

Mejoramiento del proceso de control, monitoreo y trazabilidad de datos de una celda de manufactura flexible, mediante el uso de base de datos orientada a grafos.

Isabella Ceballos Sánchez.

Juan José Restrepo Rosero.

Director:

Juan David Contreras Pérez.

Co-Directora:

Dr. María Constanza Pabón Burbano.



Pontificia Universidad Javeriana de Cali

Facultad de Ingeniería y Ciencias

Programa de Ingeniería Electrónica

Anteproyecto de Grado

2023 - 1

Santiago de Cali, 18 de febrero de 2023.

Señores,
Pontificia Universidad Javeriana - Cali
Dr. Luis Eduardo Tobón Llanos
Director de Carrera de Ingeniería Electrónica
Cali.

Cordial Saludo,

Por medio de la presente nos permitimos informarles que los estudiantes de Ingeniería Electrónica, Isabella Ceballos Sánchez (cod: 8946329) y Juan José Restrepo Rosero (cod: 8939280), trabajan bajo nuestra dirección y co-dirección en el proyecto de grado titulado “Mejoramiento del proceso de control, monitoreo y trazabilidad de datos de una celda de manufactura flexible, mediante el uso de base de datos orientada a grafos”

Atentamente,

Juan David Contreras Pérez.

Dr. María Constanza Pabón Burbano.

Santiago de Cali, 18 de febrero de 2023.

Señores,

Pontificia Universidad Javeriana - Cali

Dr. Luis Eduardo Tobón Llanos

Director de Carrera de Ingeniería Electrónica
Cali.

Cordial Saludo.

Nos permitimos presentar a su consideración el proyecto de grado titulado “Mejoramiento del proceso de control, monitoreo y trazabilidad de datos de una celda de manufactura flexible, mediante el uso de base de datos orientada a grafos” con el fin de cumplir con los requisitos exigidos por la Universidad y optar al título Profesional en Ingeniería Electrónica.

Al firmar aquí, damos fe que entendemos y conocemos las directrices para la presentación de trabajos de grado de la Facultad de Ingeniería y Ciencias aprobadas el 26 de Noviembre de 2009, donde se establecen los plazos y normas para el desarrollo del anteproyecto y del trabajo de grado.


Atentamente,

Isabella Ceballos Sánchez
Código: 8946329

Juan José Restrepo Rosero
Código: 8939280



Resumen

Actualmente existe un interés en las empresas e industrias del país, por desarrollar métodos de optimización, control y manejo de datos, para predecir y obtener un mejor resultado en los procesos. Sin embargo, entre los desarrollos aplicados, existe una problemática en cuanto a la implementación de una base de datos robusta y sencilla que permita una fácil interpretación y buenos resultados en su escalabilidad, manejo de datos complejos y tiempos de consulta. 

Una celda de manufactura flexible, junto con los sistemas de ejecución de manufactura (MES), están orientados hacia la optimización de operaciones de un área de trabajo, con el objetivo de asegurar la ejecución efectiva y optimización del rendimiento de las operaciones de fabricación de la planta de producción, y mejorar la calidad del producto final. Estos sistemas, funcionan a partir del control y análisis de la información recopilada. Por ende, una base de datos se convierte en uno de los pilares del control de un proceso, es decir, el constante flujo de datos obtenidos debe de ser regulado y analizado de una correcta forma, para así lograr añadir diferentes tecnologías como lo son los gemelos digitales y obtener una representación digital de lo que es el proceso. Por lo tanto, es de vital importancia conocer las diferentes bases de datos que se han desarrollado actualmente, y evaluar sus diferentes desempeños en cuanto al tiempo de consulta y escalabilidad.

Así pues, en este trabajo se aborda una investigación en torno al mejoramiento de los procesos de control y trazabilidad de datos de una celda de manufactura flexible, mediante el uso de base de datos orientada a grafos.

Dado esto, en el primer capítulo de este documento se detallan los antecedentes, el planteamiento y justificación del problema. En el segundo capítulo, los objetivos que contribuyen a solucionar este problema. Como tercera parte, se mencionan los alcances y limitaciones. En el cuarto capítulo, se detallan los marcos teóricos y conceptuales, que sirven de guía para solucionar los retos presentes en la investigación. En el quinto capítulo se detalla la metodología a seguir para el desarrollo del proyecto. Por último, se establecen los recursos humanos y técnicos necesarios.

Palabras Claves: Base de datos grafos, Celdas de manufactura, Neo-4j, Cypher, OPC UA, Gemelos Digitales, Sistemas de ejecución de manufactura.

Abstract

Currently there is an interest in companies and industries in the country, to develop methods of optimization, control and data management, to predict and obtain a better result in the processes. However, among the applied developments, there is a problem regarding the implementation of a robust and simple database that allows easy interpretation and good results when scaling, complex data management and query times.

A flexible manufacturing cell,^S together with manufacturing execution systems (MES), are oriented towards optimizing operations of a work area, with the objective of ensuring the effective execution and performance optimization of the manufacturing operations of the production plant, and improving the quality of the final product. These systems work from the control and analysis of the collected information. Therefore, a database becomes one of the pillars of the control of a process, i.e., the constant flow of data obtained must be regulated and analyzed in a correct way, in order to add different technologies such as digital twins and obtain a digital representation of ~~what is~~ the process. Therefore, it is of vital importance to know the different databases that have been currently developed, and to evaluate their different performances in terms of query time and scalability.

Thus, this work addresses an investigation on the improvement of the data control and traceability processes of a flexible manufacturing cell, through the use of a graph-oriented database.

Given this, the first chapter of this document details the background, the approach and justification of the problem. In the second chapter, the objectives that contribute to solve this problem. In the third part, the scope and limitations are mentioned. In the fourth chapter, the theoretical and conceptual frameworks are detailed, which serve as a guide to solve the challenges present in the research. The fifth chapter details the methodology to be followed for the development of the project. Finally, the necessary human and technical resources are established.

Keywords: Graph database, Manufacturing cells, Neo-4j, Cypher, OPC UA, Digital twins, Manufacturing execution systems.

Índice general

1. Introducción	9
1.1. Antecedentes	9
1.2. Definición del problema	11
1.3. Justificación	12
2. Objetivos	15
3. Alcances y Limitaciones	16
3.1. Alcances	16
3.2. Limitaciones	16
4. Marco Teórico	17
4.1. Glosario	17
4.1.1. Base de datos relacionales	17
4.1.2. Celda de manufactura	17
4.1.3. CP Factory	17
4.1.4. Neo4j	17
4.1.5. OPC UA (Arquitectura Unificada de Comunicaciones de Plataforma Abierta)	18
4.1.6. Trazabilidad	18
4.2. Marco Conceptual	18
4.2.1. Base de datos orientada a grafos	18
4.2.2. Lenguaje de consulta Cypher y OpenCypher	19
4.2.3. Gemelos digitales	23
4.2.4. MES (Manufacturing Execution System)	24
5. Metodología	30
5.1. Concebir	30
5.2. Diseñar	31
5.3. Implementar	31
5.4. Operar	32
6. Presupuesto	33

Índice de figuras

4.1. Base de datos de gráficos, ilustración de una red de nodos y conexiones en Neo4j	19
4.2. Arquitectura y desafíos bases SQL y NoSQL	19
4.3. Gráfico de datos de ejemplo que muestra datos de supervisión y citas para investigadores, estudiantes y publicaciones	21
4.4. Consulta en Cypher	21
4.5. Resultado de la línea 1	21
4.6. Resultado de la línea 2-4	22
4.7. Resultado de la línea 5	23
4.8. Resultado de la línea 6 y 7	23
4.9. Gemelo digital en la industria 4.0	24
4.10. Tecnologías de la información para la industria	25
4.11. IBM arquitectura MES	25

Índice de cuadros

4.1. Ventajas del uso de gemelos digitales	24
6.1. Precio estimado de los recursos humanos del proyecto	33
6.2. Precio estimado de los equipos a emplear	34
6.3. Precio estimado de los software a emplear	34
6.4. Rubro final	34

Capítulo 1

Introducción

1.1. Antecedentes

Las crecientes demandas personalizadas de los clientes en Industria 4.0, plantean un gran desafío para las empresas manufactureras en términos de flexibilidad y capacidad de respuesta. Hoy en día, se han propuesto diversos enfoques de programación dinámica efectivos para que los sistemas de fabricación respondan rápidamente a los cambios en las demandas de los usuarios, sin embargo, sigue siendo un desafío el cual requiere de una gran cantidad de trabajo y tiempo para ajustar un esquema de programación que pueda responder a las necesidades de los procesos de manufactura actuales.

Como respuesta a esta problemática, en (1) se propone una arquitectura de control adaptativo de dos niveles habilitada por un método de programación automática e integrada en una celda de fabricación de gemelos digitales (DTMC), teniendo como objetivo **mapear automáticamente** un esquema de programación de tareas de entrada con un lote de trabajos en un grupo de programas de control a través de una red de modelos de comportamiento y un conjunto de modelos de eventos integrados en DTMC.

