Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), noviembre-diciembre 2024,
Volumen 8, Número 6.

https://doi.org/10.37811/cl_rcm.v8i6

ESTRATEGIA TECNOLÓGICA PARA ANÁLISIS DE DATOS EN LA TOMA DE DECISIONES BASADA EN BIG DATA UTILIZANDO EL ENTORNO DE HADOOP Y POWER BI

TECHNOLOGY STRATEGY FOR DATA ANALYTICS IN BIG DATA-BASED DECISION MAKING USING THE HADOOP ENVIRONMENT AND POWER BI

Jose Adan Silva Hernandez
Universidad Autónoma de Coahuila, México

Lydia Aldape Rivera
Tecnológico Nacional de México, México

DOI: https://doi.org/10.37811/cl_rcm.v8i6.15231

Estrategia Tecnológica para Análisis de Datos en la Toma de Decisiones Basada en Big Data utilizando el Entorno de Hadoop y Power BI

Jose Adan Silva Hernandez[1]
adansilva@uadec.edu.mx
https://orcid.org/0000-0002-7252-1927
Universidad Autonoma de Coahuila
Escuela de Sistemas Prof. Marcial Ruiz Vargas
Mexico

Lydia Aldape Rivera
laldape@cdacuna.tecnm.mx
https://orcid.org/0000-0001-9305-0311
Tecnologico Nacional de Mexico
ITS de Ciudad Acuña
Mexico

RESUMEN

En la siguiente investigación se considera el diseño y prueba de la estrategia técnica la cual ejerza una dinámica más eficiente dentro de las empresas con volúmenes altos de manejo de datos mediante técnicas de análisis tales como Big Data (S Rani, 2023), con sistemas auxiliares como Hadoop (APACHE, 2023), en combinación con Power BI (BI m. , 2023), para el manejo de datos relacionales existentes en Bases de Datos en ACCESS (msAccess, 2023). Con el propósito de sustentar la toma de decisiones dentro de las empresas. Con el soporte de sistemas "Pantalla de trabajo Apache Hadoop" en conjunto con sistemas auxiliares de análisis. El siguiente documento proporciona una investigación del diseño en estrategia tecnológica centrada en sistemas para análisis de datos que puedan soportar la toma de decisiones dentro de las industrias. Para el caso de estudio se ha utilizado una base de datos que muestra la información de la producción de motores para la industria automotriz, con 15 tablas representando cada uno aproximadamente más de 7000 registros de datos.

Palabras clave: big data, mineria de datos, *IoT*

[1] Autor principal
Correspondencia: adansilva@uadec.edu.mx

Technology Strategy for Data Analytics in Big Data-Based Decision Making Using the Hadoop Environment and Power BI

ABSTRACT

This project aims to design and test a technological strategy that enables competitiveness in companies through data analysis technologies such as Big Data (S Rani, 2023), through the software tool Hadoop (APACHE, 2023) with Power BI (BI m. , 2023), for the analysis of relational data resident in the database. of data from ACCESS (msAccess, 2023). This could provide support to the decision-making processes in companies. The Apache Hadoop Framework reference framework combined with analysis tools has been used for implementation Several software tools have been used for this project being the main Spark and Hadoop as a framework to support the subsequent analysis with other software such as Hive and R language. The activities of extraction, transformation and loading (ETL) of data have been considered for their subsequent visualization that supports the decision making of the companies. This document is the result of the research project of design and implementation of a technological strategy based on software for data analysis that supports the decision-making processes of companies and helps them to be more competitive. For the case study, a database has been used that shows the engine production information for the automotive industry.

