



BIG DATA ESPACIAL EN LA GESTIÓN DE DATOS DE LA COSECHA DE CAÑA DE AZÚCAR.

SPATIAL BIG DATA IN THE MANAGEMENT OF SUGAR CANE HARVEST DATA

MSc. Mavelyn Batule Domínguez. (1)

Dr.C. Rafael Cruz Iglesias. (2)

Dr.C. José Luís Capote Fernández. (3)

Dr.C. Guillermo González Suárez. (4)

1. GEOCUBA UCT- IC. Agencia GeoMIX. mbatule@geomix.geocuba.cu.
2. GEOCUBA UCT- IC. Agencia GeoMIX. rcruz@geomix.geocuba.cu.
3. GEOCUBA UCT- IC. Agencia GeoMIX. capote@geomix.geocuba.cu.
4. GEOCUBA UCT- IC. Agencia GeoMIX. guille@geomix.geocuba.cu.

RESUMEN: La transformación digital en la agricultura cubana busca modernizar y aumentar la eficiencia del sector a través de tecnologías digitales, enfrentando diversos desafíos. Un avance significativo es la Agricultura de Precisión (AP), que optimiza la producción, especialmente en la caña de azúcar. Este artículo presenta una plataforma que utiliza *Big Data* Espacial para procesar información de las cosechadoras Case IH A8000 obtenidas mediante el software *SMS Ag Leader*. Los datos se integran en la Infraestructura de Datos Espaciales de AZCUBA (IDE-AZCUBA), facilitando el intercambio de información y la toma de decisiones informadas. La investigación resalta la importancia de estas herramientas en la gestión de datos agrícolas, promoviendo una agroindustria más sostenible y colaborativa.

Palabras Claves: Big Data Espacial; Datos de cosecha; Estrategia; Gestión.

ABSTRACT: Digital transformation in Cuban agriculture aims to modernize and increase the efficiency of the sector through digital technologies, facing various challenges. A significant advancement is Precision Agriculture (PA), which optimizes production, especially in sugarcane. This article presents a platform that uses Spatial Big Data to process information from Case IH A8000 harvesters obtained through SMS Ag Leader software. The data is integrated into the Spatial Data Infrastructure (IDE-

AZCUBA), facilitating information exchange and informed decisionmaking. The research highlights the importance of these tools in agricultural data management, promoting a more sustainable and collaborative agribusiness.

Keywords: Spatial Big Data; Harvest data; Strategy; Management.

Recibido: 10/08/2024

Aprobado: 06/12/2024

1. INTRODUCCIÓN

La transformación digital en la agricultura cubana está en proceso de desarrollo, buscando modernizar e incrementar la eficiencia del sector agrícola mediante la integración de tecnologías digitales. Pese a los diversos desafíos que impone esta labor, principalmente relacionados con la conectividad, acceso a internet e infraestructura, se han comenzado a implementar diferentes políticas encaminadas a lograr la digitalización de la sociedad en general. Uno de los avances más significativos se ha observado en la AP, donde se utilizan tecnologías como sensores y drones para optimizar la producción de cultivos, especialmente en la producción de caña de azúcar y arroz (Muñoz-González et al., 2022).

La industria azucarera cubana enfrenta el reto de aumentar la producción de caña de azúcar para satisfacer una demanda creciente, al tiempo que busca reducir el impacto ambiental y los costos de operación. Para abordar este tema, la AP incorpora a las maquinarias agrícolas, como se mencionaba anteriormente, dispositivos de alta tecnología, como receptores de posicionamiento, sensores y computadoras de a bordo (Gil, 2003). Lo que permite cuantificar y manejar diferencialmente la variabilidad natural de las parcelas. Esto permite planificar las tareas en base a información obtenida en operaciones previas. En el caso específico de la cosecha de caña, los datos recolectados pueden utilizarse para generar mapas de rendimiento o fertilidad del suelo, extendiendo la aplicación del concepto de AP a todas las etapas del proceso productivo. La introducción de maquinaria especializada, como las cosechadoras *Case IH A8000*, ha sido un paso importante para mejorar la eficiencia de la cosecha y ha marcado un hito en la adopción de tecnologías de AP en ese sector. Estas máquinas, equipadas con avanzados sistemas de monitoreo y control, han permitido aumentar significativamente la productividad y reducir la contaminación (Gradaille, 2017). Los datos operativos registrados con detalles georreferenciados sobre las operaciones de cosecha obtenidos por dichos monitores pueden analizarse mediante softwares especializados como *SMS Advanced* (IT Admin, 2012). El personal técnico ha sido capacitado en

