

Big Data Architectures for the Climate Change Analysis: A Systematic Mapping Study

A. Cravero, S. Sepúlveda and L. Muñoz

Abstract— Despite the volume of data generated, scientists cannot accurately predict how climate change will manifest itself locally and what measures should be applied to mitigate it effectively. On the other hand, Big Data is a new technology that faces the challenge of collecting, characterizing and analyzing a large amount of data, taking into account data from multiple sources, multiple variables and multiple scales with different spatial and temporal attributes. To do this, we review and synthesize the current state of research of Big Data architectures that help solve the problems caused by climate change in health (16%), agriculture(8%), biodiversity(16%), energy(8%), water resources(4%) and clima(48%). To achieve the objective, we have carried out a systematic mapping study, which includes four research questions, including 25 studies, published from 2013 to 2019. The architectures found have been classified according to their use, which can be for statistical analysis, monitoring and simulations; helping researchers to integrate knowledge into the practical use of Big Data in the context of climate change.

Index Terms—Big Data, Climate Change, Architectures, systematic mapping.

I. INTRODUCCIÓN

El desarrollo de los sistemas de observación del clima, en particular la observación meteorológica satelital y los modelos climáticos de alta resolución, ha originado el rápido crecimiento en el volumen de datos climáticos. En busca de nuevas ideas y métodos para el desarrollo continuo de la predicción del clima, la aplicación de Big Data ha proporcionado una solución real al procesamiento del volumen de datos almacenado [1]. Por otro lado, la rápida integración, el almacenamiento en la nube, la computación en la nube y el análisis de muestras completas de datos climáticos masivos ha hecho posible comprender los estados climáticos y su evolución de manera más objetiva [2].

Analizar los impactos que ocasiona el cambio climático varía según el punto de vista como, el aumento del nivel del mar [3], las composiciones atmosféricas, y oceánicas [4], la aparición de nuevas enfermedades [5], entre otros. Según Scholes (2016), estos impactos cubren tres áreas amplias: los efectos directos

sobre los seres humanos, sus empresas y sus activos; efectos sobre los sistemas naturales; y efectos sobre los humanos a través de sistemas naturales [6]. Sin embargo, hasta ahora, los científicos no pueden predecir con total precisión cómo se manifestará el cambio climático a nivel local y qué medidas se deben aplicar para mitigarlo con efectividad en las condiciones locales. Puede que la situación ya no sea de profunda incertidumbre, pero sí de profunda complejidad [7].

La investigación aplicada a Big Data para el análisis del Cambio Climático, combina muchas áreas de investigación y disciplinas, por lo que requiere no solo una amplia gama de datos obtenidos a través del monitoreo y análisis dinámico en tiempo real, sino también de información sobre las relaciones y leyes que conectan la gran cantidad de datos de observación de la Tierra [8].

Este artículo hace referencia a las "Arquitecturas de Big Data" (ABD) para abarcar tanto los conjuntos de datos como el conjunto de componentes y tecnologías relacionadas, que son utilizadas para el análisis del Cambio Climático hoy. Debido al poco conocimiento sobre los beneficios de aplicar una ABD en dominios ambientales [9] [10], algunos autores han generado preguntas que pueden ayudar a identificar los temas de investigación. Algunas de ejemplo son: ¿la naturaleza necesita del apoyo de Big Data? ¿Puede Big Data ayudar a comprender el medio ambiente?, ¿cuáles son las implicaciones de Big Data para ayudar con la sostenibilidad ambiental? [2].

Este artículo introduce sobre los tipos de problemas de Cambio Climático que se resuelven a través de una ABD, describiendo los beneficios obtenidos al aplicar esta nueva tecnología, identificando las fuentes de datos utilizadas. Nos centramos en el diseño de las arquitecturas y las tecnologías utilizadas, con el fin de orientar a los futuros investigadores en la aplicación de Big Data. Esto se logra a través de un Mapeo Sistemático de Estudios que identifica 24 trabajos que se utilizan para solucionar problemas ocasionados por el cambio climático.

El documento se estructura de la siguiente manera. La sección II presenta los conceptos básicos utilizados en el trabajo. La sección III describe el planteamiento del problema

Fecha envío: 06/01/2020

Este trabajo ha sido financiado por la Universidad de La Frontera a través del Proyecto DIUFRO DI-200099.

A. Cravero, es Directora del Departamento de Cs. de la Computación e Informática, Universidad de La Frontera, Temuco, Chile.

(email: ania.cravero@ufrontera.cl)

S. Sepúlveda, es Académico del Departamento de Cs. de la Computación e Informática, Universidad de La Frontera, Temuco, Chile.

(email: samuel.sepulveda@ufrontera.cl)

L. Muñoz, es Académica del Departamento de Sistemas de Información, Control y Evaluación de Recursos Informáticos, Universidad Tecnológica de Panamá, Panamá.

(email: lilia.munoz@utp.ac.pa)

y las actividades desarrolladas en el mapeo sistemático. La sección IV describe los resultados y análisis comparativo de los estudios seleccionados. La sección V presenta los Trabajos Relacionados. Finalmente, la sección VI concluye el estudio.

II. CONCEPTOS BÁSICOS

A. Big Data

Existen distintas definiciones aportadas por los estudios del tema de Big Data. Entre las cuales se tiene, una nueva generación de tecnologías y arquitecturas diseñadas para extraer económicamente el valor de volúmenes muy grandes de una amplia variedad de datos al permitir la captura de alta velocidad, el descubrimiento y/o análisis [11]. Una definición incluye las propiedades conocidas como las 5V: Volumen, Velocidad, Variedad, Valor y Veracidad [12]. Explican que Big Data no es sólo una base de datos o un problema Hadoop, sino que es todo el conjunto de componentes, o arquitectura, para almacenar, procesar, visualizar y entregar resultados para orientar las aplicaciones [13]. Según los mismos autores, este conjunto de componentes interrelacionados se puede definir como el Ecosistema Big Data que se ocupa de la evolución de los datos, los modelos e infraestructura de apoyo durante todo el ciclo de vida de Big Data [14]. Por otro lado definen un Marco de Arquitectura de Big Data, que incluye 5 componentes que abordan distintos aspectos del ecosistema: (1) modelo de datos, estructuras y tipos; (2) administración de Big Data;

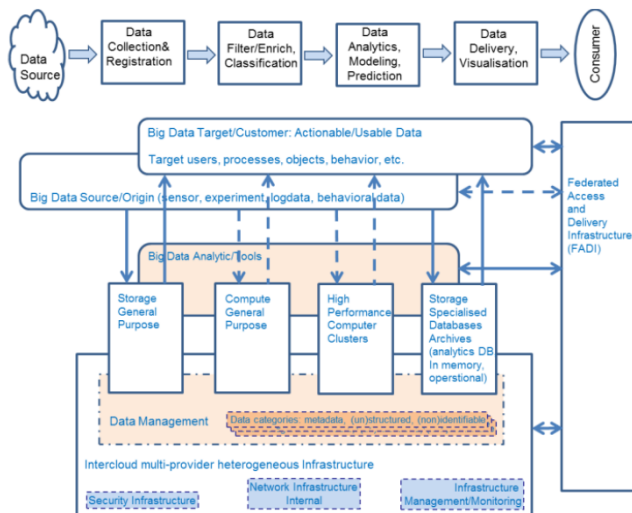


Fig. 1. Arquitectura general de Big Data [14].

(3) herramientas de análisis; (4) Infraestructura; y (5) seguridad de Big Data. La Fig. 1 muestra la relación entre los componentes.

Otras definiciones de Big Data se pueden encontrar en [15][16].

De acuerdo a la Fig. 1, el proceso de Big Data se inicia con la identificación de las fuentes desde donde se extraen los datos útiles [17]. A continuación, los datos son almacenados en alguno de los modelos de datos diseñados, de acuerdo a si son

datos estructurados o no. En el siguiente paso, los datos se clasifican y se filtran de acuerdo al tipo de análisis que se desee realizar. Los datos clasificados son analizados utilizando herramientas adecuadas para ello; ejemplo de ello, son las de minería de datos [18], las de OLAP [19], y las de ciencias de datos en general [20]. Los datos obtenidos deben ser presentados con alguna herramienta de visualización.

Finalmente los datos son analizados por los responsables de tomar decisiones [14].

B. Cambio Climático

Cambio Climático [21], alude a una variación del clima del planeta Tierra generada por la acción del ser humano. Es el efecto invernadero que provoca el llamado calentamiento global, que es el aumento paulatino y constante de la temperatura de la superficie terrestre, producido por la utilización de los combustibles fósiles en la combustión de gases por miles de años [22].

Por otro lado, el IPCC del 2013, explica que el Cambio Climático es el cambio ambiental global acentuado en los últimos años por la acción del hombre, el aumento de la población humana, además del uso inadecuado del suelo, las alteraciones en los ciclos del agua, alteraciones en la biodiversidad [23].

Según los autores en [24], el Cambio Climático se ha convertido en el principal problema de la humanidad, debido a sus consecuencias devastadoras. Debido a ello, es necesario encontrar soluciones a través de nuevas Tecnologías de Información que permitan analizar un gran volumen de datos, complejo y variable en el tiempo [25]. En este sentido, Big Data promete ser un gran aporte [26].

