguiente enlace.
[2] La página web del proyecto está disponible en el siguiente enlace.

1. Un marco de diseño de alto nivel. Permite seleccionar diferentes herramientas de análisis de datos y de aprendizaje automático a partir de una malla de servicios acoplados en pipelines de procesamiento. El sistema Xelhua implementa un servicio de diseño impulsado por datos (figura 2, desarrollo), permitiendo crear pipelines de big data de alto nivel, lo que produce grafos acíclicos dirigidos (DAG[3] , por sus siglas en inglés).
2. Un nuevo modelo de procesamiento de Extracción-Transformación-Carga (ETL[4] , por sus siglas en inglés) recursivo. Permite al sistema Xelhua convertir automáticamente los diseños de pipelines en estructuras de software independientes a la infraestructura, basándose en el DAG producido en la fase de diseño. ETL, es un proceso de integración de datos que extrae, transforma y carga datos de múltiples fuentes a un almacén de datos o a otro repositorio de datos unificado. Este modelo se encarga de encapsular las aplicaciones analíticas de datos, en imágenes de software genéricas agnósticas a la infraestructura, denominadas ABox, que incluyen las dependencias, bibliotecas y sistemas operativos requeridos por las aplicaciones analíticas para ser ejecutadas en una plataforma de contenedores virtuales (figura 2, desarrollo). Estas imágenes también incluyen interfaces de entrada/salida para interconectar diferentes estructuras ABox, creando los pipelines.
3. Un modelo de orquestación para gestionar de forma transparente la entrega y recuperación de datos a lo largo de cada fase de los pipelines de procesamiento. Este modelo garantiza que el intercambio de datos se orqueste siguiendo la dirección de las aristas del DAG (figura 2, ejecución)).
4. Un modelo descentralizado que enmascara automáticamente los incidentes de indisponibilidad de los servicios para reducir los efectos secundarios del bloqueo del proveedor relacionados con los cortes o la indisponibilidad de los datos y la infraestructura. Este modelo se basa en esquemas de gestión de datos y eventos, implementados en un software que se incrusta en las estructuras ABox. Estos esquemas de gestión de eventos cumplen con los requisitos no funcionales (NFR , por sus siglas en inglés) para garantizar el funcionamiento continuo de los servicios de big data mediante la creación de redes P2P (figura 2, operación).

[3] Direct Acyclic Graph
[4] Extraction-Transformation-Load

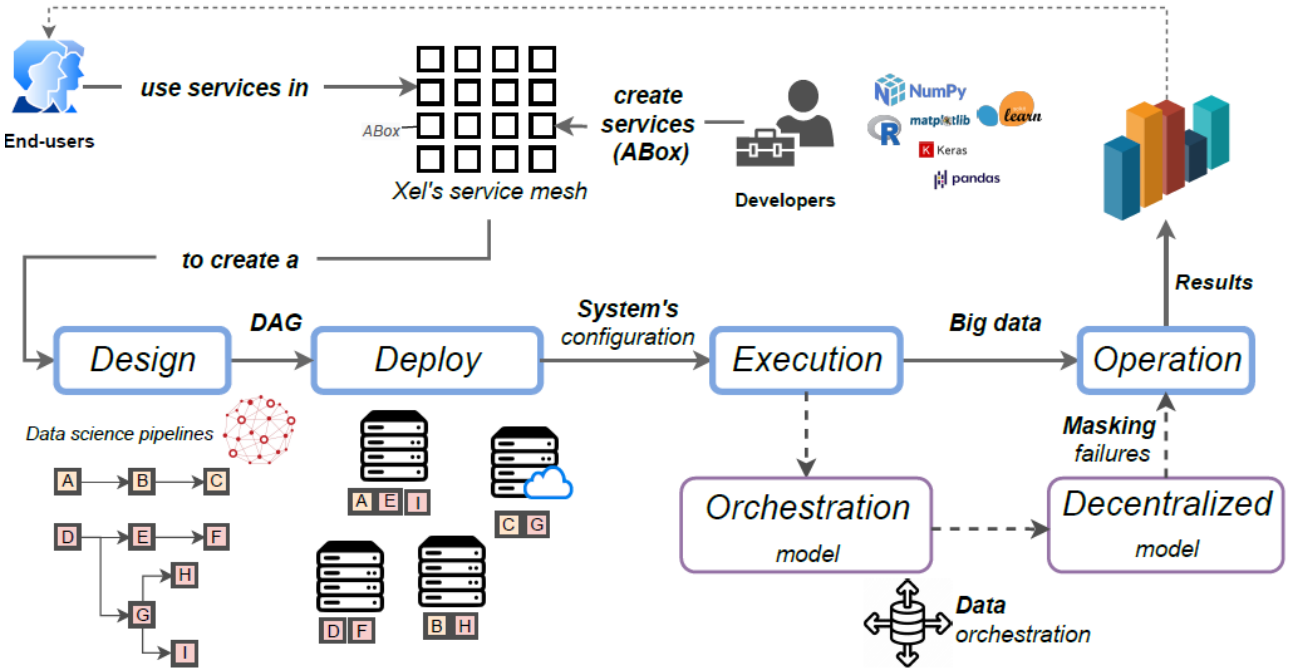

Figura 2. Representación gráfica de la metodología diseñada para el sistema Xelhua

El sistema Xelhua permite construir soluciones de alto nivel a través de un esquema impulsado por el diseño, convirtiendo automáticamente los diseños de pipelines en servicios de ciencia de datos agnósticos y de alta disponibilidad en la nube, desplegables en múltiples infraestructuras para hacer frente a los efectos secundarios del bloqueo del proveedor. En tiempo de ejecución, un motor de orquestación crea flujos de datos continuos, mientras que un modelo descentralizado garantiza las operaciones continuas de estos servicios de big data enmascarando los fallos detectados, como la indisponibilidad de aplicaciones y datos.

5.1.1. Principios de diseño del modelo de construcción de Xelhua para construir sistemas de ciencia de datos.

El sistema Xelhua define y está basado en la estructura lógica denominada *Agnostic Box* (ABox, de aquí en adelante), inspirada en la tecnología de contenedores de software, considerados el estándar para el despliegue de microservicios y aplicaciones en la nube. Una estructura ABox representa una aplicación analítica a través de la metainformación descriptiva del servicio proporcionado por dicha aplicación. La estructura ABox, al ser un contenedor de software, proporciona los componentes de software necesarios para la ejecución de la aplicación analítica, así como la configuración necesaria para el correcto despliegue del servicio proporcionado.

Las características principales del sistema Xelhua son las siguientes:

- **Agnosticismo**. El sistema Xelhua proporciona una serie de aplicaciones analíticas que integran servicios de big data, los cuales son independientes de la plataforma de software y la infraestructura de cómputo donde se ejecute . Para cumplir con lo anterior, una estructura ABox cuentan con las siguientes características:

- ***Conectividad***: Una estructura ABox puede interconectarse con otras estructuras ABox a través de interfaces de entrada y/o salida siguiendo un esquema de acoplamiento para crear cadenas de aplicaciones analíticas.
- **Manejabilidad**: Un grupo de estructuras ABox pueden ser gestionadas por un administrador descentralizado basado en un API Gateway (AG). Una AG provee funciones de acceso a servicios publicados por otras estructuras ABox.
- ***Portabilidad***: Una estructura ABox es un elemento autónomo con la capacidad de ser ejecutada en cualquier infraestructura con soporte para plataformas de contenedores.
- ***Usabilidad***: El sistema Xelhua proporciona una interfaz de usuario intuitiva, así como un proceso de implementación y administración guiados.

- **Disponibilidad**. Una estructura ABox de alta demanda[5] puede ser replicada o clonada para hacer más eficiente el servicio proporcionado por la estructura ABox, evitando la saturación de este y por tanto el tiempo de espera de cada usuario. Lo anterior provee una mejor experiencia de servicio para usuario final.
- **Confiabilidad**. El sistema Xelhua incluye una estrategia para el control de los fallos, basada en un esquema de consenso descentralizado el cual coordina la gestión y el enmascaramiento de los fallos. De esta forma, si el servicio proporcionado por una estructura ABox presenta un fallo, las peticiones a ese servicio son redireccionadas a una instancia réplica del servicio.

5.1.2. Construcción de una malle de servicios independiente de la plataforma e infraestructura de cómputo.

El sistema Xelhua proporciona diferentes aplicaciones analíticas para big data a través de la publicación de diferentes servicios. Los servicios proporcionados están contenidos en unas estructuras ABox, disponibles en un repositorio (R), basada en la tecnología de contenedores de software con las características de conectividad, manejabilidad, portabilidad y usabilidad. Las siguientes subsecciones detallan cada una de las características mencionadas.

5.1.2.1. Conectividad

El sistema Xelhua agrupa las estructuras ABox en una malla de servicios, denominada Big data service (BDServ). La BDServ está basada en una arquitectura de software para facilitar la comunicación servicio a servicio entre las estructuras ABox e incluso entre otros tipos de servicios o microservicios.

La construcción de un BDServ está basada en el acoplamiento de estructuras ABox, formando un pipeline. El BDServ integra una red de big data P2P (BD-P2P) para la gestión de la comunicación entre los elementos del pipeline a través de un entorno de malla de servicios (SME[6], por sus siglas en inglés) utilizando conexiones punto a punto (P2P).

[5] Una estructura ABox es de alta demanda si muchos usuarios hacen uso del servicio que proporciona.
[6] Service Mesh Environment

La BD-P2P representa un DAG de estructuras ABox interconectadas a través de conexiones P2P. Las interconexiones son representadas como tuplas (x, y) donde ambos elementos de la tupla son estructuras ABox, mientas que las aristas (A) del grafo representan las conexiones P2P o Internal Gateways (IG) entre estructuras ABox. Las IG representa una puerta trasera para conectar estructuras ABox del mismo BD-P2P o de una estructura ABox perteneciente a otro BD-P2P.

