

Universidad de Concepción
Departamento de Ingeniería Informática y Ciencias de la Computación

**PLATAFORMA DE ANÁLISIS DE INTERVENCIÓN HUMANA
EN DATOS HISTÓRICOS PARA LA PLANTA CELULOSA
ARAUCO VALDIVIA**

Patricio Ignacio Inostroza Araneda

Memoria presentada para la obtención del título de Ingeniero Civil Informático

Patrocinante: Marcela Varas C.
Comisión evaluadora: José Fuentes S. - Guillermo Cabrera V.

Septiembre de 2023

Agradecimientos

Me gustaría dedicar este espacio para expresar profundamente el agradecimiento a mi familia. Gracias por darme la oportunidad de estudiar y por proporcionarme todas las herramientas necesarias para crecer y progresar. Su confianza en mí ha sido mi mayor motivación, y cada logro que alcance en este viaje será un reflejo de su gran sacrificio y dedicación. Mi éxito es también su éxito, y este logro también les pertenece. Los amo con todo mi corazón.

A mis amigos de la carrera y futuros colegas, Ivonne, Leonardo, Felipe y Matías, les agradezco de todo corazón las infinitas risas, tardes de diversión y las largas noches de trabajo que compartimos durante estos años. Esta amistad ha sido una fuente importante de apoyo y alegría en mi periodo universitario. Juntos hemos superado diferentes desafíos, celebrado triunfos y creado recuerdos que serán inolvidables. Gracias por ser parte fundamental de esta etapa.

A mis amigas de Chillán, Anita, Sofi y Cami, con quienes he compartido desde la época del colegio hasta el día de hoy, y con quienes el contacto no se ha perdido, quiero expresar mi sincero agradecimiento. Han sido una parte integral de mi vida y un pilar fundamental en mi crecimiento. Nuestra amistad ha perdurado a lo largo de los años, y estoy agradecido por cada momento que hemos compartido juntos.

También quiero expresar mi agradecimiento a la empresa Arauco por brindarme la invaluable oportunidad de aplicar mis conocimientos durante estos meses y por confiar en mis capacidades. Agradezco profundamente la buena disposición y el tiempo que dedicaron para apoyarme en mi crecimiento profesional. Esta experiencia ha sido fundamental para mi desarrollo y aprendizaje en un entorno nuevo y desconocido.

Índice

1. Introducción	6
1.1 Antecedentes generales del problema	6
1.3 Objetivo general	8
1.4 Objetivos específicos	8
1.5 Metodología	8
1.6 Estructura del informe	9
2. Antecedentes	10
2.1 Estado del arte	10
2.2 Conceptos previos	13
3. Diseño de la Solución	16
3.1 Arquitectura propuesta	16
3.2 Extracción de información	16
3.3 Desarrollo y configuración de la base de datos	17
3.4 Carga y Transformación de información	20
3.5 Análisis y selección de datos	23
3.6 Identificación de KPI	27
3.7 Visualización de KPI e implementación	27
3.8 Proceso de carga web	33
3.9 Eliminación de datos	34
4. Resultados	35
4. Conclusión	38
5. Trabajo futuro	39
6. Bibliografía	40
7. Anexos	41
7.1 Historias de usuario	41

Índice de Figuras

Figura 1: Cuadrante mágico de Gartner 2023 para plataformas de análisis e inteligencia empresarial.	11
Figura 2: Arquitectura de sistemas.	16
Figura 3: Esquema relacional de los archivos MDF que emite Exploring Delta V.	17
Figura 4: Menú desplegable de base de datos en el servidor.	18
Figura 5: Selección de archivo origen para la base de datos.	19
Figura 6: Estructura de tabla "Journal".	19
Figura 7: Proceso para adjuntar base de datos.	20
Figura 8: Creación de usuario y permisos en la nueva base de datos.	21
Figura 9: Eliminación de datos duplicados en tabla Journal para la base de datos "testdb".	21
Figura 10: Asignación de clave primaria.	21
Figura 11: Inserción de datos a la tabla de áreas y módulos.	22
Figura 12: Inserción de datos a tabla Journal en DB004.	22
Figura 13: Separación de base de datos "testdb".	22
Figura 14: Eliminación de log y transferencia de archivo a otra carpeta.	23
Figura 15: Resultado de transformación de datos.	23
Figura 16: Esquema de relaciones Event_type - Category 1.	25
Figura 17: Esquema de relaciones Event_type - Category 2.	25
Figura 18: Menú de inicio de plataforma web.	28
Figura 19: Selección de área y fechas.	28
Figura 20: Resultados de consulta.	29
Figura 21: Detalle de módulo consultado.	30
Figura 22: Resultado de cambios.	31
Figura 23: Detalle de eventos de un usuario.	31
Figura 24: Resultados de consulta en dashboard.	32
Figura 25: Selección de archivo para cargar mediante plataforma.	33
Figura 26: Página de eliminación de información para bases de datos.	34
Figura 26: Línea temporal de cambios en periodo T3-2022 - T4-2022 - T1-2023 de izq. a der.	35
Figura 27: Módulos más frecuentes para los periodos T3-2022 - T4-2022 - T1-2023 de izq. a der.	36
Figura 28: Eventos y alarmas más frecuentes para los periodos T3-2022 - T4-2022 - T1-2023 de izq. a der.	36
Figura 29: Usuarios más frecuentes para los periodos T3-2022 - T4-2022 - T1-2023 de izq. a der.	37

Índice de Tablas

Tabla 1. Ventajas y desventajas en herramientas BI.	12
Tabla 2: Contenido filtrado para variables Category y Event_type.	26

1. Introducción

1.1 Antecedentes generales del problema

Arauco es una empresa de producción de celulosa y productos derivados de la madera, con presencia en varios países, entre ellos Chile, Argentina, Brasil, Uruguay, Estados Unidos y Europa. En Chile, la empresa cuenta con 6 plantas, incluyendo Licancel, Nueva Aldea, MAPA, Arauco, Constitución y Valdivia. Esta última es la única que produce celulosa de pulpa textil a partir del eucalipto, que se utiliza como materia prima para la fabricación de telas de viscosa o rayón para prendas de vestir y tapices [1].

La empresa tiene una larga trayectoria en la industria, partiendo en 1979 bajo la fusión de las compañías Celulosa Arauco S.A. y Celulosa Constitución S.A. [1], desde entonces se ha enfocado en la investigación, innovación y la sustentabilidad para mantener su posición de liderazgo en el mercado. Entre las iniciativas de sustentabilidad de Arauco se encuentra la protección de los bosques nativos, la promoción de la reutilización y el reciclaje de residuos, y la implementación de tecnologías más limpias en sus procesos productivos. Además, la empresa ha invertido en el desarrollo de nuevas tecnologías y en la diversificación de sus productos, lo que le ha permitido adaptarse a las demandas cambiantes del mercado.

Arauco está enfocada en la implementación de la Industria 4.0 en sus operaciones, esto implica la integración de tecnologías avanzadas como el internet de las cosas, analítica de big data, inteligencia artificial, computación en la nube y el aprendizaje automático en las instalaciones de producción y en todas las áreas de la empresa. La implementación de estas tecnologías permite a las empresas optimizar los sistemas de fabricación, acortar ciclos de desarrollo de nuevos productos, reducir costos de fabricación y contar con procesos productivos totalmente automatizados e integrados [2].

En la empresa, el área de confiabilidad de procesos se encarga de asegurar la confiabilidad de los procesos productivos, maximizando la producción y optimizando los recursos disponibles bajo los estándares de medio ambiente, seguridad y salud ocupacional. Esta área cuenta con diversos puestos de control que permiten supervisar los procesos operacionales y equipos dedicados a la producción de las 6 plantas ubicadas en Chile. Además, integra un equipo de expertos en automatización que tiene como uno de sus objetivos la implementación de la Industria 4.0, para así lograr procesos operacionales más automatizados.

Con el fin de alcanzar sus objetivos, el área de confiabilidad requiere de diversos indicadores para conocer el rendimiento y el nivel de automatismo que han conseguido en las diversas áreas de las plantas de celulosa. Indicadores que, hasta la fecha, no existen y que pueden permitir entregar conocimiento del avance en estos aspectos y las diversas mejoras que se pueden efectuar.

Específicamente para la Planta de Celulosa Valdivia, el área de Confiabilidad cuenta con información histórica de los eventos que ocurren en los sistemas de control a los cuales tiene acceso un operador de la planta. Este historial, que contiene información desde el año 2019 hasta abril de 2023, es extraído físicamente desde los centros de control, lo que permite tener una amplia cantidad de información del comportamiento operacional.

La problemática surgida a partir de esto, dada la cantidad de información y su formato, es que el área no dispone de una herramienta que permita visualizar y analizar los datos históricos. Es por esto que se plantea como hipótesis que el desarrollo de una plataforma de análisis de los archivos históricos permitirá conocer el grado de automatización de la planta de celulosa, entendiéndose automatismo como la capacidad que tienen los equipos de operar en modo automático y que sus funciones no son intervenidas por los operadores. Con esta plataforma, se espera identificar los cambios que realiza un operador en los equipos de producción, así como la periodicidad de dichos cambios y la detección de posibles eventos de fallos en las alarmas. Todo esto con el propósito de que el área de confiabilidad pueda comprender el comportamiento de los operadores en las distintas áreas y tener la oportunidad de incorporar mejoras en los eventos que son más frecuentemente modificados por los funcionarios de la planta.

La solución propuesta a este problema se dará a conocer con más detalle en el próximo capítulo, al igual que los alcances y limitaciones que serán presentadas en su desarrollo.

1.2 Solución propuesta y alcances

1.2.1 Solución propuesta

Para determinar el nivel de automatismo presente en la planta de celulosa Valdivia, se ha propuesto el desarrollo de una plataforma web que permita mostrar, mediante gráficos y tablas distintos indicadores clave que permitan conocer la actividad de los procesos en base a la información extraída a partir de la base de datos, agrupadas por cada una de las áreas de la planta de celulosa Valdivia. Esta herramienta ofrecerá una comprensión más detallada de los tipos de movimientos y acciones que se llevan a cabo, permitiendo identificar a los responsables de realizar estos cambios y comprender las consecuencias que pueden derivarse de las modificaciones realizadas.