Por otra parte, el desarrollo de gemelos digitales (DT) es ampliamente utilizado en varios sectores industriales, para la optimización y mantenimiento de activos físicos, sistemas y procesos de fabricación. En (2) aplican el concepto de DT al analizar los efectos de una red de control de un robot, en donde se concluye, que a partir de configuraciones de bajo retardo, la simulación proporciona resultados muy precisos. Del mismo modo, en (3) se presenta el desarrollo y la implementación de un DT bidireccional en un laboratorio de industria 4.0, en donde la solución resultante se basa en el establecimiento de conexión OPC UA entre dos robots y además, utiliza herramientas como NX Siemens, sistema SCADA para la comunicación entre ellos. Esto permitió que el DT replicara el sistema físico y fuera impulsado por entradas de la celda de ensamblaje.

Es así como, se logra evidenciar que un DT es un concepto innovador que se está volviendo cada vez más relevante en la industria 4.0. En (4) se proporciona un análisis profundo de los campos de trabajo de DT, enfocándose principalmente en las características de la formación del modelo multifísico y los detalles de traba-

jar con grandes conjuntos de datos heterogéneos. Por otro lado, en (5) se describe una solución estructurada a partir de DT, algoritmos de aprendizaje automático e internet de las cosas (IoT), con el fin de detectar la presencia de anomalías en el funcionamiento de los sistemas industriales. Como conclusión, en este trabajo se obtuvo que dos de los tres algoritmos de aprendizaje automático son lo suficientemente efectivos para predecir anomalías, lo que sugiere su implementación para mejorar la seguridad de los empleados que trabajan en plantas industriales.

Para finalizar con los DT, en (6) se presenta un sistema ciberfísico de fabricación impulsado por gemelos digitales (MCPS) para el control paralelo de talleres inteligentes bajo el paradigma de individualización masiva. Los autores concluyeron que es posible establecer una conexión ciberfísica a través de modelos gemelos digitales descentralizados y formar varios recursos de fabricación como un sistema dinámico autónomo para crear productos personalizados. Adicionalmente, en (7) los autores proponen un sistema ciberfísico basado en DT, el cual puede predecir si es posible fabricar un producto con base al cronograma solicitado por un cliente en una línea de producción de carrocerías de automóviles, donde ocurren escenarios anormales. Como resultado, el sistema propuesto logró un desempeño de predicción promedio del 96.83 % para el plan de producción real.

Actualmente, en los procesos de celda de manufactura no sólo se evidencia la implementación de DT, si no también de las aplicaciones de procesamiento y almacenamiento de datos mediante un sistema de gestión de bases de datos (DBMS). En (8) se describen implementaciones comunes de DBMS relacionales y No-SQL, y se comparan sus capacidades mediante un ejemplo práctico, teniendo como resultado que las bases de datos relaciones tienen un rendimiento deficiente en cuanto a su escalabilidad y proceso de algunos tipos de datos. Del mismo modo, en (9) propone un método de almacenamiento de Big Data (BD) para redes de distribución inteligente de un sistema de energía basado en la base de datos de grafos Neo4j. Los resultados de la simulación en los sistemas de prueba estándar IEEE muestran que el método de almacenamiento de BD propuesto puede mejorar de manera efectiva el almacenamiento de datos y aumentar la eficiencia del análisis de la red de distribución inteligente.

En particular, las consultas sobre modelos de grafos han cobrado relevancia en los últimos años debido a su aplicación en áreas como el análisis de datos biológicos, las redes sociales y la web semántica. Por otra parte, no es nada raro encontrar hoy en día, que el ofrecer sistemas de información que cuenten con herramientas que faciliten el acceso y consultas de datos, es un gran reto que se sigue enfrentando desde diferentes perspectivas. En (10), la autora propone un lenguaje visual de consulta sobre un modelo de grafos, enfocado en el usuario final, que ofrece mayor expresividad que las herramientas de exploración de grafos sin trasladar a una notación visual los elementos de un lenguaje basado en texto. Como resultado de este trabajo, se obtuvo que el lenguaje facilita la formulación de consultas ad hoc (no conocidas con anticipación) que pueden ser complejas, mostrando en las pruebas que es una opción muy viable, teniendo en cuenta el tiempo de ejecución, comparada con otras bases.

1.2. Definición del problema

En los últimos años, con el aumento en la demanda de productos en el sector manufacturero y la aparición de la industria 4.0, el paradigma de la fabricación ha experimentado una serie de cambios holísticos, impulsando el uso de tecnologías como IoT, base de datos NoSQL, gemelos digitales y computación en la nube que involucra sistemas físicos cibernéticos (CPS). Sin embargo, la transformación y digitalización del sector manufacturero implica una transición del paradigma industrial actual, en donde el mayor de los problemas se centra en encontrar maneras de lograr la interoperabilidad entre el mundo físico y el mundo digital de los sistemas de fabricación para poder cubrir las demandas y mantener un ritmo de rápido crecimiento económico (11).

Actualmente, la implementación de un CPS trae consigo ciertas pautas y mejoras en la descentralización, interoperabilidad, virtualización, capacidad en tiempo real, orientación al servicio y modularidad. El concepto de virtualización, se puede definir como la capacidad del CPS para crear una copia virtual del sistema físico, proporcionando una conexión entre los sistemas reales y virtuales para recopilar datos, lo que impactará en la simulación del modelo. Una de las principales capacidades de esta copia digital es monitorear los procesos del mundo físico (12).

Un sistema de ejecución de manufactura flexible (MES), es un software que permite el control y monitoreo de procesos industriales, media la gestión de datos recopilados. Este sistema, se centra en obtener una ejecución efectiva y optimización en el proceso, mejorando la calidad del resultado. En el caso de estos sistemas de manufactura flexible, es común encontrar que, para manejar bajas cargas volumenes de datos, se implementan DBMS relacionales. Estas bases, aunque logran un buen procesamiento, almacenamiento y tiempo de respuesta de consulta, su nivel de complejidad aumenta, a medida que el sistema tiende a escalar (8). Por ende, surge la necesidad de expansión e instalación de una base de datos más robusta.

En adición, las celdas de manufactura se encuentran en constantes modificaciones, como, por ejemplo: la añadidura de un nuevo proceso, la eliminación o adición de una máquina, el replanteamiento de las funciones a desarrollar y entre otros. Esto con lleva, a que una base de datos relacionales que implementa consultas mediante el lenguaje SQL, al requerirse un cambio en el sistema de información, así sea mínimo, se debe ajustar todo el modelo en la base de datos y replantear las consultas, ya sea en la mayoría o en su totalidad. Esto hace que utilizar DBMS relacionales en un sistema altamente flexible y reconfigurable sea poco eficiente, provocando que nazca la necesidad de la búsqueda de otro tipo de DBMS que brinde flexibilidad y a la vez grandes resultados frente al manejo y cambios de datos.

Hoy en día, existen las bases No-SQL, que al comparar sus capacidades con las DBMS relacionales, han demostrado que se obtiene un mejor resultado en cuanto escalabilidad, constante cambios, procesamiento de datos y entendimiento de ella, adicionalmente, las bases de datos No-SQL también facilita escribir nuevas consultas sin vincular diferentes tablas. Por esta razón, se decidió investigar más a fondo este tipos de bases No-SQL, y se encontró con publicaciones como (9), (13),(14) , (10)

y entre otras donde se evidencia que la base de datos No-SQL que más se ajusta a las necesidades de la implementación de una celda de manufactura flexible, es la orientada a grafos. Esto, debido a que se centra en la teoría de grafos. Esto permite que, la representación de los datos y sus relaciones, el entendimiento, análisis y desarrollo de procesos, sea mucho más simple que otras bases.

La arquitectura de una base de datos orientada a grafos (GDB), debido a su estilo de red ~~neuronal~~, permite el fácil entendimiento para aquellas personas que no controlan o conocen el tema a profundidad, además que este tipo de base de datos permite representar la topología de los datos de una manera más flexible. En (15) los autores hacen un seguimiento de los sistemas de gestión de bases de datos de grafos (GDBMS) y exploran su aplicación en el dominio biomédico. Estos concluyen que, aunque los sistemas de administración de bases de datos relacionales (RDBMS) y otros motores NoSQL aún son útiles en cuestión de escalabilidad, realización de consultas y posibilidad de adición de nuevos datos, su rendimiento puede verse afectado en conjuntos de datos densamente conectados con relaciones de muchos a muchos. En estos casos, los GDBMS pueden ser beneficiosos al proporcionar un modelado más natural de las múltiples relaciones y ofrecer medios más intuitivos para la realización de consultas.

Teniendo en cuenta lo anterior, surge la siguiente pregunta: ¿Cómo se puede mejorar los procesos de control y trazabilidad de datos de una celda de manufactura flexible, mediante el uso de una base de datos que otorgue escalabilidad y robustez al proceso? En otras palabras, como respuesta a esta problemática, se propone un sistema de ejecución de manufactura para una celda de manufactura flexible mediante el uso de bases de datos orientada a grafos, de tal manera que nos permite mejorar los procesos de control y trazabilidad de datos, obteniendo así un mejor entendimiento de sus iteraciones y recopilación de datos, y, por otro lado, la facilidad de permitir escalar y realizar cambios al proceso de celda de manufactura flexible brindando una robustez al proceso.