Keywords: big data, data mining, *IoT*

Artículo recibido 02 noviembre 2024

Aceptado para publicación: 28 noviembre 2024

INTRODUCCIÓN

En la actualidad industrial en la frontera norte de Mexico, la estandarizacion de los productos produce una gran oportunidad para innovar en los procesos de produccion en cada industria con el proposito de lograr satisfacer a la demanda del mercado global, en este contexto es inportante generar soluciones en los servicos de fabricacion en cada proceso de manufactura y/o fabricacion de manera que tengan la flexibilidad en la adaptacion de nuevas estrategias en la mejora continua del producto, Debido a la alta cantidad de informacion que se genera en la linea de pruduccion de productos electronicos se vuelve complicado el diagnostico de los defectos presentados y por consecuente se complica la solucion para reducir las cantidades de rechazos generados en el turno, con el aumneto de su demanda este proceso duplico su produccion y a su vez duplico la cantida de informacion generada, resaltando que los datos generados son cotejados de forma manual y llevado su registro en fisico, Diferentes tipos de empresas y organizaciones medianas o grandes alrededor de nuestra localidad, sin importar su actividad económica y el capital que manejan, están expuestos día a día, a enfrentar diferentes retos que deben ser atendidos de manera acelerada para poder competir de forma exitosa. Un tema prioritario en la agenda económica de Coahuila viene siendo la innovación de las empresas, siendo la industria de manufactura un pilar económico de la Región Centro de Coahuila (RCC), por lo que las acciones que fortalezcan la innovación para incrementar tenderán a incrementar la competitividad en este sector. Los reportes del Instituto Nacional de Estadística, Geografía e Informática (INEGI) colocan al Estado en el 4o lugar nacional con mayor crecimiento del PIB durante 2017, con un incremento del 5 por ciento, cuando el promedio nacional fue de 2 por ciento., por lo que son prioritarias las soluciones que tiendan a incrementar su productividad y competitividad. (Coahuila, 2019). Como lo menciona Trujillo, 2021, “los desafíos focalizados dentro de la industria automotriz en México han aumentado a lo largo de los años, teniendo como principal desafío la implementación en su totalidad de la industria 4.0 en la mayor parte de sus procesos o bien en su totalidad, si bien, la implementación que se ha logrado en la actualidad es importante, ya que se han optimizado tareas como el ensamble, pintura, soldadura, fabricación, manipulación y moldura, las cuales han beneficiado a la industria, ya sea una implementación total o parcial de equipo de maquinaria autónomo”. (Trujillo, 2021).

Los pilares que forman esta revolución industrial los presenta Habid, 2022 de la UANL, y propone en su investigación que son la robótica, la simulación, la inteligencia artificial y la manufactura aditiva, de igual forma concibe en que se debe reforzar y adecuar la educación en temas como la realidad aumentada, la ciberseguridad y la nube, así como la incorporación de estudios y análisis en el tratamiento d datos con sistemas como Big Data, y IoT. (Habid, 2022). Mukherjee 2022 desarrollo su investigacion en base a la conectividad de dispositivos sensoriales conectados al internet de los cuales menciona que son utilizados en la industria para recabar grandes volumenes de informacion en infraestructuras y ciudades inteligentes impactando grandemente en la toma de desiciones, desarrolla la aplicación de algoritmos para el analisis de Big Data en tecnologia 5G-Driven IoT los cuales detallan la taxonomia de sistemas analiticos con el proposito de entender los comportamientos de los datos al ser transferidos en procesos de stream y resalta la importancia de los protocolos de seguridad y privacidad. (Mukherjee, 2022). Como lo menciona Garces-Giraldo 2022 el Big Data puede ser una herramienta de apoyo en su empleacion transversal en las empresas gracias a su gran variedad de aplicaciones que puede desarrollar (Garces-giraldo, 2022). Zhong comenta que los metodos de analisis de Big Data y mineria de datos son usados en la mejora de los sistemas del IoT y transformar estos datos en conocimiento invaluable (Zhong, 2022). Dentro del contexto de la optimizacion del proceso bajo el amumento de demanda del producto espesifico se considero la implementacion con herramental propio de optimizar el proceso de produccion iniciando por la identificacion y reduccion de los altos volumens de rechazos y scrap que se generaban durante los turnos diurno y nocturno, de igual forma mejorar los procesos de ensamble en las estaciones de trabajo de la linea de produccion. Con la hipotesis de mejorar el proceso mediante recursos propios y conocicimentos del personal tecnico se plantea el reducir un veinte por ciento los rechazos y retrabajos en la produccion.

