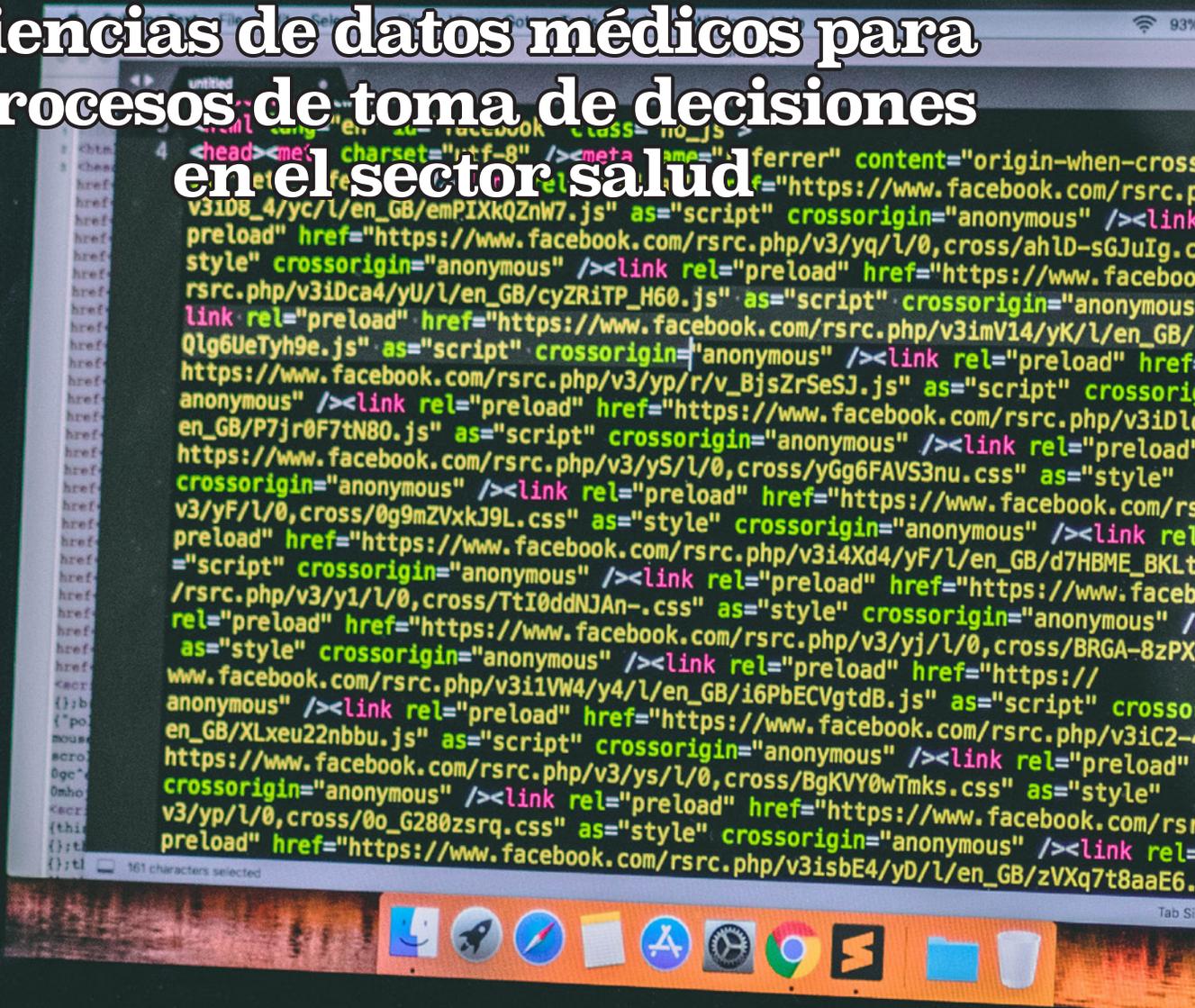


Muyal-Nez, una plataforma de construcción de sistemas de ciencias de datos médicos para procesos de toma de decisiones en el sector salud



Dr. Dante D. Sánchez Gallegos
M. C. Diana Carrizales Espinoza
Dr. José Luis González Compean
Dr. Ricardo Marcelín Jiménez
Universidad Autónoma Metropolitana. Unidad Iztapalapa
Departamento de Ingeniería Eléctrica.

Resumen

Los sistemas de ciencia de datos y big data son piezas tecnológicas complejas que, en los años recientes, se han comenzado a utilizar para analizar enormes cantidades de datos médicos, tales como radiología de diagnóstico o datos de sensores. Lo anterior, se realiza para producir conocimiento e información útil que soporte procesos de toma de decisiones, tales como la prevención, diagnóstico, pronóstico y/o tratamiento de enfermedades como el cáncer. Estas tecnologías disruptivas han comenzado a revolucionar tanto el ámbito médico-científico como las áreas de la gestión de salud pública. Por ejemplo, organizaciones del sector salud mexicano como el Instituto Nacional de Rehabilitación (INR) deben manejar, almacenar, y analizar hasta 45 millones de imágenes médicas (43 TB de datos) para ayudar a los médicos en la toma de decisiones. Sin embargo, construir un sistema de ciencia de datos médicos representa un gran desafío para este tipo de instituciones, no solo por manejar grandes volúmenes de datos o por interconectar y desarrollar complejas aplicaciones de inteligencia artificial, tales como aprendizaje de máquina o minería de datos, sino porque estos sistemas deberían cumplir con estrictas medidas de seguridad para el manejo de datos de los pacientes. En este artículo, se presentan las experiencias y lecciones aprendidas al diseñar, implementar y desplegar sistemas seguros de ciencia de datos médicos en la plataforma Muyal-Nez, que es un conjunto de herramientas para el diseño y despliegue de sistemas de ciencias de datos, a la medida de las necesidades de las instituciones de salud que pudieran requerirlo.

Palabras clave

Ciencia de datos, diagnóstico por imágenes, cómputo en la nube, big data.