1.3. Justificación

En los últimos años, se ha presentado la necesidad de generar cambios en los procesos manufactureros dado las crecientes demandas diversos enfoques. Esto provoca que se propongan diversas soluciones en cuanto a programación dinámica, monitoreo, control, flexibilidad y capacidad de respuesta del proceso.

La aplicación de un DT en el sector de manufactura afecta la forma en que se diseñan, fabrican y mantienen los productos. En un alto nivel, el DT puede evaluar las decisiones de producción, acceder al rendimiento del producto, comandar y reconfigurar máquinas de forma remota, manejar el equipo de resolución de problemas de forma remota y conectar sistemas/procesos para mejorar el monitoreo y optimizar su control [9]. Un contexto de esto, la empresa CNH Industrial implementó un DT para una línea de producción que combina la simulación con el análisis de los datos recopilados. Este DT fue el encargado de establecer una conexión ciber física a través de modelos de DT descentralizados con el propósito de formar múltiples re-

cursos de fabricación dinámicos y más eficientes en las operaciones de la empresa [10].

Por otra parte, el DT también se puede aplicar para control de procesos, monitoreo, mantenimiento predictivo, capacitación de operadores, desarrollo de productos, soporte de decisiones, análisis en tiempo real y simulación de comportamiento. Para realizar el control de procesos, el DT utiliza los datos históricos y en tiempo real para alimentar el sistema de soporte de decisiones virtual que ayudará al usuario a tomar decisiones estratégicas u operativas o implementar directamente el ajuste en la operación del sistema (11). Por ejemplo, en (16) los autores emplearon la base de datos Neo4j en donde a través de la sintaxis de Cypher generaron el grafo de conocimiento de la relación entre objetos, organización y carácter de datos generados y procesados en Python. De lo anterior, se concluyó que La herramienta de generación de gráficos de conocimiento de datos no estructurados puede resolver el problema de la relación lógica que no se puede resolver en la base de datos tradicional y también es posible visualizar los datos facilitando el análisis y encontrar intuitivamente la relación correspondiente entre los nodos.

Del mismo modo, la aplicación de un MES en la industria, es otra forma de controlar y monitorear mediante un software, la información recopilada ya sea mediante el DT, u otro intermediario, que permita tener la información en tiempo real y así documentar y gestionar los procesos de producción en entornos industriales. Dos aplicaciones en procesos de producción se evidencian en (17) y (18), en donde la primera los autores concluyen que gracias al diseño basado en DT, el sistema resulta ser flexible y puede ser fácilmente ampliable y extrapolable a otros dominios de aplicación en la automatización industrial. Por otro lado, en la segunda aplicación, el MES-DT integrado se utiliza para la toma de decisiones y gestionar estados de error, como los de montaje de baja calidad.

En este orden de ideas y teniendo en cuenta las mejoras de rendimiento y eficiencia de procesos de manufactura que traen consigo la implementación de los DT en la industria, se propone diseñar y aplicar una base de datos orientada en grafos que permita controlar la ejecución de una celda de manufactura flexible y así tener una trazabilidad de la información que se recolecta para llevar un registro del rendimiento de la máquina. Abordando el problema, se planea utilizar la base de datos Neo4j por medio de la sintaxis Cypher para generar el gráfico de conocimiento y observabilidad de los datos para posteriormente producir el reporte de desempeño.

En el ámbito económico, un sistema MES al actuar como un sistema de control y de monitoreo, que gestiona los procesos de producción, permite que se asegure la ejecución efectiva y máxima optimización ya sea en recursos o rendimiento de las operaciones, del mismo modo, obteniendo una mejora en la calidad del producto final. Esto en cuanto a resultados de costos, se verá reflejado en la disminución de materiales o productos. A su vez, al obtener un excelente producto de calidad, esto generará incrementos en los ingresos, llevando a los sistemas de planificación de recursos empresariales (ERP, por sus siglas en inglés), que tienen como foco al cliente, intervenir en las decisiones de negocio.

Esto permite obtener una alta estandarización, debido a la facilidad de ajusta-

miento o acople a cualquier modelo de información que presenta la base de datos orientada a grafos. Esto en términos de resultados, a partir del planteamiento de la celda de manufactura flexible basada en una base de grafos, se genera oportunidades en la analítica de datos debido a su similitud o su facilidad de plantearse como una red neuronal.

Dado esto, se pueden realizar análisis en cuanto a la planificación de la cadena de suministros de un sistema, es decir, el sistema podrá identificar una serie de patrones, y a partir de esto, podrá representar mejor el entorno y, por ende, tomar decisiones más óptimas. Además, el sistema a partir de la analítica de datos, puede predecir qué espera en el futuro, es decir, el sistema podrá averiguar qué patrones están surgiendo y qué riesgos podría estar enfrentando.

Principalmente, se espera que, al implementar una base de datos orientada a grafos en una celda manufacturera flexible, se obtenga inicialmente un mejor análisis y entendimiento en las relaciones del proceso, una mejor escalabilidad y manejo de datos, mejores resultados en cuanto a los cambios repentinos del proceso debido a que nos otorgan sencillez. Por otro lado, junto a la implementación de un DT y de un MES, ampliaremos el campo de beneficios, en cuanto al control y monitoreo en tiempo real del proceso industrial llevado a cabo, sin necesidad de presentar una celda física gracias a la digitalización del proceso. Además, con la implementación en conjunto, se busca asegurar de manera efectiva, la optimización del rendimiento, mejorar la calidad del proceso y la trazabilidad y análisis de la toma de datos.

Capítulo 2

Objetivos

Objetivo General

Desarrollar un sistema de ejecución de manufactura basado en base de datos orientadas a grafos para una celda de manufactura flexible.

Objetivos Específicos

- Definir los requerimientos asociados al control de la celda de manufactura flexible del CAP.
- Diseñar un modelo de **información** que permita el control de la celda de manufactura flexible a partir de una base de datos orientada a grafos.
- Evaluar el modelo propuesto para el control del sistema en un escenario de fabricación sobre pedido.
- Desarrollar una aplicación de visualización que permita presentar reportes e indicadores del sistema.



Capítulo 3

Alcances y Limitaciones

Este proyecto está enfocado en el diseño e implementación de una base de datos orientadas a grafos en el control de la ejecución de una celda de manufactura flexible, con el fin de mejorar el rendimiento operacional a través de la trazabilidad de los datos. Con base a lo anterior, se plantean los siguientes alcances y limitaciones:

3.1. Alcances

- El diseño del modelo se generará conforme a un entorno académico, es decir, se dará uso de herramientas de libre acceso.
- La implantación del modelo estará centrado en el funcionamiento dentro de una red local.
- Se evaluará el modelo propuesto mediante el funcionamiento en un escenario de fabricación sobre pedido en el entorno del Centro de Automatización de procesos CAP.
- El sistema contará en un futuro con la facilidad de aplicarse a procesos de control más robustos.

3.2. Limitaciones

- Se propondrá un modelo enfocado únicamente al proceso de celda de manufactura implementado en el CAP. Debido al difícil acceso a un proceso de manufactura en una empresa.
- No se pretende que el sistema cuenta con un alto nivel de seguridad de la información, dado que estamos limitados a un entorno académico y a uso de licencias gratuitas para el diseño.
- Inicialmente no se considera agregar nuevas capacidades de gestión de la información a los equipos de manufactura, se pretende trabajar con los sistemas actualmente implementados.

Capítulo 4

Marco Teórico

4.1. Glosario

4.1.1. Base de datos relacionales

Una base de datos relacional es aquella que almacena y proporciona acceso a puntos de datos relacionados entre sí, por medio de tablas de manera intuitiva y directa de representar datos en tablas. En una base de datos relacional, cada fila en una tabla es un registro con una ID única, llamada clave. Las columnas de la tabla contienen los atributos de los datos y cada registro suele tener un valor para cada atributo, lo que simplifica la creación de relaciones entre los puntos de datos (19).

4.1.2. Celda de manufactura

Las celdas de manufactura son un tipo de tecnología que está orientado hacia la optimización de operaciones de un área de trabajo, con el objetivo de eliminar de manera consistente y progresiva aquellas actividades que no poseen ningún valor agregado en el proceso de manufactura. Esta consiste en un almacén que se utiliza para alojar tanto piezas por procesar como piezas procesadas, un robot cartesiano que permite alcanzar las distintas localidades del almacén, un sistema de transporte que permite trasladar las piezas hacia el robot y de regreso hacia el almacén, y un manipulador robótico articulado que realiza un procesamiento (20) (21).

4.1.3. CP Factory

Conocido como los sistemas ciberfísicos, es un método de transición y representación a la fabricación digital, este incorpora todo tipo de tecnologías relevantes de la mecatrónica y la automatización. La implementación de ello, permite generar un modelo completo de fábrica industrial 4.0, reduciendo así diversos aspectos de la cadena de valor añadido (22).