C. Mapeo Sistemático de Estudios

La técnica de mapeo sistemático (o systematic mapping) define un proceso y una estructura de informe que permite categorizar los resultados que han sido publicados hasta el momento en un área determinada [27].

El objetivo de un mapeo sistemático está en la clasificación, y está por tanto dirigido al análisis temático y a la identificación de los principales foros de publicación [28]. Permite responder preguntas genéricas como ¿Qué es lo que se ha hecho hasta el momento en el campo X? Como limitación, este tipo de estudios no toma en consideración la calidad de los estudios incluidos [29] [30].

El proceso de mapeo sistemático consiste en las siguientes etapas: (1) definición de las preguntas de investigación, (2) ámbito de la revisión, (3) ejecución de la búsqueda, (4) selección de estudios, (5) filtrado de estudios, (6) esquema de clasificación, (7) extracción de datos, (8) y mapa sistemático [31].

III. PLANTEAMIENTO DEL PROBLEMA Y ACTIVIDADES DESARROLLADAS EN EL MAPEO SISTEMÁTICO

A continuación se describe las actividades principales desarrolladas en el mapeo sistemático.

A. Definición de las Preguntas de Investigación

Las siguientes Preguntas de Investigación (PI) se definieron de acuerdo a la técnica indicada por [31]:

PI1: ¿Cuál es el tipo de problema provocado por el Cambio Climático que se intenta resolver a través de Big Data?

PI2: ¿Para qué se utilizan las arquitecturas de Big Data encontradas?

PI3: ¿Cuáles son las tecnologías utilizadas en las arquitecturas encontradas?

PI4: ¿Qué resultados reportan los estudios seleccionados?

Las PI nacen a partir de algunas hipótesis formuladas como: (1) Big Data provee las herramientas necesarias para disponer de un sistema de monitoreo y simulación en el contexto del Cambio Climático; (2) Big Data permite enfrentar los desafíos del Cambio Climático en distintas situaciones problemáticas como Salud, Energía, Agricultura, y la Biodiversidad.

Estas hipótesis permitirán a los futuros investigadores decidir si es conveniente utilizar sistemas de Big Data en otras áreas problemáticas que se enfrentan debido al Cambio Climático.

Por otro lado, los mismos podrán identificar las herramientas y arquitecturas adecuadas para cada caso.

B. Ejecución de la Búsqueda

Las palabras claves utilizadas para la búsqueda fueron "Big Data", Climate y "Climate Change" según corresponda. Como requisito de búsqueda es que el concepto Climate o "Climate Change" debe estar presente en el título y "Big Data" en las palabras claves definidas por el autor. Las fuentes donde se aplicó la búsqueda son: Google Scholar, Scopus, IEEE Digital Library, ACM Digital Library, Elsevier y Springer.

Para el caso de Scopus, la cadena de búsqueda fue TITLE ("CLIMATE") AND INDEXTERMS ("BIG DATA"), que entregó como resultado 92 artículos a revisar. Por otro lado, en IEE Digital Library se utilizó la cadena ("Document Title":"Climate") AND "Index Terms":"Big Data"), entregando como resultado 17 trabajos a revisar. Búsquedas similares se llevaron a cabo en las otras fuentes.

C. Selección de Estudios y Proceso de Filtrado

La selección de los estudios se ha formulado basada en los siguientes criterios de inclusión/exclusión:

1) Inclusión

Trabajos de revistas o congresos escritos en inglés. Por otro lado se incluyen trabajos que utilizan arquitecturas de Big Data. El objetivo del trabajo debe estar relacionado con la mejora de los problemas debido al Cambio Climático.

2) Exclusión

Documentos, y trabajos que sean reportes técnicos o estudios del estado del arte. Por otro lado, no se permiten trabajos incompletos o resúmenes.

El proceso de selección consta de dos iteraciones realizadas por dos colaboradores. En la primera iteración, cada revisor aplicó los criterios de inclusión y exclusión para el título, resumen y palabras clave para los estudios obtenidos en los motores de búsqueda (ver fila 1 en la Tabla I). Una vez obtenida la lista de estudios, los revisores compararon los resultados con

el fin de asegurar que ambos aplicaron dichos criterios con rigurosidad (ver resultados en fila 2). Para el caso de los trabajos que generaba dudas, los revisores debieron leer la introducción y conclusiones para luego debatir su pertinencia para el trabajo.

De esta manera se obtuvo un total de 25 estudios (fila 3), de los cuales solo 14 de ellos describe los componentes de la

TABLA I
RESULTADOS DE LA BÚSQUEDA Y FILTRADO

Motor de Búsqueda	IEEE	ACM	Springer	Elsevier	Scopus	Scholar Google
Resultados de la búsqueda	17	45	198	26	92	44
Candidatos	12	7	15	9	18	26
Trabajos seleccionados	4	1	5	5	8	25

arquitectura de Big Data utilizada. En este caso, 8 de los estudios se encuentran en Scopus. De la lista obtenida, se determinó que 3 trabajos pertenecen a IEEE, 5 a Elsevier, 1 a ACM, y 5 a Springer.

La Tabla II presenta la lista de los estudios seleccionados de acuerdo a los criterios de inclusión y exclusión. Estos son los estudios relacionados con Arquitecturas de Big Data y el Cambio Climático. Por otro lado, la Tabla III lista artículos que solo describen el uso de Big Data para dicha solución.

Hay varios estudios relevantes que no cumplen con los criterios de inclusión y exclusión, pero que sí aportan en otros aspectos al cambio climático. Entre ellos están: [32], [33], [34], [35], [36], [37], [38], [39], [40].

Por otro lado, se encontraron estudios de estado del arte, que consideran características importantes de la implantación de Tecnología de Big Data en el ámbito del Cambio Climático. Estos trabajos no incluyen un análisis de las arquitecturas provistas. Entre los más destacados están: [41], [42], [43], [44].

D. Esquema de Clasificación

El esquema de clasificación se crea a partir de las preguntas de investigación y los conceptos claves que se obtienen de las mismas. Los conceptos han sido clasificados como sigue:

1) Tipo Problema

El tipo de problema se refiere al área en donde se aplica el Sistema de Big Data para el análisis del Cambio Climático. Se encontraron trabajos que describen problemas en: (i) *Salud*: cuando se estudia o evalúa el impacto del Cambio Climático en la incidencia de enfermedades, (ii) *Biodiversidad*: cuando se estudia o evalúa el impacto en la diversidad de especies vegetales y animales que viven en un espacio determinado; (iii) *Agricultura*: cuando se estudia o determina soluciones a los problemas, como la inseguridad alimentaria mundial provocada por el cambio climático, o para disponer de información sobre los cultivos alimentarios que pueden cultivarse en una región específica; (iv) *Recursos hídricos*: para estudios que visualizan, identifican, detectan y rastrean riesgos y desastres relacionados con el agua; (v) *Energía*: para estudios del impacto del uso y

TABLA II
ARTÍCULOS SELECCIONADOS CON ABD

	Autores	Título	Año	Citas	Revista
T1	D Lopez, G Sekaran	Climate change and disease dynamics-a Big Data perspective	2016	38	Elsevier
T2	G Manogaran, D Lopez	Spatial cumulative sum algorithm with Big Data analytics for climate change detection	2018	40	Elsevier
T3	Verma y M Poonam	Integrated Big Data Analysis of Climate Change and Disease Dynamics	2018	0	Revista IJMTE
T4	W Hallgren, L Beaumont, A Bowness...	The biodiversity and climate change virtual laboratory: where ecology meets Big Data	2016	24	Elsevier
T5	S Fiore, M Mancini, D Elia, P Nassisi...	Big Data analytics for climate change and biodiversity in the EUBrazilCC federated cloud infrastructure	2015	9	ACM
T6	ZA Sabeur, G Correndo, G Veres...	EO Big Data connectors and analytics for understanding the effects of climate change on migratory trends of marine wildlife	2017	3	Springer
T7	M Ramya, C Balaji, L Girish	Environment Change Prediction to Adapt Climate-Smart Agriculture Using Big Data Analytics	2015	4	Revista IJARCE T
T8	YJ Moon, WW Cho, J Oh, JM Kim, SY Han...	Forecasting Cultivable Region-Specific Crops Based on Future Climate Change Utilizing Public Big Data	2017	0	Springer
T9	Woodring J., Sell M., Fukuda M., Asuncion H., Salathe E.	A Multi-agent Parallel Approach to Analyzing Large Climate Data Sets	2017	0	IEEE
T10	Li Z., Huang Q., Carbone G.J., Hu F.	A high performance query analytical framework for supporting data-intensive climate studies	2017	5	Elsevier
T11	Wang X., Song L., Wang G., Ren H., Wu T., Jia X., Wu H., Wu J.	Operational climate prediction in the era of Big Data in China: Reviews and prospects	2016	1	Springer
T12	Ramos M.P., Tasinaffo P.M., De Almeida E.S., Achite L.M., Da Cunha A.M., Dias L.A.V.	Distributed systems performance for Big Data	2016	1	Springer
T13	Schnase J.L.	Climate Analytics as a Service	2016	2	Elsevier
T14	Fang S., Xu L.D., Zhu Y., Ahati J., Pei H., Yan J., Liu Z.	An integrated system for regional environmental monitoring and management based on internet of things	2014	116	IEEE
T15	D Lopez, G Sekaran	Climate change and disease dynamics-a Big Data perspective	2016	38	Elsevier

aumento de la energía por causas del Cambio Climático; y (vi) *Clima*: para estudios de datos climáticos que ayudan a mejorar predicciones.