Abstract

In recent years, data science systems and big data itself have turned into complex technological assets that have started showing their possibilities to work with massive amounts of medical information, to support decision making processes such as prevention, detection, diagnosis, and prognosis, for instance, during cancer treatments. These groundbreaking technologies have triggered a revolution in scientific environments, but also in public health management. For example, a public health organization such as the Instituto Nacional de Rehabilitación (INR), in Mexico, has a collection of 45 million of medical images amounting to 43 Terabytes (TB) that require storage, transportation, and process, to help physicians during decision making processes. Nevertheless, the construction of data science system represents a huge challenge for this type of institutions as they do not only have to manage large volumes of data, or interconnect and develop complex artificial intelligence applications (such as machine learning or data mining applications), but also, they are compelled to accomplish a very strict set of regulations concerning security issues during the management of personal data. In this paper, we present the experiences and learnings obtained from the design, implementation of Muyal-Nez, which is a set of tools oriented to the design and deployment of data science infrastructure, that fits with the particular requirements of any public health organization on the aforementioned conditions.

Keywords

Data science, diagnostic imaging, cloud computing, big data

Introducción

En México, existe un problema de saturación de los servicios médicos debido a

la alta dispersión y separación geográfica de las fuentes de datos, los SECEs (sistemas de expedientes clínicos electrónicos), los pacientes y los profesionales de la salud. En promedio, los países miembros de la Organización para la Cooperación y el Desarrollo Económico (OCDE) cuentan con 3.5 médicos por cada 1,000 habitantes, mientras que, en México, solo se cuenta con 2.4 médicos por cada 1,000 habitantes. Lo anterior representa un déficit en el número de médicos, lo cual genera una sobrecarga de trabajo para estos (Heinze-Martin, y otros 2018). Además, en el país existe un desbalance del número de médicos que existen por entidad federativa. Por ejemplo, la Ciudad de México cuenta con aproximadamente 5.05 médicos por 1,000 habitantes, mientras que Chiapas cuenta con únicamente 0.35 médicos por cada 1,000 habitantes. Aunado a esto, el número de especialistas en el país es menor a lo requerido. Por ejemplo, en 2017 solamente se contaban con 876 cirujanos oncólogos, 151 ginecólogos oncólogos, 610 oncólogos médicos y 238 radio-oncólogos. Lo anterior, ocasiona que el sistema de salud mexicano se sature, así como que las personas tengan que migrar para recibir atención médica (Heinze-Martin, y otros 2018).

En este sentido, tecnologías disruptivas tales como el internet de las cosas (IoT, por sus siglas en inglés), big data, así como los algoritmos de análisis de datos y aprendizaje de máquina, se han empezado a emplear en el proceso de toma de decisiones en el ámbito médico (Scott Kruse y Beane 2018). Dichos sistemas, mejoran radicalmente el acceso y la eficiencia de los servicios de salud tradicionales, y son utilizados para adquirir datos de los pacientes (por ejemplo, signos vitales o actividad) y ayudar a los profesionales de la salud en la toma de decisiones. Además, tanto profesionales de

la salud como organizaciones médicas pueden obtener hallazgos importantes al procesar los datos adquiridos desde dispositivos de IoT (Rose, Eldridge y Chapin 2015), así como del procesamiento de imágenes médicas, estudios, expedientes clínicos y registros históricos de datos (Lee 2019). Por ejemplo, organizaciones del sector salud mexicano como el Instituto Nacional de Rehabilitación (INR) deben manejar, almacenar, y analizar hasta 45 millones de imágenes médicas (43 TB de datos) para ayudar a los médicos en la toma de decisiones. Para encontrar estos hallazgos en los datos, las organizaciones construyen sistemas de ciencias de datos y big data que realizan el tratamiento de la información a través de una secuencia que normalmente incluye la adquisición de los datos, su preprocesamiento y procesamiento, así como su posterior preservación y visualización. Lo anterior, se realiza para producir información útil que pueda generar conocimiento que ayude a los médicos en procesos de toma de decisiones, tales como prevención, diagnóstico, pronóstico y/o tratamiento de enfermedades como el cáncer.

Durante la adquisición de los datos desde la fuente (e. g., tomógrafos, electrocardiogramas, rayos X, etc.), se emplean herramientas de extracción, transformación y carga de datos, las cuales hacen disponibles los datos en crudo para su posterior preprocesamiento. En el preprocesamiento los datos son preparados para adaptarlos a los algoritmos de minería de datos y big data que son utilizados en la siguiente etapa (procesamiento) (Lee 2019). Para ello, se usan diversas técnicas de preprocesamiento de datos, técnicas tales como la transformación (e.g., transformar una imagen en formato DICOM (Yan 2018) a formato PNG), la limpieza y corrección en caso de errores durante su captura, y la integración, por mencio-

nar algunas. Una vez preparados los datos, estos son procesados con algoritmos y herramientas de big data, minería de datos o aprendizaje de máquina, con el objetivo de transformar los datos de entrada en información que pueda ser empleada por tomadores de decisiones (e.g., médicos) para realizar un diagnóstico o generar nuevo conocimiento (Sanchez-Pinto, Luo y Churpek 2018). En este sentido, los tomadores de decisiones pueden acceder a la información generada mediante técnicas y esquemas de visualización de datos. Además, de forma transversal se ejecutan procesos tales como el manejo de metadatos, la documentación de procesos, y el manejo de la calidad de datos (verificando su veracidad y validez), así como el manejo de respaldo, almacenamiento y la seguridad de estos.

Sin embargo, la construcción de estos sistemas de ciencia de datos no es una tarea sencilla debido a que no solo se requiere la construcción de complejas aplicaciones de inteligencia artificial (IA) y el procesamiento de grandes volúmenes, sino que

además las organizaciones deben manejar y acoplar un conjunto de aplicaciones en cada una de las etapas del ciclo de vida de los datos, y además deben de considerar aplicaciones para cumplir con estrictas medidas de seguridad para el manejo de datos de los pacientes. Además, los participantes en la creación de sistemas de ciencia de datos deben tener conocimiento del dominio con el que están tratando para que sus soluciones sean aplicables para resolver los problemas que se presentan en dicho dominio (e.g., manejo de datos médicos) (Chang y Grady 2019).