4.1.4. Neo4j

Es una base de datos orientada a grafos, destinada a almacenar y consultar datos gráficos en vez de tablas como lo realizan las bases de datos relacionales. Es una de las bases de datos de gráficos de propiedades más populares que almacena gráficos de forma nativa en el disco y proporciona un marco para atravesar gráficos y ejecutar

operaciones gráficas. Por lo tanto, el lenguaje se usa en cientos de aplicaciones de producción en muchos dominios verticales de la industria. (23)

4.1.5. OPC UA (Arquitectura Unificada de Comunicaciones de Plataforma Abierta)

Es un protocolo de comunicación de datos para la comunicación en operaciones de automatización industrial (comunicación de máquina a máquina o de PC a máquina), que se caracteriza por ser extensible e independiente del fabricante o proveedor del sistema de la aplicación, el lenguaje de programación en el que se programó y el sistema operativo en el que funciona la aplicación (24).

4.1.6. Trazabilidad

La trazabilidad se entiende como la manera de encontrar y seguir el rastro de un proceso en cada una de sus etapas de ejecución e identificar las condiciones que lo rodean a lo largo de la cadena de logística y detectar el origen de una incidencia con facilidad (25).

4.2. Marco Conceptual

4.2.1. Base de datos orientada a grafos

Una base de datos orientada a grafos, se centra en la teoría de los grafos, donde existen un conjunto de objetos (vértices y aristas) que permiten la representación de datos interconectados. Estas bases de datos es parte de la familia de bases NoSQL, por ende, pueden usarse en contextos variados y con fines muy distintos, permiten analizar información y entender, evaluar y aprovechar los procesos y las relaciones (26).

Uno de los principales lenguajes implementados para el uso de GoDB es Cypher, un lenguaje de consulta declarativo que se basa en los conceptos básicos y cláusulas de las bases de datos SQL, pero con el plus de generar una funcionalidad adicional, facilitando la implementación de grafos. Este lenguaje es usualmente encontrado en el programa Neo4j.

En la imagen 4.1, se muestra un ejemplo de una base de datos orientada a grafos implementada en el programa de desarrollo Neo4j.

Actualmente las bases más implementadas son las relacionales (SQL), estas se basan en modelos relaciones, en donde se implementa una forma intuitiva y directa de representar los datos en tablas. Pero SQL se enfrenta a diferentes desafíos en cuanto su escalabilidad y rendimiento en sus consultas, por ejemplo, las consultas SQL largas requieren más tiempo para ejecutarse y es muy probable que incluyan errores de codificación humanos debido a su complejidad. Por otro lado, las bases orientadas a grafos nos otorgan sencillez y resultados en tiempo real, lo que hace más sencillo la interpretación y la generación de consultas.

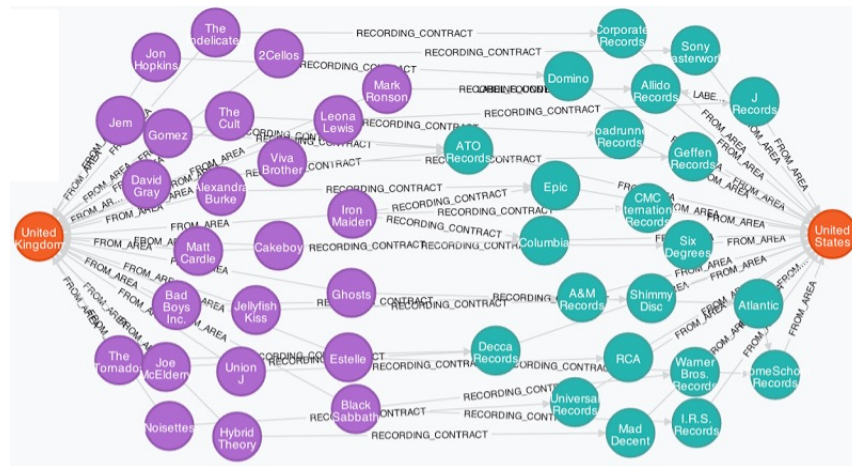


Figura 4.1: Base de datos de gráficos, ilustración de una red de nodos y conexiones en Neo4j

En la imagen 4.2, se muestra una comparación frente a la arquitectura y ventajas de las bases SQL y NoSQL.

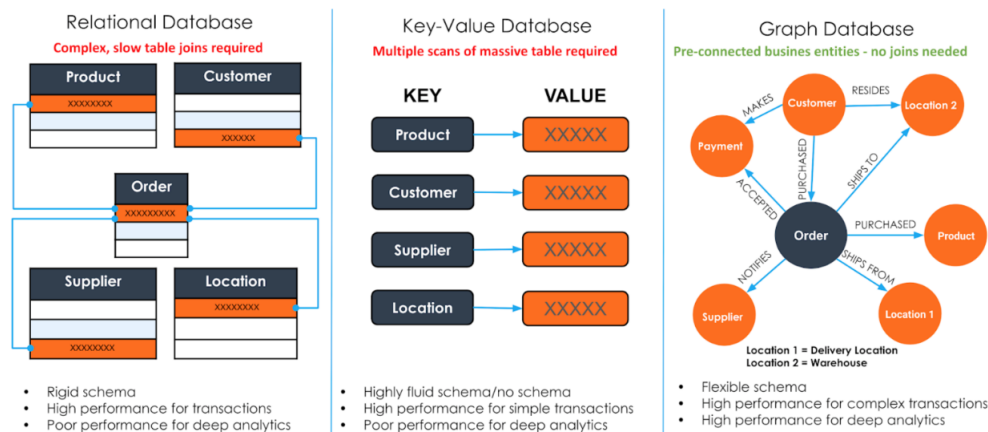


Figura 4.2: Arquitectura y desafíos bases SQL y NoSQL

4.2.2. Lenguaje de consulta Cypher y OpenCypher

El lenguaje de consulta de gráficos de propiedades Cypher es un lenguaje en evolución, originalmente diseñado e implementado como parte de la base de datos de gráficos Neo4j, y actualmente es utilizado por varios productos de bases de datos comerciales e investigadores.(23)

El modelo de datos de Neo4j que utiliza Cypher es el de gráficos de propiedades. El modelo comprende nodos, que representan entidades (como personas, cuentas bancarias, departamentos, etc.), y relaciones (sinónimo de bordes), representando las conexiones o relaciones entre las entidades.(23)

- **Consultas lineales:** Una consulta Cypher toma como entrada un gráfico de propiedades y genera una tabla. Esto permite a los usuarios pensar en el procesamiento de consultas como si comenzara desde el principio del texto de la consulta y luego progresara linealmente hasta el final. Cada cláusula en una consulta es una función que toma una tabla y genera una tabla que puede expandir el número de campos y agregar nuevas tuplas. Toda la consulta es entonces la composición de estas funciones.

El flujo lineal de consultas en Cypher se extiende a la composición de consultas. El **WITH** cláusula permite las mismas proyecciones que **RETURN**, incluidas las agregaciones. Además de esta forma lineal de componer consultas, Cypher también admite subconsultas anidadas como **UNION**.(23)

- **Modificación de datos:** Cypher admite un lenguaje de actualización enriquecido para modificar el gráfico. Las cláusulas de actualización reutilizan el lenguaje de patrón gráfico visual y proporcionan el mismo modelo semántico simple y de arriba hacia abajo que el resto de Cypher. Las cláusulas básicas para las actualizaciones incluyen **CREATE** para crear nuevos nodos y relaciones, **DELETE** para eliminar entidades, y **SET** para actualizar propiedades. Además, Cypher proporciona una cláusula llamada **MERGE** que intenta hacer coincidir el patrón dado y crea el patrón si no se encuentra ninguna coincidencia.