Además, permitirá medir si las mejoras realizadas por el área de confiabilidad de procesos han generado una reducción en las intervenciones realizadas por parte de los operadores, y en qué áreas dentro de la planta se deben ir realizando mejoras en base a la toma de decisiones informada a través de la plataforma.

1.2.2 Alcances y limitaciones

El área confiabilidad de procesos es un área no vinculada con el área TI de la organización por lo que el desarrollo de cualquier plataforma web no podrá ser vinculada con la

red corporativa de Arauco dadas las restricciones de seguridad. Por lo mismo todo desarrollo se realizará localmente sin conexiones a internet.

Además, este desarrollo no cuenta con financiamiento por parte de la organización, por lo que no se han destinado recursos económicos para la compra de nuevo equipamiento hardware para el proyecto en cuestión. En su lugar, se hará uso de los dispositivos ya disponibles para llevar a cabo el desarrollo.

Por último, es importante mencionar que el área de confiabilidad de procesos no posee experiencia en el desarrollo de software ni en el diseño de proyectos. Por lo tanto, todo este proceso representa un nuevo desafío.

1.3 Objetivo general

Desarrollar una herramienta informática que permita medir la cantidad y el nivel de intervenciones humanas en la Planta de Celulosa Arauco Valdivia.

1.4 Objetivos específicos

- Identificar datos relevantes para lograr los objetivos organizacionales.
- Desarrollar una base de datos a partir de información histórica de Planta Celulosa.
- Seleccionar información fundamental de la base de datos sobre la base de los KPIs e información crítica identificada.
- Diseñar plataforma web que permita realizar búsquedas y acceder al contenido de la base de datos.
- Desarrollar un panel tipo dashboard que entregue información resumida de distintos KPIs.
- Generar analítica de datos que permita identificar patrones y generar conocimientos para informar y automatizar decisiones.
- Dejar operativo el sistema con al menos un usuario clave capacitado.

1.5 Metodología

Para el desarrollo de este proyecto, se utilizó la metodología incremental, que forma parte de las metodologías tradicionales de desarrollo de software. En un principio, se realizó una planificación basada en los requerimientos y objetivos de la organización, estableciendo los hitos principales y definiendo las funcionalidades básicas a implementar.

Posteriormente, se fueron realizando reuniones presenciales cada 2 semanas con los principales actores involucrados, estas reuniones tenían como objetivo dar seguimiento al avance del proyecto, discutir algunos problemas y tomar decisiones de mejora o desarrollo de nuevas funcionalidades a futuro.

Además, se estableció un modelo de trabajo híbrido en el cual se acudía a la oficina durante tres días a la semana para realizar el desarrollo del proyecto de forma presencial, mientras que los otros dos días se trabajaba de manera remota destinados a investigación y aprendizaje.

1.6 Estructura del informe

En el capítulo 2 se presentará el estado de arte, en el que se conocerá qué otras alternativas existen para la visualización y análisis de datos, al igual que las ventajas y desventajas entre ellas. Luego se presentan los conceptos previos con sus respectivas definiciones para los términos que serán utilizados a lo largo del trabajo.

El diseño de la arquitectura propuesta se muestra en el capítulo 3, donde se detallan los pasos de desarrollo desde la extracción de información a la visualización con los distintos KPI que se identifican durante el proyecto.

Por último, el capítulo 4 contiene la conclusión que comenta si los objetivos generales y específicos se pudieron lograr, al igual que la satisfacción de parte de los usuarios para luego comentar en el capítulo 5 las posibles mejoras e implementaciones en el trabajo futuro.

2. Antecedentes

2.1 Estado del arte

En el último tiempo la cantidad de datos creados y consumidos en internet han ido aumentando considerablemente llegando a triplicar periodos anteriores y se prevé que esta tendencia se mantenga así para los próximos años [3]. Esta creciente cantidad de datos ha generado la necesidad de desarrollar herramientas y técnicas que permitan gestionar y aprovechar de manera efectiva la gran cantidad de información. Es en este contexto donde se enmarca el campo de la Inteligencia de Negocio, que se ha convertido en una disciplina fundamental para las organizaciones.

La Inteligencia de Negocio o también conocida como BI, se refiere al conjunto de metodologías, procesos, herramientas y tecnologías que permiten recopilar, analizar y transformar los datos en información valiosa para la toma de decisiones empresariales. Esta disciplina se basa en la extracción de conocimiento a partir de los datos almacenados en los sistemas de información de una organización, permitiendo a los tomadores de decisiones identificar patrones, tendencias y oportunidades de mejora.

Uno de los aspectos clave en la Inteligencia de Negocio es la visualización de datos. Esta es la representación gráfica de la información con el objetivo de facilitar su comprensión y análisis. Mediante el uso de gráficos, tablas y otras herramientas visuales, es posible presentar de manera clara y concisa la información contenida en grandes conjuntos de datos, lo que facilita la identificación de patrones y tendencias, la detección de anomalías y la toma de decisiones informadas. La visualización de datos desempeña un papel fundamental en la Inteligencia de Negocio, ya que proporciona una forma intuitiva y efectiva de explorar y comprender los datos. Al utilizar técnicas de visualización adecuadas, es posible representar la información de manera interactiva y dinámica, lo que permite a los usuarios interactuar con los datos y descubrir información valiosa.

Según el cuadrante de Gartner en la figura 1, una de las consultoras tecnológicas más prestigiosas que realiza análisis de proveedores y tecnologías de TI, las empresas líderes en inteligencia de negocios son Microsoft, Tableau y Qlik.



Source: Gartner March (2023)

Figura 1: Cuadrante mágico de Gartner 2023 para plataformas de análisis e inteligencia empresarial. [4]

A continuación, se muestra en la tabla 1 que presenta las ventajas y desventajas de las principales herramientas de BI.

Tecnología	Ventajas	Desventajas
Microsoft PowerBI	<p>Integración con aplicaciones de Microsoft</p> <p>Análisis y comprensión de consultas impulsado por la IA con Copilot.</p> <p>Información actualizada en tiempo real.</p> <p>Variedad de gráficos y visualizaciones disponibles</p> <p>Trabaja con datos en entorno local y en la nube.</p>	<p>Compatibilidad solo con Windows</p> <p>La versión gratuita es limitada y para acceder a más funcionalidades y almacenamiento se debe conseguir la licencia Pro o Premium</p>
Tableau	<p>Experiencia de usuario intuitiva y fácil de usar para usuarios no técnicos.</p>	<p>Alto coste, especialmente en comparación con otras herramientas de BI,</p>

		partiendo desde los 70 USD/mes.
Qlik	<p>Información en tiempo real.</p> <p>Integración con diversas fuentes de datos.</p> <p>Capacidades de visualización rápidas.</p>	Análisis en memoria, la memoria RAM del equipo establece un límite por cada carga de conjuntos de datos pesados.

Tabla 1. Ventajas y desventajas en herramientas BI. [5] [6] [7]

Estas tecnologías líderes en BI ofrecen diversas ventajas y características que las hacen atractivas para diferentes tipos de empresas. Microsoft, Tableau y Qlik proporcionan soluciones flexibles, potentes y escalables que se adaptan a las necesidades empresariales, permitiendo el análisis de datos en tiempo real, generación de informes personalizados, visualización intuitiva y colaboración entre equipos.

Por otro lado, es importante mencionar que, aunque estas herramientas de BI ofrecen funcionalidades avanzadas, su despliegue implica costos significativos y las licencias gratuitas ofrecidas poseen funcionalidades limitadas. Además de la suscripción a servicios de alto valor, se debe considerar la inversión en personal especializado y el tiempo dedicado a la configuración y capacitación.

Una alternativa es optar por un desarrollo propio, esto puede tener la ventaja de la gran personalización que se puede conseguir para las necesidades de la organización, además de integrar información de sistemas existentes y un mayor control al acceso de datos sensibles. Además, el desarrollo interno brinda la oportunidad de diseñar una interfaz y una experiencia de usuario altamente intuitivas y adaptadas a la forma en que la organización desea interactuar con los datos. Esto puede mejorar la eficiencia y la productividad al permitir a los usuarios acceder y analizar la información de manera más rápida y efectiva.

Sin embargo, es importante tener en cuenta que el desarrollo interno de una solución de BI requerirá una gran cantidad de recursos, incluido tiempo y experiencia en desarrollo de software y análisis de datos. Además, la organización tendrá que mantener y actualizar la solución a lo largo del tiempo.

Después de evaluar las distintas alternativas existentes, junto con la organización se decide optar por un desarrollo propio en el ámbito de BI. Esta elección se fundamenta en la búsqueda de una solución altamente personalizada y adaptada a las necesidades y requerimientos específicos de la organización.

2.2 Conceptos previos

2.2.1 Industria 4.0 [2]

La Cuarta Revolución Industrial, también conocida como Industria 4.0, representa un cambio en la forma en que se produce y fabrica la producción industrial. Implementa la convergencia de tecnologías digitales, físicas y biológicas para transformar las fábricas tradicionales en inteligentes y conectadas.

Algunas de las principales tecnologías utilizadas en la Industria 4.0 incluyen:

- Internet de las cosas (IoT): Facilita el intercambio de información en tiempo real al conectar y comunicarse entre dispositivos, máquinas y sistemas.
- Análisis de datos: El procesamiento y el análisis de grandes cantidades de datos producidos por sistemas y sensores permiten obtener información importante para la toma de decisiones.
- La computación en la nube (Cloud computing): Es el almacenamiento y el procesamiento de datos en servidores remotos, lo que permite el acceso inmediato a los recursos informáticos y facilita la colaboración y el intercambio de datos.
- El aprendizaje automático y la inteligencia artificial (IA) automatizan procesos, encuentran patrones y optimizan la producción.
- Robótica avanzada: el uso de robots y sistemas automatizados permite la realización de tareas repetitivas y peligrosas, lo que aumenta la eficiencia y la seguridad en el entorno de trabajo.
- Realidad aumentada (AR) y Realidad virtual (VR): Las simulaciones de procesos, la visualización y la capacitación mejoran la productividad y reducen los errores.