el uso de este software con vistas a lograr la gestión de estos datos a nivel de campo, visualizando labores realizadas, productividad, consumo de combustible, entre otros (Pérez-García et al., 2022). El intercambio de información agrícola entre los diversos actores del sector es fundamental para toda la comunidad productiva. La agroindustria azucarera cubana ha adoptado una Infraestructura de Datos Espaciales (IDE) para integrar sus datos y metadatos con componentes geoespaciales. La IDE-AZCUBA, agrupa diferentes iniciativas de almacenamiento, procesamiento, cálculos de índices de importancia agrícola y publicación de información relativa a las áreas cañeras (Delgado Mora et al., 2023; Delgado-Mora et al., 2023; Soca-Muñoz et al., 2020).

Incorporar la información recolectada por las cosechadoras a la IDE constituye un enfoque más integrado y colaborativo en la gestión cañera, ya que mejora la organización, disponibilidad y acceso compartido de los datos. En este contexto, el uso de tecnologías avanzadas se convierte en un factor clave para garantizar la recolección y análisis de datos en tiempos cercanos al real desde las cosechadoras.

Este artículo presenta una plataforma que soporta la información generada por el computador de a bordo de las cosechadoras, la cual es procesada mediante el software *SMS Ag Leader*. Esta plataforma, desplegada a través de tecnologías de *Big Data* geoespacial, se encuentra en el Centro de Información Geoespacial de GEOCUBA (CIG). Su implementación no solo facilita la gestión de datos agrícolas, sino que también permite a los productores tomar decisiones informadas basadas en análisis precisos y actualizados.

A medida que avanzamos hacia una agricultura más digitalizada, es esencial explorar y comprender las herramientas que pueden transformar la manera en que se gestionan las operaciones agrícolas. En este sentido, el artículo busca contribuir al conocimiento sobre la aplicación de tecnologías de *Big Data* en la agroindustria, destacando su potencial para mejorar la productividad y sostenibilidad del sector agrícola.

2. METODOLOGÍA

El CIG ha estado a la vanguardia en la implementación de tecnologías de *Big Data* geoespacial (Fernández & Iglesias, 2020). Este centro se ha enfrentado a retos significativos relacionados con el manejo de grandes volúmenes de información, la variedad de datos y la velocidad requerida para su análisis. Para abordar estos desafíos, se han evaluado y adoptado tecnologías como *GeoMesa*, *GeoTrellis* y *MrGeo*, que permiten gestionar datos tanto vectoriales como ráster.

2.1 Tecnologías

La solución implementada está integrada dentro de las instancias que conforman el CIG. Sus componentes tecnológicos principales incluyen *Apache Hadoop* (Shvachko et al., 2010), *Apache Accumulo* (Sen et al., 2013), *GeoMesa* (Hughes et al., 2015) y *GeoServer* (Deoliveira, 2008). Estos elementos pertenecen a un ecosistema de código abierto, lo que garantiza su extensibilidad y mantenibilidad a largo plazo. Los componentes de desarrollo propios, también han sido implementados utilizando lenguajes de código abierto como *Java (Spring Boot)* y *JavaScript (Leaflet)*. Esta elección de tecnologías favorece a la adaptabilidad del sistema a futuras necesidades, asegurando un soporte continuo y mejoras en la funcionalidad.

2.2 Modelo de datos

La estructura del modelo de datos para la información de las variables de las cosechadoras se ha diseñado a partir del mapeo entre los datos originales en formato CSV, obtenidos de la exportación del software, y los equivalentes disponibles en *GeoMesa*. La geometría de los datos se representa mediante elementos puntuales georreferenciados en el sistema de coordenadas WGS 84. Para optimizar el rendimiento de las consultas, se han creado índices para los atributos que representan el identificador del equipo, la fecha del dato y la posición del mismo dentro del conjunto de datos. Esta estrategia de indexación permite una recuperación eficiente de la información, facilitando el análisis y la toma de decisiones basada en los datos recopilados por las cosechadoras. El modelo de datos resultante es flexible y escalable, adaptándose a las necesidades cambiantes del sector agrícola y permitiendo la integración de nuevas fuentes de información en el futuro.

3. RESULTADOS Y DISCUSIÓN

La solución incluye un flujo de gestión integral que abarca desde la ingestión de datos en el clúster de *Big Data* hasta su publicación como servicio web.