METODOLOGÍA

La siguiente investigacion es del tipo aplicada con el proposito de dar solucion a la problemática actual, explicando la determinacion de las causas que probocan el alto indice de rechazos y sus consecuencias ante la alta demanda del producto, utilizando el metodo experimental en etapas de recoleccion analisis y tratamiento de datos, siendo el analisis del tipo cuantitativo midiendo los efectos en los cambios aplicados en el fenomeno experimentando los cambios efectuados al proceso, siguiendo un metodo

inductivo con el proposito de obtener conclusiones optimas que afecten positivamente a las variables presentes en la operatividad del proceso, utilizando el anailisis de manera transversal centrandonos en la comparacion de los resultado aplicados en el mismo periodo de prueba para obtener datos enteros positivos a los esperados. El estudio es desarrollado en el area de productos electronicos de control de energia, durante el proceso de ensamble y prueba de rendimiento asi como en los momentos de aceptacion y reparacion.

Para la transicion de tecnologia analogica a I4.0 (Bellantuono, 2021), se utilizaron modulos digitales de medicion de Voltaje, corriente, velocidad (Schneider, 2022), para la recoleccion de datos se utilizaron interfaces ADC , en la transformacion de datos analogos se utilizaron los sistemas SNORT (Snort, 2023), HADOOP (APACHE, 2023), para el almacenamiento de los datos convertidos se utilizaron los sistemas de base de datos MSACCESS (msAccess, 2023), y para la interpretacion y analisis de los mismos se utiliza el sistema MSPOWER BI (BI m. , 2023) asi como las interfaces de lectura o terminales de operador (rockwellautomation, 2023) en las cuales por medio de colorimetria se identificara el estado del producto.

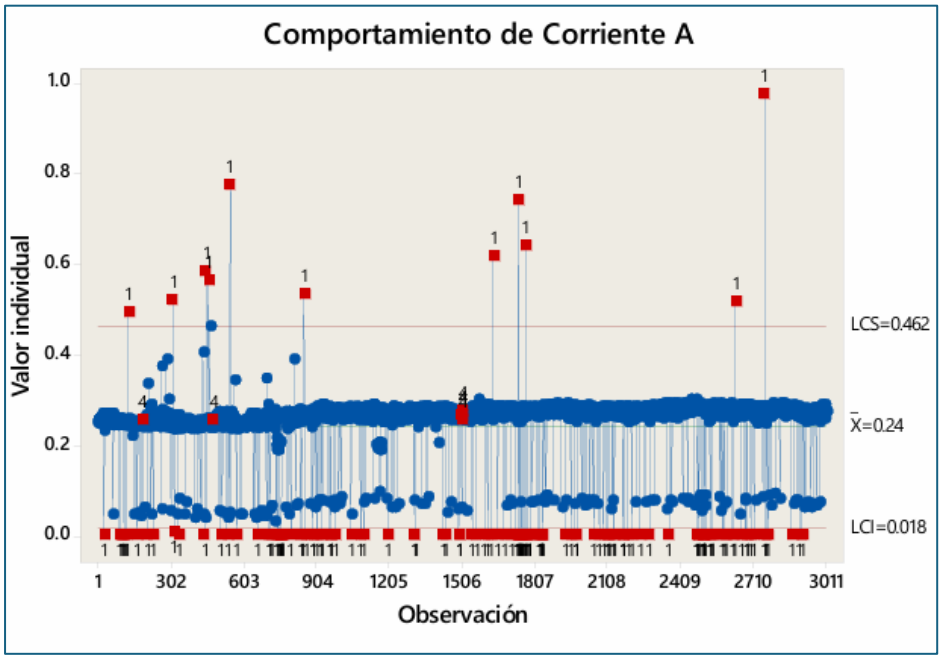
En la etapa final del proceso de ensamble de las tablillas encodificadoras se les somete a pruebas electricias para la verificacion de la operatividad del componente, en este momento el equipo aplica los elementos de voltaje, corriente a la tablilla la cual realiza sus operaciones y arroja los valores de potencia y revoluviones por minuto y son registrados en la base de datos, para luego ser procesados y se ponen a disposicion para su analisis estadistico, de igual forma conforme a la programación de la espesificacion del componente el equipo compara los valores obtenidos y emite un resultado el cual es mostrado a traves de colorimetria en la terminal de operador, quien visualiza el color verde para componentes aceptados o rojo para componentes con defectos.