TABLA III
ARTÍCULOS SELECCIONADOS SIN ABD

	Autores	Año	Título
T16	Y Zhu	2017	Global Climate Change Studying Based on Big Data Analysis of Antarctica
T17	MF Abdullah, MZM Amin, MF Mohamad, MM Ideris...	2018	N-HyDAA-Big Data Analytics for Malaysia Climate Change Knowledge Management
T18	A Bagdonavicius, A Kaklauskas, L Garliauskaite	-	Big Data and Decision Support System for Climate Change and Resilience Management of Built Environment
T19	JMC Loaiza, G Giuliani...	2017	Big-Data in Climate Change Models—A Novel Approach with Hadoop MapReduce
T20	S Mayor, ME Andrew, S Elmendorf...	2013	Mismatches in Phenology of Birds and Their Food Due to Climate Change: Big Data, Analytical Challenges, and Scale Sensitivity
T21	Sathiaraj D., Huang X., Chen J.	2019	Predicting climate types for the Continental United States using unsupervised clustering techniques
T22	Nik V.M.	2016	Making energy simulation easier for future climate - Synthesizing typical and extreme weather data sets out of regional climate models (RCMs)
T23	OP Popoola, NN Nuamah	2018	New Trends in Modelling Climate Change in the era of Big Data
T24	JP Gouveia, P Palma	2019	Harvesting Big Data from residential building energy performance certificates: retrofitting and climate change mitigation insights at a regional scale
T25	V Mjimba, G Sibanda	2019	Biomimicry, Big Data and artificial intelligence for a dynamic climate change management policy regime

2) Objetivo de la Arquitectura

Las arquitecturas se han clasificado de acuerdo al objetivo o al uso que se le da, que puede ser en: (i) *Monitoreo*: cuando es utilizada para monitorear los cambios estacionales en los datos climáticos, que pueden ser en la lluvia, la precipitación, la

temperatura máxima, la temperatura mínima, la humedad, la velocidad del viento y la energía solar; (ii) *Simulación*: utilizada para proyección del cambio climático (series temporales); (iii) *Análisis Estadístico*: utilizada para el análisis de datos en base a indicadores.

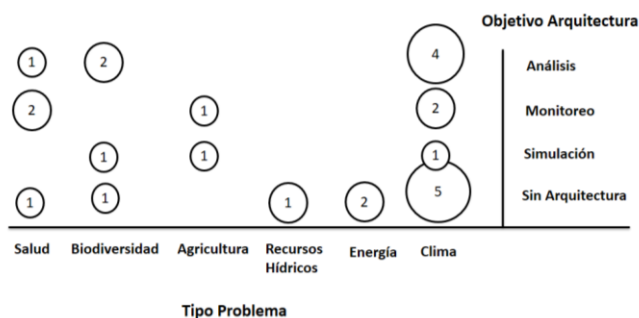


Fig. 2. Mapa Sistemático. Visualización según el tipo de problema a resolver, versus el objetivo de la arquitectura a implementar

3) Tecnologías

Se creó una lista de tecnologías a partir de la lectura de los trabajos seleccionados.

4) Resultados

Son todos los beneficios que reporten los trabajos seleccionados. Estos fueron analizados desde el punto de vista de las áreas problemas identificados.

E. Extracción de datos y Mapeo Sistemático

Tras definir el esquema de clasificación, el último paso del mapeo sistemático consiste en la extracción de datos y el proceso de mapeo de las distintas dimensiones. La Fig. 2, representa un diagrama de dispersión XY. El tamaño de cada burbuja es proporcional al número de artículos que están en el par de categorías que corresponden a las coordenadas en las que se ubica la burbuja.

IV. RESULTADOS, ANÁLISIS COMPARATIVO Y DISCUSIÓN

A continuación damos respuesta a las preguntas de investigación formuladas en la sección III a través de los resultados obtenidos.

A. PII. ¿Cuál es el tipo de problema provocado por el Cambio Climático que se intenta resolver a través de Big Data?

A continuación se describe el problema a resolver por cada tipo.

1) Big Data para la Salud

Desde la Fig. 2 se observa que se han implementado 2 ABD de monitoreo para el área de Salud, y una de ellas es para el análisis de datos.

La toma de decisiones en el área de la salud se ha vuelto difícil, debido a que las fuentes de generación de datos se han incrementado de manera exponencial, por lo que requiere herramientas y técnicas avanzadas para procesar un volumen tan grande de datos provenientes de registros clínicos [45]. Por otro lado, se requiere del desarrollo de modelos para una

predicción precisa [46] y receptiva basada en datos propietarios para proporcionar estadísticas vitales sobre la vigilancia en la salud de las personas y la hospitalización [47]. Otro requisito son los depósitos de datos geoespaciales y de salud para generar estadísticas, que permitan proporcionar información valiosa sobre los determinantes espaciotemporales de la salud y el bienestar. Ejemplo de ello, son las enfermedades infecciosas emergentes que se pueden desencadenar por las olas de calor y los períodos de frío [47].

Con el fin de ayudar en estos problemas, investigadores de tecnologías, bioinformática y modelos climáticos han desarrollado plataformas de Big Data que ayudan en el monitoreo constante de los datos obtenidos [46].

2) Big Data para la Agricultura

Se observa (ver Fig.2) que solo se encontraron 2 ABD que son utilizados para enfrentar problemas en la Agricultura. Uno de ellos usa la arquitectura para monitoreo y el otro para simulación.

En [48] reportaron mejoras a los problemas, como la inseguridad alimentaria mundial provocada por el Cambio Climático, para predecir y mitigar el impacto de los fenómenos meteorológicos extremos en las finanzas mundiales. Utilizaron una ABD para monitorear el estado de las siembras con el fin de permitir aumentar la productividad de los agricultores a través de la mejora en la toma de decisiones. La toma de decisiones en la agricultura es relevante, ya que el crecimiento de la población, junto con los factores socioeconómicos, se han asociado históricamente a la escasez de alimentos [49].

Por otro en [50] utilizaron una ABD para simular cultivos alimentarios para un pequeño sector. En el futuro permitirá a los agricultores planificar su producción. Big Data permite enfrentar los crecientes desafíos de la producción agrícola generando oportunidades de negocio [51]. A pesar de estas mejoras, Big Data no se ha aplicado ampliamente en la agricultura [52].

3) Big Data para la Biodiversidad

De acuerdo a la Fig. 2, se encontraron solo 3 ABD para enfrentar problemas de la Biodiversidad. Dos de ellas son utilizadas para realizar análisis estadísticos de los datos, y una es para la simulación.

En [53] y [54], utilizan las ABD para realizar análisis de varios tipos de datos y productos que vienen en una variedad de formatos, escalas y extensiones. Fiore investiga las tendencias de comportamiento animal en regiones marinas de interés específicas. Por otro lado Hallgren analiza datos a través de un Laboratorio Virtual de Biodiversidad y Cambio Climático basado en la nube (BCCVL), que brinda acceso a numerosas herramientas de modelado de distribución de especies; una colección de conjuntos de datos biológicos, climáticos y otros ambientales; y datos experimentales sobre el impacto del cambio climático en la biodiversidad.

Por otro lado en [55] utilizan una ABD para llevar a cabo procesos de simulación. Los datos sirven para investigar las tendencias de comportamiento animal en regiones marinas de

interés específicas como el mar mediterráneo [55].

4) Big Data para el análisis de Recursos Hídricos

En la Fig.2, se observa que solo hay un trabajo que utiliza Big Data para problemas del Cambio Climático en Recursos Hídricos. Sin embargo, no incluye la descripción de la arquitectura utilizada.

Abdullah (2018) utiliza Big Data para visualizar, identificar, detectar y rastrear riesgos y desastres relacionados con el agua. Utiliza datos desde un modelo de hidrocima regional de NAHRIM, desde el modelo atmosférico de mesoescala y desde el modelo de hidrología ambiental de una cuenca regional [56].

La demanda de agua ha aumentado [57] debido al crecimiento de la población como consecuencia del desarrollo económico, mientras que en varias regiones sufren de inundaciones y de sequía, lo que lleva a la mala gestión de los recursos hídricos [58]. El cambio climático ejerce un gran impacto en los sistemas de agua causando grandes cambios en los recursos hídricos debido a sus efectos directos sobre los procesos hidrológicos como la precipitación, evaporación y humedad [59].

5) Big Data para el análisis de la Energía

Desde la Fig. 2 se observa que se encontraron solo 2 trabajos que mejoran problemas debido al Cambio Climático. En estos trabajos se utilizan sistemas de Big Data, pero no se detallan aspectos de su arquitectura.

En [60] utilizan Big Data para realizar estudios de modernización de la energía y las evaluaciones de mitigación del cambio climático, intentando establecer el terreno para el trabajo futuro en modernización de edificios, implementación de medidas de eficiencia energética, mitigación del cambio climático, confort térmico y estudios de pobreza energética [60].