2.2.2 Business intelligence [8]

La Inteligencia de Negocios (BI) es un conjunto de tecnologías y estrategias que se utilizan para analizar datos históricos o actuales, y convertirlos en información útil e importante para la toma de decisiones dentro de una organización. Las empresas pueden utilizar informes, paneles, gráficos y tablas para comprender mejor su rendimiento y tomar decisiones más informadas y precisas en tiempo real.

El BI implica la integración de varias fuentes de datos, la creación de modelos y el uso de herramientas analíticas para descubrir patrones y tendencias, lo que permite a las empresas obtener una visión más completa de su negocio y del entorno competitivo.

2.2.3 KPI [9]

Los KPI (Indicadores Clave de Desempeño) son medidas cuantificables y específicas que se utilizan para evaluar el rendimiento de una organización en relación con sus objetivos estratégicos previamente definidos. Estos son fundamentales para medir y monitorear el progreso hacia el logro de los objetivos, así como para identificar oportunidades de mejora y optimizar el desempeño en algunas áreas.

2.2.4 Dashboard [10]

Un dashboard es una solución a la administración de información y visualización de datos por medio del análisis de los datos con elementos interactivos como gráficos e informes en una sola pantalla, con el fin de resumir y presentar información de manera clara y concisa. De esta manera los usuarios pueden conocer el panorama general al instante, esto les permite tomar decisiones informadas y mejorar el rendimiento de la organización. Además, los dashboards permiten ser personalizados para adaptarse a las necesidades específicas de cada usuario.

2.2.5 Sistema de control distribuido (DCS) [11]

El Sistema de Control Distribuido (DCS) es un sistema informático utilizado en la industria para controlar y supervisar de manera centralizada las operaciones de una planta o proceso industrial. Este sistema proporciona un control preciso y eficiente de los procesos industriales al integrar una variedad de tecnologías y componentes, como controladores, sensores, actuadores, interfaces de usuario y redes de comunicación.

2.2.6 ELT [12]

El proceso ELT es un proceso alternativo al de ETL (extracción, transformación y carga de datos), la diferencia que presenta es el orden de la ubicación de la transformación, los datos se transforman utilizando las capacidades de procesamiento del almacén de datos de destino en lugar de un motor de transformación independiente. Los procesos ELT se suelen utilizar para el manejo de grandes volúmenes de datos o datos en tiempo real.

A continuación, se describe cada parte del proceso.

- **Extracción:** Los datos se copian o exportan a un área de ensayo durante el proceso de extracción de datos. El conjunto de datos puede estar compuesto por una variedad de tipos de datos y puede provenir de casi cualquier fuente estructurada o no estructurada.
- **Carga:** Los datos sin procesar van a la zona de almacenamiento, como un almacén de datos o data lake, este proceso suele ser automatizado, continuo y por lotes.

- Transformación: Esta etapa final transforma los datos, a menudo usando scripts SQL. Dentro de los procesos de transformación se pueden realizar filtros, eliminación, obtención de cálculos, dar formato, etc.

2.2.7 Flask [13]

Flask es un framework escrito en Python utilizado como backend para el desarrollo de creación de páginas web utilizando el modelo MVC, este modelo permite diferenciar y separar lo que es la vista, el modelo de datos, y el controlador. Flask es compatible con el motor de plantilla llamado Jinja2 que se utiliza para crear HTML, XML u otros formatos de marcado que se devuelven a los usuarios a través de una respuesta HTTP.

2.2.8 Microsoft SQL Server [14]

SQL Server es el sistema de bases de datos profesional de Microsoft, cuenta con una variedad de características y herramientas que se pueden utilizar para desarrollar y administrar bases de datos y soluciones de todo tipo basadas en ellas. De sus entornos integrados y herramientas el más utilizado es SQL Server Management Studio, que permite configurar y administrar todos los componentes de SQL Server y está diseñado para desarrollar y administrar objetos de bases de datos relacionales.

3. Diseño de la Solución

3.1 Arquitectura propuesta

La arquitectura presente en la figura 2 es la que se utilizará para el desarrollo del proyecto, consta de un Dispositivo NAS (Network-Attached Storage) donde se almacenará la base de datos, en el cual se almacenará toda información entregada para su posterior análisis. También cuenta con un equipo utilizado como servidor, este se encargará de ser el host y ejecutar el software especializado que permite realizar consultas, análisis y otras operaciones en los datos almacenados en el NAS. La conexión entre estos dos dispositivos se realiza a través de una conexión inalámbrica utilizando un router y para poder acceder al servidor es necesario estar conectado a la red local.

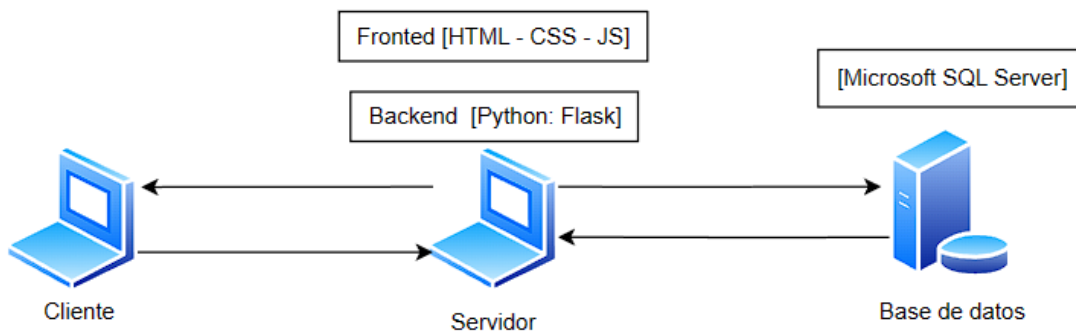


Figura 2: Arquitectura de sistemas. (Elaboración propia)

En los siguientes capítulos se verán con más detalle la implementación de los sistemas en cada uno de los equipos hardware, así como sus características y las conexiones realizadas.

3.2 Extracción de información

La planta de Valdivia dispone del software “Exploring Delta V”, el cual permite el monitoreo, configuración y visualización del funcionamiento de sus equipos desplegados en las diversas áreas de la planta de celulosa Valdivia. En esta aplicación se generan automáticamente archivos MDF con datos históricos de eventos, según la configuración que se haya escogido previamente, estos pueden ser que se alcanzó el tamaño límite del archivo ó que se cumplió con un lapso de tiempo de un mes, semana o día.

Los archivos MDF (Master Data File) son un tipo de archivo que almacena datos y objetos principales de la base de datos, como tablas, vistas, procedimientos almacenados y más. La estructura que presentan los archivos se puede ver en la figura 3, que muestra las tablas contenidas en el archivo y sus respectivos atributos, que son descritos más adelante en el capítulo 3.5.

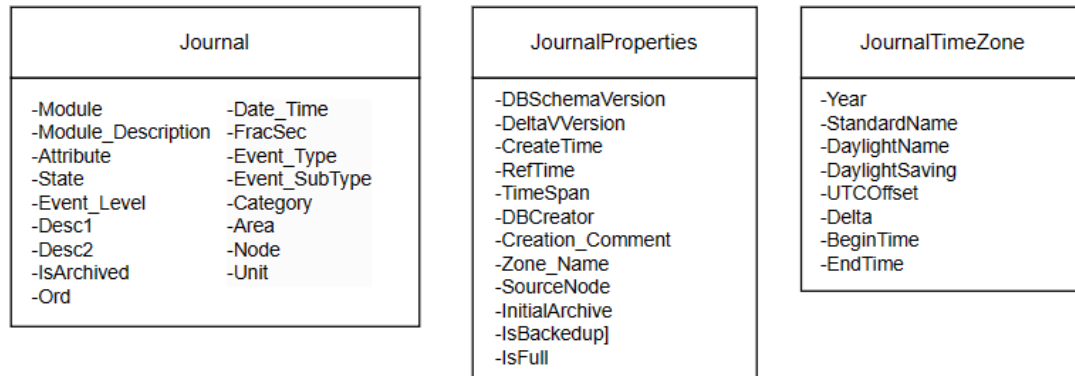


Figura 3: Esquema relacional de los archivos MDF que emite Exploring Delta V.
(Elaboración propia)

Los archivos son almacenados en un equipo dedicado al almacenaje de archivos históricos para su posterior extracción, este mecanismo es realizado de manera presencial en la planta de Valdivia mediante un dispositivo USB tipo pendrive o disco duro para que posteriormente sea almacenado en el equipo NAS, el proceso es realizado de esta forma puesto que la arquitectura del DCS no presenta una conexión directa a internet para su acceso desde una fuente externa y la protección de estos dispositivos ante los ciberataques es fundamental para el correcto funcionamiento.

3.3 Desarrollo y configuración de la base de datos

Para el almacenamiento de los datos se cuenta con un dispositivo NAS Dell NX430 (Network Attached Storage) que opera como almacenamiento para los datos extraídos desde los dispositivos de DCS e integra un sistema gestor de bases de datos Microsoft SQL Server.

El equipo cuenta con las siguientes características de hardware:

- Procesador: Intel Xeon CPU-E3-1220 v6 3.00GHz
- RAM: 8 GB 2666 MHz
- Espacio disco: 10 TB HDD
- OS: Microsoft Windows Storage Server 2016 Workgroup

El primer paso que se realizó fue definir un directorio dentro del disco que contendrá los archivos y la base de datos, este es *D:/DB/DATA*. Luego se hizo la instalación del sistema gestor de bases de datos Microsoft SQL Server Management Studio, al momento de la configuración de la aplicación el equipo ya tenía una configuración de servidor local al que se podía acceder y con las credenciales creadas.

Para el acceso al servidor se realizó una configuración de protocolos TCP/IP indicando el puerto 1433, puerto utilizado por Microsoft SQL Server para la comunicación entre el servidor

de base de datos y los clientes, además de la habilitación del servicio *MSSQLSERVER01*, instancia que permite la ejecución del servicio SQL Server. Así se puede acceder al servidor desde un equipo externo con el nombre del equipo y las credenciales.

Los archivos MDF son almacenados manualmente en el directorio *D:/DB/DATA*, este proceso en primera instancia es realizado por el autor, luego se le indicará a un integrante del equipo de automatización el proceso a realizar, acompañado de un instructivo para administración. El periodo de traspasar los archivos dependerá de la frecuencia con la que se exporten los datos mencionados en el capítulo 3.1, por otro lado, si llegase a ocurrir omisiones o errores estos se verán con más detalle en el capítulo 3.8.