Una ves que los datos entran a la base de datos son canalizados hacia el sistema de Power BI (BI m. , 2023), en el cual de manera grafica presenta estadisticas como paretos (MH, 2023) empleados para determinar la frecuencia de los numeros de parte mas solicitados, graficas de anillo (BI Z. , 2023), para verificar los estados de aceptacion y rechazo, y a su vez registros de corrimiento en tiempo real los cuales ayudan a visualizar el estado del producto en prueba.

RESULTADOS Y DISCUSIÓN

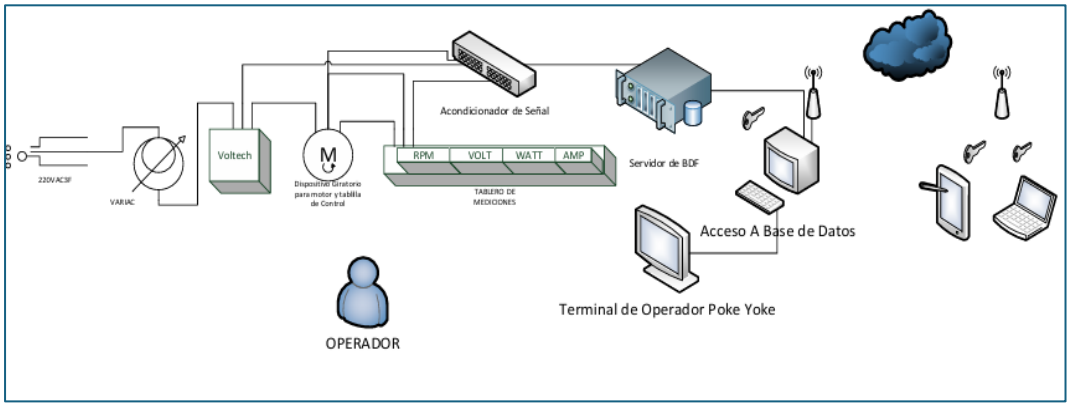
En el Gráfica 1 se muestra la alteración en los parámetros de corriente (A) respectivamente, El estadístico muestra las piezas por ID que presentan fallas en el parámetro de corriente lo cual nos dice que las piezas fallan en alcanzar su valor de corriente A en el tiempo ciclo, siendo que esta pieza se rechaza y se manda a inspección. Con el propósito de disminuir los índices de rechazo y retrabajo se procedió a desarrollar un modo de lectura más versátil y con el propósito de potencializar la trazabilidad del producto, se optó por facilitar el proceso de lectura que a su vez arroje información en tiempo real a los controles de producción y/o calidad.

Grafico 1. Comportamiento de la corriente por pieza



Se diseñó el diagrama y protocolos de la estación de trabajo conectada al nodo de transferencia de información con el propósito de mejorar la lectura quedando como en la Fig. 1.