Por otro lado, debido a las incertidumbres significativas en los modelos climáticos, en [61] sugieren que se debe considerar varios escenarios climáticos en la evaluación de impacto en el uso de energía, lo que aumenta significativamente la cantidad de simulaciones y el tamaño de los conjuntos de datos complicando la evaluación y la toma de decisiones [61]. El autor sugiere un método fácil de usar para disminuir la cantidad de simulaciones considerando un número mínimo de conjuntos de datos meteorológicos, asegurando no perder la calidad y los detalles de los escenarios climáticos futuros originales. La aplicación del método se evalúa para un edificio de oficinas en Ginebra y el parque de viviendas en Estocolmo.

La gran cantidad de datos sobre energía proporciona una nueva forma de analizar y comprender el comportamiento del consumo de energía de las personas, y así mejorar la eficiencia energética y promover la conservación de la energía [62].

6) Big Data para el análisis del Clima

En la Fig. 2 se observa que 7 trabajos mejoran problemas de análisis del clima a través de una ABD. Sin embargo, se consideraron 5 trabajos que no explican la arquitectura implementada, pero que son un aporte importante para la comunidad.

En [63], [64] y [65], utilizan una ABD para el análisis de datos estadísticos. En general utilizan una amplia gama de análisis climáticos, desde la agregación espaciotemporal básica hasta la detección de anomalías más sofisticada. Uno de los trabajos utiliza una ABD para generar una plataforma de servicio de análisis del clima debido al Cambio Climático.

Por otro lado, en [66] y [67], proponen ABD para monitorear el estado del clima en un contexto de Cambio Climático.

En [1] utilizan una ABD para simular las operaciones de predicción del clima utilizando datos de información geográfica, datos del clima, modelos climáticos y datos satelitales.

En [68], [69], [70], [71] y [72], describen estudios para ayudar a comprender el tipo de clima de una región, con análisis en las consecuencias de la salud pública, el medio ambiente, las ciencias actuariales, los seguros, la agricultura e ingeniería.

B. PI2. ¿Para qué se utilizan las arquitecturas de Big Data encontradas?

Las ABD encontradas se clasificaron de acuerdo al uso que se le da, que puede ser para el análisis de los datos, el monitoreo o para simulaciones.

1) Arquitecturas para el Análisis Estadístico

Una arquitectura utilizada para el análisis de datos climáticos

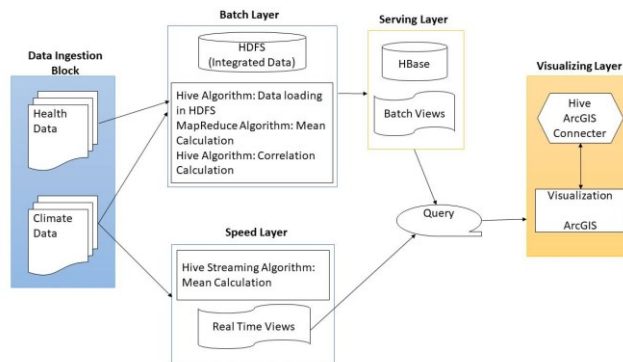


Fig. 3. Arquitectura de Big Data Lambda usada en Salud [47]

y salud es Lambda (ver Fig. 3), ya que es escalable y confiable para aplicaciones en tiempo real y se enfoca en el manejo de enormes cargas de trabajo y casos de uso [47]. Utiliza el sistema de archivos local de Hadoop para implementar el bloque de ingreso de datos. Cinco capas están asociadas a la arquitectura lambda: bloque de ingreso de datos, capa de procesos por lotes, capa de velocidad, capa de servicio y capa de visualización. Apache Hadoop, MapReduce y Apache Hive se utilizan para implementar la capa de proceso por lotes. Apache HBase se usa para implementar la capa de servicio y ArcGIS 10.2 se usa en este estudio para visualizar los resultados. La arquitectura Lambda de Big Data ayuda en el análisis estadístico debido a que se centra principalmente en calcular la temperatura máxima promedio mensual, la temperatura mínima, la precipitación, el viento, la humedad relativa, la energía solar y el cálculo del coeficiente de correlación de Pearson entre los parámetros climáticos y el número de infecciones por dengue.

Otra arquitectura es la propuesta por [54], que es utilizada

para el análisis del clima y la biodiversidad (ver Fig. 4). Los datos utilizados en el análisis son: LiDAR (o *Light Detection*

resultados a través de múltiples fuentes de datos heterogéneas en una sola plataforma integrada.

2) Arquitecturas para la Simulación

La ABD EO4wildlife diseñado por [55] está alojado en una plataforma SparkInData, que ofrece un conjunto de servicios centrales para el ingreso de datos de observación de la tierra, la integración de procesos, la ejecución y el descubrimiento de datos a través de la simulación (ver Fig. 5).

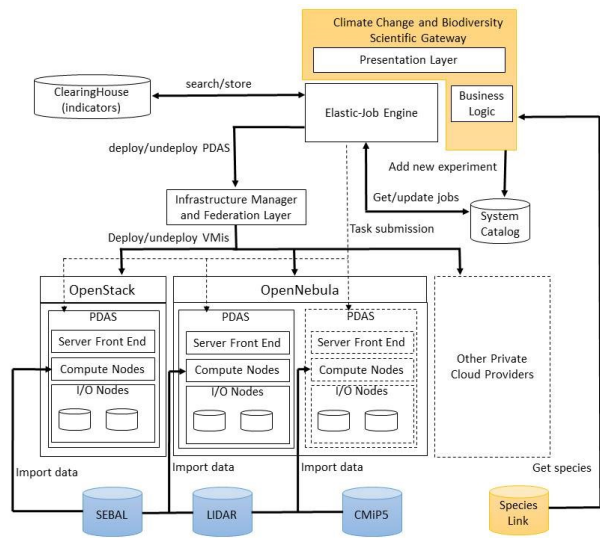


Fig. 4. ABD para el análisis del impacto del cambio climático en la biodiversidad [54]

and Ranging), imágenes hiperpectrales, imágenes satelitales, sensores a nivel del suelo, datos meteorológicos y de biodiversidad. El análisis de imágenes de sensores remotos proporciona información en 3D sobre la estructura de la vegetación, como la distribución de biomasa dentro del dosel del bosque y los patrones de densidad de la brecha forestal, lo que mejora los indicadores de biodiversidad como el balance energético y la evapotranspiración.

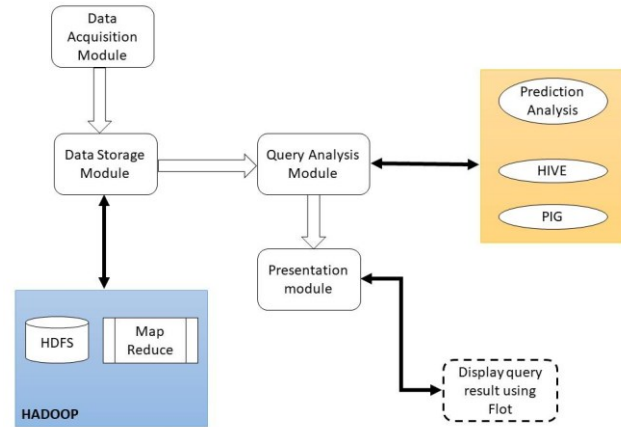


Fig. 6. ABD para en análisis del Cambio Climático en la Agricultura [48]

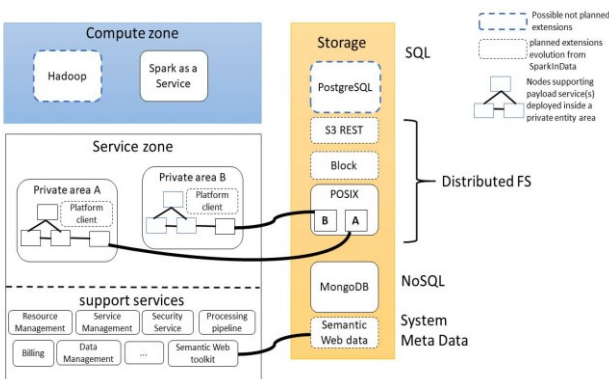


Fig. 5. ABD utilizado para la simulación de los datos del cambio climático y la biodiversidad [55]

La arquitectura de la Fig.4 está compuesta por 6 bloques: aplicación de usuario, administración de infraestructura, capa de federación, recursos de infraestructura, análisis como servicio y seguridad. Según los autores, el mayor impacto en la comunidad de usuarios finales está relacionado con el entorno integrado, debido a la posibilidad de realizar integración de datos, procesamiento, análisis estadístico y visualización de

La plataforma SparkInData, también conocida como Smart Elastic Enriched Earth Data (SEEED), es una plataforma genérica con una infraestructura y servicios en la nube.

La ABD de la Fig. 5 contiene un bloque de Semantic Web Toolkit, que garantiza el almacenamiento de datos vinculados, el acceso a recursos RDF y los mecanismos que definen las políticas de control de acceso para los almacenes de gráficos.

Las ontologías desarrolladas dentro de este bloque apuntan a cubrir la brecha, por medio la simulación, entre los productores de datos (por ejemplo, Argos, Copernicus o los dueños de los datos de animales) y los consumidores de datos (por ejemplo, los científicos que desarrollarán los flujos de trabajo).