En este artículo, se describe la experiencia adquirida durante la construcción de sistemas de ciencias de datos para el procesamiento de datos no estructurados utilizando el servicio Moyal-Nez, nombrado así en honor a Nezahualcōyotl, quien fue arquitecto en la antigua Tenochtitlan (<http://www.adaptivez.org.mx/e-SaludData/nez.html>).

Moyal-Nez se compone de un conjunto de servicios que permiten la construcción de

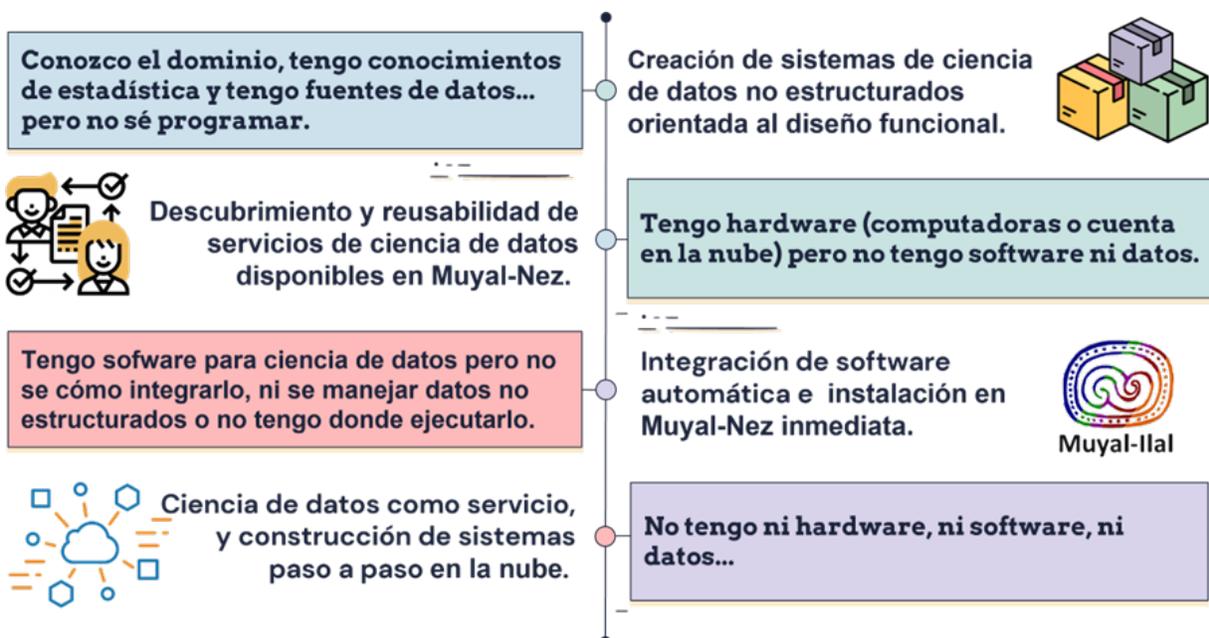


Figura 1. Retos que surgen durante la creación de sistemas de e-salud para ciencias de datos.

sistemas de ciencias de datos para e-salud. Durante la creación de sistemas de ciencia de datos, las organizaciones deben hacer frente a diversos retos, los cuales han motivado el desarrollo de estos servicios. En la Figura 1 muestra una descripción gráfica de dichos retos.

El primer reto que se enfrenta durante la creación de sistemas de ciencia de datos es aquel en el que los miembros de una organización (e.g., médicos y especialistas tales como oncólogos y radiólogos) son expertos en el dominio a tratar, y además cuentan con diferentes fuentes de datos (e.g., datos de mamografías, tomografías, expedientes clínicos, datos de electrocardiogramas); sin embargo, no cuentan con conocimientos de informática y/o programación. Por lo anterior, estos deben realizar una inversión para contratar recursos humanos afines a estas áreas. En este sentido, Muyal-Nez implementa mecanismos orientados al diseño, los cuales reducen la brecha existente entre los especialistas en el dominio y el desarrollo de sistemas de ciencias de datos no estructurados.

El segundo reto por enfrentar se centra en la utilización de los recursos disponibles. Comúnmente, las organizaciones centran sus esfuerzos en adquirir el hardware (computadoras personales, servidores privados, o la nube pública) requerido para la preservación y compartición de sus datos. Sin embargo, en escenarios de ciencia de datos, para poder aprovechar completamente todos los recursos disponibles, es necesario instalar diversos sistemas y servicios específicos que permitan realizar el manejo de los datos, lo cual puede representar un alto costo (tanto en valor monetario, como en tiempo). Tomando en cuenta este reto, Muyal-Nez cuenta con un catálogo de servicios de adquisición, pre-

procesamiento, procesamiento, y visualización de datos que permite a las organizaciones construir diferentes sistemas de ciencias de datos en su infraestructura sin requerir hardware especializado. Lo anterior, es importante para organizaciones que ya cuentan con recursos disponibles para efectuar el procesamiento de datos, ya que no requiere la adquisición de nuevos equipos o el arrendamiento de máquinas virtuales en la nube, lo cual produce costos económicos no contemplados.

El tercer reto, se da cuando las organizaciones cuentan con aplicaciones de ciencias de datos previamente adquiridas o desarrolladas, pero no saben cómo realizar el manejo automático de los datos a través de estas aplicaciones, por lo cual la intervención humana se vuelve necesaria para trasladar los datos. Para hacer frente a este reto, Muyal-Nez cuenta con un mecanismo de encapsulación de aplicaciones en contenedores virtuales (Randal 2020) que permite el acoplamiento de estas con otras aplicaciones de forma automática e inmediata para crear un sistema de ciencia de datos.

Finalmente, las organizaciones se pueden enfrentar a un escenario donde no cuentan con los recursos de hardware y software para construir un sistema de ciencia de datos, además de no tener datos disponibles con los cuales probar ese tipo de sistemas. En este sentido, Muyal-Nez se puede desplegar en un servicio en la nube, desde donde las organizaciones puedan construir sus sistemas de ciencia de datos utilizando el catálogo de aplicaciones y datos disponibles en Muyal-Nez.