5.1.2.2. Manejabilidad

Para la gestión, ejecución y acoplamiento de aplicaciones, las estructuras ABox dependen de un conjunto de entidades denominadas Black Box (BBox). Una entidad BBox representa una abstracción de una aplicación analítica, por ejemplo, el servicio de una aplicación analítica (ABox) de clasificadores que ofrece diferentes opciones de clasificación de acuerdo con el tipo de clasificador deseado: supervisado (BBoxa), no supervisado (BBoxb), por mencionar un ejemplo. En este sentido, una estructura ABox se puede definir como sigue:

$$ABox = (server, \{BBox\}, client)$$

donde:

- **BBox**, representa un conjunto de entidades BBox, donde cada entidad BBox es una abstracción de una aplicación analítica con parámetros específicos, esto se representa como: $BBox = \{BBox_1, \ldots, BBox_2, \ldots, BBox_m\}$
- **server**, representa la fuente de datos de entrada para la estructura ABox
- **client**, representa la salida de datos de la estructura ABox.

Derivado de la definición anterior, y considerando que una entidad BBox es una abstracción de una estructura ABox, una entidad BBox se define como:

$$BBox = \{server, client\}$$

donde:

- **server**, representa la fuente de datos de entrada para la entidad BBox.
- **client**, representa la salida de datos de la entidad BBox.

5.1.2.3. Portabilidad y elasticidad

La portabilidad de una estructura ABox se consigue a través de la metainformación referente a la aplicación analítica contenida en dicha estructura. Los metadatos integran información sobre las dependencias de software relacionadas a la aplicación analítica asociada a la estructura ABox (por ejemplo, sistema operativo, bibliotecas y variables de entorno), dando lugar a una estructura ABox genérica autocontenida.

Una estructura ABox, así como las entidades BBox asociadas, sigue un modelo de Extracción-Transformación-Carga (ETL). ETL, es un proceso de integración de datos que extrae, transforma y carga datos de múltiples fuentes a un almacén de datos o a otro repositorio de datos unificado. El procesamiento realizado por una aplicación analítica,

encapsulada en una estructura ABox, sigue las etapas de ETL, donde el primer paso es la recolección de los datos de entrada (E) para la aplicación analítica, después el servicio proporcionado por la estructura ABox (la aplicación analítica) transforma (T) los datos y carga (L) los resultados para almacenarlos y, si es el caso, enviar los resultados a otra aplicación analítica para repetir el proceso hasta alcanzar el resultado final. La figura 3 ilustra el proceso de ETL, donde S representa la etapa de extracción (E), la estructura $Abox_i$ representa la etapa de transformación (T) y C representa la etapa de carga (L).

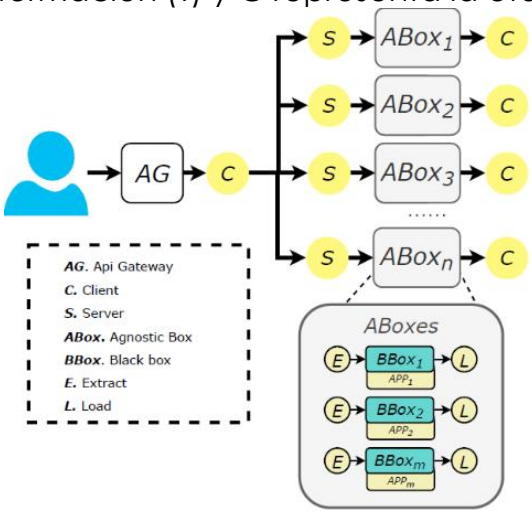

Figura 3. Representación gráfica de los componentes del sistema Xelhua.

En la figura 3, la estructura $ABox_n$ ilustra el concepto de una estructura ABox con diferentes entidades BBox, en otras palabras, la estructura $ABox_n$ define diferentes entidades BBox de acuerdo con los parámetros de la aplicación analítica que representa. Las etapas de extracción (E) y carga (L) de cada entidad BBox corresponden con las etapas de extracción (S) y carga (C) de la estructura ABox que la contiene. La entidad BBox para la etapa de transformación (T) dependerá de los parámetros definidos por el usuario con respecto al servicio proporcionados por la estructura ABox.

En este sentido, la etapa de transformación (T) puede estar asociada a un sólo servicio o a un conjunto de servicios definidos por las entidades BBox.

En función de la relación entre los datos de entrada y de salida en una entidad BBox, se definen dos tipos de relaciones:

- Relación paralela (PR). Las entidades BBox aceptan los mismos datos de entrada, realizan una transformación completa de los mismos y producen resultados distintos. Por ejemplo, supongamos que la entidad $BBox_a$ y la entidad $BBox_b$ ofrecen un servicio de clasificación con algoritmos diferentes; tanto la entidad $BBox_a$ como la entidad $BBox_b$ reciben el mismo conjunto de datos como entrada, sin embargo, cada servicio genera resultados distintos. La relación PR se representa de la siguiente manera:

$$PR = (E_j = E_h) \wedge (L_j \neq L_h) \mid (j \neq h) \wedge (BBox_j, BBox_h \in BBoxes_i)$$

- Relación secuencial (SR). Las entidades BBox pueden o no aceptar los mismos datos de entrada, realizan una transformación completa de los mismos y el resultado puede (o no) ser utilizado por otra entidad BBox, o por la misma entidad BBox, con un servicio complementario. Por ejemplo, supongamos que la entidad $BBox_j$ y la entidad $BBox_h$ contienen algoritmos de eliminación de valores atípicos. En este contexto, los datos de entrada son transformados por el proceso de limpieza del servicio proporcionado por la entidad $BBox_j$ y el resultado puede pasar (o no) como dato de entrada para el servicio proporcionado por la entidad $BBox_h$. De esta manera, es posible encadenar diferentes entidades BBox, creando pipelines de procesamiento dentro de una estructura ABox. La relación SR se representa de la siguiente manera:

$$SR = (L_j = E_h) \mid (j \neq h) \wedge (BBox_j, BBox_h \in BBoxes_i)$$

5.1.2.4. Usabilidad

El enfoque ETL es clave en el sistema Xelhua para el desarrollo de soluciones de big data. Un servicio de big data se construye mediante un conjunto de servicios de aplicaciones analíticas que se ejecutan en una determinada secuencia, siguiendo los pasos de ETL para la extracción de datos, su transformación y la carga del resultado en una fuente de datos especifica.

El uso del enfoque ETL provee al sistema Xelhua de la característica de generalidad, ya que las etapas de extracción, transformación y carga son independientes entre cada estructura ABox o entidad BBox, siendo posible acoplar los resultados a través de la malla de servicios (BDServ) para dar solución a un problema de big data particular. La característica de generalidad también permite expandir una malla de servicios añadiendo nuevos servicios (estructuras ABox) a la malla utilizando los componentes de extracción y carga.

A partir de lo anterior, una BDServ puede ser considerada como una serie de transformaciones (TR) sobre una determinada secuencia de datos, que son extraídos y/o almacenados en una fuente de datos (source) específica a través de tuberías, entonces:

$$BDServ = \{pipe_1, pipe_2, pipe_3, \dots, pipe_v\}$$
$$pipe = source \rightarrow TR_1 \rightarrow TR_2 \rightarrow \cdots \rightarrow TR_s \rightarrow sink \mid TR_i \in R$$

Los pipelines comienzan el procesamiento a partir de una fuente de datos (source) común. El proceso de transformación se realiza a través de cada una de las estructuras ABox hasta producir un resultado final (sink). Los pipelines de un BDServ pueden utilizar una estructura ABox la cantidad de veces necesarias, ya sea dentro de un mismo pipeline, o entre diferentes pipelines. Por ejemplo:

$$source_1 \rightarrow TR_1 \rightarrow TR_2 \rightarrow TR_3 \rightarrow TR_4 \rightarrow sink_1$$
$$=$$

$$source_1 \rightarrow ABox_1 \rightarrow ABox_3 \rightarrow ABox_1 \rightarrow ABox_2 \rightarrow sink_1$$

El flujo de datos comienza desde la fuente 1 ($source_1$), donde los datos son transformados por la estructura $ABox1$, cargando el resultado en la estructura $ABox_3$. La estructura $ABox_3$ carga su resultado en la estructura $ABox_1$, y, finalmente, la estructura $ABox_1$ carga el resultado en la estructura $ABox_2$. Debido a las características de las estructuras ABox, éstas les permiten conectarse entre sí, construyendo así estructuras complejas para dar solución a un problema de big data específico.

5.1.2.5. Disponibilidad y fiabilidad

En los escenarios del mundo real, varios usuarios pueden hacer peticiones a un conjunto de servicios para procesar datos. En este contexto, si uno de los servicios falla, los usuarios no pueden acceder a él. Una solución a este problema es utilizar la redundancia de servicios, por ejemplo, hacer copias de un servicio. El sistema Xelhua, considerando las características de las estructuras ABox, permite la implementación de redundancia de servicios debido a que estas estructuras pueden ser "clonadas" y desplegadas en múltiples recursos informáticos. Si una copia de la estructura ABox falla, las demás siguen estando disponibles para procesar los datos.

En escenarios donde no se implementa la redundancia de servicios para el control sobre los fallos, como es el caso de las infraestructuras/plataformas externalizadas (por ejemplo, la nube y serverless) comunes en los escenarios de big data. En estos escenarios, es necesario recopilar información sobre el estado actual de los servicios proporcionados y gestionar la carga de trabajo distribuida a las copias de tales servicios. Para lograr este tipo de gestión compleja, se diseñó una nueva versión serverless para la tolerancia a fallos y la alta disponibilidad del protocolo Paxos en combinación con un nuevo balanceo de carga y la gestión de la carga de trabajo basada en tablas hash distribuidas y un balanceo de carga probabilístico inspirado en el enfoque de dos-opciones (two-choices, por su nombre en inglés).