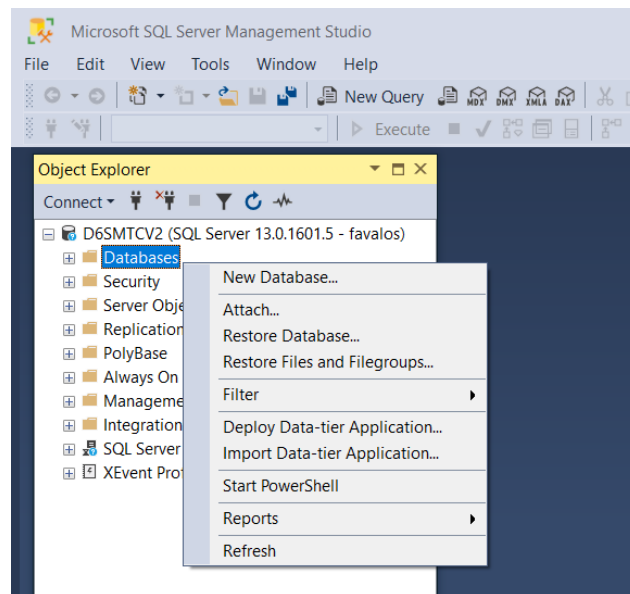


Figura 4: Menú desplegable de base de datos en el servidor. (Elaboración propia)

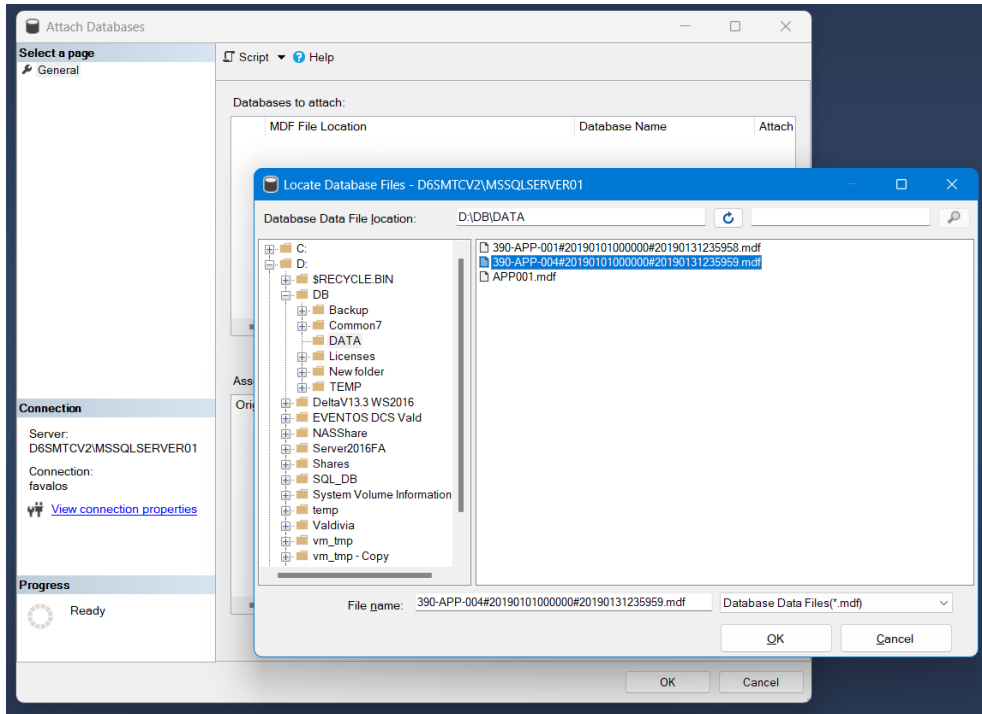


Figura 5: Selección de archivo origen para la base de datos. (Elaboración propia)

La figura 4 muestra las acciones permitidas para el servidor NAS, para este caso se utiliza “New Database...”. Por convención se decide escoger el archivo más antiguo que corresponde al seleccionado en la figura 5, para que luego en base a este mismo archivo origen sean insertados los otros datos de los siguientes archivos. La ventaja existente de este método es que Microsoft SQL Server permite la lectura de los archivos con la extensión MDF, por lo que no es necesario una modificación o transformación previa.

El nombre asignado a la base de datos es “DB004”, y esta será la base de datos con la cual se trabajará para el desarrollo de la solución.

Date_Time	Event_Ty...	Category	Area	Node	Unit	Module	Module_...	Attribute	State	Event_Level	Desc1	Desc2	Ord
2019-01-01 03:00:00...	ALARM	PROCESS	346_LAVADO	346-MDC-R-01	SCREEN_...	346PDI442	SCREEN ...	HI_ALM	INACT/...	07-ADVISORY	HIGH	High Alarm Va...	141394807
2019-01-01 03:00:00...	EVENT	PROCESS	346_LAVADO	346-MDC-R-03	FILTRATE_...	346FIC1160LG	Control M...		ACTIVE	4-INFO	Error Cle...	I/O Output Fail...	141394808
2019-01-01 03:00:00...	ALARM	PROCESS	372_SECADO	372-MDC-R-02		372PI636A	Lump Bre...	LO_ALM	INACT/...	11-WARNING	LOW	Low Alarm Val...	141394809
2019-01-01 03:00:01...	ALARM	PROCESS	372_SECADO	372-MDC-R-01		372M543	CIRCULA...	FAIL_ALM	ACT/U...	11-WARNING	FAILED	STOP Confirm...	141394810
2019-01-01 03:00:01...	ALARM	PROCESS	385_EFLUE...	385-MDC-R-02		385C11639	Conductiv...	HI_ALM	INACT/...	11-WARNING	HIGH	High Alarm Va...	141394811
2019-01-01 03:00:01...	EVENT	PROCESS	346_LAVADO	346-MDC-R-03	FILTRATE_...	346FIC1160LG	Control M...		ACTIVE	4-INFO	346FIC1...	I/O Output Fail...	141394812
2019-01-01 03:00:01...	ALARM	PROCESS	372_SECADO	372-MDC-R-02		372PI636A	Lump Bre...	LO_ALM	ACT/U...	11-WARNING	LOW	Low Alarm Val...	141394813
2019-01-01 03:00:02...	ALARM	PROCESS	372_SECADO	372-MDC-R-01		372M543	CIRCULA...	FAIL_ALM	INACT/...	11-WARNING	FAILED	Clear	141394814
2019-01-01 03:00:02...	ALARM	PROCESS	346_LAVADO	346-MDC-R-01	SCREEN_...	346PDI442	SCREEN ...	HI_ALM	ACT/ACK	07-ADVISORY	HIGH	High Alarm Va...	141394815
2019-01-01 03:00:02...	EVENT	PROCESS	346_LAVADO	346-MDC-R-03	FILTRATE_...	346FIC1160LG	Control M...		ACTIVE	4-INFO	Error Cle...	I/O Output Fail...	141394816
2019-01-01 03:00:02...	ALARM	PROCESS	372_SECADO	372-MDC-R-02		372PI636A	Lump Bre...	LO_ALM	INACT/...	11-WARNING	LOW	Low Alarm Val...	141394817
2019-01-01 03:00:03...	ALARM	PROCESS	346_LAVADO	346-MDC-R-01	SCREEN_...	346PDI442	SCREEN ...	HI_ALM	INACT/...	07-ADVISORY	HIGH	High Alarm Va...	141394818
2019-01-01 03:00:03...	EVENT	PROCESS	346_LAVADO	346-MDC-R-03	FILTRATE_...	346FIC1160LG	Control M...		ACTIVE	4-INFO	346FIC1...	I/O Output Fail...	141394819
2019-01-01 03:00:04...	EVENT	PROCESS	346_LAVADO	346-MDC-R-03	FILTRATE_...	346FIC1160LG	Control M...		ACTIVE	4-INFO	Error Cle...	I/O Output Fail...	141394820
2019-01-01 03:00:05...	EVENT	PROCESS	346_LAVADO	346-MDC-R-03	FILTRATE_...	346FIC1160LG	Control M...		ACTIVE	4-INFO	346FIC1...	I/O Output Fail...	141394821
2019-01-01 03:00:06...	ALARM	PROCESS	347_BLANQ...	346-MDC-R-04		347FIC1093	Lower Cir...	LO_ALM	INACT/...	11-WARNING	LOW	Low Alarm Val...	141394822

Figura 6: Estructura de tabla “Journal”. (Elaboración propia)

Por último, se decide hacer un análisis de la información antes de cargar los archivos restantes. Durante el proceso, se pudo observar que la tabla carecía de una clave primaria, así que se decide la creación de una clave, a través de una asignación a una de las columnas de la tabla, la tabla presenta una estructura como se muestra en la figura 6. La elección de la clave debería ser tal que el dato no se repita dentro de la columna, es decir que sea único y que además el valor no fuese nulo o vacío. El atributo “Ord” es el designado como clave primaria dada su naturaleza al ser un valor entero positivo incremental por cada fila existente.

También, se decide eliminar columnas que no entreguen una información a utilizar o que tienen campos vacíos, como lo son Event_SubType y IsArchived, así en las inserciones futuras no se consideran estos campos.

3.4 Carga y Transformación de información

La carga de los datos es la continuación del proceso de ELT, después de llevar a cabo la extracción de los archivos y creación de la base de datos, se procede a la selección de los archivos. Para realizar esta carga, se utiliza un procedimiento almacenado llamado “Events_Copy”, el cual requiere el nombre del archivo que se desea ingresar a la base de datos.

El procedimiento almacenado “Events_Copy” se encarga de ejecutar diferentes subprocesos para la transformación de la información. Estos subprocesos involucran la eliminación de filas duplicadas, la actualización de tablas ya existentes y la inserción de datos que no se encuentren presentes en la tabla original.