Muyal-Nez: diseño intuitivo y despliegue automático de sistemas de e-Salud para ciencia de datos

Muyal-Nez forma parte de una platafor-

ma de servicios eficientes para la gestión, aseguramiento, intercambio, trazabilidad, y almacenamiento de grandes volúmenes de datos médicos. En esta plataforma, las instituciones de salud pueden producir servicios seguros de ciencias de datos, los cuales analizan la información médica (sensores, notas médicas, imágenes, o bases de datos) para apoyar los procesos de toma de decisiones (diagnósticos asistidos y predicciones de riesgo). Además, la plataforma implementa servicios que permiten visualizar información, así como servicios para verificar/asegurar que los SECEs cumplan con los protocolos (DICOM/HL7) y las normas nacionales (NOM-024-SSA3-2012) e internacionales (NIST, ISO 27001:2013 y COBIT 5) para el manejo de datos/contenidos sensibles (por ejemplo, expediente clínico personal, dirección, edad, entre otros).

Los servicios producidos con Muyal-Nez permiten a las organizaciones y la comunidad científica manejar el ciclo de vida de datos médicos de forma segura y eficiente,

lo anterior se realiza sin depender de herramientas producidas por terceros sobre las cuales no se tiene control con respecto al manejo y procesamiento de los datos. Con Muyal-Nez las organizaciones son capaces de crear sistemas de ciencia de datos siguiendo un paradigma guiado por el diseño. Es decir, en caso de que los diseñadores del sistema no cuenten con conocimientos avanzados de programación o informática, la creación de este no se verá afectada debido a que los sistemas en Muyal-Nez son diseñados mediante interfaces gráficas intuitivas. Estas interfaces están basadas en un carrito de compras, donde los diseñadores eligen las aplicaciones que procesarán sus datos (requerimientos funcionales), las características que desean agregar a sus datos (requerimientos no funcionales tales como seguridad, trazabilidad, integridad o confidencialidad) y la fuente de sus datos. Finalmente, el sistema será generado de acuerdo con las especificaciones de las aplicaciones, características y fuentes previamente seleccionadas (ver Figura 2).

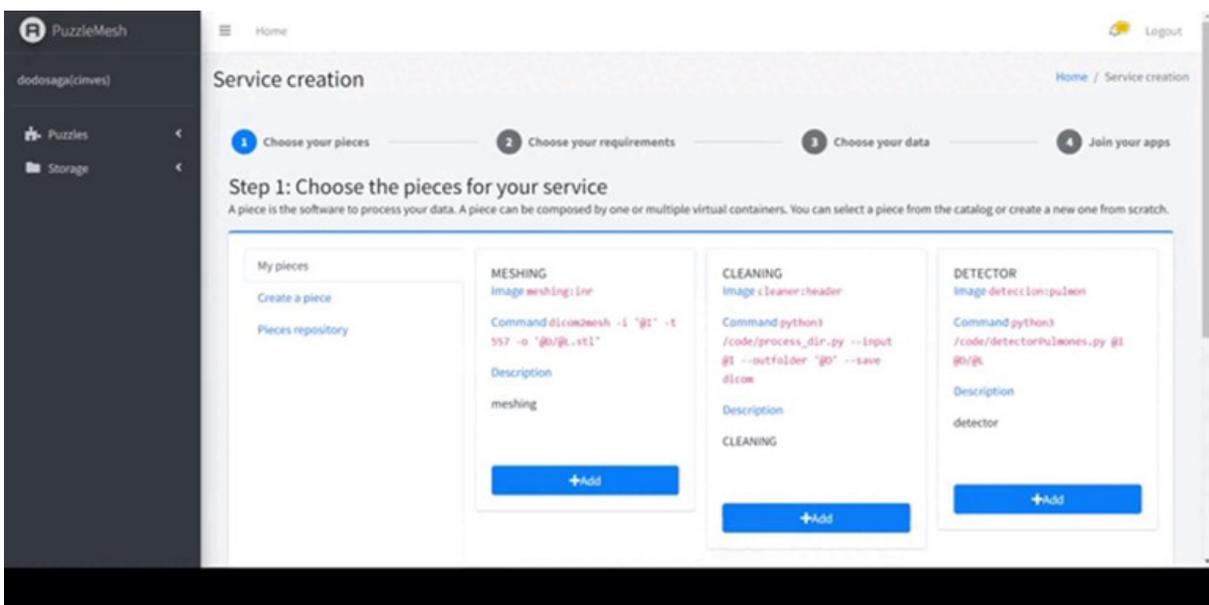


Figura 2. Interfaz gráfica de Muyal-Nez.

Los sistemas de e-salud para ciencia de datos construidos con Muyal-Nez cuentan con las siguientes características: i) modularidad, ii) software propio y autónomo, iii) eficiencia, iv) portabilidad, y v) reusabilidad.

La modularidad se refiere a la capacidad de los sistemas de ser divididos en componentes más pequeños, los cuales puedan ser reemplazados y/o modificados de forma individual sin afectar al resto de los componentes. En Muyal-Nez, un sistema de e-salud es creado mediante la unión de múltiples componentes de software llamados bloques de construcción, los cuales son creados utilizando la tecnología de contenedores virtuales. En este sentido, un contenedor virtual se puede definir como una máquina virtual ligera, la cual encapsula una aplicación con todas sus dependencias. En Muyal-Nez, cada bloque de construcción contiene una aplicación que se emplea para procesar los datos, las dependencias de software (librerías, variables de entorno, o requerimientos), y un conjunto de estructuras para el manejo de datos de entrada y salida. Estas estructuras permiten la interconexión de cada bloque con otros para producir un sistema de ciencias de datos para e-salud.