El acoplamiento de las estructuras ABox en el sistema Xelhua se realiza a través de redes P2P descentralizadas. El esquema descentralizado no sólo permite a las estructuras ABox establecer un consenso entre las entidades participantes de una red de servicios P2P, sino que también mantiene un grado de tolerancia a los fallos a través de la redundancia del servicio, así como el control del intercambio de datos entre las estructuras ABox.

En el esquema descentralizado, se utilizan entidades con diferentes roles como nodos, proponentes, aceptantes y aprendices. El rol de nodo realiza tareas de equilibrio de carga, asignación y distribución de la carga de trabajo. Los proponentes procesan cada solicitud que llega a un servicio (SV). Los aceptadores reciben y procesan cada solicitud y finalmente responden afirmativa o negativamente a los proponentes. Cuando los aceptadores reciben una solicitud, la envían a los aprendices, que, por consenso, deciden reenviar la solicitud a la estructura ABox correspondiente. El protocolo permite que, aunque fallen algunas entidades, las peticiones puedan ser atendidas por el resto de las entidades disponibles sin interrumpir la funcionalidad del servicio de big data.

Siguiendo este mismo proceso, los nodos permiten establecer controles sobre las réplicas de las estructuras ABox. Esto significa que en escenarios de fallos de una estructura ABox, los nodos realizan interacciones entre las estructuras ABox disponibles.

Nota: Para más información acerca de los servicios de construcción de soluciones de big data, ver el reporte técnico **"Xelhua: sistema agnóstico en la nube para la construcción de soluciones de big data basada en el diseño de servicios de ciencia de datos de alta disponibilidad y tolerancia"** en el **Anexo A.1.**

5.2. *Muyal-NEZ*: Servicio de construcción de sistemas e-salud

Muyal-Nez es un conjunto de servicios que permite a las organizaciones de salud y la comunidad científica crear sistemas de e-Salud agnósticos de la infraestructura para el procesamiento y manejo de grandes volúmenes de datos médicos.
Muyal-Nez permite crear sistemas de e-Salud de forma automática mediante interfaces gráficas y sin tener conocimientos avanzados de programación. Dichos sistemas de e-Salud se crean mediante el encadenamiento de dos o más aplicaciones para el procesamiento y manejo de datos médicos. Además, los servicios de e-Salud pueden ser manejados internamente por una organización (servicio de e-Salud intra-institucional) o por múltiples organizaciones (servicio de e-Salud inter-institucional).

Los servicios de e-salud construidos con *Muyal-Nez* tienen las siguientes características:

- **Modularidad**: los sistemas de e-salud se encuentran construidos utilizando abstracciones conocidas como bloques de construcción. Dichos bloques de construcción son autónomos e independientes, por lo que no afectan el funcionamiento de otros bloques, por lo tanto, pueden ser remplazados por otros bloques de construcción sin necesidad de construir un sistema de e-salud nuevo.
- **Agnosticidad**: tanto los sistemas de e-salud, como los bloques de construcción que conforman estos sistemas, son agnósticos. Es decir, estos sistemas pueden ser desplegados en diferentes infraestructuras sin necesidad de hacer cambios en el código de las aplicaciones. Los sistemas de e-salud pueden ser desplegados en diferentes plataformas y sistemas operativos, por ejemplo, en sistemas Windows o Linux, ya sea en una computadora personal o en la nube.
- **Eficiencia**: los sistemas de e-salud construidos integran patrones de paralelismo implícitos, los cuales mejoran la eficiencia para procesar datos reduciendo los tiempos de respuesta de las aplicaciones.
- **Portabilidad**: los sistemas de e-salud construidos pueden ser trasladados de una infraestructura a otra sin requerir hacer cambios en el sistema o las aplicaciones.
- **Reusabilidad**: mediante el repositorio y catálogo de sistemas y aplicaciones, es posible que una organización reutilicé un sistema de e-salud (o parte de él).

Muyal-Nez incluye los siguientes productos:

- Un esquema de construcción de cripto-contenedores de datos y cripto-contenedores de aplicaciones.
- Un esquema de despliegue de e-Servicios independientes de la infraestructura.

- Un esquema de bloques de construcción de flujos de trabajo y servicios de e-Salud basado en mapas de microservicios y nanoservicios.

5.2.1. Esquema de bloques de construcción de flujos de trabajo y servicios de e-Salud basado en mapas de microservicios y nanoservicios

El proceso de construcción de sistemas de e-salud se compone de tres etapas principales, las cuales se encuentran ilustradas en la Figura 4.

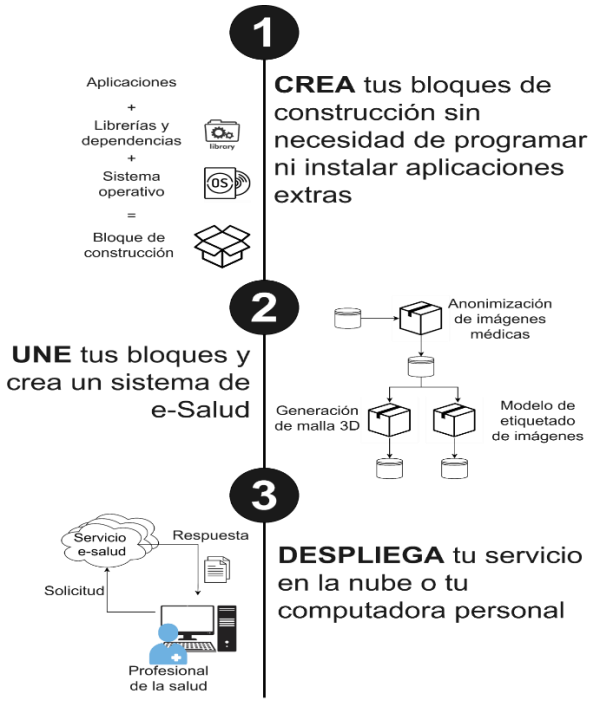

Figura 4. Construcción automática de sistemas e-salud.

Estas etapas son:

i. Creación de los bloques de construcción:
Este proceso consiste en encapsular en contenedores virtuales el código fuente o ejecutables de la aplicación que será ejecutada en el bloque de construcción. Además, dentro del contenedor virtual son colocadas las dependencias de software (librerías, frameworks, etc.) de la aplicación encapsulada, así como estructuras de control utilizadas durante tiempo de ejecución para invocar al bloque de construcción.

ii. Diseño de la estructura de procesamiento del sistema de e-salud:
En esta etapa, los bloques de construcción son organizados por el diseñador, de tal manera que la salida de una etapa es la entrada de la siguiente etapa en la estructura. La idea básica de esta etapa es que el diseñador construya un flujo de trabajo[1] interconectado cada uno de los bloques de construcción y eligiendo las fuentes de datos a procesar. En este sentido, diferentes patrones de procesamiento pueden ser diseñados. Algunos de los patrones básicos que es posible construir en un flujo de trabajo se describen a continuación:

- **Secuencial:** Una actividad en el proceso del flujo de trabajo es ejecutada inmediatamente después de que la actividad anterior ha sido completada.
- **Paralelo:** En algún punto del flujo de trabajo se pueden ejecutar diferentes hilos para ejecutar actividades en paralelo.
- **Sincronización:** Se utiliza cuando dos actividades paralelas convergen en una tercera actividad.
- **Elección exclusiva:** Se da cuando se debe de elegir cuál es la siguiente actividad en ser ejecutada a partir de un conjunto de actividades candidatas. Para ello se implementa un conjunto de condicionales en el flujo de trabajo.
- **Mezcla simple:** Sucede cuando en algún punto del flujo de trabajo dos o más actividades, que no se ejecutan paralelamente, se unen sin sincronización.

iii. Despliegue de la estructura de procesamiento:
Para ello se realiza el despliegue automático de los bloques de construcción (contenedores virtuales) en la infraestructura especificada. En este sentido, la estructura de procesamiento puede ser desplegada en un solo equipo o en múltiples equipos.

5.2.2. Esquema de construcción de cripto-contenedores de datos y cripto-contenedores de aplicaciones

Muyal-Nez implementa un esquema de construcción de cripto-contenedores de datos y aplicaciones. Su objetivo es cumplimentar, en forma automática y transparente, las normas oficiales (NOM-024-SSA3-2010 y NOM-004-SSA3-2012) e internacionales (ISO-270001-13, COBIT5, NIST) para garantizar la privacidad, confidencialidad, integridad, disponibilidad de los contenidos, tolerancia a fallas de servicios/servidores y trazabilidad.

Para distribuir los datos entre diferentes infraestructuras, *Muyal-Nez* está conectado con *Muyal-Chimalli* para crear redes de cripto-contenedores de datos y blockchain que permitan verificar que los sistemas de e-salud intercambien dato de forma segura.
Muyal-Nez contempla los requerimientos no funcionales de seguridad, confiabilidad, y eficiencia.

Los servicios de seguridad se agregan para resolver problemas que surgen cuando los datos e información son manejados y compartidos con múltiples usuarios a través de ambientes no controlados. En este contexto, la integridad y confidencialidad de datos, así como el control de acceso de usuarios son aspectos de seguridad importantes considerados en este servicio.

Los servicios de confiabilidad ayudan a solucionar problemas relacionados con fallas en la infraestructura donde los datos son procesados y almacenados. Este requerimiento resulta clave para evitar que los usuarios y organizaciones sufran de los efectos de no disponibilidad de los datos.