3.1 Recolección de Datos

Este flujo se inicia con la recolección de datos por parte de los monitores de rendimiento de las cosechadoras *Case IH A8000*, que generan entre 7,000 y 8,000 registros georreferenciados por turno. Estos datos, que incluyen información sobre ubicación, rendimiento y condiciones operativas de las máquinas, son exportados utilizando el software *SMS Ag Leader*.

El personal técnico del central azucarero es responsable de configurar adecuadamente las cosechadoras para garantizar la captura precisa de los datos de cosecha. Además, deben seguir la

metodología desarrollada por (Pérez García, 2016) para la extracción y procesamiento de los datos utilizando el software *SMS Ag Leader*.

Una vez que el software SMS genera los archivos CSV, se ha desarrollado una aplicación web para ingestar estos datos en el clúster de *Big Data*. Esta aplicación web permite a los usuarios autorizados cargar los archivos de tramas de las cosechadoras y subirlos a la plataforma de *Big Data* de manera segura y controlada.

Durante el desarrollo de esa aplicación, se estableció un mecanismo para la entrada de metadatos junto a cada archivo que se introduzca en la plataforma. Este proceso se llevó a cabo como parte de la ingestión de datos, con el objetivo de caracterizar y describir cada archivo procesado. Los metadatos generados se insertaron en el catálogo de metadatos de la IDE, utilizando el proyecto *GeoNetwork*, concebido específicamente para este propósito.



Figura 1: Flujo de gestión de la solución.

3.2 Almacenamiento y Procesamiento

Para el almacenamiento y procesamiento de datos, se utilizan tecnologías de *Big Data* como *Apache Hadoop*, que permiten gestionar grandes volúmenes de información a lo largo del tiempo y facilitan su análisis mediante procesamiento distribuido. Además, se implementa *AccumuloGeoMesa* para manejar datos históricos geoespaciales, lo que posibilita consultas eficientes y análisis espaciales complejos.

Los datos clave para evaluar el rendimiento de los cultivos y la eficiencia operativa incluyen:

- Toneladas cosechadas por hectárea: Evalúa la productividad de cada campo, permitiendo ajustar prácticas agrícolas para maximizar la eficiencia y reducir costos.
- Consumo de combustible: Indica la eficiencia operativa de la maquinaria.

- Condiciones del suelo y clima: Afectan la calidad y cantidad de la cosecha.

Con estos registros y las capacidades tecnológicas de la plataforma, se pueden realizar análisis para extraer información relevante sobre el rendimiento de los cultivos y las condiciones del terreno, impactando en la toma de decisiones informadas y optimizando recursos para prácticas agrícolas más sostenibles.

En trabajos futuros, se podrán cruzar estos datos con fuentes externas relevantes para generar análisis y visualizaciones más completas sobre rendimiento y productividad. Además, se planea extender el procesamiento para aplicar modelos estadísticos y algoritmos de inteligencia artificial (IA), lo que permitirá predecir rendimientos futuros y detectar problemas potenciales, como plagas o enfermedades en los cultivos.

3.3 Publicación y Acceso a la Información

Utilizando la extensión de *GeoMesa* y los datos de las cosechadoras almacenados en *Accumulo*, se creó una capa en el servidor de mapas a través de una interfaz WMS. Este servicio se integró en el visor de la IDE-AZCUBA, permitiendo el acceso a la capa "Cosechadoras" para múltiples usuarios. Para su representación, se pueden emplear fondos de capas satelitales o imágenes de drones, así como combinarla con áreas de cultivo y datos sobre el tipo de cultivo, objetivos de rendimiento y otras variables disponibles en la IDE.

La capa se complementa con dos estilos visuales: uno para la presión de corte y otro para la velocidad, utilizando una escala de colores que indica el rendimiento específico de cada punto en estas variables. Esta representación visual no solo mejora la comprensión de los datos, sino que también facilita decisiones informadas en la gestión de cosechadoras y optimización de la producción agrícola.

La elaboración de estos mapas puede ampliarse a diversas variables proporcionadas por las cosechadoras. Es fundamental seguir desarrollando estilos de representación más efectivos para los mapas temáticos de cada variable de cosecha y optimizar los parámetros de interpolación, con el objetivo de crear mapas comparables a los generados por el sistema SMS, tanto para QGIS como para la IDE.