Los pasos que sigue el procedimiento almacenado son los siguientes:

1. Adjuntar: En la figura 7 el archivo es adjunto como una base de datos mediante el uso del procedimiento almacenado “sp_attach_single_file_db”, esta tiene el nombre de “testdb” y es necesario indicar el nombre del archivo.

```
declare @route_file nvarchar(100)
SET @route_file = 'D:\DB\DATA\' + @filename
EXEC sp_attach_single_file_db @dbname= 'testdb' , @physname = @route_file
PRINT ' Database Temporal : ' + @filename + '  atachada  '
```

Figura 7: Proceso para adjuntar base de datos. (Elaboración propia)

2. Permisos de usuario: En la figura 8 se crea el usuario y se le asignan permisos de “db_owner” ante cualquier modificación que se quiera realizar para la nueva base de datos.

```

SET @execstring = 'USE testdb;
CREATE USER favalos
FOR LOGIN favalos;
exec sp_addrolemember 'db_owner', 'favalos';'
EXEC (@execstring)
PRINT ' Usuario creado para la DB temporal '

```

Figura 8: Creación de usuario y permisos en la nueva base de datos.
(Elaboración propia)

3. Datos duplicados: En la figura 9 se ejecuta una query para la base de datos “testdb” que elimina los datos duplicados para la tabla Journal bajo el atributo “Ord”.

```

select @execstring = 'WITH CTE AS
(SELECT *, ROW_NUMBER() OVER (PARTITION BY hh.Ord ORDER BY hh.Ord) ROWNUM
FROM [testdb].[dbo].[Journal] as hh)
DELETE FROM CTE WHERE ROWNUM>1 '
EXEC (@execstring)
PRINT 'Se han eliminado duplicados de Ord en la DB temporal'

```

Figura 9: Eliminación de datos duplicados en tabla Journal para la base de datos “testdb”. (Elaboración propia)

4. Clave primaria: En la figura 10 se asigna como clave primaria el atributo “Ord” a la base de datos “testdb” para poder realizar las operaciones de comparaciones y eliminaciones de manera eficiente.

```

SET @execstring = 'ALTER TABLE [testdb].[dbo].[Journal]
ADD CONSTRAINT PK_JOURNAL_Ord PRIMARY KEY (Ord) '
EXEC(@execstring)
PRINT 'Primary key asignada'

```

Figura 10: Asignación de clave primaria. (Elaboración propia)

5. Actualizar tabla de áreas: En la figura 11 se ejecuta la sentencia que actualiza la tabla “AreaModules” en la base de datos “DB004” mediante el uso de la sentencia *left join* sólo si “testdb” presenta nuevas áreas y módulos. De no ser así, se pasa al siguiente proceso.

```

select @execstring = ' INSERT INTO AreaModules(area,module,module_description)
SELECT DISTINCT t.area, t.module, t.module_description
FROM [testdb].[dbo].[Journal] AS t
LEFT JOIN [DB004].[dbo].[AreaModules] AS j
ON t.module = j.module
WHERE j.module IS NULL '

EXEC (@execstring)

```

Figura 11: Inserción de datos a la tabla de áreas y módulos. (Elaboración propia)

- Actualizar tabla journal: La sentencia de la figura 12 inserta los datos de “testdb” en la tabla Journal de “DB004” solo si son nuevos datos usando la sentencia *left join*.

```

select @execstring = 'SET IDENTITY_INSERT [DB004].[dbo].[Journal] ON
INSERT Into [DB004].[dbo].[Journal]
([Date_Time],[FracSec],[Event_Type],[Event_SubType]
,[Category],[Area],[Node],[Unit],[Module],[Module_Description]
,[Attribute],[State],[Event_Level],[Desc1],[Desc2],[IsArchived],[Ord] )

SELECT TA.[Date_Time], TA.[FracSec],TA.[Event_Type],TA.[Event_SubType]
,TA.[Category],TA.[Area],TA.[Node],TA.[Unit],TA.[Module],TA.[Module_Description]
,TA.[Attribute],TA.[State],TA.[Event_Level],TA.[Desc1],TA.[Desc2],TA.[IsArchived],TA.[Ord]
FROM [testdb].[dbo].[Journal] AS TA
LEFT JOIN [DB004].[dbo].[Journal] as TB
ON TB.Ord = TA.Ord
WHERE TB.Ord IS NULL '

```

Figura 12: Inserción de datos a tabla Journal en DB004. (Elaboración propia)

- Separar: En la figura 13 el proceso quita la base de datos “testdb” del servidor, pues toda su información ya se encuentra en “DB004”.

```

select @execstring = 'USE [master];
ALTER DATABASE testdb
SET SINGLE_USER
WITH ROLLBACK IMMEDIATE;
EXEC master.dbo.sp_detach_db @dbname = N''testdb'''

EXEC (@execstring)
PRINT ' Base de Datos Temporal Desatachada!'

```

Figura 13: Separación de base de datos “testdb”. (Elaboración propia)

- Eliminar: La figura 14 muestra el proceso de eliminar el archivo de transacciones “testdb.ldf”, así cuando sea ejecutado nuevamente el procedimiento almacenado, no se obtenga un error por un archivo ya existente. Luego el archivo que ya fue cargado es trasladado a otra carpeta para que no aparezca visible en el proceso de carga.

```

DECLARE @cmd NVARCHAR(MAX) = 'xp_cmdshell ''del "D:\DB\DATA\testdb_log.ldf"''';
EXEC (@cmd)
PRINT ' Log temporal borrado.'

DECLARE @command NVARCHAR(1000)
SET @command = 'move /Y D:\DB\DATA\' + @filename + ' D:\DB\DATA\LoadedFiles\' + @filename
EXEC xp_cmdshell @command

```

Figura 14: Eliminación de log y transferencia de archivo a otra carpeta.
(Elaboración propia)

Ya finalizado los pasos anteriores, se operará bajo dos tablas, Journal y AreaModules como se observa en la figura 15, en donde la tabla Journal contendrá toda la información y la tabla AreaModules funcionará como un diccionario para conocer las áreas disponibles, módulos y su respectiva descripción. Esta separación ayuda a que la descripción del módulo sea única, pues ocurría que existían módulos que presentaban 2 o más descripciones asociadas.

Journal		AreaModules
-Module	-Date_Time	-Area
-Module_Description	-Event_Type	-Module
-Attribute	-Category	-Module_Description
-State	-Area	
-Event_Level	-Node	
-Desc1	-Unit	
-Desc2		
-Ord		

Figura 15: Resultado de transformación de datos. (Elaboración propia)

La ejecución del procedimiento almacenado debe llevarse a cabo individualmente para cada archivo que se desee incluir en la base de datos. El tiempo necesario para ejecutar este procedimiento puede variar en función del tamaño del archivo, pudiendo requerir más tiempo para archivos de mayor tamaño. En el caso de que falle la ejecución del procedimiento almacenado u ocurra un desperfecto técnico que impida que se ejecute la totalidad del procedimiento almacenado, basta con ejecutar el procedimiento nuevamente con el mismo archivo, ya que no se generará duplicidad de la información.

3.5 Análisis y selección de datos

Con la base de datos creada y completa se procede al análisis de las variables presentes en las tablas, identificar dependencias y relaciones entre los datos, esto con el fin de ayudar a comprender la estructura para más adelante realizar consultas de información más precisas y detalladas.

Previo al desarrollo de este análisis, la organización comentó el principal objetivo de información que les gustaría poder observar que es conocer la cantidad de intervenciones que ha realizado un operador en un área específica. Con esta información ya se puede deducir que

dentro de los datos a analizar es relevante el área y si los eventos corresponden a un cambio o modificación realizada.

El primer paso es conocer qué información está contenida dentro de la tabla principal, esta tabla está compuesta por 16 columnas y su información pudo ser obtenida consultando los datos y de ser necesario consultar con personal

- Date_time: Fecha con formato AAAA-MM-DD HH:MM:SS en la que se registra la actividad.
- Event_type: Tipo de evento, estos pueden ser Change, Alarm, Download, Event.
- Category: Tipo de categorías en la que se clasifican los eventos, estos son Device, Hardware, Instrument, Process, System, User.
- Area: Nombre de las áreas en la planta de celulosa.
- Node: Nombre del dispositivo DCS que captura la información del evento.
- Unit: Nombre de equipos industriales.
- Module: El módulo presenta una nomenclatura compuesta por el número del área de la planta, una abreviatura para identificar el instrumento y un número correlativo que puede tener relación o no con otro equipo conectado directamente.
- Module_Description: Breve descripción a lo que se refiere el módulo
- Attribute: Nombre de alarmas, variables de un controlador.
- State: Estado de la alarma o variable dependiendo del tipo de evento.
- Event_Level: Niveles de advertencia.
- Desc1: Presenta una combinación del módulo y un atributo, o nombre de un operador, este valor depende del tipo de evento.
- Desc2: Corresponde a una descripción con más detalle de la columna "Desc1", cambios en valores de alarmas y cambios en valores de variables de un controlador.
- Ord: Número incremental ascendente por cada evento.

Para la identificación de los valores se hicieron diferentes filtros utilizando sentencias de SQL en la base de datos para ver los distintos valores que podía contener un atributo y con qué otros valores podría estar vinculado, para así establecer relaciones entre los datos de *Event_Type* y *Category*.

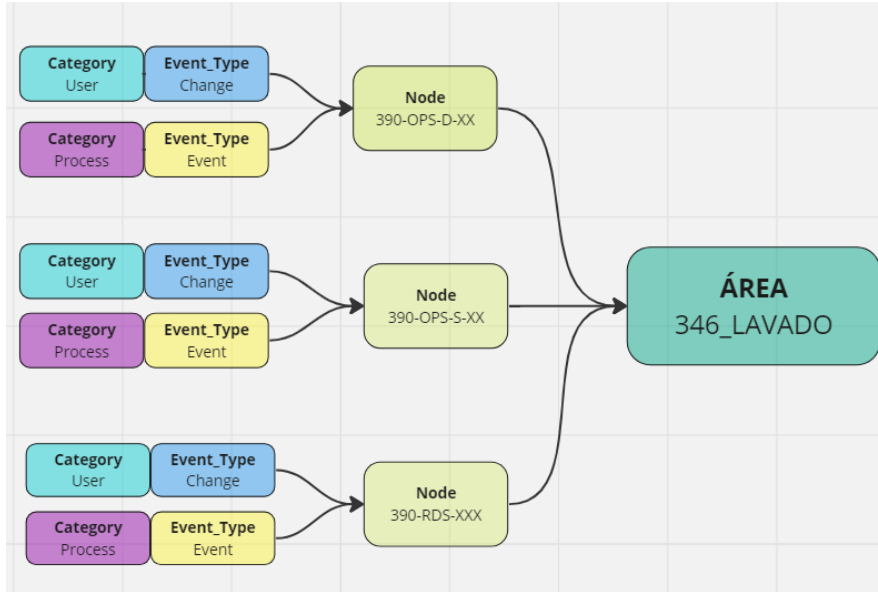


Figura 16: Esquema de relaciones Event_type - Category 1.
(Elaboración propia)

La figura 16 muestra cómo están relacionados algunos de los valores, se parte en la búsqueda con el filtro del Area: 346_LAVADO y luego con la selección de unos Nodos. Con la finalidad de partir desde lo más general a lo más específico. De la figura se puede apreciar que Category: *User* está vinculada con Event_type: *Change* al igual que Category: *Process* con Event_type: *Event*.