La construcción de sistemas basada en bloques de construcción de Muyal-Nez, permite que las organizaciones, diseñadores, y desarrolladores reutilicen los bloques previamente construidos. Lo anterior resulta ser de gran importancia para las organizaciones, debido a que estas pueden extender el tiempo de vida del software con el que cuentan sin la necesidad de realizar una inversión económica para la adquisición de nuevo software. Los sistemas creados con Muyal-Nez son generados utilizando software propio, autónomo y portable que no se encuentra

ligado a aplicaciones/sistemas de terceros, y/o a una infraestructura/plataforma en específico (por ejemplo, a un proveedor en la nube como Amazon Web Services). En este sentido, las organizaciones no necesitan modificar el código de las aplicaciones para desplegarlas en diferentes equipos de cómputo (por ejemplo, en la nube, computadoras personales, servidores de alto rendimiento, etc.) con diferentes sistemas operativos (Windows o sistemas Linux). Lo anterior, se debe a que la construcción de los sistemas está basada en la tecnología de contenedores virtuales que permite empaquetar aplicaciones junto con sus dependencias (por ejemplo, librerías como TensorFlow), y lenguajes de programación (por ejemplo, C/C++, Java, o Python).

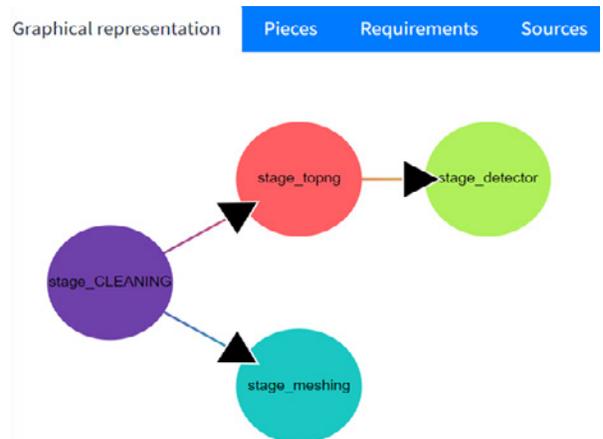


Figura 3. Ejemplo de un servicio de e-salud construido con Muyal-Nez representado como un grafo.

La construcción de un sistema de e-salud en Muyal-Nez se realiza mediante el encadenamiento de aplicaciones en la forma de un grafo. En la Figura 3 se muestra un ejemplo de un sistema de e-salud construido como un grafo utilizando Muyal-Nez, en donde diferentes aplicaciones (nodos) son conectadas con otras aplicaciones para procesar los datos a través de

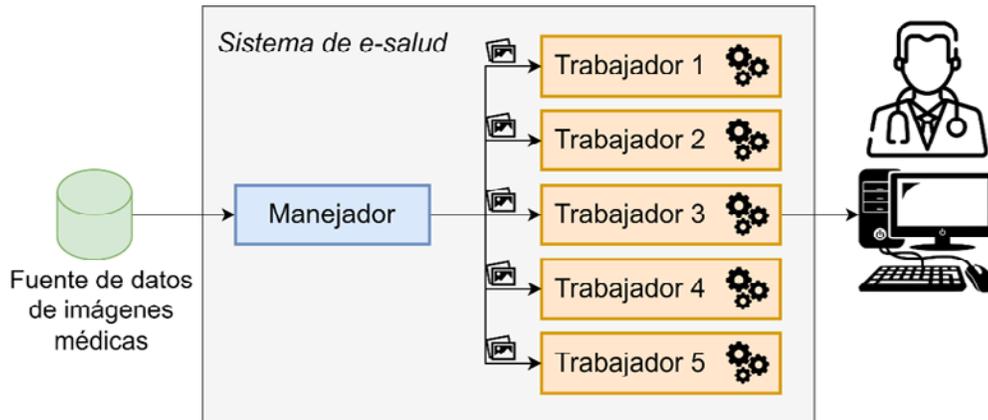


Figura 4. Ejemplo de un patrón de paralelismo construido con Muyal-Nez.

este grafo. Este grafo fue construido para el manejo y procesamiento de imágenes médicas en formato DICOM (Yan 2018), en el cual se consideran cuatro nodos. El primer nodo anonimiza las imágenes y las entrega a dos nodos para la transformación de las imágenes DICOM a PNG, así como la generación de una representación 3D a partir de las imágenes. Finalmente, las imágenes en formato PNG son entregadas por un nodo para la detección de tumores, el cual automáticamente etiqueta la zona de la imagen identificada como un tumor.

En escenarios de tomas de decisiones asistidos por sistemas de ciencias de datos, es crucial que los tomadores de decisiones (médicos y especialistas) reciban, en el menor tiempo posible, los resultados del procesamiento de los datos que utilizarán para tomar una decisión (dar un diagnóstico). Para ello, Muyal-Nez implementa estructuras paralelas que permiten reducir el tiempo de procesamiento en sistemas de e-salud para ciencia de datos. En la Figura 4, se muestra un ejemplo de un sistema paralelo construido con Muyal-Nez. En este, una aplicación es clonada en instancias conocidas como trabajadores que reciben los datos a procesar desde una

instancia conocida como manejador. Los trabajadores procesan los datos en paralelo, lo cual reduce el tiempo requerido para obtener resultados que puedan ser utilizados por los tomadores de decisiones (Ortega-Arjona 2010).

Sistemas de e-salud intra/inter-institucionales

Los servicios de e-Salud creados con Muyal-Nez pueden ser manejados internamente por una organización (servicio de e-Salud intra-institucional), o por múltiples organizaciones (servicio de e-Salud inter-institucional).

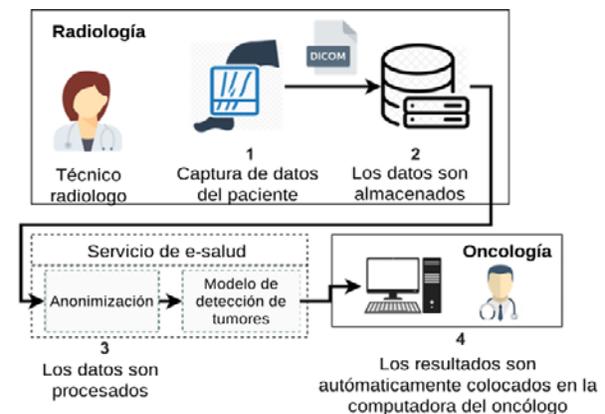


Figura 5. Ejemplo de un sistema de e-salud intra-institucional.