5.2.3. Esquema de despliegue de e-Servicios independientes de la infraestructura.

Una vez que se han construido los flujos de trabajo mediante los bloques de construcción y diseñado el sistema de e-salud, el siguiente paso es desplegarlos en la infraestructura en donde será ejecutado el sistema. En este sentido, el sistema y sus bloques pueden ser desplegados en un solo equipo en la nube o en una computadora personal, o de forma distribuida en un clúster de cómputo de alto desempeño dentro de una organización o desplegando los bloques en infraestructuras de diferentes organizaciones y/o la nube. Dos tipos de sistemas de e-salud pueden ser desplegados: i) intra-institucionales, e ii) inter-institucionales.

Los sistemas de e-salud *intra-institucionales* son aquellos permiten a las instituciones y organizaciones de salud procesar e intercambiar datos entre profesionales de la salud y departamentos dentro la organización u hospital.

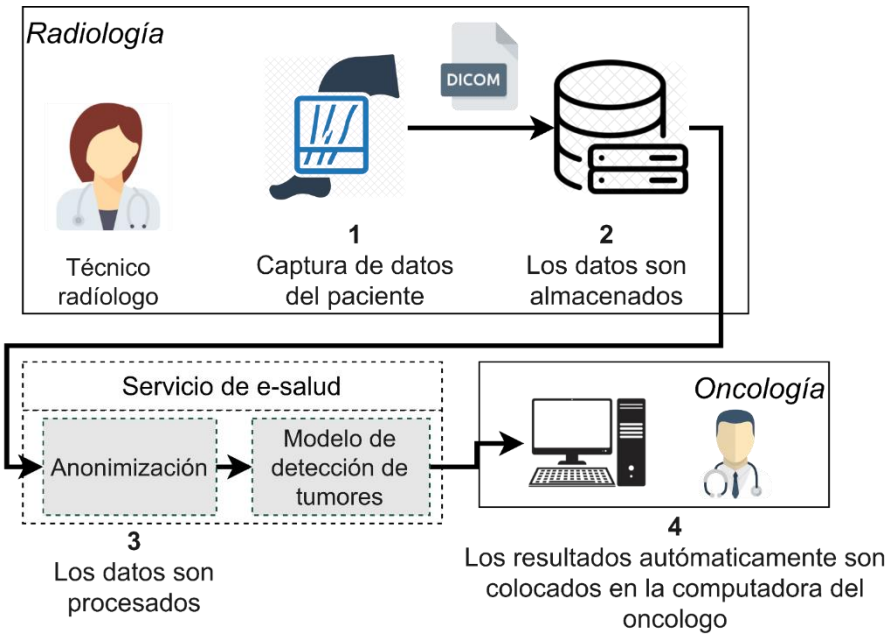

Figura 5. Ejemplo de un sistema de e-salud intra-institucional.

Un ejemplo de este tipo de sistema se muestra en la Figura 5. En este ejemplo, un técnico radiólogo realiza la captura de tomografías del paciente en el Instituto Nacional de Rehabilitación. Las imágenes son almacenadas, y el radiólogo las comparte con oncólogo del instituto para que las valore y de su diagnóstico. En este sentido, las imágenes antes de ser compartidas con el oncólogo son procesadas mediante un servicio de e-salud automático construido para anonimizar las imágenes y etiquetarlas utilizando un modelo de detección de tumores. Finalmente, los resultados son entregados al oncólogo de forma automática en su computadora.

Los sistemas de e-salud inter-institucional son creados para compartir datos entre diferentes instituciones, organizaciones u hospitales. En este tipo de sistemas, los datos son distribuidos utilizando una red segura de distribución de contenidos, la cual

automáticamente distribuye los datos a los participantes de las organizaciones que tenga autorización para acceder a los datos.

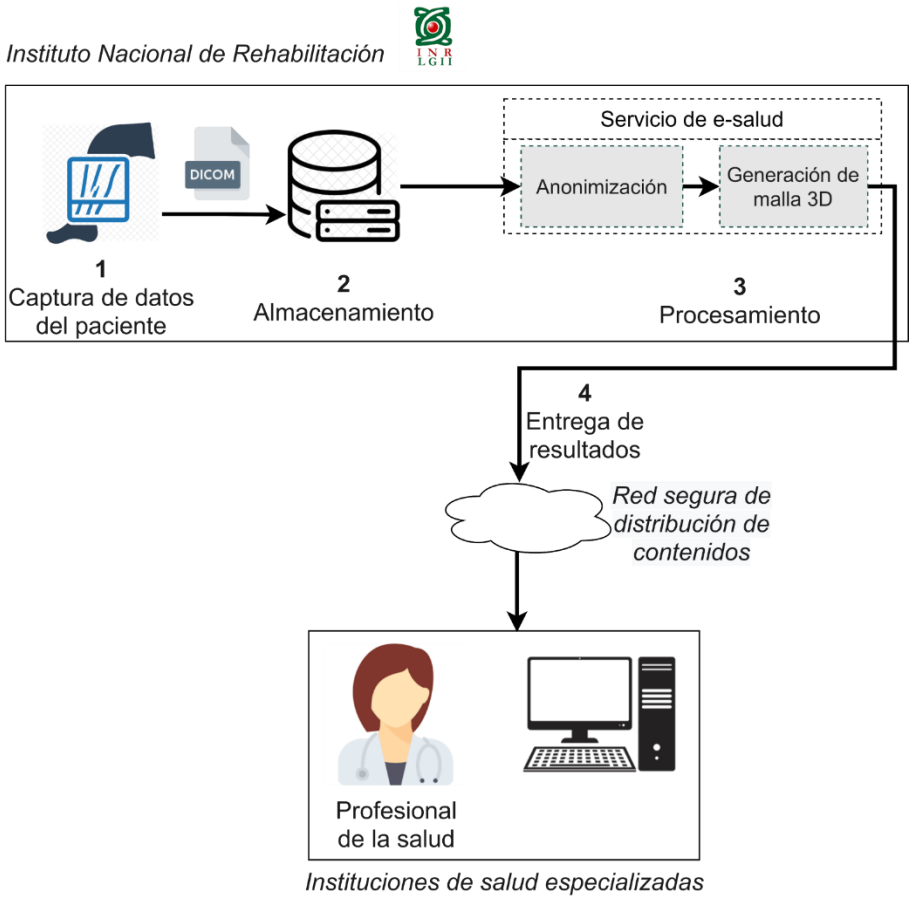

Figura 6. Ejemplo de un sistema de e-salud inter-institucional.

La Figura 6 muestra un sistema de e-salud inter-institucional creado para intercambiar datos entre el Instituto Nacional de Rehabilitación (INR) y una Institución de salud especializada. Los datos son capturados en el INR y almacenados en su infraestructura. Posteriormente, los datos son inyectados automáticamente en un sistema de procesamiento de e-salud el cual anonimiza los datos y realiza la generación de una malla 3D de las imágenes DICOM. Los resultados, son entregados a la red segura de distribución de contenidos, y posteriormente son automáticamente colocados en las computadoras de los profesionales de salud con acceso a los datos.

Nota: Para más información acerca de los servicios de construcción de sistemas de e-salud, ver el reporte técnico **"Nez: servicios de construcción de sistemas de e-salud"** en el **Anexo A.1.**

5.3. *Muyal-Chimalli*: Servicio que permite a las instituciones de salud, profesionales de la salud, pacientes y/o comunidad científica acceder a los servicios de e-salud y/o sistemas de analítica

Muyal-Chimalli es un servicio que permite a las instituciones de salud, profesionales de la salud, pacientes y/o comunidad científica acceder de forma segura a los servicios de

e-salud y/o sistemas de analítica. Para ello, cuenta con un sistema de almacenamiento eficiente y tolerante a fallos, y una red de distribución segura de contenidos sensibles. *Muyal-Chimalli* garantiza que los datos y los tomadores de decisiones sean aptos para realizar procesos de análisis y/o expuestos. Permite validar y registrar de eficientemente las operaciones realizadas en el servicio de e-salud, ya que cada operación o modificación a los datos es registrada en la blockchain de forma automática. Además, gestiona y verifica automáticamente contratos inteligentes y transacciones.

Muyal-Chimalli permite cumplir en un 70% las normas ISO 27001-13, NISTy COBIT para el transporte e intercambio de datos sensibles. Sin estos servicios, este porcentaje baja al 20% considerando solo tolerancia a fallos. Muyal-Chimalli asegura el anonimato de los datos, así como la confidencialidad mediante el cifrado de los datos entrantes y salientes de los sistemas de e-Salud. Además, permite detectar alteraciones en los datos. También permite la gestión automática de contratos inteligentes, la gestión automática de transacciones y la verificabilidad de transacciones de forma confidencial. Chimalli provee a los datos de las características de confiabilidad, eficiencia, integridad, confidencialidad y seguridad.

Muyal-Chimalli transporta y almacena los datos de forma eficiente mejorando los tiempos de respuesta hasta 10.22x veces. También reduce costos de almacenamiento en la nube hasta en un 70%, y permite compartir información sensible con profesionales de la salud externos a la organización principal.

Este servicio agrega seguridad a los sistemas de e-salud de forma automática para:

- Cumplir normas internacionales: sin programación ni extensos conocimientos sobre computación.
- Asegurar la confidencialidad de los datos: solo aquellas personas autorizadas tienen acceso a los datos.
- Evitar fugas de datos sensibles: los datos almacenados y transportados son anónimos

Muyal-Chimalli incluye los siguientes productos:

1. Servicios de preparación y recuperación de datos médicos configurables que incluya los requerimientos de seguridad, trazabilidad, integridad y eficiencia.
2. Mecanismos de trazabilidad de datos basados en blockchain.
3. Mecanismos de control de acceso de usuarios.
4. Servicios de validación de normas oficiales mexicanas y protocolos DICOM/HL7.
5. Un servicio que permite la utilización de técnicas de criptografía de siguiente generación para la transformación de datos en objetos seguros.