Una implementación de Cypher puede usar primitivas de sincronización de bases de datos, como el bloqueo, para garantizar que los patrones coincidan con **MERGE** son únicos dentro de la base de datos. (23)

- **Pragmático:** Cypher es intencionalmente similar a SQL para ayudar a los usuarios a realizar la transición entre los dos idiomas. Sigue la misma estructura de sintaxis de cláusulas e implementa la semántica establecida para muchas funciones. Cypher tiene soporte incorporado para los parámetros de consulta, lo que facilita la eliminación de los problemas de inyección de consultas. La sintaxis para agrupar y agregar es simple, y el lenguaje de expresión incluye potentes características como división de listas y comprensión de listas, subconsultas existenciales y trabajo con rutas.(23)
- **La coincidencia de patrones:** Los patrones en Cypher se expresan de forma visual como ".arte ASCII", como (a)-[r]->(b).(23)
- **Implementación de Neo4j:** La ejecución de consultas en Neo4j sigue un modelo convencional, descrito por Volcano Optimizer Generator. La planificación de consultas en Neo4j se basa en el algoritmo IDP, usando un modelo de costos. La compilación final de la consulta utiliza un modelo de ejecución simple basado en un iterador tupla a la vez, o compila la consulta en código de bytes de Java con un modelo de ejecución basado en push.(23)

A continuación, en la figura 4.3 se muestra un gráfico de datos G formado por investigadores, estudiantes y publicaciones. Para cada investigador, mostramos los

estudiantes que supervisan y las publicaciones que han escrito, y para cada publicación, mostramos qué otras publicaciones cita.(23)

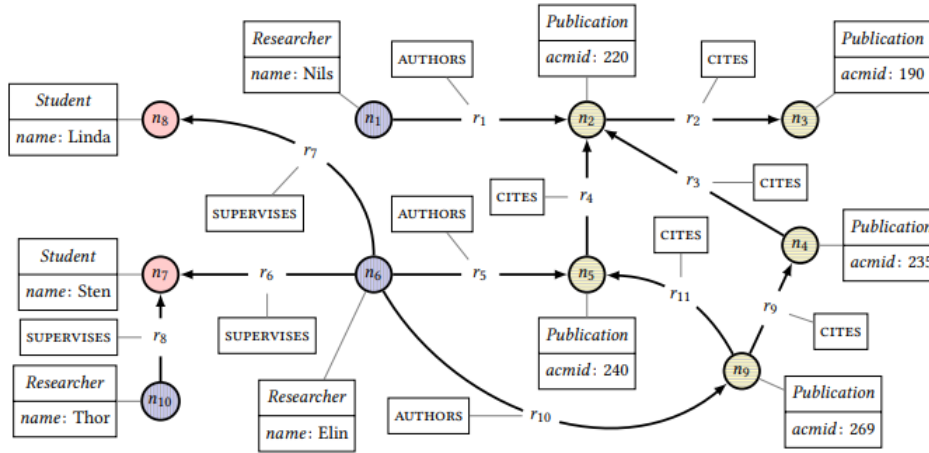


Figura 4.3: Gráfico de datos de ejemplo que muestra datos de supervisión y citas para investigadores, estudiantes y publicaciones

La siguiente consulta de Cypher devuelve el nombre de cada investigador en G, la cantidad de estudiantes que supervisan actualmente y la cantidad de veces que una publicación de la que son autores ha sido citada, tanto directa como indirectamente, por otras publicaciones.

```

1 MATCH (r:Researcher)
2 OPTIONAL MATCH (r)-[:SUPERVISES]->(s:Student)
3 WITH r, count(s) AS studentsSupervised
4 MATCH (r)-[:AUTHORS]->(p1:Publication)
5 OPTIONAL MATCH (p1)-[:CITES*]->(p2:Publication)
6 RETURN r.name, studentsSupervised,
7        count(DISTINCT p2) AS citedCount

```

Figura 4.4: Consulta en Cypher

r	s	r	studentsSupervised
n1	null	n1	0
n6	n7	n6	2
n6	n8	n10	1
n10	n7		

(a)

(b)

Figura 4.5: Resultado de la línea 1

El patrón dado en la primera cláusula **MATCH** de la línea 1 coincide con todos los investigadores; es decir, los nodos con la etiqueta Investigador. Esto produce tres vínculos para la variable r, a saber, n1, n6 y n10 representados como tres filas en una tabla con un único atributo r, que corresponden a los investigadores Nils, Elin y Thor.

La cláusula **MATCH** tiene una variante opcional: **OPTIONAL MATCH**, que es análoga a la construcción outer join de SQL. Esta cláusula produce filas para todas las coincidencias del mismo modo que **MATCH**, siempre que se encuentre

el patrón completo en el gráfico de datos. Sin embargo, en los casos en que no se encuentre ningún dato que coincida con todo el patrón, se producirá una única fila en la que los enlaces de todas las variables introducidas en **OPTIONAL MATCH** se establecerán en null.

La cláusula **OPTIONAL MATCH** de la línea 2 empareja a todos los estudiantes supervisados por cada investigador. Esto produce un enlace de la variable recién introducida *s* para cada valor al que *r* estaba enlazado por la cláusula **MATCH** de la línea 1. Cuando *r* está vinculada a *n1* (Nils, que no supervisa a ningún estudiante), la vinculación correspondiente para *s* es nula.

La cláusula **WITH** de la línea 3 se utiliza tanto para proyectar un subconjunto de las variables actualmente en el ámbito y sus ligaduras a la parte de la consulta que sigue a **WITH**, como para calcular una agregación. La cláusula **WITH** tiene dos expresiones, la segunda de las cuales es una agregación que funciona de forma muy similar a SQL. La primera expresión, *r*, es una expresión no agregadora y, por lo tanto, actúa como una clave de agrupación implícita para la función agregadora `count(s)`. Tomando los resultados mostrados en la tabla anterior, contamos todos los valores no nulos de *s* para cada enlace único de *r*, con el alias `estudiantesSupervisado`.

La cláusula **WITH** proyectará todas las vinculaciones producidas para *r* y `estudiantesSupervisado`. Observamos que la variable *s* ya no está en el ámbito después de la línea 3, ya que no fue proyectada por **WITH**, y ya no puede utilizarse en el resto de la consulta. Esta tabla actúa ahora como tabla impulsora de la cláusula **MATCH** de la línea 4. Las variables de enlace producidas por la línea 4 incluyen a todos los investigadores que son autores de al menos una publicación, el número de estudiantes que supervisan y la publicación de la que son autores:

<i>r</i>	<code>studentsSupervised</code>	<i>pl</i>
<i>n1</i>	0	<i>n2</i>
<i>n6</i>	2	<i>n5</i>
<i>n6</i>	2	<i>n9</i>

Figura 4.6: Resultado de la línea 2-4

La cláusula **OPTIONAL MATCH** de la línea 5 empareja, para cada publicación de la que es autor uno de los investigadores de *G*, todas las publicaciones que lo citan, tanto directa como indirectamente. Esto se consigue mediante el uso de un patrón que contiene la relación de longitud variable `CITES`, que indica que deben recorrerse una o varias relaciones `CITES`. Los resultados obtenidos son:

r	studentsSupervised	p1	p2
n_1	0	n_2	n_4
n_1	0	n_2	n_9 †
n_1	0	n_2	n_5
n_1	0	n_2	n_9 †
n_6	2	n_5	n_9
n_6	2	n_9	null

Figura 4.7: Resultado de la línea 5

Cuando p1 está ligado a n9, que no es citado por ninguna publicación, la ligadura correspondiente para p2 es nula. Además, observamos que hay dos filas idénticas, indicadas con †. La existencia de estas filas duplicadas se debe al patrón de relación de longitud variable: n9 es alcanzable desde n2 a través de los nodos intermedios n5 y n4.

RETURN es la última cláusula de la consulta (líneas 6 y 7) y, al igual que la cláusula **WITH** de la línea 3, calcula y proyecta expresiones. Se proyecta el valor de la propiedad name de cada investigador, junto con el valor de studentsSupervised. La expresión de agregación, a diferencia de la de la línea 3, cuenta los valores distintos de p2 (excluyendo los valores nulos) y aliasa los resultados como citedCount, que denota el número de publicaciones que citan una publicación cuyo autor es el investigador. La tabla formada por las filas que contienen el resultado de las expresiones se proyecta mediante **RETURN**:

r.name	studentsSupervised	citedCount
Nils	0	3
Elin	2	1

Figura 4.8: Resultado de la línea 6 y 7

4.2.3. Gemelos digitales

El concepto de gemelo digital (DT) aparece en 1991 en una publicación de David Gelernter. Sin embargo, en 2002 se dió la primera representación y desarrollo de un DT por el Dr. Michael Grieves. Aunque esta aplicación estuvo centrada en el área aeroespacial, hoy en día, su mayor campo de acción se encuentra en la industria.

Un DT se define como una representación virtual de un objeto o sistema que abarca su ciclo de vida. Esta representación se actualiza a partir de datos en tiempo real y utiliza simulación, machine learning y razonamiento para apoyar en la toma de decisiones sobre el sistema. Por ejemplo, en la figura 4.9 se visualiza la representación digital de un robot que incluye modelos físicos, aplicaciones de machine learning y resultados de simulaciones, entre otras.

En el cuadro 4.1 se puede evidenciar las ventajas y beneficios de la aplicación de los DT a ambientes industriales:

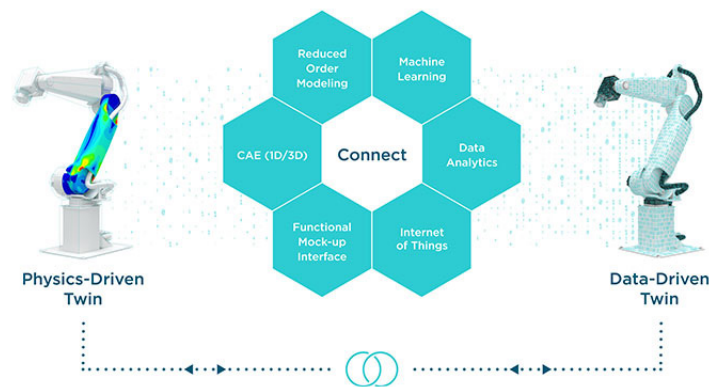


Figura 4.9: Gemelo digital en la industria 4.0

Característica	Ventaja
Mejor Investigación y Desarrollo	Investigación y diseño de un producto más eficaz, permitiendo el refinamiento de los productos a partir de los datos obtenidos.
Mayor eficiencia	Permiten ayudar a reflejar y monitorear los sistemas de producción, manteniendo la máxima eficiencia durante la fabricación.
Fin de vida del producto	Se determinan qué materiales de los productos se pueden recolectar a través del reciclaje u otras medidas.