Los sistemas de e-salud intra-institucio-

nales, son aquellos que permiten a las instituciones y organizaciones de salud procesar e intercambiar datos entre profesionales de la salud y departamentos dentro una organización u hospital. En la Figura 5, se muestra un ejemplo de un sistema de e-salud intra-institucional. En este ejemplo, un técnico radiólogo realiza la captura de tomografías del paciente en el Instituto Nacional de Rehabilitación (INR). Posteriormente, las imágenes son almacenadas y el radiólogo las comparte con un oncólogo del instituto para que este las valore y emita su diagnóstico. En este sentido, antes de compartir las imágenes con el oncólogo, estas pueden procesarse mediante un servicio automático de e-salud construido para anonimizar las imágenes y etiquetarlas utilizando un modelo de detección de tumores. De esta manera, los resultados son entregados al oncólogo de forma automática en su computadora.

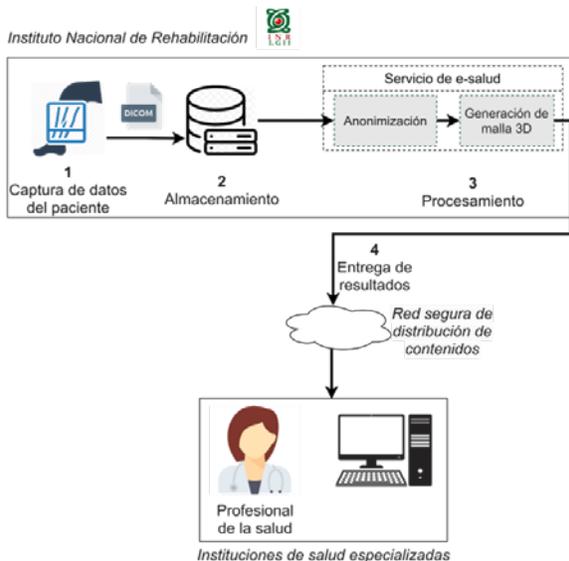


Figura 6. Ejemplo de un sistema de e-salud inter-institucional.

Los sistemas de e-salud inter-institucional, por otra parte, son creados para compartir datos entre diferentes institu-

ciones, organizaciones, u hospitales. En este tipo de sistemas, los datos son distribuidos empleando una red segura de distribución de contenidos, la cual distribuye automáticamente los datos a los participantes de las organizaciones que tengan autorización para acceder a ellos.

Por ejemplo, en la Figura 6 se muestra una representación conceptual de un sistema de e-salud inter-institucional generado para intercambiar datos entre el Instituto Nacional de Rehabilitación (INR) y otra 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 permite anonimizar los datos y realizar 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 colocados de forma automática en las computadoras de los profesionales de salud autorizados para consultarlos.

Estudio de caso: sistema de diagnóstico de cáncer de hueso largo asistido por inteligencia artificial

Actualmente, los sistemas de diagnóstico asistido por computadora son cruciales en los procesos de toma de decisiones en hospitales. Estos sistemas ayudan a los médicos a tener una interpretación de los datos que reciben, por ejemplo, etiquetando zonas donde podría estar presente un tumor en una imagen médica. Estos sistemas, comúnmente se apoyan en técnicas de aprendizaje supervisado, en donde, de forma similar a como aprenden los humanos, un especialista le enseña a un robot mediante el uso de ejemplos. Como resultado, el algoritmo aprende a identi-

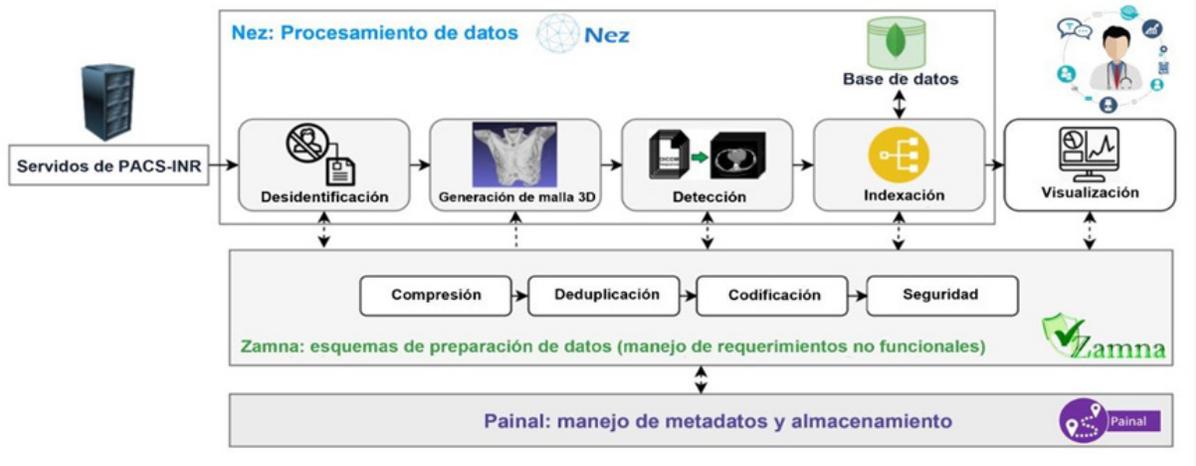


Figura 7. Estudio de caso basado en el manejo y procesamiento de imágenes médicas para el diagnóstico de cáncer.

ficar objetos en imágenes, distinguiendo así los patrones que componen el objeto en estudio. Este aprendizaje se realiza en dos etapas: i) entrenamiento y ii) prueba o implementación. En el entrenamiento, se genera un modelo a partir de las muestras previamente etiquetadas por un especialista. Por ejemplo, tomografías de pulmón en donde se etiqueten zonas con nódulos. Posteriormente, en la etapa de prueba se utilizan muestras sin etiquetar para que el robot etiquete las zonas que, por ejemplo, tienen una alta probabilidad de ser nódulos en una tomografía.