3) Arquitectura para el Monitoreo

En [73] integran grandes repositorios de datos geospaciales y de salud derivados de flujos tradicionales como estadísticas vitales, vigilancia y hospitalización, y fuentes no tradicionales, incluidas las redes sociales, proporcionan información valiosa sobre los determinantes espacio-temporales de la salud y el bienestar. Los datos sobre enfermedades infecciosas afectadas por vectores (malaria) se obtienen de varios centros de salud públicos y privados, y los datos meteorológicos diarios se obtienen del Centro meteorológico regional, Chennai. El enfoque sugerido por el autor se implementa como un sistema de Big Data utilizando la arquitectura lambda y el modelo de procesamiento de datos MapReduce.

Por otro lado, en [48] se propuso un modelo de predicción

impacto de los fenómenos meteorológicos extremos en las finanzas mundiales. El mayor beneficio de la aplicación, es que utilizan una nueva técnica de predicción con alta precisión que ha sido diseñada por los autores. Además, en futuras mejoras, esta aplicación de predicción se puede usar en diferentes aplicaciones, como alertas de comunicación de emergencia en peligros naturales como ciclones, terremotos y tsunamis.

Por otro lado, el 2017 [50] propone una ABD para predecir cultivos en terrenos específicos de una región basados en el cambio climático futuro. En su estudio utilizó factores regionales de temperatura, factores de precipitación regional, acidez de la tierra, radiación solar, cantidad de nubes y factores climáticos apropiados para cada cultivo. La ABD permite extraer la información de cada pequeña región, como los tipos de cultivos que se cultivan actualmente, los tipos de cultivos alimentarios regionales, los tipos de frutas regionales cultivables, los tipos de cultivos medicinales regionales, los tipos de vegetales regionales y las tendencias cambiantes de cada uno. El mayor beneficio es que la ABD permite a los agricultores aumentar las ganancias futuras al proporcionar información sobre cultivos medicinales, cultivos alimentarios, verduras y frutas que se pueden producir en cada granja regional. También es posible presentar futuros cultivos recomendados a operadores comerciales individuales utilizando estos grandes datos públicos, para sugerir la necesidad de desarrollo e investigación sobre cultivos que puedan cultivarse en cada región, y sugerir planes de mercadeo para los cultivos actuales y futuros.

Algunas ventajas del uso de Big Data que hemos deducido de los estudios son: (1) monitoreo de plagas y enfermedades causadas por el cambio climático en tiempo real, (2) monitoreo del crecimiento de las plantas y su calidad, con respecto al clima (3) procesamiento de datos heterogeneos como imágenes, datos de sensores, datos del clima, índice de vegetación, teledetección (4) el análisis histórico del clima de los sectores en donde de cultiva, (5) predecir el comportamiento del clima, (6) predecir y mitigar el impacto de los fenómenos meteorológicos extremos en diversos cultivos.

3) Resultados para la Biodiversidad

En [54] estudian las interacciones desde escalas microscópicas a macroscópicas, y en escalas (genómicas, taxonómicas, de ecosistemas) del nivel de especies de plantas y animales individuales. Por otro lado, los cambios de monitoreo en la biodiversidad en un paisaje se abordan a través del análisis temporal de indicadores seleccionados (análisis de cambios) y, por ejemplo, mediante el uso de técnicas de modelado de nichos ecológicos para comprender los cambios en la distribución geográfica de las especies. Los beneficios obtenidos son estudios del impacto del Cambio Climático en regiones de gran interés como Amazonia brasileña y las regiones semiáridas de Caatinga en Brasil, y la conservación de la biodiversidad.

En [53] presentan el Laboratorio Virtual de Biodiversidad y Cambio Climático (BCCVL) basado en la nube, que proporciona acceso a numerosas herramientas de modelado de distribución de especies; una gran y creciente colección de conjuntos de datos biológicos, climáticos y ambientales, así

como una variedad de tipos de experimentos para realizar investigaciones sobre el impacto del cambio climático en la biodiversidad. Los usuarios de la ABD pueden cargar y compartir conjuntos de datos, lo que podría aumentar la colaboración y la fertilización cruzada de ideas e innovación entre la comunidad de usuarios. La retroalimentación confirma que se están cumpliendo los objetivos de BCCVL de reducir los requisitos técnicos para el modelado de distribución de especies y reducir el tiempo dedicado a dicha investigación. Los autores presentan un caso de estudio que ilustra la utilidad de BCCVL como una herramienta de investigación que puede aplicarse al problema de las enfermedades transmitidas por vectores y la probabilidad de que el cambio climático altere su distribución futura en Australia. Utilizaron el BCCVL para hacer una proyección futura de estos modelos con climas futuros basados en dos escenarios de emisiones extremas, durante varios años.

El beneficio de BCCVL es que puede asimilar los datos de los usuarios y, por lo tanto, utilizarse para realizar modelos de distribución de especies y otros análisis de biodiversidad en cualquier parte del mundo: la investigación actual que utiliza el BCCVL se ha ampliado al análisis de especies en otros países, lo que confirma la utilidad del BCCVL para realizar investigaciones de biodiversidad a nivel internacional. Además, el BCCVL también puede servir como una plantilla para otros países que desean configurar su propio Laboratorio Virtual. El uso de BCCVL tiene muchas ventajas: permite a los investigadores realizar experimentos de modelado y análisis relacionados de manera mucho más eficiente y efectiva, disminuyendo el tiempo de preparación asociado al modelado de distribución de especies.

En [55] los autores diseñan una ABD cuyos componentes se apoyan en el descubrimiento de comportamientos migratorios de la vida silvestre marina, algunos de los cuales pueden ser una consecuencia directa de los recursos variables del océano y el cambio climático [55]. El proyecto lo nombraron EO4wildlife, que se enfoca específicamente en la implementación de servicios web de analítica avanzada que es accesible a una amplia comunidad de usuarios para investigar las tendencias del comportamiento animal en torno a regiones marinas de interés.

Por otro lado, están desarrollando servicios de minería de datos para la extracción de conocimiento sobre hábitats de especies y tendencias de comportamiento temporal, para pronosticar el comportamiento de la fauna marina con incertidumbres estimadas. Estos se prueban y demuestran en escenarios temáticos específicos utilizando una plataforma de Big Data y recursos en la nube.

Las características de Big Data permitirán a los científicos (1) utilizar datos de distintos tipos provenientes desde las imágenes satelitales, (2) estudiar una amplia gama de organismos, ubicaciones, procesos y métodos, que cubren temas más amplios relacionados con los observatorios, (3) documentar los cambios con respecto al clima de la población para una especie en peligro de una manera expedita, (4) analizar datos heterogéneos como las de imágenes, seres vivos, y hábitat, (5) predecir el comportamiento del cambio climático en un hábitat específico.

4) Resultados para Recursos Hídricos

En Malasia existe un Portal de Conocimiento de Cambio Climático que es un Acelerador de Análisis de Datos Hidroclimáticos (N-HyDAA) de NAHRIM. Se ha desarrollado principalmente para proporcionar datos, información, conocimiento y tecnología relacionados con el cambio climático y agua, que es crucial para actividades empresariales relacionadas con el agua presentes y futuras, prácticas de ingeniería y medio ambiente [56]. Cuenta con ocho módulos de clima hidroclimático, que incluyen, entre otros, lluvia, inundaciones, sequías y condiciones de estrés hídrico utilizando la tecnología Big Data por medio de herramientas integrales de análisis y visualización interactiva. N-HyDAA puede rastrear, detectar, identificar y visualizar problemas futuros del agua asociados con los impactos adversos del cambio climático en Malasia. El mayor beneficio es que N-HyDAA ayuda a las entidades comerciales, operadores de agua, ingenieros, planificadores y tomadores de decisiones en el diseño, planificación y desarrollo de programas relacionados con el agua y la gestión de riesgos en la lucha contra el impacto del cambio climático, ya sea acciones de mitigación o adaptación.

N-HyDAA permite encontrar patrones climáticos, correlaciones desconocidas y proyectar tendencias del clima. Por otro lado, mejora la gestión de riesgos, fortaleciendo la adaptación al cambio climático por parte de los responsables de tomar decisiones. El sistema permite visualizar, identificar, detectar y rastrear riesgos y desastres relacionados con el agua utilizando más de 1000 simulaciones entre los años 1970-2100.

5) Resultados para la Conservación de la Energía

Los estudios no reportan los beneficios del uso de Big Data. Sin embargo, el estudio de [61] explica que el mayor beneficio que se puede obtener es la reducción del tiempo de análisis y del volumen de datos a almacenar, ya que es un aspecto importante para llevar a cabo las simulaciones con los datos [61].

6) Resultados para el análisis del Clima

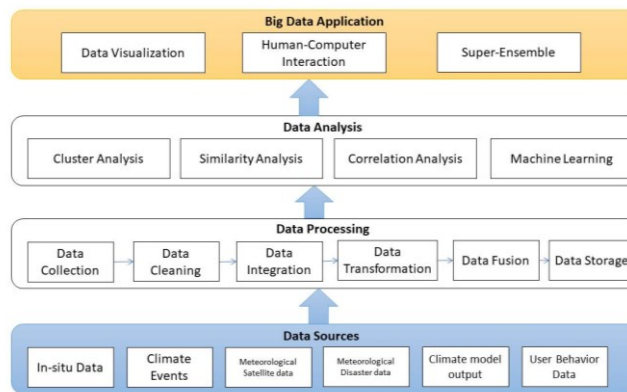
Fig. 8. ABD propuesta por Wang (2016) para el análisis del Cambio Climático [1]

En cuanto al Cambio Climático y el problema de inundaciones, los autores proponen un sistema basado en tecnología de Big Data que nombraron REGIA. El propósito de REGIA es crear condiciones favorables para la toma de decisiones basadas en la geografía y la Gestión de la Resiliencia del Ambiente, facilitando el intercambio de información.