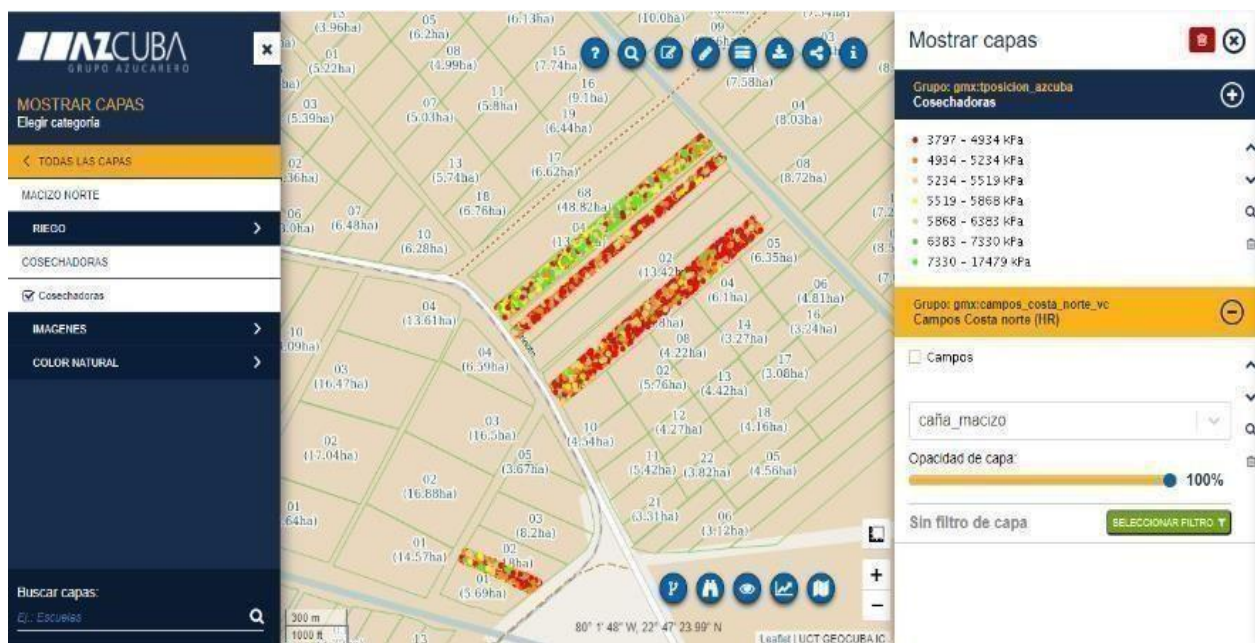


Figura 2 Capa cosechadoras en el visor de la IDE-AZCUBA.

Además, es esencial incorporar los principales indicadores identificados por AZCUBA en la IDE y avanzar en la identificación de formatos de datos generados por los computadores de a bordo de las máquinas. Esto facilitará el desarrollo de nuevas aplicaciones que utilicen o conviertan dichos formatos, enriqueciendo la toma de decisiones en el sector agrícola.

Con estos avances, se logra un manejo más efectivo de los datos, permitiendo comparaciones simultáneas y análisis de mapas de rendimiento, con desplazamientos temporales y personalizaciones en los estilos de visualización por parte de todos los usuarios competentes. Es necesario implementar un programa de capacitación para el personal involucrado en el uso de la plataforma, asegurando que cuenten con las habilidades necesarias para maximizar el potencial de las herramientas disponibles.

3.4 Contribución a una estrategia de desarrollo sostenible

El uso de la plataforma desarrollada y de las tecnologías que la componen, no solo mejora la gestión de datos agrícolas, sino que también se alinea con una estrategia de desarrollo sostenible para la agroindustria cubana. A continuación, se detallan las formas en que los resultados obtenidos contribuyen a esta estrategia:

- Optimización de Recursos

La plataforma desarrollada permite un uso más eficiente de los recursos agrícolas, como agua, fertilizantes y pesticidas. Al proporcionar datos precisos sobre las condiciones del cultivo y el rendimiento de las

cosechadoras, los agricultores pueden aplicar insumos de manera más específica y ajustada a las necesidades reales del cultivo. Esto no solo reduce costos operativos, sino que también minimiza el impacto ambiental asociado al uso excesivo de productos químicos.