Cuadro 4.1: Ventajas del uso de gemelos digitales

4.2.4. MES (Manufacturing Execution System)

El Manufacturing Execution System (MES), es un software que actúa como un sistema de control y de monitoreo de la información para gestionar procesos de producción en entornos industriales, documentando la gestión de la planta. Estos sistemas tienen como objetivo asegurar la ejecución efectiva y optimización del rendimiento de las operaciones de fabricación de la planta de producción, y mejorar la calidad del producto final.

El sistema MES requiere de sistemas de control de supervisión y adquisición de datos (sistema SCADA), con el fin de gestionar y controlar los equipos implementados. Del mismo modo, es fundamental integrar soluciones ERP (planificación de recursos empresariales) para transferir información sobre el rendimiento de la producción, el consumo de materiales y entre otros.

A continuación, en la figura 4.10 se muestran las tecnologías de la información para la implementación en una industria, en donde se aprecia el orden y conexión de los sistemas SCADA, MES y ERP.



Figura 4.10: Tecnologías de la información para la industria

Del mismo modo, en la figura 4.11 se presenta la arquitectura que presenta un sistema MES según IBM. Esta arquitectura de referencia representa la generalización de experiencias prácticas de cientos de compromisos en torno a la digitalización de la fabricación por parte de IBM, Red Hat y socios de todo el mundo. Basado en los principios de la nube híbrida, combina prácticas y lecciones para la operar las plantas en la nube (27).

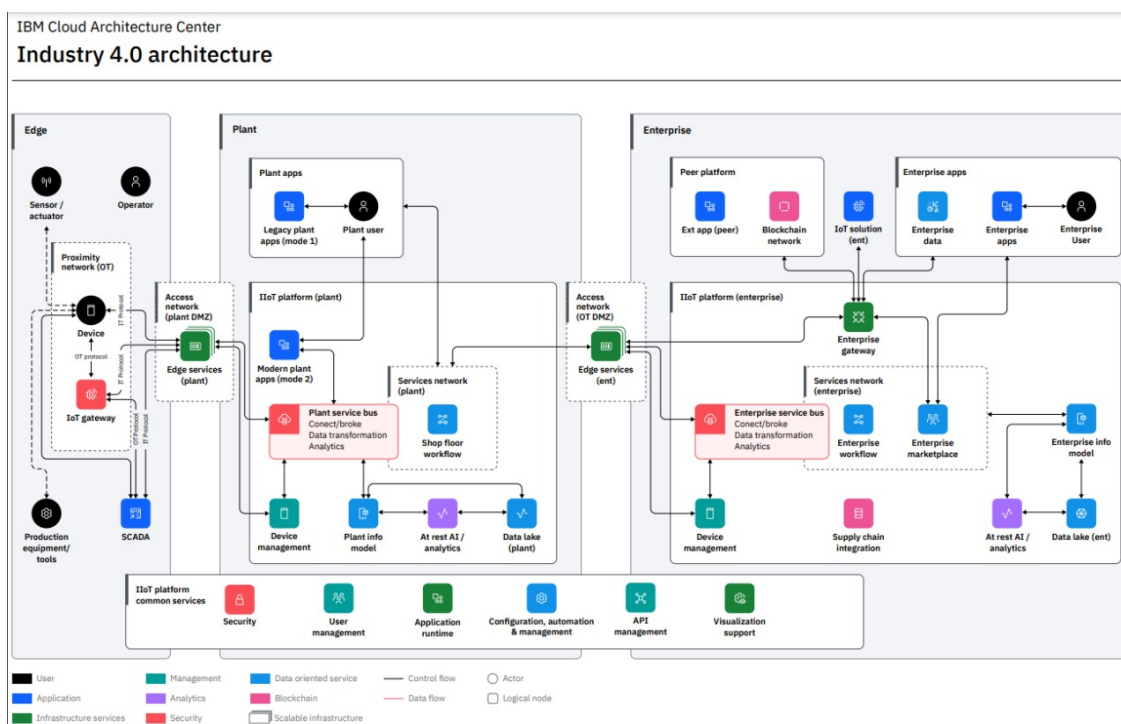


Figura 4.11: IBM arquitectura MES

La arquitectura de referencia distribuida de 3 capas es la representación continua de la experiencia de IBM en la aplicación de tecnología y la ejecución de proyectos de TI en entornos de fabricación con estas características clave:

1. Abierta
2. Basado en estándares

3. Modular
4. Reemplazable
5. Modular
6. Específico de la planta
7. Independiente del proveedor

Con base en esta propuesta de arquitectura, es posible identificar los componentes y sus funciones requeridas y necesarias para la digitalización de un entorno de fabricación. Algunos de estos componentes son:

- **Dispositivo de control:** El dispositivo o controlador lógico programable (PLC) es la representación y control de una entidad física o sensor que proporciona información del dispositivo, la cual es accesible mediante el uso de protocolos industriales. La integración de red para el dispositivo es parte de la parte cibernética del dispositivo. El PLC puede emplear protocolos de tecnología operativa (OT, por sus siglas en inglés) como MODBUS/TCP, S7 y ProfiBUS/ProfiNET.
- **IoT gateway:** Es el responsable de conectar los sistemas de control de supervisión y adquisición de datos, las puertas de enlace de sensores y los dispositivos a la plataforma IoT. El gateway, recibe datos del dispositivo mediante el uso de protocolos industriales y del dispositivo, formatea los datos según sea necesario y utiliza un protocolo de TI para comunicarse con la plataforma. Por lo general, implementa el patrón Adapter y convierte los protocolos de fabricación heredados, como ProfiNet/ProfiBus, MODBUS, Open Platform Communications (OPC DA), entre otros.
- **Gateway de Planta o Bus de Servicio de Planta:** Implementa el patrón de bus de servicio dentro del taller al proporcionar funciones de enrutamiento, mediación y transformación a nivel de fábrica o planta. El bus de servicio asume la lógica de integración de la aplicación, las máquinas y los dispositivos de la planta al reducir la complejidad de la conectividad punto a punto. Este es complementario a un sistema de ejecución de fabricación (MES), lo que permite la separación de preocupaciones entre la implementación del proceso de fabricación.
- **Gateway Empresarial o Bus de Servicio Empresarial:** Proporciona funciones de bus de servicio clásicas para la integración desacoplada de sistemas y máquinas, asumiendo la lógica de integración, los sistemas y las máquinas. Además, la puerta de enlace empresarial o bus de servicio empresarial proporciona funciones de bus de servicio clásicas para la integración desacoplada de sistemas y máquinas, asumiendo la lógica de integración, los sistemas y las máquinas.
- **Operador:** Un operador generalmente completa tareas como la operación del equipo, la configuración y las tareas de apoyo a las operaciones. El operador interactúa con dispositivos o interfaces de usuario específicas de operaciones en sistemas que se integran con la solución de IoT mediante el uso de datos de IoT y operaciones recopilados.

- **Sensor o actuador:** sensores que proporcionen medidas de variables significativas como a ubicación, la temperatura o la humedad. Los actuadores, como las válvulas, también están representados por esta entidad.
- **Herramientas de equipo de producción:** Estas herramientas son objetos o máquinas que son de interés y están sujetas a mediciones de sensores y comportamiento real. Estas herramientas pueden ser un robot, un contenedor, una máquina, un horno o un AGV.
- **SCADA:** El control de supervisión y la adquisición de datos (SCADA) es una arquitectura de sistema de control que utiliza computadoras, comunicaciones de datos en red e interfaces gráficas de usuario para la gestión de supervisión de procesos de alto nivel. En esta arquitectura, SCADA representa diferentes cosas en diferentes dominios. Por ejemplo, en el dominio de gestión de edificios, SCADA representa un sistema de gestión de edificios o un sistema de control de bajo nivel que controla los procesos en el edificio, como un refrigerante
- **Modelo de información de la planta:** El modelo de información IoT es el componente lógico responsable de los diferentes modelos de información que se emplean en la plataforma, incluidas las jerarquías de gemelos digitales.