REGIA se basa en el mapa catastral, con la posibilidad de incorporar diferentes datos de registros [69]. El mayor beneficio es que REGIA fue desarrollada para las autoridades locales: su gente, los funcionarios públicos y las empresas que operan en ella, lo que pueden trabajar en el entorno de servicio, pueden crear y administrar sus propias capas de datos, almacenar y administrar información o documentos, crear y proporcionar servicios basados en datos georreferenciados.

En [70] desarrollaron un software para el análisis de la biogeoquímica en el Mar Mediterráneo, producida en el marco de los Servicios de Monitoreo del Medio Marino de la UE

Copernicus.



En [71] clasifican el tipo de clima para las regiones de los Estados Unidos [71]. Los autores explican que la comprensión del tipo de clima de una región tiene aplicaciones en salud pública, medio ambiente, ciencias actuariales, seguros, agricultura e ingeniería. Incluso el trabajo se puede ampliar para incorporar futuros escenarios de modelos climáticos.

En [65] crean un servicio llamado CAaaS, para explotar los recursos escalables habilitados por la computación en la nube y los servicios web [65]. Los datos que utiliza este servicio son temperatura, precipitación, dirección y velocidad del viento, humedad relativa, presión atmosférica, entre otros datos que provienen desde el sistema Earth Observing System Distributed Information System (EOS DIS) de la NASA; además de datos públicos para investigación desde Earth System Grid

Federation (ESGF).

Finalmente en [66] presentan un sistema que combina Internet de las cosas (IoT), Cloud Computing, Geoinformatics (que consiste en sensores remotos (RS), sistema de información geográfica (GIS) y sistema de posicionamiento global (GPS)), y lo combina con e-Science para el monitoreo ambiental y su gestión. Aplica un estudio de caso sobre el cambio climático regional y sus efectos ecológicos.

Los ABD descritas aquí se centran exclusivamente en la mejora de los problemas relacionados con el Cambio Climático que afectan directamente a las personas o al medio ambiente.

Sin embargo, hay mucho que mejorar, porque la mayoría de los gobiernos no tienen las políticas o el presupuesto para superar los problemas de los efectos del Cambio Climático.

En China, [1] presenta las tecnologías clave para analizar el cambio climático mundial. Los datos climáticos se derivan en gran medida de observaciones del sistema terrestre y simulaciones numéricas, que incorporan niveles inherentemente altos de incertidumbre.

Como se observa en la Fig.8 [1], el sistema climático global es un sistema altamente complejo que consta de cinco esferas: la atmósfera, la hidrosfera, la criosfera, la litosfera, la biosfera y su interacción. Es esencial comprender la historia de la atmósfera, los océanos y la Tierra, así como describir su estado actual con precisión para predecir el clima. Por lo tanto, en la predicción climática operativa, la ABD debe incluir datos de observación de estaciones meteorológicas, a altas resoluciones temporales y espaciales, así como datos satelitales, datos de

radar, datos sobre eventos climáticos históricos, datos de modelos climáticos y otros. Además, para hacer una predicción basada en el impacto, también es muy importante recopilar datos sobre el comportamiento del usuario y datos sobre desastres climáticos.

V. TRABAJOS RELACIONADOS

En esta sección se presenta un conjunto de trabajos relacionados que consolidan y evalúan la evidencia disponible en el tema de Big Data y Cambio Climático. Se han identificado 4 trabajos relacionados en los últimos 5 años.

Las ventajas y consecuencias de usar la computación en la nube para abordar Big Data en dominios científicos se plantean en [62]. Este trabajo presenta un conjunto de innovaciones y una agenda de investigación para la computación en la nube que respalda la transformación de las 5 Vs como elementos valiosos de Big Data para aplicaciones científicas con impacto local y global.

El trabajo de [75], plantea una revisión de iniciativas para el cambio climático en México. Si bien este trabajo no menciona explícitamente el Big Data, se considera, pues presentar medidas acorde a la realidad latino americana. La debilidad que declara el autor es el tipo de documentos revisados, lo cual puede afectar la calidad de la evidencia y conclusiones obtenidas.

Un estudio de la literatura existente sobre el Cambio Climático realizada por investigadores egipcios es lo que se presenta en [38]. Si bien el trabajo no aborda de forma explícita el uso del Big Data, plantea un enfoque respecto de la investigación realizada por los científicos de dicho país. Los resultados muestran que una gran parte de los trabajos se publican en congresos y revistas locales, por lo cual se espera que estas publicaciones tengan un impacto muy limitado en la comunidad científica.

Una revisión de los desafíos planteados por el Cambio Climático a las organizaciones, así como las implicancias de la gestión del Big Data se presenta en [51]. Los resultados muestran que las organizaciones responden a la crisis climática invirtiendo en investigación y desarrollo, así como a través de la cooperación con otras organizaciones.

Por último, un resumen del estado del arte en cuanto a la aplicación del Big Data para el estudio del Cambio Climático se presenta en [41].

En resumen, nuestra propuesta comparte con todos los trabajos presentados las líneas temáticas de Big Data y Cambio Climático y la metodología usada. Por otro lado, además del aumento del horizonte temporal el trabajo realizado agrega los siguientes aspectos: (i) una categorización del tipo de problema ocasionado por el Cambio Climático (PI1), (ii) la revisión de arquitecturas específicas para el desarrollo de Big Data en el análisis del Cambio Climático (PI2), (iii) un primer reconocimiento de las tecnologías usadas para el desarrollo de soluciones (PI3), y finalmente (iv) los resultados obtenidos en cada una de las categorías identificadas para los problemas ocasionados por el Cambio Climático (PI4).

VI. CONCLUSIONES

Este documento presenta una Mapeo Sistemático de Estudios sobre arquitecturas de Big Data (ABD) que ayudan a resolver problemas que ocasiona el Cambio Climático. Los problemas han sido clasificados en el área de salud, agricultura, biodiversidad, energía y recursos hídricos. Se seleccionaron 25 estudios, de los cuales 15 (60%) describen los componentes de las ABD.

De los trabajos seleccionados, 12 (48%) son estudios que explican diversos problemas sobre el Clima y que son resueltos utilizando tecnologías de Big Data. Por otro lado, 4 (16%) de los estudios intentan resolver problemas en salud, y el mismo porcentaje en biodiversidad. Son pocos los trabajos que describen una posible solución para enfrentar problemas de energía, agricultura y recursos hídricos. Las razones pueden ser diversas, entre las cuales se tiene la falta de recursos de datos idóneos, la falta de financiación para la investigación, la comunicación deficiente entre grupos de investigadores con diferentes habilidades, la incertidumbre en la información disponible, entre otros factores.

Las arquitecturas encontradas han sido clasificadas según su uso, que puede ser para análisis estadístico, monitoreo y simulaciones; ayudando a los investigadores a integrar conocimiento en el uso práctico de Big Data en el contexto del Cambio Climático. Para el caso del análisis estadístico Big Data ayuda a integrar información histórica del Clima en diversos sectores, incluyendo aquellos más complejos, como es el caso de sequías e inundaciones; para luego clasificar los datos con respecto a problemas de salud y biodiversidad.

Por otro lado, los componentes de Big Data, poseen la capacidad de procesar grandes volúmenes de datos de imágenes, sensores y de fuentes de datos heterogéneas, con el fin de proveer herramientas para el monitoreo y simulación del Cambio Climático. En el área de la salud monitorean el comportamiento de enfermedades, y en la agricultura el proceso de crecimiento de las plantas y las posibles plagas. La simulación es utilizada para descubrir problemas que genera el Cambio Climático en hábitat y en la agricultura. Esto se consigue generando distintos escenarios con herramientas inteligentes.

Una de las tendencias de investigación identificadas entre las arquitecturas analizadas es la computación en la nube, que proporciona una mejor solución para los requisitos de almacenamiento, procesamiento y análisis de datos. También vale la pena destacar la integración de IoT, el almacenamiento de grandes datos complejos en bases de datos no estructuradas, el uso de técnicas analíticas o de minería de datos más avanzadas, y el uso de algoritmos de procesamiento avanzado como MapReduce.

REFERENCIAS

- [1] X. Wang, L. Song, G. Wang, H. Ren, T. Wu, and X. Jia, "Operational climate prediction in the era of Big Data in china: Reviews and prospects," *J. Meteorol. Res.*, vol. 30, no. 3, pp. 444--456, 2016.