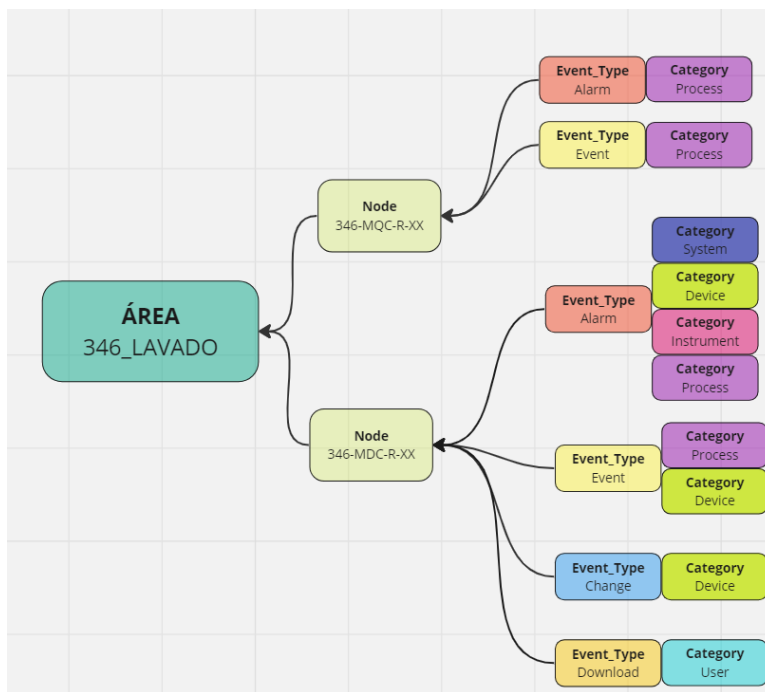


Figura 17: Esquema de relaciones Event_type - Category 2.
(Elaboración propia)

Utilizando la misma área de búsqueda, podemos ver en la figura 17 que Event_type: *Alarm* está relacionado con Category: *Process, Instrument, System, Device* y que Event_type: *Change* ahora también tiene una relación con Category: *Device*

Bajo estas búsquedas y relaciones se puede determinar algunas combinaciones que resultan útiles para analizar y conocer su contenido, por lo que se procede a filtrar nuevamente los datos siguiendo los parámetros de la tabla 2, cada una de las búsquedas permite generar una breve descripción sobre la información que presentan.

Category	Event_type	Descripción
USER	CHANGE	Cambios realizados por un usuario para alarmas y variables de un controlador.
USER	DOWNLOAD	Eventos de actualización de lógica para los módulos realizados por administrador.
DEVICE	CHANGE	Modificaciones de parámetros en instrumentos.
PROCESS	EVENT	Fallas en los dispositivos, avisos de subida de corriente.
PROCESS	ALARM	Alarmas de advertencias por valores bajo un valor o por sobre un valor.
DEVICE	ALARM	Fallas de dispositivos y de comunicación, errores de calibración.
SYSTEM	ALARM	Duración de alarma.
INSTRUMENT	ALARM	Alarmas de fallas en entradas y salidas de un dispositivo, alarmas de válvulas.

Tabla 2: Contenido filtrado para variables Category y Event_type.
(Elaboración propia)

Con esto en conocimiento, ya se puede emplear la construcción de las queries y poder identificar los KPI útiles y relevantes para la organización.

3.6 Identificación de KPI

La selección de los KPI es de vital importancia ya que permite medir el rendimiento y progreso de acuerdo a los objetivos organizacionales, así la organización puede identificar y comprender dónde se encuentran los problemas y qué mejoras son necesarias. Esto permite a la organización tomar decisiones informadas sobre cómo abordar esas áreas problemáticas y asignar recursos de manera efectiva para impulsar el cambio y el crecimiento.

El principal objetivo de la organización es conocer qué tan automatizada se encuentra la planta de Celulosa Valdivia, esto es, poder conocer qué cantidad de intervenciones por parte de los operadores son realizadas, estas intervenciones son el resultado de realizar modificaciones en el funcionamiento de maquinarias mediante el uso del centro de monitoreo.

De acuerdo al objetivo organizacional, a las reuniones con los usuarios y a la información obtenida de las búsquedas del capítulo 3.5, se pueden seleccionar como indicadores:

- Módulos más frecuentes con cambios de los operadores.
- Usuarios y cantidad de cambios que han realizado los operadores.
- Total de cambios de usuario por día.
- Módulos más frecuentes de alarmas y eventos.

Estos indicadores tendrán como valores de entrada el área y un periodo de tiempo determinado, además para algunos indicadores se podrá desplegar más información al usuario para tener un detalle específico de los cambios y modificaciones realizadas. Con estos KPI identificados se puede proceder a la visualización y el despliegue de información.

3.7 Visualización de KPI e implementación

Se diseñó una plataforma web utilizando el framework Flask del lenguaje de programación Python en Backend para mostrar los KPIs. Además, se utilizó Plotly, HTML, Jinja2 y Bootstrap para el frontend. Flask permite la creación ágil de aplicaciones web y Plotly cuenta con variadas herramientas de visualización de datos.

En la implementación, se creó una página para cada KPI, permitiendo la visualización y análisis específicos para cada uno mostrando la información relevante del KPI correspondiente de manera clara y concisa. Además, se implementó una página tipo Dashboard que permite ver de forma gráfica cada uno de los KPI identificados en un mismo lugar.

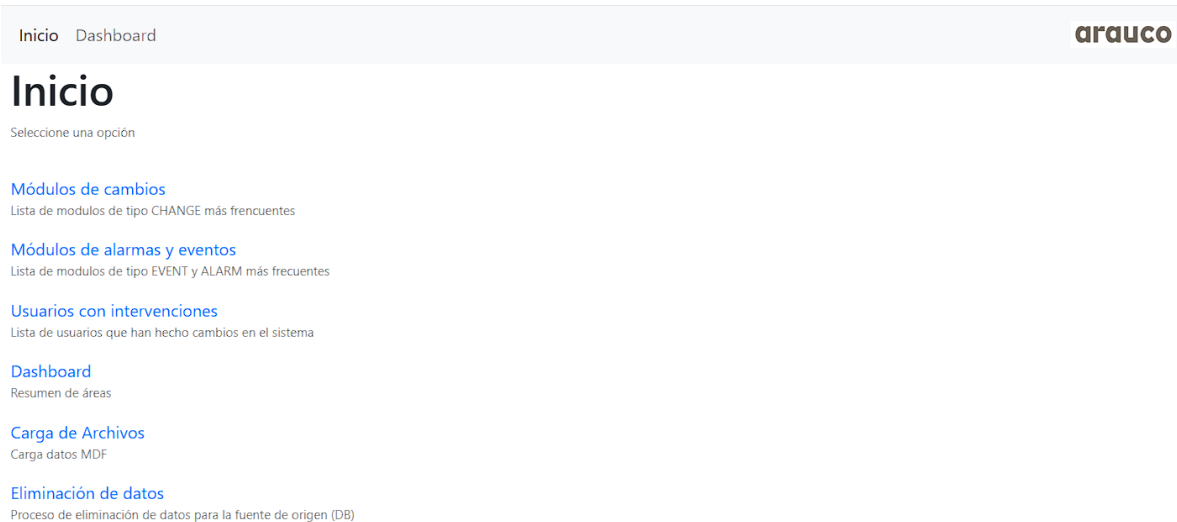


Figura 18: Menú de inicio de plataforma web.
(Elaboración propia)

El menú principal permite el acceso a las páginas creadas en la figura 18 se puede apreciar que el diseño cuenta con una barra de navegación que está disponible en todos los sitios y con el logo corporativo de la organización. Para la consulta de la información de los KPI se cuenta con un selector del área deseada como se ve en la figura 19, las áreas indicadas son las contenidas en la base de datos y un rango de fecha que indica el intervalo disponible de datos.



Figura 19: Selección de área y fechas.
(Elaboración propia)

Luego de indicar el área y las fechas, se presiona el botón *Buscar* para realizar la consulta. El tiempo de espera para los resultados depende directamente del rango de fechas

que se interese conocer. Ya con los resultados obtenidos estos son desplegados como se puede ver en la figura 20, en el centro se ve el detalle de lo consultado, luego en la zona izquierda se muestra una gráfica de barras indicando la cantidad de cambios que se realizaron y el nombre del módulo asociado. El lado derecho indica más información como una breve descripción del módulo y un selector para tener más detalle de las modificaciones.