5.3.1. Servicios de preparación y recuperación de datos médicos configurables que proveen seguridad, trazabilidad, integridad y eficiencia a los datos

En los procesos de preparación de datos para escenarios reales de gestión de datos, diferentes requerimientos no funcionales (RNFs) (por ejemplo, seguridad, eficiencia, y confiabilidad) deben ser considerados debido a las normas de gestión de la salud (por

ejemplo, las normas oficiales mexicanas NOM-024-SSA3-2010 y NOM-004-SSA3-2012) y las leyes impuestas por los gobiernos y organizaciones [1], [2].

En este sentido, el esquema de preparación de datos que añade propiedades no funcionales a los datos. La preparación de los datos se realiza antes del transporte de estos a través de los flujos de datos (cargados para su almacenamiento o transmitidos utilizando entornos no controlados como Internet y la nube [3]). Primero se describe la estructura de procesamiento de tuberías definida en Chimalli para crear los esquemas de preparación, y más tarde se describen los RNFs elegidos para ser añadidos a la tubería.

La estructura de los esquemas de preparación de datos se encuentra construido en forma de una tubería, la cual se modelo con base en un gráfico acíclico dirigido (DAG, por siglas en inglés de Directed Acyclic Graph). En el DAG, los nodos representan a los algoritmos de los RNFs, y las aristas representan la entrada requerida por los algoritmos y los resultados producidos por ellos. Por lo tanto, una tubería puede incluir tantos algoritmos de NFRs como sea necesario para cumplir con las normas y leyes internacionales para el almacenamiento e intercambio de datos sensibles. De este modo, la ejecución secuencial de los RNFs crea una tubería de procesamiento.

Los algoritmos para cumplir con los RNFs en los flujos de datos son computacionalmente costosos. Esto se debe a que los algoritmos añaden retrasos a cada etapa de la tubería definida por un DAG. Para mitigar este impacto en el rendimiento de la preparación, los esquemas consideran bifurcaciones para producir paralelismo de datos y procesamiento de tareas concurrentes en cada etapa de la tubería.

Las etapas invocan al gestor de paralelismo y crean clones de los algoritmos de RNFs para asignarles carga de trabajo, lo que convierte a estos nodos en trabajadores. Además, el gestor de paralelismo despliega un balanceador de carga para mejorar el rendimiento del procesamiento de datos/tareas. Este modelo de procesamiento produce un paralelismo implícito que permite la ejecución del algoritmo de los RNFs de forma paralela y/o concurrente. Esto reduce el tiempo necesario para preparar los datos antes de transportarlos a través de los flujos de datos.

5.3.2. Mecanismos de trazabilidad de datos basados en blockchain

El proceso de trazabilidad juega un papel importante dentro de los flujos trabajo debido a que brindan la posibilidad de identificar el origen y las distintas etapas por las que ha pasado un producto a lo largo de todo su ciclo de vida (proceso productivo, distribución y logística, hasta llegar a un consumidor final).

Este proceso cumple una parte fundamental dentro del proyecto debido a que permitirá a cualquiera de las entidades involucradas (personal médico y usuarios finales) acceder a la información de cada una de las etapas por las cuales ha pasado el contenido digital, verificando si este cumple con las acciones pactadas y que ha sido procesado por las entidades correctas.

Lo anterior posibilita aceptar o rechazar el expediente digital basado en la información del flujo del producto (traza) apoyando de esta manera la toma de decisiones y mejorando la confianza en el resultado obtenido.

El mecanismo de trazabilidad de Chimalli provee las características de trazabilidad y verificabilidad a cada uno de los productos que se procesan en cualquiera de los servicios construidos a través de la plataforma de e-Salud. Este mecanismo permite realizar trazabilidad tanto interna como externa (dentro de una misma institución, así como colaboraciones entre varias de ellas) de los productos digitales que son procesados y manejados a través de los sistemas de e-Salud.

Además, la tecnología de blockchain permite almacenar y transmitir información de forma transparente, distribuida y segura sin un órgano central de control. Esta tecnología utiliza una base de datos segura compartida por diferentes usuarios autorizados para que todos puedan comprobar la validez de los procesos realizados en el flujo de trabajo en cada uno de sus bloques. Debido a las características previamente mencionadas, la blockchain tiene muchas ventajas para el sector de cadenas de suministros (enfoque utilizado en el actual proyecto de flujos de tareas en salud).

5.3.3. Mecanismos de control de acceso de usuarios

Para poder cumplimentar con el requisito de seguridad se creó un repositorio de bloques de seguridad. Dicho repositorio cuenta con un conjunto de bloques de seguridad disponibles para que los usuarios finales puedan incluirlos en sus cripto-contenedores.

Este repositorio incluye bloques de seguridad como criptosistemas simétricos (estándar de cifrado avanzado, AES [4]), cifrado basado en emparejamiento (firmas cortas) y cifrado basado en atributos (CP-ABE) [5], así como un bloque de trazabilidad, que se utiliza para registrar cada transacción de intercambio de información en una cadena de bloques privada [6].

En este contexto, CP-ABE [7] se utiliza para hacer cumplir criptográficamente el *control de acceso*. Los criptosistemas basados en emparejamiento que producen firmas cortas se utilizan para servicios de autenticación, no repudio e integridad. Estos criptosistemas

pueden proporcionar cualquiera de los niveles de seguridad equivalentes a 128, 192 o 256 bits, que cumplen con la mayoría de los estándares (por ejemplo, NIST [8], [9]).

Tanto CP-ABE como las firmas cortas utilizan un emparejamiento bilineal (asimétrico) computable eficiente: e: $\boldsymbol{G_1} \times \boldsymbol{G_2} \rightarrow \boldsymbol{G_T}$, con $\boldsymbol{G_1} - \langle g_1 \rangle$, $\boldsymbol{G_2} - \langle g_2 \rangle$ y $\boldsymbol{G_T}$, los cuales son grupos cíclicos de orden r. En la práctica, $\boldsymbol{G_1}, \boldsymbol{G_2}$ son subgrupos de una curva elíptica definida sobre un campo finito F_q , y $\boldsymbol{G_T}$ es el grupo multiplicativo del campo de extensión $\boldsymbol{F_{qk}}$, con k referido como el grado de incrustación de la curva elíptica, que es el número entero positivo más pequeño tal que r divide a $q^k - 1$. Los procedimientos de firma y verificación se implementan utilizando la instancia de la función hash SHA. La combinación de estos algoritmos y funciones da como resultado la adición de diferentes propiedades de seguridad a la información mediante un cripto-contenedor.

5.3.4. Servicios de validación de normas oficiales mexicanas y protocolos DICOM/HL7

En los procesos de preparación, transporte, compartición y almacenamiento de datos para escenarios reales de gestión de datos sensibles, diferentes requerimientos no funcionales (NFRs) (por ejemplo, seguridad, eficiencia, y confiabilidad) deben ser considerados debido a las normas de gestión de la salud (por ejemplo, las normas oficiales mexicanas NOM-024-SSA3-2010 y NOM-004-SSA3-2012) y las leyes impuestas por los gobiernos y organizaciones [1], [2].

Para asegurar el cumplimiento de estas normas y leyes, es necesario contar con un servicio que permita su validación. En este sentido, en este proyecto se desarrolló un servicio que permite determinar el grado de cumplimiento de los flujos de trabajo generados en las soluciones de e-Salud con base en los marcos de referencia de normas internacionales (NIST, ISO 27001:2013 y COBIT 5) y nacionales (NOM-024-SSA3-2010) para el transporte y almacenamiento de datos sensibles.

El programa desarrollado cumple con las siguientes características:

- Funcionales:
    - Preprocesamiento de los datos para leer y manipular los archivos desde el código.
    - Identificación de los flujos de trabajo que conforman los servicios de e-Salud.
    - Consulta de fuentes de información utilizando APIs para obtener datos y características contextuales de los contenedores, las cuales representan las tareas que ejecuta el contenedor.
    - Búsqueda de palabras clave para determinar el cumplimiento de las normas internacionales y nacionales.

- Obtención del porcentaje de cumplimiento, y generación de un reporte donde se visualizan los resultados.
- Descubrimiento de los flujos de trabajo asociados a los archivos de configuración del servicio de e-salud, y generación de una representación de este en un grafo acíclico dirigido.

- No funcionales:

  - Norma nacional que se verifica: NOM-024-SSA3-2010.
  - Normas internacionales que son verificadas: NIST, ISO 27001:2013 y COBIT.
  - Eficiencia y eficacia en la realización de tareas.

5.3.5. Servicio para la transformación de datos en objetos seguros mediante el uso de técnicas criptográficas de siguiente generación

El cómputo en la nube se está convirtiendo en el nuevo núcleo de aplicaciones y servicios. Se espera que, en los próximos años, el 49 % de los datos se almacenen en la nube [10]. Al igual que el almacenamiento en la nube, los servicios de entrega de contenido se han convertido en piedras angulares para que las organizaciones, los usuarios finales y los trabajadores participen en cualquiera de los flujos de trabajo organizativos en línea, trabajo remoto parcial/total u oficina en casa [11].

Para evitar incidentes o mitigar riesgos que aún surgen en la nube, como alteraciones de datos [12], privacidad, violaciones de confidencialidad y accesos no autorizados, las organizaciones deben entregar o recuperar información a/de socios o usuarios finales de forma segura y transparente [13]. Para ello, *Chimalli* cuenta con un servicio de patrones paralelos que permite construir sistemas de seguridad eficientes para que las organizaciones compartan, intercambien y rastreen información confidencial en la nube.