Figura 1 Cambios en la estación de prueba física



Esta muestra los diferentes puntos que se optimizaron con el propósito de disminuir el índice de rechazo en las piezas fabricadas, a continuación, se describen los cambios y modificaciones: En la toma de lectura de parámetros medidos se instaló un módulo de entradas y salidas con el propósito de convertir estas señales a un protocolo digital, al obtener estas señales se instaló un servidor para el tratamiento conversión proceso y almacenamiento de información mediante la base de datos Access. De igual forma se instaló un procesador con periféricos para la manipulación y programación de lectura e interpretación de datos, a continuación, se instaló una terminal de operador la cual de manera visual (por Color) avisa al operador del resultado de la prueba y a su vez la comunica al sistema de trazabilidad del producto.

En cuanto al tiempo de proceso en la estación de trabajo se obtuvieron los resultados favorables los cuales se presentan en la siguiente comparación de tiempos en la tabla 1 y tabla 2.

Tabla 1. Tiempos de operación en estación de prueba

Tiempo (s)		
Ensamble	Ciclo	Comparación
30	20	40

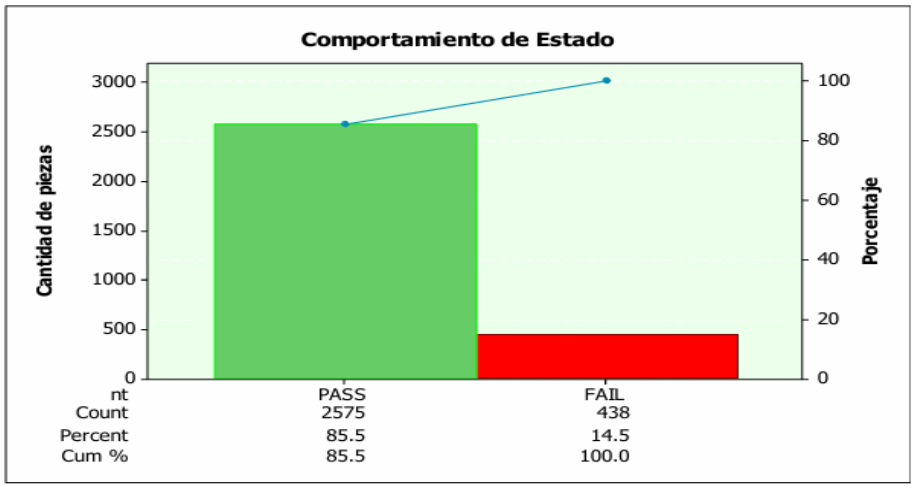
Tabla 2. Tiempos de operación en estación de prueba cambios

Tiempo (s)		
Ensamble	Ciclo	Comparación
30	20	15

Se observa que el tiempo de comparación se redujo en un treinta y ocho por ciento (de cuarenta segundos a quince segundos) logrando que la pieza sea pasada de manera instantánea a la estación de retrabajo el cual de igual forma tiene el parámetro fallido y se repara de manera más eficiente.

Con respecto al índice de rechazo en la estación, con base a los datos recabados se encontró la reducción a los catorce puntos cinco por ciento lo cual nos integra dentro de los parámetros permitidos por el área de calidad, como lo muestra el Gráfico 2.

Gráfico 2. Comportamiento de estado de partes



Con lo anterior, una vez expuesto el análisis y mejora que se ha implementado en la empresa del caso de estudio, derivado del análisis y procesamiento de grandes volúmenes de datos, se puede afirmar que las herramientas de big data para el procesamiento de datos son efectivas y se pueden implementar en las empresas de manufactura apoyados por una estrategia de big data (S Rani, 2023).

CONCLUSIONES

se ha propuesto el desarrollo de una estrategia tecnológica basada en el software Hadoop para el manejo de grandes volúmenes de datos y el posterior análisis en el software Power BI, aplicado a una empresa de manufactura, es importante en todo momento mantener el orden del desarrollo del proyecto para que se pueda lograr tener avances significativos, cada fase está diseñada para que se puedan desarrollar las herramientas necesarias y no divagar en la selección de ellas, el proyecto presenta oportunidades de mejora al momento de encontrar los tiempos de generación de reporte de producción con respecto a la producción física del componente, cada estadístico empleado fue alineando el proceso de la información para encontrar las variables que se presentaban fuera de control, o inestables con respecto a la especificación de cliente, cada fase con su herramienta logro formar la oportunidad para hacer el cambio de tecnología a una del tipo I4.0, y así poder ser competitiva con respecto al mercado mundial, actualmente se monitorean las variables críticas en tiempo real con los respectivos accesos mediante las comunicaciones por wifi, hacia los dispositivos inteligentes asignados.