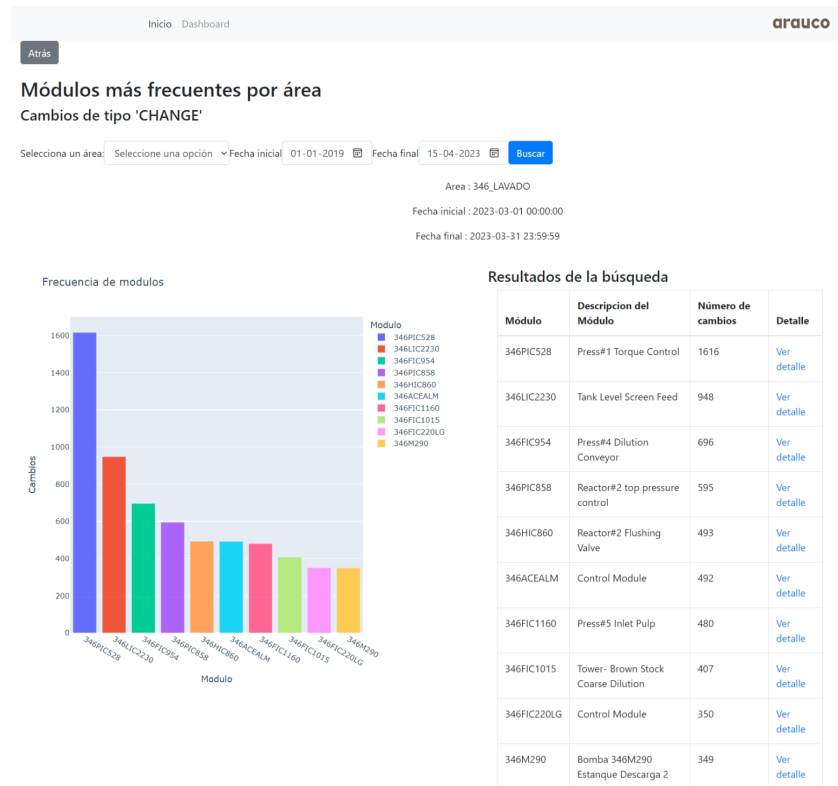


Figura 20: Resultados de consulta. (Elaboración propia)

Al presionar el botón *Ver detalle* se abre una nueva ventana como la de la figura 21, esta indica el tipo de atributo y con qué frecuencia fue modificado este valor. Además, está la posibilidad de poder descargar como CSV los datos del módulo consultado en la fecha determinada, para así poder realizar un análisis más detallado y conocer a qué hora específica los valores de los eventos fueron modificados.

[Atrás](#)

Detalles del módulo 346PIC528

Fecha: 2023-03-01 00:00:00 hasta 2023-03-31 23:59:59

[Descargar CSV](#)

Atributo	Número de eventos
PID1/OUT.CV	1177
PID1/SP.CV	373
PID1/MODE.TARGET	66

Figura 21: Detalle de módulo consultado.
(Elaboración propia)

La página para visualizar el KPI de “módulos frecuentes de alarmas y eventos” es similar en estructura, se selecciona el área y el intervalo de fechas. El resultado se despliega como el de la figura 21 y permite visualizar con más detalle los eventos ocurridos con la posibilidad de descargar el reporte en formato CSV.

Para el KPI de “usuarios y cantidad de cambios” la consulta es análoga, se selecciona el área y fechas, luego se despliega el resultado de la consulta como se ve en la figura 22. Un gráfico circular indicando el porcentaje asociado a la cantidad de eventos realizados por cada uno de los usuarios, y una tabla que indica el nombre del usuario, la cantidad de cambios y la opción de ver el detalle de ese usuario. El detalle del usuario se abre en una nueva ventana, como el de la figura 23, e indica el nombre del usuario, área, fechas y en una tabla indica el nombre y la cantidad de modificaciones con la opción de descargar como CSV los eventos del usuario.

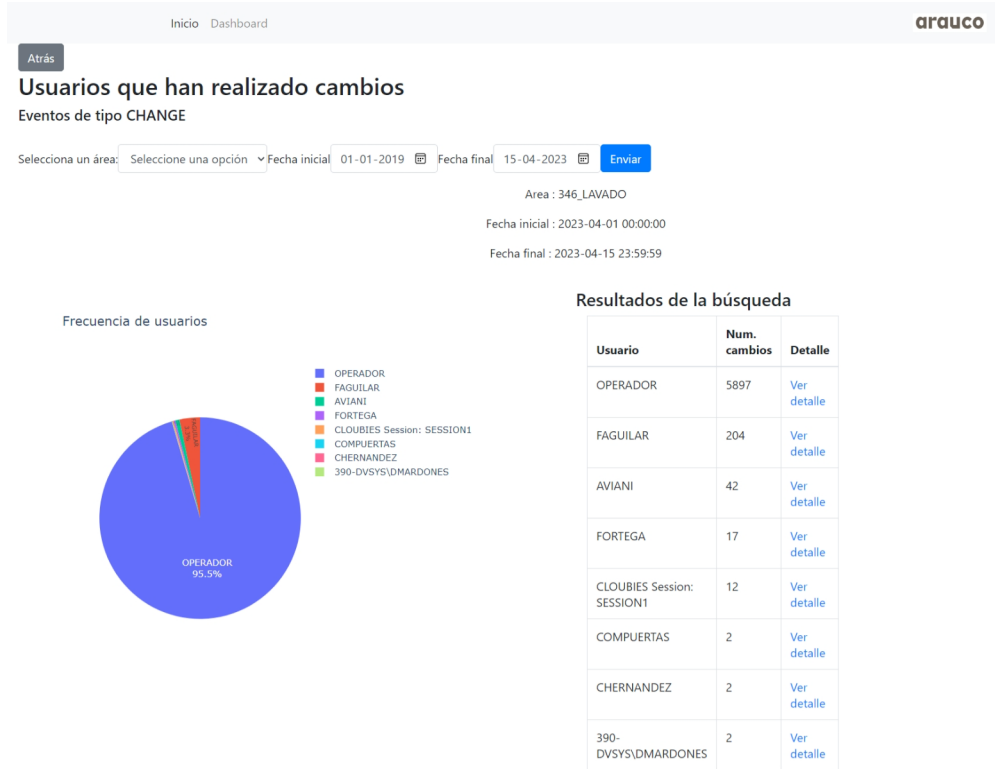


Figura 22: Resultado de cambios. (Elaboración propia)

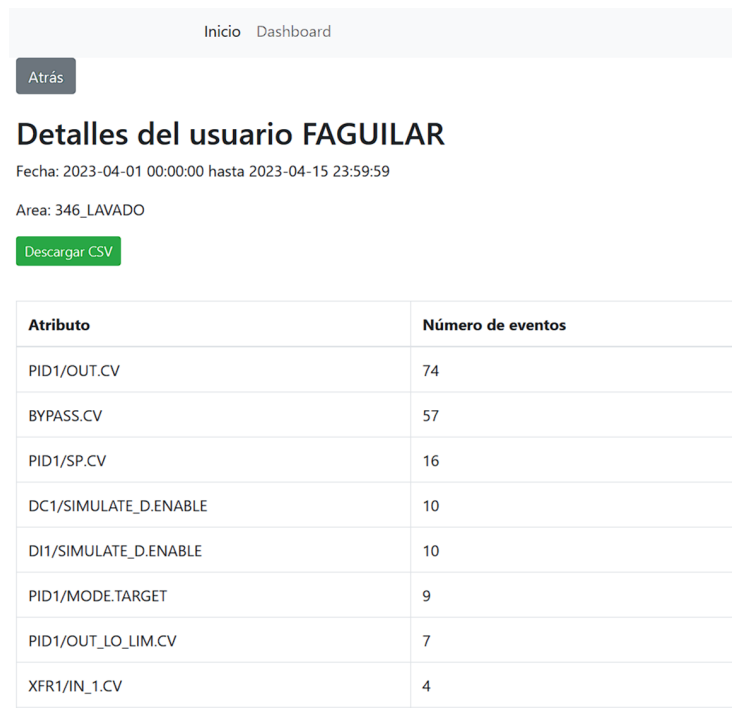


Figura 23: Detalle de eventos de un usuario. (Elaboración propia)

El Dashboard proporciona una visión general y panorámica de los eventos y actividades que ocurren en un área específica de la planta. Esta herramienta permite a los usuarios obtener rápidamente una comprensión de los indicadores clave de rendimiento, el estado de los eventos y cualquier situación relevante que requiera atención.

Los KPI presentes en el dashboard de la figura 24 corresponden a los módulos de cambios más frecuentes, módulos de alarmas y eventos más frecuentes, usuarios con más cambios, y una línea temporal que indica la cantidad de cambios por día del operador.

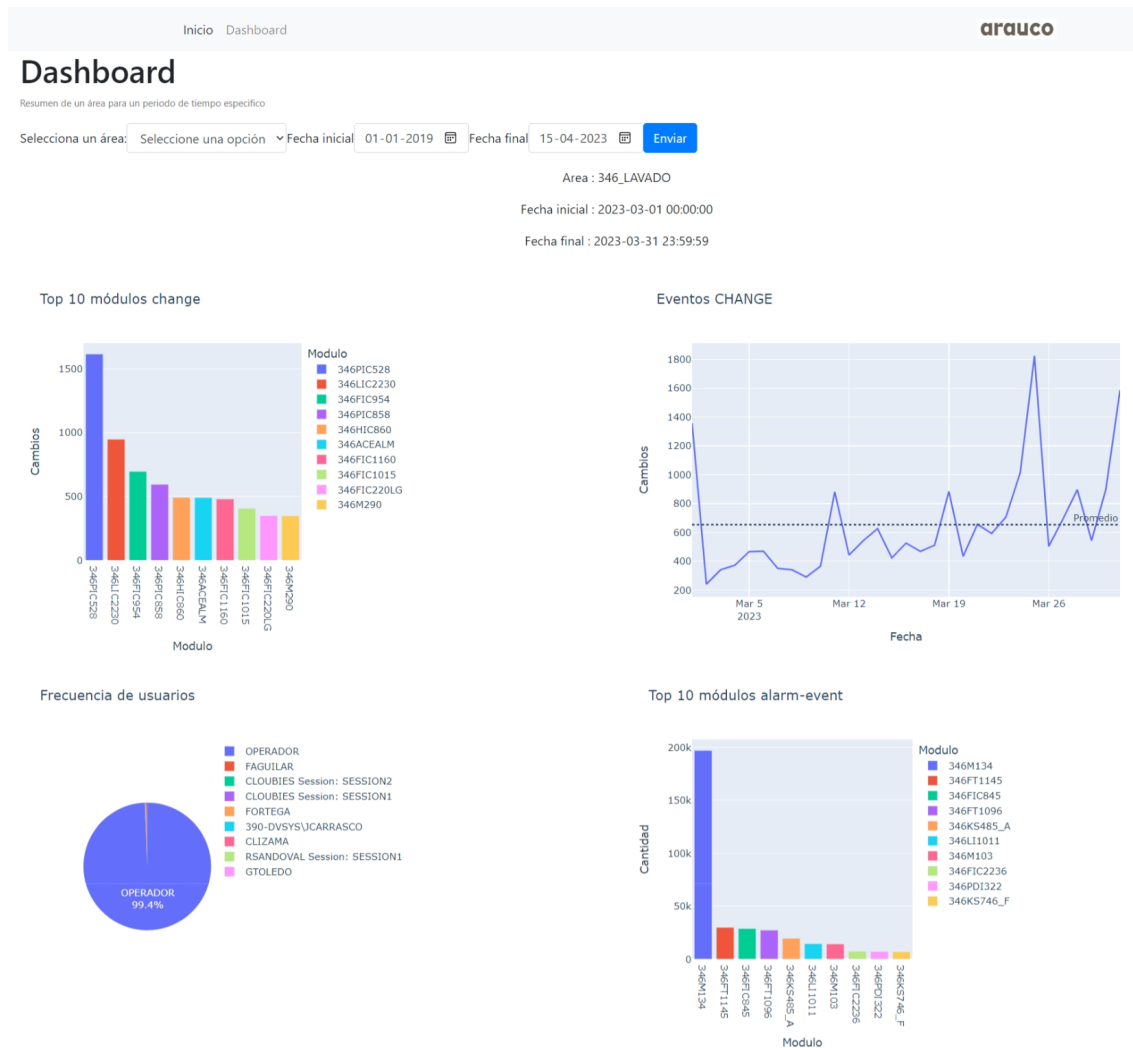


Figura 24: Resultados de consulta en dashboard.
(Elaboración propia)