- Mejora de la Toma de Decisiones

La integración de datos en la IDE-AZCUBA no solo facilita el acceso a información crítica para la toma de decisiones, sino que también permite a los agricultores analizar tendencias históricas y patrones, anticipando problemas y adaptando sus prácticas agrícolas. Al combinar tecnologías avanzadas con datos precisos, se promueve una gestión más eficiente de los recursos, lo que resulta en un aumento significativo en la productividad y sostenibilidad del sector. Esta sinergia entre tecnología y gestión es fundamental para aumentar la resiliencia del sector ante cambios climáticos y otros desafíos, permitiendo a los agricultores tomar decisiones más informadas y específicas en un entorno agrícola en constante cambio.

- Fomento de la Colaboración

La plataforma promueve un enfoque colaborativo entre los diferentes actores del sector agrícola, incluyendo productores, investigadores y entidades gubernamentales. Al compartir datos y experiencias, se pueden desarrollar estrategias conjuntas que aborden problemas comunes y optimicen la producción a nivel territorial. Esta colaboración es clave para construir un sistema agroindustrial más cohesionado y eficiente.

- Sostenibilidad Económica y Social

Al mejorar la eficiencia productiva y reducir costos, se contribuye a la sostenibilidad económica del sector agrícola. Además, al fomentar prácticas agrícolas más responsables y sostenibles, se promueve un entorno social más saludable para las comunidades rurales. La capacitación en el uso de estas tecnologías también genera nuevas oportunidades laborales y fortalece las capacidades locales.

- Innovación Continua

La adopción de tecnologías digitales en la agricultura cubana representa un paso significativo hacia la innovación continua en el sector. La investigación y el desarrollo en agro-informática no solo mejoran las prácticas actuales, sino que también abren la puerta a nuevas soluciones tecnológicas que pueden ser desarrolladas en el futuro. La plataforma implementada es altamente adaptable y está diseñada para crecer, permitiendo la integración de herramientas avanzadas como la IA para optimizar aún más los procesos agrícolas. Esta capacidad de evolución asegura que los agricultores

puedan beneficiarse de análisis predictivos y automatización, mejorando la eficiencia y sostenibilidad de sus operaciones.

Además, la cultura del trabajo en las IDEs juega un papel crucial en este proceso de innovación. La integración de información proveniente de diversas fuentes permite una visión holística del entorno agrícola, facilitando la creación de políticas informadas que respondan a las necesidades del sector. Al fomentar el intercambio de datos y la colaboración entre diferentes actores, se establece un marco propicio para el desarrollo de estrategias integrales que aborden los desafíos actuales. Así, la innovación se convierte en un motor clave para transformar el sector agropecuario, garantizando su competitividad y resiliencia ante los desafíos del entorno, mientras se promueve una gestión más eficiente y sostenible de los recursos agrícolas.

4. CONCLUSIONES

Este trabajo propone una estrategia integral para la gestión de datos en la recolección de caña de azúcar en entornos IDE. El flujo de gestión abarca desde la ingestión de datos en un clúster de *Big Data* hasta su publicación como servicio web, garantizando la creación de una cadena de valor agregado sobre los datos geoespaciales primarios. Además, facilita la integración de estos datos en la IDE-AZCUBA, promoviendo un enfoque colaborativo en la gestión de información.

La mejora en la transparencia de los datos sienta las bases para futuras investigaciones y desarrollos en el ámbito de la agroindustria, destacando el potencial transformador de las tecnologías de *Big Data* en el sector. La investigación subraya la importancia de la evaluación continua y adaptación de estas tecnologías para enfrentar los desafíos actuales, asegurando que la información esté disponible para todos los usuarios que la requieran en sus actividades laborales e investigativas. En resumen, este trabajo establece un camino hacia una agroindustria más eficiente y colaborativa, capaz de responder a las demandas del futuro mediante:

- Una estrategia integral de gestión de datos en entornos IDE.
- La creación de una cadena de valor agregado sobre datos geoespaciales primarios.
- La promoción de un enfoque colaborativo en la gestión de información a través de la IDE-AZCUBA.
- El destaque del potencial transformador de las tecnologías de Big Data en la agroindustria.
- La importancia de la evaluación continua y adaptación de estas tecnologías para enfrentar desafíos actuales.

- La disponibilidad de información para todos los usuarios en actividades laborales e investigativas.