En este servicio, los criptosistemas de seguridad y el software de blockchain se convierten en servicios en la nube independientes y autónomos llamados *cripto-contenedores*. Para mejorar la eficiencia de los servicios de seguridad y la experiencia del servicio del usuario final, se agrega un patrón paralelo implícito junto con el balanceo de carga a los cripto-contenedores. Este servicio permite a las organizaciones respaldar patrones de intercambio de información en línea entre múltiples participantes al acoplar conjuntos de cripto-contenedores para cumplir con múltiples combinaciones de requisitos de seguridad (por ejemplo, confidencialidad, integridad, no repudio, autenticación y trazabilidad). Este servicio de *Chimalli* cuenta con dos características principales:

- ***Flexibilidad*** para integrar, sobre la marcha y bajo demanda, tantas aplicaciones de seguridad (criptosistemas) como inquietudes expresadas por las organizaciones

(para las etapas), por los participantes de cada flujo de trabajo organizacional, en un único sistema de seguridad integral.

En este método, los criptosistemas de seguridad y el software de blockchain se convierten en servicios en la nube independientes y autónomos llamados cripto-contenedores. Los conjuntos de cripto-contenedores se combinan para crear sistemas de seguridad en la nube para admitir flujos de trabajo organizacional en línea que incluyen a varios participantes. Un framework basado en este método, crea sistemas de seguridad en la nube que cumplen con múltiples combinaciones de requisitos de seguridad, como confidencialidad, integridad, no repudio, autenticación y trazabilidad.

- ***Eficiencia*** utilizando un modelo de programación paralela de gestión de datos basado en la combinación de patrones y esquemas de balanceo de carga, que están integrados en los cripto-contenedores. El modelo es utilizado para crear dos esquemas de patrones paralelos llamados "*Pipeline*" y "*Overlapped*".

    - El patrón *Pipeline* incluye dos tipos de patrones: un *pipe&filter* (tuberías y filtros) en combinación con el patrón manejador/*trabajador*. El primer patrón organiza los criptosistemas en forma de tuberías, mientras que el segundo despliega estas tuberías como trabajadores para ejecutarlos en paralelo. Este esquema fue diseñado para codificar/decodificar conjuntos de archivos/tareas pequeñas (por ejemplo, asegurar archivos pequeños con un tamaño de clave pequeño de 128 bits) en paralelo.

    - El patrón *Overlapped* acopla criptosistemas independientes para que se ejecuten de manera superpuesta, mientras que los criptosistemas que incluyen un tipo de dependencia se acoplan en forma de tubería. Todos los criptosistemas se ejecutan como una tubería, que también se gestionan como trabajadores (en un patrón de *manejador/trabajador*); como resultado, las tuberías Overlapped también se ejecutan en paralelo. Este esquema fue diseñado para codificar/decodificar conjuntos de grandes contenidos/tareas (por ejemplo, proteger archivos con un tamaño de clave grande de 192 y 256 bits) en paralelo.

Este servicio tiene dos componentes principales:

1. Un método de seguridad múltiple en la nube para crear servicios de gestión de seguridad de la información confidencial flexibles, integrales y eficientes.

2. Dos nuevos patrones paralelos eficientes e implícitos como ***PFP*** y ***Overlapped*** para mejorar significativamente el rendimiento de los sistemas de seguridad en la nube, así como la experiencia de servicio de los usuarios finales.

Nota: Para más información acerca del servicio para el acceso seguro a servicios de e-salud y/o sistemas de analítica, ver el reporte técnico **"Chimalli**: **Servicio de acceso seguro a servicios de e-salud y/o sistemas de analítica"** en el **Anexo A.2.**

5.4. Muyal-Painal: Servicio para el intercambio seguro y confiable de datos médicos

Painal es un conjunto de servicios y sistemas desarrollados para que las organizaciones de salud y la comunidad científica puedan i) almacenar, distribuir y localizar sistemas o servicios de procesamiento a través de catálogos de servicios ii) generar soluciones que permitan brindar usabilidad costo-beneficio del almacenamiento y transporte de datos y iii) almacenar, publicar y transmitir repositorios de datos de manera local (intrainstitucional) y federada (interinstitucional) utilizando un modelo de publicación/suscripción.

Painal permite crear catálogos de servicios en las cuales las organizaciones pueden colocar sus sistemas, servicios o aplicaciones para que otras instituciones de la federación puedan descargar y utilizar.

Por otro lado, para almacenar los resultados producidos por sus servicios y aplicaciones es necesario generar sistemas de almacenamiento que consideren las características de costo-beneficio del almacenamiento y transporte de datos. Painal ofrece una arquitectura de malla para generar el almacenamiento de datos considerando los recursos disponibles en la infraestructura de la organización que la está utilizando.

Por último, Painal permite crear catálogos de datos para el almacenamiento y compartición de datos entre múltiples organizaciones de forma segura. Dentro de los catálogos de datos se pueden almacenar los datos sin procesar o los resultados obtenidos por algún tipo procesamiento. Cuenta con un mecanismo de publicación/suscripción que permite generar catálogos de datos, agregar nuevos datos a dichos catálogos y descargar el contenido de estos. Todo el proceso de publicación y suscripción se basa en la utilización de tokens de acceso que permiten verificar la entidad de las organizaciones de la federación y los permisos que tienen para acceder a los catálogos.

Painal se compone por los siguientes productos:

1. Servicio manejador de catálogos para la carga y descarga de servicios.
2. Mecanismo para el almacenamiento en la nube que brinda usabilidad costo-beneficio para el almacenamiento y transporte de datos.
3. Servicio de publicación suscripción para el manejo de catálogos, fuentes y repositorios de datos.

5.4.1. Servicio para el manejo, carga y descarga de servicios desde un catálogo/repositorio de servicios

En esta sección se presenta el diseño de un sistema de distribución de contenidos federado (FCDS, por sus siglas en inglés) desarrollado para Painal y empleado para crear servicios de sincronización de datos médicos que son tecnológicamente agnósticos de la infraestructura, lo cual permite a las organizaciones desplegar sus servicios en cualquier nube privada, pública o híbrida. La Figura 7 muestra una representación conceptual del FCDS de Painal, que permite compartir datos (p. ej., tomografías, mamografías y resonancias magnéticas) de forma sincrónica entre tres hospitales a través de una Red de Entrega de Contenidos (CDN, por sus siglas en inglés) [14] [15]. La motivación pare el uso de la federación de servicios es que las organizaciones puedan tener un gobierno sobre sus servicios, datos, infraestructura y los requisitos no funcionales mediante la inclusión de métodos para cumplirlos.

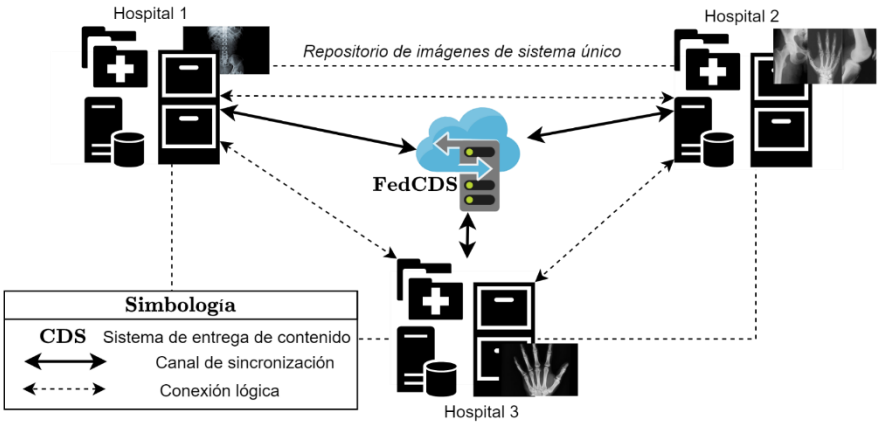

Figura 7. Arquitectura del sistema.

5.4.2. Mecanismo de usabilidad costo-beneficio para el almacenamiento y transporte de datos

En esta sección se presenta una arquitectura de malla para el almacenamiento de datos que permite crear y operar servicios de almacenamiento configurables, fiables y flexibles para infraestructuras heterogéneas, la cual permite crear servicios de almacenamiento confiables sin servidor (SeRSS) en Painal.

La arquitectura de malla para el almacenamiento de datos se basa en una estructura de almacenamiento, que se asocia para gestionar nodos de almacenamiento (SN, por sus siglas en inglés). Los SN son sistemas de almacenamiento tradicionales que incluyen sistemas como la entrega/recuperación de archivos, balanceadores de carga, distribución de datos y los sistemas de colocación, así como los nano servicios para atender los requisitos no funcionales (NFR, por sus siglas en inglés) basados en el algoritmo de dispersión de la información (IDA) [16], [17] para esquemas tolerantes a fallos, esquemas de preparación para resolver problemas de seguridad, así como patrones paralelos para mejorar la eficiencia de los componentes de los SN.

La idea básica es que las organizaciones puedan elegir los parámetros de almacenamiento a alto nivel (es decir, el número de nodos de almacenamiento necesarios para una solución determinada) y los parámetros NFR (por ejemplo, n, m o los tipos de seguridad y eficiencia requeridos por los usuarios finales). SeRSS utiliza esta información para construir automáticamente un sistema de almacenamiento fiable en forma de sistema P2P sin servidor que puede ser consumido por los usuarios finales como un servicio, que puede ser desplegado en infraestructuras múltiples y heterogéneas.

5.4.3. Servicio de publicación/suscripción (pub/sub) para el manejo de catálogos, fuentes y repositorios de diferentes tipos de datos clínicos

En esta sección, se describe el servicio publicación/suscripción (pub/sub) sobre un almacenamiento en la nube diversificado. El servicio de pub/sub de Painal divide la entrega de contenidos en dos capas.