- [2] J. Pollard, T. Spencer, and S. Jude, "Big Data approaches for coastal flood risk assessment and emergency response," *Wiley Interdiscip. Rev. Clim. Chang.*, vol. 9, no. 5, p. e543, 2018.
- [3] R. J. Nicholls, S. E. Hanson, J. A. Lowe, R. A. Warrick, X. Lu, and A. J. Long, "ea-level scenarios for evaluating coastal impacts," *WIREs Clim. Chang.*, vol. 5, no. 129--150, 2014.
- [4] M. Rummukainen, "Changes in climate and weather extremes in the 21st century.," *Wiley Interdiscip. Rev. Clim. Chang.*, vol. 3, no. 2, pp. 115--129, 2012.
- [5] I. Oluigbo, O. Nwokonkwo, G. Ezeh, and N. Ndukwe, "Revolutionizing the healthcare industry in nigeria: The role of internet of things and Big Data analytics.," 2017.
- [6] R. Scholes, "Climate change and ecosystem services.," *Wiley Interdiscip. Rev. Clim. Chang.*, vol. 7, no. 4, pp. 537--550, 2016.
- [7] B. Nerlich, N. Koteyko, and B. Brown, "Theory and language of climate change communication.," *Wiley Interdiscip. Rev. Clim. Chang.*, vol. 1, no. 1, pp. 97--110, 2010.
- [8] H.-D. Guo, L. Zhang, and L.-W. Zhu, "Earth observation Big Data for climate change research.," *Adv. Clim. Chang. Res.*, vol. 2, no. 2, pp. 108--117, 2015.
- [9] J. Gabrys, "Practicing, materialising and contesting environmental data.," *Big Data Soc.*, vol. 3, pp. 1--7, 2016.
- [10] A. Keeso, "Big Data and environmental sustainability: A conversation starter," *Work. Pap. Ser. 14-04, Oxford, Engl. Smith Sch.*, vol. 34, 2014.
- [11] J. Gantz and D. Reinsel, "Extracting Value from Chaos, IDC IVIEW," <http://www.emc.com/collateral/analystreports/idc-extracting-value-from-chaos-ar.pdf>, 2011.
- [12] Y. Demchenko, P. Membrey, P. Grosso, and C. De-Laet, "Addressing Big Data Issues in Scientific Data Infrastructure," *First Int. Symp. Big Data Data Anal. Collab. (BDDAC). Part 2013 Int. Conf. Collab. Technol. Syst. (CTS 2013). San Diego, California, USA*, 2013.
- [13] A. De Mauro, M. Greco, and M. Grimaldi, "A formal definition of Big Data based on its essential features," *Libr. Rev. Emerald Gr. Publ. Ltd.*, vol. 65, no. 3, pp. 122--135, 2016.
- [14] Y. Demchenko, C. De-Laet, and P. Membrey, "Defining architecture components of the Big Data ecosystem.," *Int. Conf. Collab. Technol. Syst.*, pp. 104--112, 2014.
- [15] P. Ylijoki, "Perspectives to Definition of Big Data: A Mapping Study and Discussion," *ournal Innov. Manag. JIM*, vol. 4, 2016.
- [16] P. Mikalef, I. Pappas, J. Krogstie, and M. Giannakos, "Big Data analytics capabilities: a systematic literature review and research agenda," *Inf. Syst. E-bus. Manag.*, vol. 16, pp. 547--578, 2018.
- [17] M. Santos *et al.*, "A Big Data analytics architecture for industry 4.0," *World Conf. Inf. Syst. Technol. Springer*, pp. 175--184, 2017.
- [18] R. Sowmya and K. Suneetha, "Data mining with Big Data," *2017 11th Int. Conf. Intell. Syst. Control (ISCO). IEEE*, pp. 246--250, 2017.
- [19] C. Ordóñez, C. García-Alvarado, and I.-Y. Song, "Special issue on DOLAP 2015: Evolving data warehousing and OLAP cubes to Big Data analytics.," *Inf. Syst.*, vol. 68, pp. 1--2, 2017.
- [20] I.-Y. Song and Y. Zhu, "Big Data and data science: what should we teach?," *Expert Syst. Wiley Online Libr.*, vol. 33, no. 4, pp. 364--373, 2016.
- [21] M. Kahn, "The climate change adaptation literature," *Rev. Environ. Econ. Policy. Oxford Univ. Press*, vol. 10, no. 1, pp. 166--178, 2016.
- [22] Y. González, Y. Fernández, and T. Gutiérrez, "El cambio climático y sus efectos en la salud," *Rev. Cuba. Hig. y Epidemiol. 1999, Editor. Ciencias Médicas*, vol. 51, no. 3, pp. 331--337.
- [23] I. P. I. de Cambio Climático, "Cambio climático," *Base Defic. física. Suiza IPCC*, 2013.
- [24] M. Molina, J. Sarukhán, and J. Carabias, "El cambio climático: causas, efectos y soluciones," *Fondo Cult. Econ.*, 2017.
- [25] W. Pearce, S. Niederer, Sabine and Ózkula, and N. Sánchez Querubín, "The social media life of climate change: Platforms, publics, and future imaginaries," *Wiley Interdiscip. Rev. Clim. Chang. Wiley Online Libr.*, vol. 10, no. 2, p. e569, 2019.
- [26] J. Ford *et al.*, "Opinion: Big Data has big potential for applications to climate change adaptation," *Proc. Natl. Acad. Sci. Natl. Acad. Sci.*, vol. 113, no. 39, pp. 10729--10732, 2016.
- [27] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," *EASE'08 Proc. 12th Int. Conf. Eval. Assess. Softw. Eng. Br. Comput. Soc. Swint.*, pp. 68--77, 2008.
- [28] N. Salleh, F. Mendes, and E. Mendes, "A Systematic Mapping Study of Value-Based Software Engineering," *2019 45th Euromicro Conf. Softw. Eng. Adv. Appl. (SEAA). IEEE*, pp. 404--411, 2019.
- [29] A. Sepúlveda, Samuel and Rivera, "Systematic Mapping Protocol Feature Modeling Tools," *arXiv Prepr. arXiv1907.08076*, 2019.
- [30] G. Matturro, F. Raschetti, and C. Fontán, "A Systematic Mapping Study on Soft Skills in Software Engineering," *J. UCS*, vol. 25, no. 1, pp. 16--41, 2019.
- [31] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," *Technical Rep. EBSE'07*, 2007.
- [32] M. Auffhammer, "Climate adaptive response estimation: Short and long run impacts of climate change on residential electricity and natural gas consumption using Big Data," *Natl. Bur. Econ. Res.*, 2018.
- [33] E. Lacey, "Climate change, collections and the classroom: using Big Data to tackle big problems," *Evol. Educ. Outreach. BioMed Cent.*, vol. 10, no. 1, p. 2, 2017.
- [34] A. D'Anca, L. Conte, C. Palazzo, S. Fiore, and G. Aloisio, "A Big Data approach for climate change indicators processing in the CLIP-C project," *EGU Gen. Assem. Conf. Abstr.*, vol. 18, 2016.
- [35] S. Fiore, D. Williams, A. D'Anca, P. Nassisi, and G. Aloisio, "The Ophidia Stack: Toward Large Scale, Big Data Analytics Experiments for Climate Change," *AGU Fall Meet. Abstr.*, 2015.
- [36] K. Miller, "Big Data for Small Parks: Examining Regional Vegetation Patterns to Assess the Current Condition and Vulnerability of Eastern National Parks to Climate Change," 2018.
- [37] J. Levy and R. Prizzia, "From Data Modeling to Algorithmic Modeling in the Big Data Era: Water Resources Security in the Asia-Pacific Region under Conditions of Climate Change," *Asia-Pacific Secur. Challenges. Springer*, pp. 197--220, 2018.
- [38] M. Hassaan, "Using scholarly Big Data in assessing contribution of national expertise to climate change knowledge; case study: Egypt," *J. Data Anal. Inf. Process.*, 2018.
- [39] E. Lindquist, "The convergence of climate change and Big Data in an urban and regional context: a policy perspective," *AGU Fall Meet. Abstr.*, 2018.
- [40] C. Kalaitzidis, I. Gitas, V. Ambrosia, T. Katagis, and M. Lateb, "New trends in forest fire research incorporating Big Data and climate change modeling," *South-Eastern Eur. J. Earth Obs. Geomatics*, vol. 8, no. 1, pp. 1--163, 2019.
- [41] H. Hassani, X. Huang, and E. Silva, "Big Data and climate change," *Big Data Cogn. Comput. Multidiscip. Digit. Publ. Inst.*, vol. 3, no. 1, p. 12, 2019.
- [42] Z. Zhang and J. Li, "Big Data Mining for Climate Change," *Elsevier*, 2019.
- [43] J. Minx, "Tracking adaptation to climate change using Big Data," *nercdtp.leeds.ac.uk*.
- [44] B. Knusel *et al.*, "Applying Big Data beyond small problems in climate research," *Nat. Clim. Chang.*, vol. 9, no. 3, p. 196, 2019.
- [45] T. Murdoch and A. Detsky, "The inevitable application of Big Data to health care," *Jama*, vol. 309, no. 13, pp. 1351--1352, 2013.
- [46] G. Manogaran and D. Lopez, "Spatial cumulative sum algorithm with Big Data analytics for climate change detection," *Comput. & Electr. Eng.*, vol. 65, pp. 207--221, 2018.
- [47] P. Verma and M. Poonam, "Integrated Big Data analysis of climate change and disease dynamics," *Int. J. Manag. Technol. Eng.*, vol. 8, no. 8, 2018.
- [48] M. Ramya, C. Balaji, and L. Girish, "Environment change prediction to adapt climate-smart agriculture using Big Data analytics.," *Int. J. Adv. Res. Comput. Eng. & Technol.*, vol. 4, 2015.
- [49] P. Slavín, "Climate and famines: A historical reassessment," *Wiley*