Por otra parte, existen ciertos requisitos no funcionales que las arquitecturas de Industria 4.0 deben satisfacer, como los siguientes:

Rendimiento

- **Función de descarga al borde:** Se requiere una configuración centralizada con descarga al borde para la toma de decisiones basada en reglas y el análisis de transmisión.
- **Diseño modular:** Una orientación a microservicios permite el diseño modular de soluciones IoT.
- **Diseño modular:** Una orientación a microservicios permite el diseño modular de soluciones IoT.
- **Datos no estructurados y estructurados** El sistema debe manejar datos estructurados, incluidos mensajes en tiempo real y transferencias masivas, y datos no estructurados, como archivos acústicos, archivos de imágenes, grabaciones y documentos de texto.
- **Tiempo de respuesta del usuario:** El sistema debe proporcionar tiempos de respuesta aceptables a los usuarios, independientemente del volumen de datos que se almacenen y los análisis que se realicen en segundo plano.
- **Comunicaciones en tiempo real:** Se necesita soporte para comunicaciones bidireccionales en tiempo real. Este requisito está relacionado con el requisito de soporte de protocolos industriales o de dispositivos en el perímetro. Es posible que las respuestas deban ocurrir en menos de 1 segundo según el caso de uso. Este requisito también impulsa la necesidad de análisis de nivel de borde y toma de decisiones para esa clase de casos de uso.

- **Granularidad de captura de series temporales:** Se requiere una marca de tiempo de al menos una milésima de segundo para que, incluso si los datos llegan demasiado tarde para el control en tiempo real, el análisis de actualización puede ver la secuencia de eventos.

Escalabilidad

- **Escalabilidad horizontal:** El sistema debe manejar las necesidades de carga y retención de datos en expansión en función de la ampliación del alcance de la solución, como la adición de instalaciones de fabricación y edificios.

Mantenibilidad

- **Adaptables y flexibles:** Debe ser posible adaptar rápidamente el sistema para cambiar el proceso, la información o los participantes que se intercambian.
- **Mantenibilidad y tiempo de actividad:** Debe ser posible completar el mantenimiento del sistema sin violar los acuerdos de nivel de servicio para el tiempo de actividad. En un entorno de 24 horas al día, 7 días a la semana, debe poder realizar el mantenimiento sin desmantelar el sistema por completo.

Disponibilidad

- **Alta disponibilidad:** Algunas soluciones y dominios de IoT exigen sistemas de alta disponibilidad para operaciones 24x7. Este tipo de sistema no es una aplicación de producción crítica, lo que significa que las operaciones o la producción no se detienen si la solución IoT se detiene. Debe confirmar la expectativa real para el sistema.

Seguridad

- **Seguridad del dispositivo:** Los dispositivos deben poder registrarse y comunicarse de forma segura, por ejemplo, mediante TLS. No se permiten dispositivos no autorizados.
- **Seguridad del usuario:** El inicio de sesión del usuario en cualquier dispositivo debe ser seguro y validado para su función. Este requisito está conectado al LDAP o registro de usuarios que utilizan los clientes.
- **Seguridad de la aplicación:** Los usuarios autorizados del sistema que intercambien información deberán poder hacerlo con los controles de seguridad adecuados.
- **Seguridad de datos:** Todos los datos persistentes requieren un acceso seguro.

Volumétrica

- **Grandes datos:** El sistema debe ser capaz de almacenar y analizar volúmenes de datos, tanto históricos como actuales, en escalas que comúnmente se conocen como Big Data.
- **Velocidad, capacidad y accesibilidad de la plataforma:** La plataforma debe soportar grandes volúmenes de transmisiones de datos 24x7. Este requisito incluye soporte para ubicaciones remotas y fuentes de datos móviles.

Manejabilidad

- **Administración de incidentes:** El sistema debe incluir soporte para alertas, notificaciones y gestión de incidentes.
- **Gestión de soluciones:** Debe existir soporte para la administración centralizada de soluciones para que el personal de soporte del sistema pueda determinar rápidamente la causa raíz de los problemas y solucionarlos para evitar el tiempo de inactividad.

Usabilidad

- **Soporte móvil:** Los usuarios deben poder interactuar en los mismos roles y en las mismas tareas en computadoras y dispositivos móviles, donde sea práctico, dadas las capacidades móviles.

Capítulo 5

Metodología

La metodología que se usará para llevar a cabo este trabajo de grado será la del modelo CDIO; una metodología usada en el campo de la ingeniería, la cual brinda las herramientas necesarias para enfrentar de manera innovadora y flexible problemas complejos de la sociedad (28). En este orden de ideas, el desarrollo de este trabajo se dividirá en cuatro etapas las cuales son: Concebir, Diseñar, Implementar y Operar. A continuación, se explicarán las tareas que se realizarán en cada una de las etapas anteriores, para poder cumplir con los objetivos específicos planteados en un principio.

5.1. Concebir

En esta primera etapa es donde se concibe la idea del problema a resolver, en donde esta se realiza un estudio bibliográfico acerca de los antecedentes y aplicaciones relacionadas con sistemas de ejecución de manufactura basados en bases de datos orientadas en grafos para una celda de manufactura flexible, para luego, poder definir los requerimientos asociados al control de la celda de manufactura flexible del CAP. Con base en lo anterior, se proponen las siguientes tareas para cumplir con la etapa de concepción:

- **Tarea 1:** Realizar un levantamiento bibliográfico sobre los temas de interés (Gemelo Digital, Base de Datos Orientada a Grafos, Sistemas Físicos Cibernéticos) y aplicaciones en el área industrial.
- **Tarea 2:** Realizar un levantamiento de requerimientos teniendo en cuenta los alcances y limitaciones del problema a resolver.
- **Tarea 3:** Comprender y analizar las ventajas y desventajas entre las bases de datos relacionales y las no relacionales.
- **Tarea 4:** Investigar sobre la documentación acerca de la herramienta Neo4j y el lenguaje de consulta Cypher, para la creación e interacción con la base de datos orientada a grafos.
- **Tarea 5:** Llevar a cabo una revisión de la base de datos ya existente en la celda de manufactura para determinar parámetros importantes a tener en cuenta para el diseño de la nueva base basada en grafos.

Finalizando esta etapa, se espera tener bien identificado y comprendido el problema, al igual que haber obtenido los conocimientos necesarios para continuar con la etapa de diseño.

5.2. Diseñar

En la segunda etapa, se pretende diseñar un modelo de información el cual permita controlar la celda de manufactura flexible a partir de una base de datos orientada a grafos. Aquí, se definen los criterios de diseño, alcances y limitaciones del modelo de información deseado.

- **Tarea 1:** Diseñar el modelo del sistema de información basado en grafos a partir de la herramienta Neo4j y del uso del lenguaje de consulta Cypher, que permitirá operar la celda de manufactura flexible existente en el CAP.
- **Tarea 2:** Identificar las variables de interés a monitorear y la frecuencia con la que se deberá entregar el reporte de desempeño de la celda de manufactura.
- **Tarea 3:** Definir alternativas de visualización de los indicadores a reportar. Revisar las opciones disponibles y escoger la mejor opción mediante una matriz de selección.
- **Tarea 4:** Validar el prototipo desarrollado con los expertos del CAP.

5.3. Implementar

La tercera etapa pretende evaluar el modelo propuesto para el control del sistema en un escenario de fabricación sobre pedido. Con base a los resultados de la evaluación, se hará una retroalimentación del rendimiento del modelo y la realización de ajustes y correcciones requeridas.

- **Tarea 1:** Implementar el modelo de información orientado a una base de datos grafos al escenario de fabricación sobre pedidos.
- **Tarea 1:** Implementar la visualización de acuerdo a los indicadores escogidos, mediante un Dashboard. Esto, con el fin de realizar un análisis y trazabilidad de datos, frente al resultado del escenario de fabricación sobre pedidos.
- **Tarea 1:** Realizar una evaluación de la respuesta del modelo de información a la hora de realizar las consultas requeridas por el usuario en el escenario de fabricación sobre pedidos.
- **Tarea 2:** Llevar a cabo los cambios necesarios basados en los resultados obtenidos y las retroalimentaciones de los expertos del CAP para cumplir con los objetivos planteados.

Finalizada esta etapa, se espera que la base de datos diseñada haya pasado por todas las pruebas requeridas y tenga la validación de los usuarios del CAP.

5.4. Operar

En la última etapa se busca que el modelo propuesto haya sido validado completamente y se encuentre en funcionamiento completo en el escenario de fabricación sobre pedidos.

Además, se evaluará el reporte de los indicadores mediante la aplicación de visualización desarrollada.

- **Tarea 1:** Evaluar el funcionamiento de la base de datos orientada a grafos, en cuanto a escalabilidad y facilidad de consulta.
- **Tarea 2:** Hacer una segunda valoración del funcionamiento y rendimiento del prototipo de base de datos, después de unas semanas y ver si los reportes de las consultas se llevan a cabo de acuerdo a como el usuario los espera.

Al culminar esta etapa se espera dar por terminado el proyecto de grado al tener la nueva base de datos instalada en la celda de manufactura flexible, con la cual se pueda someter al escenario de fabricación de pedidos y hacer las consultas en el momento requerido.

Capítulo 6

Presupuesto

Para cumplir con las etapas de concebir, diseñar, implementar y operar, es necesario adquirir y dar uso de algunas herramientas tales como bases de datos, componentes electrónicos, herramientas de laboratorio, software de simulación y entre otros.