La primera capa se basa en patrones de publicación-suscripción que permiten devolver el control de los metadatos al propietario del contenido. Esta capa establece los siguientes roles (i) Editores, usuarios que producen nuevos contenidos, (ii) Usuarios finales, clientes externos que se suscriben a los contenidos, y (iii) Editores/administradores, usuarios de la organización encargados de aceptar/rechazar tanto las publicaciones como las suscripciones. Un esquema de gestión de recursos permite a las organizaciones establecer el control de acceso necesario para mantener los flujos de metadatos, así como el flujo de los patrones pub/sub entre editores y usuarios finales.

La segunda capa se basa en la dispersión de la información [17] sobre una plataforma de almacenamiento multi-nube con la que el servicio de pub/sub consigue un uso eficiente del espacio de almacenamiento y una alta fiabilidad. En esta capa, la dispersión se realiza en el lado del editor mediante el uso de mecanismos resistentes llamados flujos de trabajo de entrega, que dividen los contenidos en un conjunto de bloques redundantes y anónimos (utilizando procesos descritos en Chimalli) que se distribuyen en múltiples

proveedores de almacenamiento. De esta manera se garantiza que un determinado proveedor no recibe suficientes bloques para reconstruir el contenido original. Por tanto, la forma de reconstruir el contenido original se mantiene en el lado del editor. Los usuarios finales se encargan de obtener los contenidos publicados mediante flujos de trabajo de recuperación, que recuperan un subconjunto de bloques y reconstruyen los contenidos en el lado del usuario final. Esto significa que los flujos de trabajo de recuperación pueden obtener contenidos incluso cuando algunas ubicaciones de almacenamiento en la nube no están disponibles. Como resultado, se reducen los riesgos de la dependencia del proveedor y permite a la organización externalizar el almacenamiento de contenidos de forma controlada.

Nota: Para más información acerca del servicio para el intercambio confiable y seguro de datos médicos, ver el reporte técnico **"Painal: Servicio para el intercambio seguro y confiable de datos médicos"** en el **Anexo A.3.**

5.5. Entregable 4.1: Sistema de e-salud para el diagnóstico asistido de cáncer de hueso largo y pulmones mediante inteligencia artificial

El Sistema de e-salud para el diagnóstico asistido de cáncer de hueso largo y pulmones mediante inteligencia artificial, proporciona un apoyo al especialista indicarle en las imágenes en las que podría existir algún tumor, y la región en la que se ubica.

Dado que los tumores presentan diferentes características dependiendo de la parte del cuerpo en la que se desarrollen, es necesario generar un modelo diferente para cada uno. Anteriormente, se utilizaron técnicas en las que se buscaban intensidades de las tomografías que pudiesen indicar, por ejemplo, algún tipo de clasificación en algunos órganos, pero en el caso particular de cáncer de hueso, estas técnicas no siempre resultan eficientes, ya que el tumor puede presentar intensidades similares al de un hueso sano. Es por ello por lo que, para este proyecto, se decidió utilizar nuevas técnicas de visión artificial que permiten distinguir diferentes patrones más allá de diferentes intensidades.

5.5.1. Flujo para detección asistida para cáncer de huesos largos y Detección de nódulos en pulmón.

Se diseñó un flujo para la detección asistida para el cáncer de huesos largos existe y nódulos en pulmón (ver Figura 8). El flujo permite la obtención de las imágenes DICOM, en donde el proceso de extracción de estas contiene un formato adecuado. Posteriormente, se realiza una partición de estas de manera aleatoria y formar un conjunto de entrenamiento, pruebas y validación, en los que regularmente se toman el 70%, 20% y 10% de la cantidad total de imágenes. Mismas que se genera un TF Records a partir de

imágenes y XML´s. Continuando con el flujo, se realiza una tubería de entrenamiento extrayendo las imágenes y entrenando el modelo, seguido de la exportación del modelo y prueba del del mismo para obtener el conjunto de pruebas para la detección asistida para cáncer de huesos largos con base en marcos de referencia de normas internacionales (NIST, ISO 27001:2013 y COBIT 5) y nacionales (NOM-024-SSA3-2010). El programa, además, descubre el flujo de trabajo asociado a los archivos de configuración.

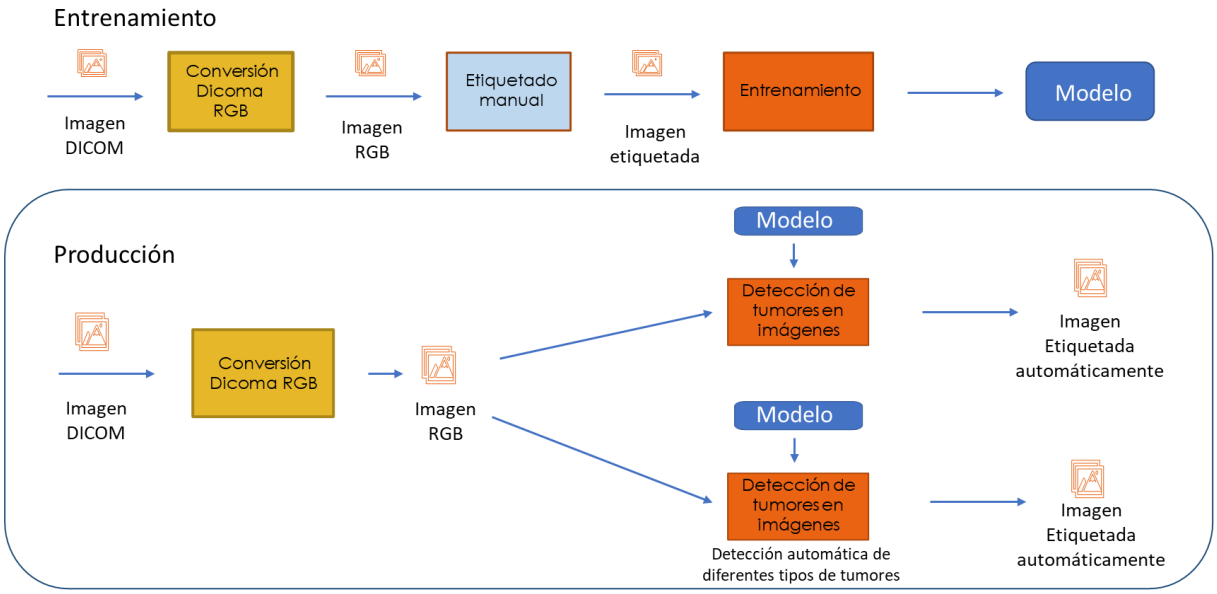

Figura 8. Representación general del flujo para la detección de cáncer de huesos largos y pulmón.

Nota: Para más información acerca del sistema de e-salud para el diagnóstico asistido de cáncer de hueso largo y pulmones mediante inteligencia artificial, ver el resumen ejecutivo **"Entregable 4.1: Sistema de e-salud para el diagnóstico asistido de cáncer de hueso largo y pulmones mediante inteligencia artificial"** en el **Anexo A.4.**

ANEXOS

A. Reportes Técnicos

A.1. Reporte técnico "Nez: servicios de construcción de sistemas de e-salud "

En años recientes, el uso de las tecnologías de la información y comunicación (TICs) ha jugado un rol importante en el ámbito médico, en la prevención, diagnóstico y tratamiento de patologías y enfermedades crónicas-degenerativas. En este sentido, organizaciones médicas construyen y administran sistemas de e-salud para transportar, procesar y almacenar datos médicos utilizando las TICs disponibles en la organización. En escenarios reales estos servicios de e-salud deben de hacer frente al reto de manejar grandes volúmenes de datos de forma eficiente. Por ejemplo, hasta 2020 el Instituto Nacional de Rehabilitación (INR) almacenaba 45 millones de imágenes médicas, lo que representa 43 TB de datos, los cuales deben de ser transportados y procesados para ayudar a médicos en la toma de decisiones. Otro reto al que deben de hacer frente las organizaciones para construir estos servicios de e-salud, es el de interconectar aplicaciones heterogéneas para el procesamiento de datos, que además pueden estar distribuidas en diferentes computadoras dentro de una organización, o incluso entre organizaciones…
Leer reporte completo…

A.2. Reporte técnico "Chimalli: Servicio de acceso seguro a servicios de e-salud y/o sistemas de analítica."

En los últimos años, la producción de datos ha crecido de forma exponencial debido a la producción continua de datos por fuentes como dispositivos de IoT (p. ej., electrocardiogramas, máquinas de rayos x y espirómetros) y dispositivos de usuarios finales (p. ej., celulares, tabletas, laptops y estaciones de trabajo). En este contexto, los métodos para acceder y gestionar los datos han cambiado para hacer frente a este crecimiento. Actualmente, los datos no se almacenan en un solo lugar, sino que son almacenados en diferentes ubicaciones durante su ciclo de vida. Lo anterior da como resultado una gestión de datos jerárquica para producir una respuesta rápida al analizar un gran volumen de datos en escenarios actuales de manejo de grandes volúmenes de datos (big data)…
Leer reporte completo…

A.3. Reporte técnico "Painal: Servicio para el intercambio seguro y confiable de datos médicos."

El volumen de datos producidos y gestionados por las organizaciones ha ido creciendo en los últimos años, esto debido a que los usuarios finales asociados a las organizaciones producen, almacenan y utilizan datos de forma constante y continua, lo que produce un efecto de acumulación de datos. Los usuarios finales y las aplicaciones consumen los servicios de almacenamiento en la nube a través de un modelo de externalización denominado pago por uso (pay-as-you-go). A pesar de que estos servicios se construyen utilizando sistemas distribuidos, una acumulación constante de datos crea gradualmente una colección centralizada de datos en los servicios de almacenamiento. Esto no solo da lugar a un único punto de fallo en los escenarios de interrupción, sino que también produce una dependencia con el proveedor de…

Leer reporte completo…

A.4. Resumen Ejecutivo "Entregable 4.1: sistema de e-salud para el diagnóstico asistido de cáncer de hueso largo y pulmones mediante inteligencia artificial."