REFERENCIAS BIBLIOGRAFICAS

APACHE, H. (05 de 06 de 2023). *https://hadoop.apache.org*. Obtenido de https://hadoop.apache.org/: https://hadoop.apache.org/

Bellantuono, N. N. (2021). Digital transformation models for the I4.0 Transition: Lessons from the change management literature. Sustainability. *MDPI*, 13(23).

BI, m. (05 de 06 de 2023). *https://www.microsoft.com/*. Obtenido de https://www.microsoft.com/: https://www.microsoft.com/en-us/power-platform/products/power-bi/downloads

BI, Z. (06 de 05 de 2023). *zebrabi.com*. Obtenido de zebrabi.com: https://zebrabi.com/guide/how-to-customize-ring-chart-in-power-bi/#:~:text=Donut%20charts%2C%20often%20referred%20to%20as%20ring%20charts%2C,the%20proportion%20of%20data%20elements%20in%20a%20chart.

Coahuila, G. d. (06 de 05 de 2019). *coahuila.gob.mx*. Obtenido de coahuila.gob.mx: https://coahuila.gob.mx/noticias/index/coahuila-es-cuarto-lugar-nacional-en-incremento-en-el-pib-11-12-18

Garces-giraldo. (2022). Uso de Big Data en contexto empresarial como . *Revista Ibérica de Sistemas e Tecnologias de Informação*, 570-585.

Habid, M. (2022). Presencia de los pilares de la industria 4.0 en la formación de ingenieros en el noreste de México. *www.scielo.org*, 9.

MH, M. (05 de 06 de 2023). *www.powerbi.tips*. Obtenido de www.powerbi.tips: https://powerbi.tips/2016/10/pareto-charting/

msAccess. (06 de 05 de 2023). *https://www.microsoft.com/es-mx/microsoft-365/access*. Obtenido de https://www.microsoft.com/es-mx/microsoft-365/access: https://www.microsoft.com/es-mx/microsoft-365/access?ef_id=_k_3d245aa4918f1544acc63dbeed742e3a_k_&OCID=AIDcmmz9xdherf_SEM__k_3d245aa4918f1544acc63dbeed742e3a_k_&msclkid=3d245aa4918f1544acc63dbeed742e3a

Mukherjee, S. (30 de 07 de 2022). *https://onlinelibrary.wiley.com*. Obtenido de https://onlinelibrary.wiley.com: https://onlinelibrary.wiley.com/doi/abs/10.1002/ett.4618

rockwellautomation. (06 de 05 de 2023). *literature.rockwellautomation.com*. Obtenido de literature.rockwellautomation.com: https://literature.rockwellautomation.com/idc/groups/literature/documents/um/2711-um014_-en-p.pdf

S Rani, P. B. (2023). *Big Data, Cloud Computing and IoT: Tools and Applications.* india: CRC Press, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-27242.

Schneider. (05 de 06 de 2022). PowerLogic PM5000 series. *Basic multi-function meters*, págs. 1-9.

Snort. (05 de 06 de 2023). *https://www.snort.org*. Obtenido de https://www.snort.org: https://www.snort.org/downloads

Trujillo, C. (2021). RETOS Y DESAFÍOS DE LA INDUSTRIA AUTOMOTRIZ EN MÉXICO. UNA VISIÓN DESDE LA INDUSTRIA 4.0, 2021. *Repositorio Institucional UNAM*, 113-129.

Zhong, Y. (2022). A systematic survey of data mining and big data analysis in internet of things. *springer link*, 1.