La descripción de estos recursos se presenta a continuación, sin embargo, cabe aclarar que estos son componentes y servicios que son potenciales. Para saber con exactitud los costos y cantidades es necesario poner en marcha las etapas de diseño, implementación y operación.

- **Recursos Humanos:** Los recursos humanos del proyecto están conformados por Isabella Ceballos Sánchez y Juan José Restrepo Rosero quienes realizarán las labores de investigación, diseño e implementación. Además, se cuenta con la orientación del profesor y coordinador del CAP, Juan David Contreras Pérez.

A continuación, se muestra la descripción de costos de los implicados.

Nombre	Valor Hora	Número Horas Semana	Javeriana		Recursos propios		Total
			Especie	Efectivo	Especie	Efectivo	
Juan David Contreras	\$36,500	1	\$0	\$730,000	\$0	\$0	\$730,000
Isabella Ceballos	\$10,000	1	\$0	\$0	\$200,000	\$0	\$200,000
Juan José Restrepo	\$10,000	1	\$0	\$0	\$200,000	\$0	\$200,000
TOTAL			\$0	\$730,000	\$400,000	\$0	\$1,130,000

Cuadro 6.1: Precio estimado de los recursos humanos del proyecto

- **Equipos** Los equipos a utilizar a lo largo del desarrollo de este proyecto pertenecen principalmente a el CAP y a los investigadores. La descripción de estos recursos se presenta a continuación.

Nombre	Valor	Cantidad	Javeriana		Recursos propios		Total
			Especie	Efectivo	Especie	Efectivo	
Computador Asus TUF	\$6,400,000	1	\$0	\$0	\$6,400,000	\$0	\$6,400,000
Computador Asus X541n	\$3,000,000	1	\$0	\$0	\$3,000,000	\$0	\$3,000,000
Celda de manufactura flexible	\$200,000,000	1	\$200,000,000	\$0	\$0	\$0	\$200,000,000
TOTAL			\$200,000,000	\$0	\$9,400,000	\$0	\$209,400,000

Cuadro 6.2: Precio estimado de los equipos a emplear

- **Software** Los software a emplear para llevar a cabo este proyecto son de carácter gratuito, por lo que no se requieren de gastos adicionales desde la parte de programas. Los software son los siguientes.

Nombre	Valor	Cantidad	Javeriana		Recursos propios		Total
			Especie	Efectivo	Especie	Efectivo	
Python	\$0	1	\$0	\$0	\$0	\$0	\$0
Neo4j	\$0	1	\$0	\$0	\$0	\$0	\$0
TOTAL			\$0	\$0	\$0	\$0	\$0

Cuadro 6.3: Precio estimado de los software a emplear

De acuerdo a las tablas de recursos técnicos y humanos, se muestra a continuación los montos finales de cada categoría y el origen de los recursos.

Rubro	Javeriana		Recursos propios		Total
	Especie	Efectivo	Especie	Efectivo	
Recurso humano	\$0	\$730,000	\$400,000	\$0	\$1,130,000
Equipos	\$200,000,000	\$0	\$9,400,000	\$0	\$209,400,000
Software	\$0	\$0	\$0	\$0	\$0

Cuadro 6.4: Rubro final

Bibliografía

- [1] C. Zhang, G. Zhou, Y. Jing, R. Wang, and F. Chang, “A digital twin-based automatic programming method for adaptive control of manufacturing cells,” *IEEE Access*, vol. 10, pp. 80784–80793, 2022.
- [2] G. Szabó, S. Rácz, N. Reider, H. A. Munz, and J. Pető, “Digital twin: Network provisioning of mission critical communication in cyber physical production systems,” in *2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, pp. 37–43, 2019.
- [3] A. Protic, Z. Jin, R. Marian, K. Abd, D. Campbell, and J. Chahl, “Implementation of a bi-directional digital twin for industry 4 labs in academia: A solution based on opc ua,” in *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 979–983, 2020.
- [4] M. Brovkova, V. Molodtsov, and V. Bushuev, “Implementation specifics and application potential of digital twins of technological systems,” *International Journal of Advanced Manufacturing Technology*, vol. 117, no. 7/8, pp. 2279 – 2286, 2021.
- [5] G. P. Tancredi, G. Vignali, and E. Bottani, “Integration of digital twin, machine-learning and industry 4.0 tools for anomaly detection: An application to a food plant,” *Sensors (14248220)*, vol. 22, no. 11, p. 4143, 2022.
- [6] J. Leng, H. Zhang, D. Yan, Q. Liu, X. Chen, and D. Zhang, “Digital twin-driven manufacturing cyber-physical system for parallel controlling of smart workshop,” *Journal of ambient intelligence and humanized computing*, vol. 10, no. 3, pp. 1155–1166, 2019.
- [7] Y. H. Son, K. T. Park, D. Lee, S. W. Jeon, and S. Do Noh, “Digital twin-based cyber-physical system for automotive body production lines,” *International Journal of Advanced Manufacturing Technology*, vol. 115, no. 1/2, pp. 291 – 310, 2021.
- [8] M. M. Maran, N. A. Paniavin, and I. A. Poliushkin, “Alternative approaches to data storing and processing,” in *2020 V International Conference on Information Technologies in Engineering Education (Inforino)*, pp. 1–4, 2020.
- [9] P. Shi, G. Fan, S. Li, and D. Kou, “Big data storage technology for smart distribution grid based on neo4j graph database,” in *2021 IEEE 4th International Conference on Electronics Technology (ICET)*, pp. 441–445, 2021.

- [10] M. C. Pabón Burbano, “Lenguaje visual de consulta basado en transformación de grafos: aplicación en el dominio médico,” 2016.
- [11] M. Tisch, C. Hertle, J. Cachay, E. Abele, J. Metternich, and R. Tenberg, “A systematic approach on developing action-oriented, competency-based learning factories,” *Procedia CIRP*, vol. 7, pp. 580–585, 2013. Forty Sixth CIRP Conference on Manufacturing Systems 2013.
- [12] M. Hermann, T. Pentek, B. Otto, *et al.*, “Design principles for industrie 4.0 scenarios: a literature review,” *Technische Universität Dortmund, Dortmund*, vol. 45, 2015.
- [13] C. H. dos Santos, G. T. Gabriel, J. V. S. do Amaral, J. A. B. Montevechi, and J. A. de Queiroz, “Decision-making in a fast fashion company in the industry 4.0 era: a digital twin proposal to support operational planning,” *International Journal of Advanced Manufacturing Technology*, vol. 116, no. 5/6, pp. 1653 – 1666, 2021.
- [14] J. Qiao, A. Zhou, and H. Qiu, “Research and implemention of system of electric business expanding installation process based on graph database,” in *2020 IEEE 3rd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, pp. 10–13, 2020.
- [15] S. Timón-Reina, M. Rincón, and R. Martínez-Tomás, “Overview of graph databases and their applications in the biomedical domain,” *Database: The Journal of Biological Databases Curation*, vol. 2021, pp. 1 – 22, 2021.
- [16] N. Ayinuer, A. Ruxianguli, and A. Yasen, “Design and research of unstructured data knowledge graph toolbased on neo4j graph database,” in *2022 11th International Conference on Communications, Circuits and Systems (ICCCAS)*, pp. 296–300, 2022.
- [17] A. S. García, E. E. Estévez, I. R. Ruano, J. G. Ortega, and J. G. García, “Plataforma código abierto para generar gemelos digitales,” in *XLIII Jornadas de Automática: libro de actas: 7, 8 y 9 de septiembre de 2022, Logroño (La Rioja)*, pp. 941–948, Servizo de Publicacións, 2022.
- [18] E. Negri, S. Berardi, L. Fumagalli, and M. Macchi, “Mes-integrated digital twin frameworks,” *Journal of Manufacturing Systems*, vol. 56, pp. 58–71, 2020.
- [19] Oracle, “¿qué es una base de datos relacional?,” 2022.
- [20] “Guide to industrial control systems (ics) security,” May 2015.
- [21] F. d. Ingenieria, “Revista ingeniería investigación y tecnología: Facultad de ingeniería,” 2019.
- [22] FESTO, “Cp factory – the cyber-physical factory,” 2022.
- [23] N. Francis, A. Green, P. Guagliardo, L. Libkin, T. Lindaaker, V. Marsault, S. Plantikow, M. Rydberg, P. Selmer, and A. Taylor, “Cypher: An evolving query language for property graphs,” in *Proceedings of the 2018 International Conference on Management of Data, SIGMOD ’18, (New York, NY, USA)*, p. 1433–1445, Association for Computing Machinery, 2018.

- [24] Paessler, “Monitor your industrial environment using opc ua,” Jun 2021.
- [25] M. Mecalux, “Software de trazabilidad: Todos los productos bajo control,” Mar 2020.
- [26] IONOS, “Graph database: Bases de datos para una interconexión eficiente,” Oct 2019.
- [27] P. Kiradjiev, M. Carvalho, J. Benke, A. Kühmichel, J. Favilla, and S. Bonnaud, “Industry 4.0 architecture for manufacturing,” Jan 2020.
- [28] E. Amaya, “Metodologías de trabajo,” 2015.