En este sentido, utilizando Moyal-Nez desarrollamos un servicio de ciencia de datos para el diagnóstico asistido por computadora. En la Figura 7 se muestra la representación conceptual de dicho sistema, el cual se encuentra conectado a un sistema PACS (sistema de archivo y comunicación de imágenes) en el Instituto Nacional de Rehabilitación Luis Guillermo Ibarra Ibarra (INR-LGII). En 2020, el INR-LGII manejaba hasta 45 millones de imágenes médicas (tomografías, radiografías, resonancias magnéticas, entre otros), las cuales tienen que ser procesadas y trans-

portadas para ayudar a los médicos en la toma de decisiones.

Para automatizar el procesamiento de las imágenes médicas, el sistema de ciencia de datos construido con Moyal-Nez incluye las siguientes etapas de procesamiento:

1. Anonimización: elimina los datos personales de las imágenes médicas en formato DICOM. Lo anterior, permite eliminar cualquier dato personal que las imágenes puedan contener.
2. Generación de malla 3D: genera una representación 3D de un estudio a partir de un conjunto de imágenes DICOM.
3. Detección de nódulos: es un sistema entrenado para identificar nódulos en huesos y pulmones. Este sistema fue implementado utilizando la red neuronal Faster R-CNN (Ren, y otros 2017).
4. Indexación: toma los metadatos de las imágenes, así como los resultados de las etapas anteriores, y los registra en una base de datos. El objetivo de esta etapa es que los datos puedan ser consultados

por los usuarios finales (médicos o enfermeras), o por otras aplicaciones.

5. Visualización: permite la exploración de los datos mediante un sistema de visualización.

Además, para asegurar que el almacenamiento y transporte de datos médicos cumpla con las normas nacionales e internacionales (norma oficial mexicana NOM-O24-SSA3-2010, ISO 270001-13, NIST y COBIT 5), Muyal-Nez se encuentra conectado con dos servicios para el manejo seguro de los datos y metadatos llamados Muyal-Zamna y Muyal-Painal. Muyal-Zamna es un servicio que se encarga de agregar características no-funcionales como confidencialidad, verificación de integridad, control de acceso, y disponibilidad. Mientras que Muyal-Painal se compone de un conjunto de servicios que permite el transporte de datos entre diferentes computadoras. Además, Muyal-Painal se encarga del almacenamiento seguro de los datos en infraestructuras tales como la nube pública o la nube privada.

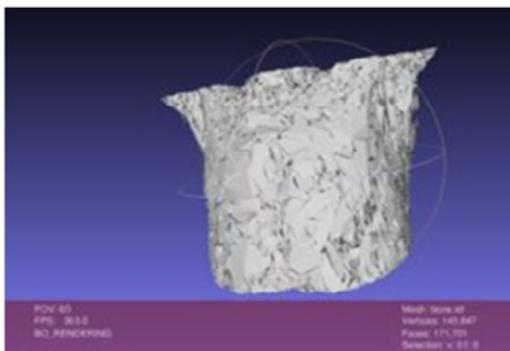
Resultados

En el ámbito médico, es crucial que el procesamiento de datos se ejecute en el menor

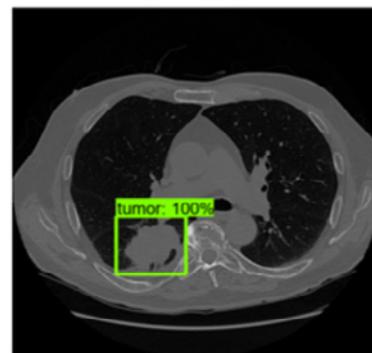
tiempo posible (eficiencia) con el objetivo de que los médicos tengan disponibles los resultados del procesamiento (información) para apoyarlos durante el proceso de toma de decisiones (diagnóstico). En este sentido, Muyal-Nez permite el despliegue inmediato de sistemas de ciencias de datos para el manejo de datos médicos. Por ejemplo, el sistema mostrado en la Figura 7 es desplegado en tan solo 6 minutos. Este despliegue es automático y el usuario no requiere tener conocimientos avanzados de programación, ingeniería computacional, y/o arquitectura de sistemas.

En la Figura 8, se muestran ejemplos de los resultados obtenidos del procesamiento de imágenes médicas con la estructura mostrada en la Figura 7. En la Figura 8a se muestra el resultado de la reconstrucción 3D de imágenes de tomografía, mientras que la Figura 8b muestra una imagen etiquetada, en donde la etiqueta indica el porcentaje de probabilidad de que esa zona sea un tumor.

Otra de las ventajas de Muyal-Nez, con respecto a otras soluciones disponibles en estado del arte, es su eficiencia durante el procesamiento de datos producida por los esquemas de paralelismo implementados,



a) Reconstrucción 3D creado a partir de un conjunto de imágenes médicas.



b) Imagen etiquetada.

Figura 8. Resultados obtenidos del procesamiento de imágenes médicas.

lo cual reduce el tiempo requerido por un médico para obtener resultados que lo ayuden a tomar una decisión. Por ejemplo, en la Figura 9 se observa una comparación del tiempo de procesamiento de Muyal-Nez contra una solución tradicional que no implementa paralelismo, y que es como comúnmente se realiza el diseño y despliegue de los sistemas de e-salud. En este sentido, Muyal-Nez procesa 25 estudios (37 GB, 70270 archivos DICOM) en 1.1 horas. En contraste, el sistema implementado como una solución tradicional sin paralelismo procesa los mismos estudios en 5.8 horas.

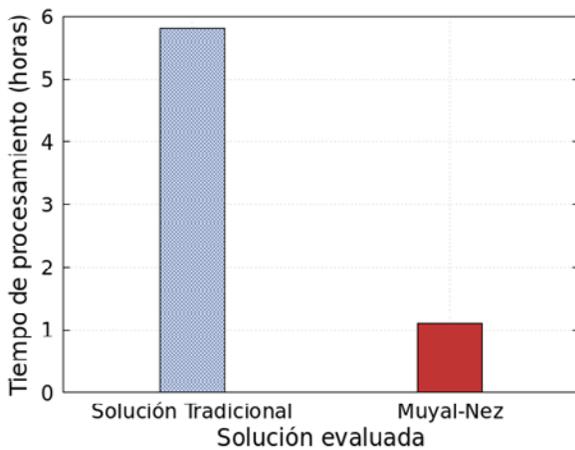


Figura 9. Comparación del tiempo de procesamiento de Muyal-Nez y una solución tradicional que no implementa paralelismo.