5. REFERENCIAS

1. Delgado Mora, I., Aday Díaz, O., Lugo Ruiz, I., Martínez, R., Cruz Iglesias, R., & Hernández. (2023). VISUALIZACIÓN Y PERFECCIONAMIENTO DE LAS RECOMENDACIONES DE LOS SERVICIOS CIENTÍFICOS TÉCNICOS EN CAÑA DE AZÚCAR. X CONVENCIÓN INTERNACIONAL DE AGRIMENSURA 2023. La Habana.
2. https://www.researchgate.net/publication/374288434_VISUALIZACION_Y_PERFECCIONAMIENTO_DE_LAS_RECOMENDACIONES_DE_LOS_SERVICIOS_CIENTIFICOS_TECNICOS_EN_CANA_DE_AZUCAR
3. Delgado-Mora, I., Aday-Díaz, O., Lugo-Ruiz, I., Martínez, R., Orozco-Morales, & Hernández-Santana, R. (2023). LA AGRICULTURA DE PRECISIÓN EN LA DETECCIÓN DE ÁREAS INFECTADAS CON ROYA EN VILLA CLARA. X CONVENCIÓN INTERNACIONAL DE AGRIMENSURA 2023, La Habana.
4. Deoliveira, J. (2008). GeoServer: Uniting the GeoWeb and spatial data infrastructures. 25-29.
5. Fernández, J. L. C., & Iglesias, R. C. (2020). Tecnologías de Big Data geoespacial en el Centro de Información Geoespacial de Geocuba. *Revista Cubana de Transformación Digital*, 1(2), Article 2.
6. Gil, E. (2003). SITUACIÓN ACTUAL Y POSIBILIDADES DE LA AGRICULTURA DE PRECISION. Universidad Politécnica de Catalunya.
7. Gradaille, L. A. D. (2017). Indicadores técnicos y de explotación de las cosechadoras de caña de azúcar CASE-IH 7000 y 8000 en la provincia de Ciego de Ávila. *Revista Ingeniería Agrícola*, 4(3), Article 3.
8. Hughes, J., Annex, A., Eichelberger, C., Fox, A., Hulbert, A., & Ronquest, M. (2015). GeoMesa: A distributed architecture for spatio-temporal fusion. 94730F. <https://doi.org/10.1117/12.2177233> IT Admin. (2012). SMS Basic/Advanced.
9. Muñoz-González, R., Ramírez-Cruz, Z. D., & Donéstevez-Sánchez, G. M. (2022, enero). Transformación digital de la agricultura en Cuba: Estado y perspectivas: Digital transformation of agriculture in Cuba: State and perspectives | Anuario Facultad de Ciencias Económicas y Empresariales. *Anuario Facultad de Ciencias Económicas y Empresariales*, 13. <https://anuarioeco.uo.edu.cu/index.php/aeco/article/view/5276>
10. Pérez García, C. A. (2016). Metodología para la extracción de datos de cosecha en cosechadoras cañeras Case IH A8000 [Tesis de Pregrado - Ingeniería Automática, Universidad Central «Marta Abreu» de Las Villas. Facultad de Ingeniería Eléctrica. Departamento de Automática y Sistemas Computacionales]. <https://dspace.uclv.edu.cu/handle/123456789/6343>
11. Pérez-García, C. A., Rodríguez-Conte, A., Hernández-Santana, L., Rodríguez-Orozco, M.

- A., Cruz-Iglesias, R., Capote-Fernández, J. L., Pérez-García, C. A., Rodríguez-Conte, A., Hernández-Santana, L., Rodríguez-Orozco, M. A., Cruz-Iglesias, R., & Capote-Fernández, J. L. (2022). Intervención en las cosechadoras Case IH A8000 de ingenio azucarero en Villa Clara, Cuba. *Revista Ciencias Técnicas Agropecuarias*, 31(1). http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2071-00542022000100005&lng=es&nrm=iso&tlng=en
12. Sen, R., Farris, A., & Guerra, P. (2013). Benchmarking Apache Accumulo BigData Distributed Table Store Using Its Continuous Test Suite. 2013 IEEE International Congress on Big Data, 334-341. <https://doi.org/10.1109/BigData.Congress.2013.51>
 13. Shvachko, K., Kuang, H., Radia, S., & Chansler, R. (2010). The Hadoop Distributed File System. 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), 1-10. <https://doi.org/10.1109/MSST.2010.5496972>
 14. Soca-Muñoz, J. L., Rodríguez-Machado, E., Aday-Díaz, O., Hernández-Santana, L., Orozco-Morales, R., Soca-Muñoz, J. L., Rodríguez-Machado, E., Aday-Díaz, O., Hernández-Santana, L., & Orozco-Morales, R. (2020). Spectral signature of brown rust and orange rust in sugarcane. *Revista Facultad de Ingeniería Universidad de Antioquia*, 96, 920. <https://doi.org/10.17533/udea.redin.20191042>