- Interdiscipl. Rev.: Clim.*, vol. 7, no. 3, pp. 433–447, 2016.
- [50] Y.-J. Moon, W. W. Cho, J. Oh, J. M. Kim, S. Y. Han, and K. H. Kim, “Forecasting cultivable region-specific crops based on future climate change utilizing public Big Data,” *Adv. Multimed. Ubiquitous Eng.*, pp. 399–404, 2017.
- [51] B. Seles, A. De-Sousa, C. Jabbour, P. De-Camargo, Y. Mohd-Yusoff, and A. Thomé, “Business opportunities and challenges as the two sides of the climate change: Corporate responses and potential implications for Big Data management towards a low carbon society,” *J. Clean. Prod.*, vol. 189, pp. 763–774, 2018.
- [52] R. Lokers, R. Knapen, S. Janssen, Y. Van-Randen, and J. Jansen, “Analysis of Big Data technologies for use in agro-environmental science,” *Environ. Model. & Softw.*, vol. 84, pp. 494–504, 2016.
- [53] W. Hallgren, L. Beaumont, A. Bowness, L. Chambers, E. Graham, and H. Holewa, “The biodiversity and climate change virtual laboratory: How ecology and Big Data can be utilised in the fight against vector-borne diseases,” *Environ. Model. & Softw.*, vol. 76, no. 182–186, 2016.
- [54] S. Fiore, M. Mancini, D. Elia, P. Nassisi, F. V. Brasileiro, and I. Blanquer, “Big Data analytics for climate change and biodiversity in the eubrazilicc federated cloud infrastructure,” *Proc. 12th ACM Int. Conf. Comput. Front.*, vol. 52, 2015.
- [55] Z. A. Sabeur, G. Correndo, G. Veres, B. Arbab-Zavar, J. Lorenzo, and T. Habib, “Eo Big Data connectors and analytics for understanding the effects of climate change on migratory trends of marine wildlife. Environmental Software Systems,” *Comput. Sci. Environ. Prot. 12th IFIP WG 5.11 Int. Symp. ISESS 2017, Zadar, Croat.*, pp. 85–94, 2017.
- [56] M. F. Abdullah, M. Z. M. Amin, M. F. Mohamad, M. M. Ideris, Z. Zainol, and N. Y. Yussof, “N-hydaa - Big Data analytics for malaysia climate change knowledge management,” *13th Int. Conf. Hydroinformatics HIC 2018 - Palermo 1-6 July.*, 2018.
- [57] & Y. Ai, P. and Z.-X., “A framework for processing water resources Big Data and application,” *Appl. Mech. & Mater.*, 2014.
- [58] D. García, J. Quevedo, V. Puig, and J. Saludes, “Water demand estimation and outlier detection from smart meter data using classification and Big Data methods,” *2nd New Dev. IT & Water Conf. 8-10 February, Rotterdam*, pp. 1–8, 2015.
- [59] A. Cravero, O. Saldana, R. Espinosa, and C. Antileo, “Big Data architecture for water resources management: A systematic mapping study,” *IEEE Lat. Am. Trans.*, vol. 16, no. 3, pp. 902–908, 2018.
- [60] J. P. Gouveia and P. Palma, “Harvesting Big Data from residential building energy performance certificates: retrofitting and climate change mitigation insights at a regional scale,” *Environ. Res. Lett. IOP Publ.*, vol. 14, no. 9, p. 095007, 2019.
- [61] V. M. Nik, “Making energy simulation easier for future climate - synthesizing typical and extreme weather data sets out of regional climate models (rcms),” *Appl. Energy*, vol. 177, pp. 204–226, 2016.
- [62] K. Zhou and S. Yang, “Understanding household energy consumption behavior: The contribution of energy Big Data analytics,” *Renew. Sustain. Energy Rev.*, vol. 56, 2016.
- [63] J. Woodring, M. Sell, M. Fukuda, H. Asuncion, and E. Salathe, “A Multi-agent Parallel Approach to Analyzing Large Climate Data Sets,” *2017 IEEE 37th Int. Conf. Distrib. Comput. Syst. (ICDCS). IEEE*, pp. 1639–1648, 2017.
- [64] Z. Li, Q. Huang, G. J. Carbone, and F. Hu, “A high performance query analytical framework for supporting data-intensive climate studies,” *Comput. Environ. Urban Syst. Elsevier*, vol. 62, pp. 210–221, 2017.
- [65] J. L. Schnase, “Climate Analytics as a Service,” *Cloud Comput. Ocean Atmos. Sci. Elsevier*, pp. 187–219, 2016.
- [66] S. Fang *et al.*, “An integrated system for regional environmental monitoring and management based on internet of things,” *IEEE Trans. Ind. Informatics*, vol. 10, no. 2, pp. 1596–1605, 2014.
- [67] M. Ramos, P. Tasinaffo, E. de Almeida, L. Achite, A. da Cunha, and L. Dias, “Distributed systems performance for Big Data,” *Inf. Technol. New Gener. Springer*, pp. 733–744, 2016.
- [68] Y. Zhu, “Global Climate Change Studying Based on Big Data Analysis of Antarctica,” *Proc. Fourth Int. Forum Decis. Sci. Springer*, pp. 39–45, 2017.
- [69] A. Bagdonavicius, A. Kaklauskas, and L. Garliauskaite, “Big Data and decision support system for climate change and resilience management of built environment.”
- [70] J. Loaiza, M. Carmona, G. Giuliani, and G. Fiameni, “Big-Data in Climate Change Models—A Novel Approach with Hadoop MapReduce,” *2017 Int. Conf. High Perform. Comput. & Simul. (HPCS). IEEE*, vol. 45–50, 2017.
- [71] D. Sathiaraj, X. Huang, and J. Chen, “Predicting climate types for the Continental United States using unsupervised clustering techniques,” *Environmetrics. Wiley Online Libr.*, vol. 30, no. 4, p. e2524, 2019.
- [72] O. Popoola and N. Nuamah, “New Trends in Modelling Climate Change in the era of Big Data,” *Ann. Comput. Sci. Ser.*, vol. 16, no. 2, 2018.
- [73] D. Lopez and G. Sekaran, “Climate change and disease dynamics-a Big Data perspective,” *Int. J. Infect. Dis.*, vol. 45, pp. 23–24, 2016.
- [74] Y. Wang, L. Kung, and T. Byrd, “Big Data analytics: Understanding its capabilities and potential benefits for healthcare organizations,” *Technol. Forecast. Soc. Chang. Elsevier*, vol. 126, no. 3–13, 2018.
- [75] J. Silva, “Climate change initiatives in Mexico: A review,” *Manag. Environ. Qual. An Int. Journal.*, 2018.



A. Cravero es Licenciada en Ciencias de la Ingeniería (1996) e Ingeniera Civil Industrial m. Informática (1997), por la Universidad de La Frontera, Temuco, Chile. Obtuvo su máster en Tecnologías de la Información, por la Universidad Politécnica de Madrid, España (2006). Ha obtenido en 2010 su Doctorado en Cs. de la Computación y Sistemas Informáticos por la Atlantic International University, EE.UU. Se desempeña como Académico en el Departamento de Ciencias de la Computación e Informática, e investigadora en el Centro de Estudios en Ingeniería de Software, Universidad de La Frontera. Participa en el comité de gestión del Centro de Excelencia de Computación Científica. Sus intereses de investigación están en el área de Modelado Bases de Datos, Ingeniería de Requisitos para Almacenes de Datos y Big Data.



S. Sepúlveda es Licenciado en Cs. de la Ingeniería (1998) e Ing. Civil Industrial m. Informática (1999), por la Universidad de La Frontera, Temuco, Chile. Obtuvo su máster en Dirección y Gestión de Sistemas de Información y TIC, por la Universitat Oberta de Catalunya, España (2006). Actualmente postula al grado de Doctor en Aplicaciones de la Informática por la Universidad de Alicante, España. Se desempeña como Académico en el Dpto. de Ciencias de la Computación e Informática, e investigador en el Centro de Estudios en Ingeniería de Software, Universidad de La Frontera. Sus intereses de investigación están en el área de Ingeniería de Requerimientos, Modelado de Líneas de Productos de Software y estudios secundarios aplicados en Ingeniería de Software.



L. Muñoz es Ingeniera de Sistemas Computacionales por la Universidad Tecnológica de Panamá. Obtuvo la Maestría en Computación con énfasis en Sistemas de Información en el Instituto Tecnológico de Costa Rica. Ha obtenido el Doctorado en Aplicaciones de la Informática por la Universidad de Alicante en el seno del grupo de Investigación Lucentia (2010).

Docente tiempo completo en la Facultad de Ingeniería de Sistemas Computacionales de la Universidad Tecnológica de Panamá. Sus áreas de interés son la calidad de software, almacenes de datos, auditoría de sistemas, bases de datos, informática aplicada a la educación, internet de las cosas. Actualmente coordina el Grupo en Tecnologías Computacionales Emergentes en la Universidad Tecnológica de Panamá, Centro Regional de Chiriquí.