Figura 10. Porcentaje de cumplimiento de normas de un sistema para el diagnóstico de cáncer que incluye componentes de seguridad y confiabilidad.

A diferencia de otras soluciones del estado del arte, las cuales solo contemplan el procesamiento de datos, Muyal-Nez se encuentra interconectado con un sistema de manejo de requerimientos de seguridad y confiabilidad llamado Muyal-Zamna. La seguridad (incluyendo confidencialidad, integridad, control de acceso y trazabilidad) es importante en ambientes donde los datos serán almacenados en infraestructuras administradas por terceros, por ejemplo, la nube pública.

De igual manera, la confiabilidad asegura que los datos se encuentren disponibles a pesar de fallas en la infraestructura de almacenamiento. En este sentido, en la Figura 10 se muestra el porcentaje de cumplimiento de diferentes normas nacionales e internacionales (norma oficial mexicana NOM-O24-SSA3-2010, ISO 270001-13, NIST y COBIT 5) del sistema mostrado en la Figura 7, el cual incluye el manejo de seguridad y confiabilidad de los datos de forma automática. Como se puede observar, el sistema cumple con el 100% de los requisitos incluidos en la norma COBIT 5, mientras que para la norma oficial mexicana NOM-O24-SSA3-2010 se cumple con el 88.9% de los requisitos. Es importante notar que, en esta evaluación, se incluyen requisitos que deben de ser manejados de manera manual por los administradores de la red y la infraestructura de una organización u hospital (es decir, procesos que deben ser realizados por humanos), por ejemplo, la generación de roles de acceso a los datos y seguridad en las redes de una organización.

Conclusiones

La ciencia de datos, en el ámbito médico, es crucial para ayudar a los médicos a que el proceso de toma de decisiones (diagnóstico) se realice en el menor tiempo posible.

En este contexto, hemos presentado el diseño y desarrollo de Muyal-Nez, el cual permite la construcción automática de sistemas de ciencia de datos para el procesamiento y manejo eficiente de la información médica. La construcción de sistemas con Muyal-Nez es sencilla debido a que los diseñadores no requieren de conocimientos avanzados de programación y/o ingeniería computacional.

Los sistemas de ciencia de datos construidos con Muyal-Nez, pueden ser tanto intra-institucionales (para conectar diferentes departamentos dentro de un hospital) o inter-institucionales (para conectar diferentes hospitales).

Además, Muyal-Nez permite a las organizaciones construir sistemas de ciencias de datos seguros y confiables, lo cual es crucial cuando estos sistemas son desplegados en ambientes como la nube, o en escenarios donde se utiliza la nube para el intercambio y almacenamiento de datos. Estos componentes de seguridad y confiabilidad de Muyal-Nez, permiten que las organizaciones cumplan, de forma automática, con un porcentaje alto de los requisitos establecidos en diferentes normas tanto nacionales, como internacionales para asegurar el intercambio y preservación de datos sensibles. Los requisitos que se cubren con Muyal-Nez, tienen que ver con aquellos que pueden ser cumplimentados sin la necesidad de intervención humana, es decir, solo aquellos procesos que pueden ser realizados por una aplicación y/o servicio como lo es la verificación de integridad o el control de acceso.

Referencias

Chang, W. y Nancy, G. *NIST Big Data Interoperability Framework: Volume 1, Definitions*. NIST Special Publication

1500-1r2. National Institute of Standards and Technology, Gaithersburg, MD, 2019, pp.1-53. Disponible en: <https://www.nist.gov/publications/nist-big-data-interoperability-framework-volume-1-definitions>. Consultado el 11 de julio de 2022.

Heinze, G., Canchola, V. H. O., Miranda, G. B., Fuentes, N. A. B., & Sánchez, D. P. G. Los médicos especialistas en México, *Gaceta médica de México*, 154[3], pp.342-351, 2018. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=7197316>. Consultado el 11 de julio de 2022.

Kruse, C. S., y Beane, A., Health Information Technology Continues to Show Positive Effect, *Journal of medical Internet research*, 20[2], pp.1-9, 2018. Disponible en: <https://www.jmir.org/2018/2/e41/>. Consultado el 11 de julio de 2022.

Lee, R. (Ed.). *Big Data, Cloud Computing, and Data Science Engineering Vol. 844*, Springer, Berlin/Heidelberg, Germany, 2019, pp.51-68. Disponible en: <https://doi.org/10.1007/978-3-030-24405-7>. Consultado el 11 de julio de 2022.

Ortega-Arjona, J. L., *Patterns for Parallel Software Design*, Wiley Publishing, Chichester, United Kingdom, 2010, pp.1-411. Randal, A., The ideal versus the real: Revisiting the history of virtual machines and containers, *ACM Computing Surveys (CSUR)*, 53[1], pp.1-31, 2020. Disponible en: <https://doi.org/10.1145/3365199>. Consultado el 11 de julio de 2022.

Ren, S., He, K., Girshick, R., y Sun, J., Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28, pp.1137-1149, 2015. Disponible en: <https://proceedings>

neurips.cc/paper/2015/file/14bfa6b-b14875e45bba028a21ed38046-Paper.pdf. Consultado el 11 de julio de 2022.

Rose, K., Eldridge, S., y Chapin, L., *The internet of things: An overview, The internet society (ISOC)*, 80, pp.1-50, 2015.

Sanchez-Pinto, L. N., Luo, Y., y Churpek, M. M., Big Data and Data Science

in Critical Care, *Chest*, 154[5], pp.1239-1248, 2018. Disponible en: <https://doi.org/10.1016/j.chest.2018.04.037>. Consultado el 11 de julio de 2022.

Yan, L., DICOM Standard and Its Application in PACS System, *Medical Imaging Process & Technology*, 1[1], pp.1-8, 2018. Disponible en: <http://dx.doi.org/10.24294/mipt.v1i1.221>. Consultado el 11 de julio de 2022.