La prevención y seguimiento juegan un rol sumamente importante para abordar el problema de las enfermedades crónico-degenerativas (ECD) en México. Para esto, es fundamental que los pacientes tengan un acceso efectivo a servicios de salud que contemplen el diagnóstico de las ECDs. Sin embargo, las barreras geográficas y económicas en México imponen retos importantes para acceder a este tipo de servicios, principalmente para el caso de zonas urbano-marginales y comunidades rurales. Se ha demostrado que las TICs tienen el potencial para mejorar el acceso a los servicios de salud por parte de la población. Debido a la escasez de profesionales de la salud altamente especializados y a la distribución no equitativa de los mismos, en años recientes se ha incrementado el interés en usar TIC para la prestación de servicios de medicina a distancia…

Leer reporte completo…

A.5. Reporte técnico "Xelhua: sistema agnóstico en la nube para la construcción de soluciones de big data basada en el diseño de servicios de ciencia de datos de alta disponibilidad y tolerante a fallos. "

Xelhua es un sistema de big data agnóstica para la construcción, asistida por el diseño, de servicios de ciencia de datos de alta disponibilidad para la toma de decisiones basada en datos. El sistema Xelhua consta de cuatro componentes principales: (i) un marco de diseño de alto nivel para la selección de herramientas analíticas y de aprendizaje automático, a través de una malla de servicios acoplados basada en pipelines de procesamiento, (ii) un nuevo modelo de procesamiento basado en el modelo de Extracción-Transformación-Carga (ETL, por sus siglas en inglés) recursivo para convertir automáticamente los diseños de pipelines en estructuras de software agnósticas a la infraestructura, (iii) un modelo novedoso de orquestación para gestionar, de forma transparente, la entrega de datos a lo largo de cada etapa ETL de los pipelines de procesamiento utilizados en los sistemas de ciencia de

datos, y (iv) un modelo descentralizado de datos para enmascarar de forma transparente la indisponibilidad de algún servicio debido a, por ejemplo, las interrupciones en la nube y la indisponibilidad de las aplicaciones o los datos. En la fase de experimentación, se generaron, a través del sistema Xelhua, servicios de ciencia de datos para el análisis de publicaciones científicas, el análisis de sentimientos a partir de una colección de publicaciones, en redes sociales, relacionadas a la pandemia provocada por el virus de covid-19 y el agrupamiento de críticas de películas. Estos servicios fueron evaluados como casos de estudio revelando la eficacia del sistema Xelhua para el diseño de soluciones en múltiples tipos de problemas de ciencia de datos basado en el diseño. Igualmente, se evalúo el enmascaramiento automático de fallas en el sistema por la falta de disponibilidad de recursos y datos en la nube. Actualmente, el sistema Xelhua es utilizado para la creación de un observatorio nacional del cáncer y de un sistema de ciencia de datos para fusionar conjuntos de datos relacionados al suicidio, salud mental y consumo de drogas, además de un conjunto de datos macroeconómicos para encontrar patrones espaciotemporales.
Leer reporte completo…

6. Referencias

[1] H. Mier y T. Delgadillo, «Regulación del acceso al expediente clínico con fines de investigación en México,» *Revista CONAMED,* vol. 22, pp. 27--31, 2018.

[2] Phillips, «International data-sharing norms: from the OECD to the General Data Protection Regulation (GDPR),» *Human genetics,* vol. 137, pp. 575--582, 2018.

[3] C. B. Tan, M. H. A. Hijazi, Y. Lim y A. Gani, «A survey on proof of retrievability for cloud data integrity and availability: Cloud storage state-of-the-art, issues, solutions and future trends,» *Journal of Network and Computer Applications,* vol. 110, pp. 75--86, 2018.

[4] J. a. R. V. Daemen, The Design of Rijndael: The Advanced Encryption Standard (AES), Springer Nature, 2020.

[5] M. Morales-Sandoval, J. L. Gonzalez-Compean, A. Diaz-Perez y V. J. Sosa-Sosa, «A Pairing-based Cryptographic Approach for Data Security in the Cloud,» *Int. J. Inf. Secur.,* vol. 17, p. 441–461, August 2018.

[6] F. B. a. S. H. a. T. Halevi, «Supporting private data on Hyperledger Fabric with secure multiparty computation,» *IBM Journal of Research and Development,* vol. 63, n° 2/3, pp. 3-8, 2019.

[7] N. a. L. J. a. Z. Y. a. G. Y. Chen, «Efficient CP-ABE Scheme with Shared Decryption in Cloud Storage,» *IEEE Transactions on Computers,* 2020.

[8] D. Giry, «NIST Report on Cryptographic Key Length and Cryptoperiod (2020),» Key length, 2020.

[9] E. a. B. E. a. B. W. a. P. W. a. S. M. a. o. Barker, «Recommendation for Key Management: Part 1-General,» National Institute of Standards and Technology, Technology Administration, 2020.

[10] F. Della Rosa, «Worldwide Software as a Service and Cloud Software Forecast, 2020–2024,» International Data Corporation (IDC), 2020.

[11] E. a. H. J. J. a. O. A. a. R. D. a. S. G. a. T. H.-Y. Brynjolfsson, «COVID-19 and remote work: An early look at US data,» National Bureau of Economic Research, 2020.

[12] J. a. M. B. a. V. W. Kelly Finnerty and Sarah Fullick and Helen Motha and Navin Shah, Cyber Security Breaches Survey 2019, Department for Digital, Culture, Media and Sport, 2019.

[13] C. H. a. W. C. a. L. Y. a. Y. D. a. S. J. a. Y. T. a. H. C. a. D. Chen, «Toward security as a service: A trusted cloud service architecture with policy customization,» *Journal of Parallel and Distributed Computing,* vol. 149, pp. 76-88, 2021.

[14] D. Higuero, J. M. Tirado, J. Carretero, F. Félix y A. de La Fuente, «HIDDRA: a highly independent data distribution and retrieval architecture for space observation missions,» *Astrophysics and Space Science,* vol. 321, p. 169–175, 2009.

[15] J. L. Gonzalez, J. C. Perez, V. J. Sosa-Sosa, L. M. Sanchez y B. Bergua, «SkyCDS: A resilient content delivery service based on diversified cloud storage,» *Simulation Modelling Practice and Theory,* vol. 54, p. 64–85, 2015.

[16] P. Morales-Ferreira, M. Santiago-Duran, C. Gaytan-Diaz, J. L. Gonzalez-Compean, V. J. Sosa-Sosa y I. Lopez-Arevalo, «A data distribution service for cloud and containerized

storage based on information dispersal,» de *2018 IEEE Symposium on Service-Oriented System Engineering (SOSE)*, 2018.

[17] R. Marcelín-Jimenez, J. L. Ramírez-Ortíz, E. R. De La Colina, M. Pascoe-Chalke y J. L. Gonzalez-Compean, «On the complexity and performance of the information dispersal algorithm,» *IEEE Access,* vol. 8, p. 159284–159290, 2020.

[18] D. Carrizales-Espinoza, D. D. Sanchez-Gallegos, J. L. Gonzalez-Compean y J. Carretero, «A Federated Content Distribution System to Build Health Data Synchronization Services,» *2021 29th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*. , 2021.

[19] D. D. Sanchez-Gallegos, J. L. Gonzalez-Compean, J. Carretero, . H. Marin-Castro, A. Tchernykh y R. Montella, «PuzzleMesh: A puzzle model to build mesh of agnostic services for edge-fog-cloud.,» *IEEE Transactions on Services Computing*..

[20] J. Gonzalez-Compean, M. Morales-Sandoval, . Rios-Barrientos y L. F. Moguel-Jiménez, «Hermes, a parallel pattern method for building efficient security systems for information sharing. ,» *IEEE Transactions on cloud computing*.

[21] J. A. Barron Lugo, J. L. Gonzalez-Compean, . Carretero, . López-Arévalo y R. Montella., «A novel transversal processing model to build environmental big data services in the cloud. ,» *Environmental Modelling and Software* .

[22] I. Lopez-Arevalo, J. L. Gonzalez-Compean, M. Hinojosa-Tijerina, C. Martinez-Rendon, R. Montella y J. L. Martinez-Rodriguez, « A WoT-Based Method for Creating Digital Sentinel Twins of IoT Devices. Sensors,,» vol. 21(16), nº 5531. , 2021.

[23] J. L. Gonzalez-Compean, V. J. Sosa-Sosa, A. Diaz-Perez, J. Carretero y R. Marcelin-Jimenez, «FedIDS: a federated cloud storage architecture and satellite image delivery service for building dependable geospatial platforms,» *International journal of digital earth,* vol. 11, p. 730–751, 2018.

[24] M. Malik, «Internet of Things (IoT) Healthcare Market by Component (Implantable Sensor Devices, Wearable Sensor Devices, System and Software), Application (Patient Monitoring, Clinical Operation and Workflow Optimization, Clinical Imaging, Fitness and Wellness Measur,» *Allied Market Research,* p. 124, 2016.

[25] C.-P. Deng, T. Wang, T. SH Teo y Q. Song, «Organizational agility through outsourcing: Roles of it alignment, cloud computing and knowledge transfer,» *International Journal of Information Management,* 2021.

[26] H. S. Gunawi, M. Hao, R. O. Suminto, A. Laksono, A. D. Satria, J. Adityatama y K. J. Eliazar, «Why does the cloud stop computing?: Lessons from hundreds of service outages.,» *In Proceedings of the Seventh ACM Symposium on Cloud Computing,*, 2016.