3.8 Proceso de carga web

Como se mencionó anteriormente en el capítulo 3.4 la carga de archivos se realiza ejecutando un procedimiento almacenado desde el sistema gestor de base de datos, para tener el acceso a la ejecución del procedimiento son necesarias las credenciales para la autenticación y esto supone un riesgo para la integridad de la información en la base de datos para los usuarios que accedan y realicen modificaciones.

Para evitar este acceso no controlado, se diseñó una página web que permite seleccionar los archivos que se desean cargar a la base de datos, por requerimientos de la organización los archivos son copiados a la ruta 'D:\DB\DATA' en el equipo NAS. Luego, cuando se ingresa a la sección de carga de la plataforma se envía una consulta para conocer qué archivos están presentes en esta carpeta, cuidando que el archivo origen que contiene la base de datos DB004 y los archivos que no poseen la extensión MDF no sean mostrados por el selector.

Los nombres de los archivos presentan una nomenclatura que se puede diferenciar del resto,

390-APP-004 # 2023 04 01 10 00 28 # 2023 04 15 17 00 16

En el primer apartado, se indica la categoría APP-004, continuado por la fecha de inicio del registro, en este ejemplo es 2023-04-01 10:00:28, y luego la fecha final del registro 2023-04-15 17:00:16.

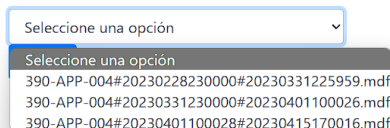
Siguiendo este patrón, el siguiente archivo a cargar debe continuar con una fecha de inicio menor o igual al de la fecha final del archivo anterior o la fecha indicada en la plataforma como se aprecia en la figura 25. Es decir, el siguiente archivo puede comenzar con 390-APP-004#20230414... siguiendo el ejemplo anterior.

Carga de archivos APP-04

Proceso de carga de archivos MDF para integrar a la base de datos

Ultima fecha ingresada APP-04: 2023-04-15

Selecciona un archivo:



Seleccione una opción

- 390-APP-004#20230228230000#20230331225959.mdf
- 390-APP-004#20230331230000#20230401100026.mdf
- 390-APP-004#20230401100028#20230415170016.mdf

Figura 25: Selección de archivo para cargar mediante plataforma.
(Elaboración propia)

Ya seleccionado el archivo, se procede a seleccionar el botón "Cargar". Esto ejecutará el procedimiento almacenado creado en el capítulo 3.3 y dejará a la página cargando hasta que el proceso finalice y la fecha de ingreso se actualice. No existe problema en que un archivo sea ingresado más de una vez, pues los datos duplicados no son ingresados, para evitar esto el archivo luego de ser cargado a la base de datos es movido a una carpeta externa para no ser desplegado en el selector de carga.

En el caso extraordinario que se realice una carga de un archivo posterior a la última fecha de datos, existirá un espacio de tiempo sin información que puede afectar el resultado de las consultas, pero cuando ya se cuente con el archivo este puede ser cargado sin problemas. Una alternativa para dar cuenta que falta información es observar la línea temporal, pues esta no tendrá datos en la fecha faltante.

3.9 Eliminación de datos

Con una gran cantidad de información el espacio del disco se ve comprometido y limitado a la integración de nuevos datos, considerando que entre los periodos de enero de 2019 y abril de 2023 el tamaño de la base de datos "DB004" es de 432,5 GB y cada archivo con datos mensuales pesa alrededor de 4 GB, esto significa que a medida que se carguen más archivos el tamaño de la base de datos crecerá considerablemente ocurriendo en un futuro la necesidad de eliminar información para liberar el espacio. Para hacer esto posible se diseñó una página web, como el de la figura 26, para poder realizar este proceso para el usuario, en el que indique el intervalo de tiempo que desee borrar para la base de datos deseada.

The screenshot shows a web application interface with a light gray header containing 'Inicio Dashboard' on the left and the 'arauco' logo on the right. The main content area is divided into two columns. The left column is titled 'Eliminación de datos APP-001' and contains the text 'Eliminación de información en base de datos APP-001'. Below this, it says 'Periodo de información APP-001:' followed by two date input fields, both set to '01-01-2019', each with a calendar icon. A red 'Eliminar' button is at the bottom. The right column is titled 'Eliminación de datos APP-004' and contains the text 'Eliminación de información en base de datos APP-004'. Below this, it says 'Periodo de información APP-004 desde: 2019-01-01 al 2023-04-15'. It also has two date input fields, both set to '01-01-2019', each with a calendar icon, and a red 'Eliminar' button at the bottom.

Figura 26: Página de eliminación de información para bases de datos.
(Elaboración propia)

La página verifica que el usuario haya escrito correctamente la fecha, siendo la fecha inicial menor que la fecha final y enviando una alerta en el navegador para asegurar que el usuario está seguro de querer ejecutar esa acción.

4. Resultados

Utilizando los KPIs diseñados se pueden conocer los resultados para las áreas de interés, en este caso se verán los resultados del área de Lavado, un área de interés para el área de confiabilidad de procesos para conocer su rendimiento. Estos resultados corresponden a 3 trimestres diferentes que fueron consultados en el Dashboard, el primer periodo 01/06/22 al 30/09/22 llamado T3-2022, 01/10/22 al 31/12/22 llamado T4-2022 y 01/01/23 al 31/03/2023 llamado T1-2023.

En la figura 26, podemos observar cómo varía el nivel de intervenciones a lo largo del tiempo, en el T3-2022 se ve una disminución en el periodo de Julio-Agosto 2022, esto concuerda con lo reportado, pues en este periodo las áreas de la planta de celulosa estuvieron detenidas para la producción y es en Septiembre 2022 que comienza su partida nuevamente.

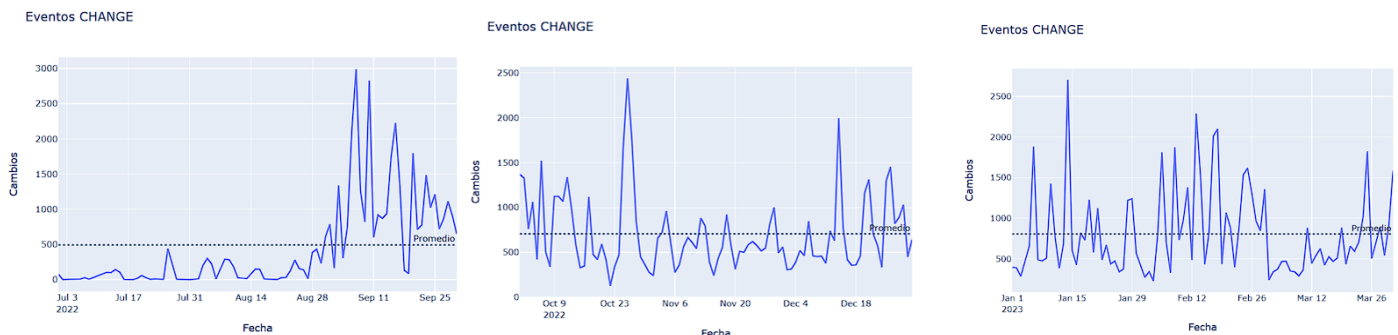
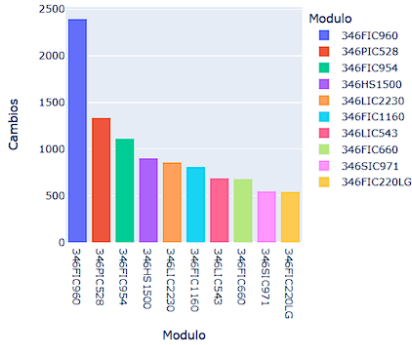


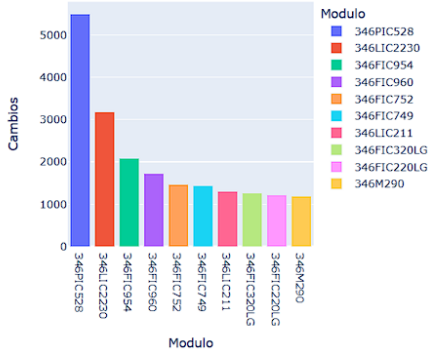
Figura 26: Línea temporal de cambios en periodo T3-2022 - T4-2022 - T1-2023 de izq. a der. (Elaboración propia)

Para la figura 27 podemos observar los módulos intervenidos por los usuarios que más suelen tener cambios en el área de lavado, esto nos permite conocer si hay módulos que se repiten en los distintos periodos de tiempo, caso que efectivamente sucede para los módulos 346FIC960, 346LIC2230, etc. Así, esta información permite dar el comienzo a investigar en más profundidad, por qué los usuarios siguen interviniendo estos módulos, o si es posible conseguir una disminución en la intervención. Lo mismo ocurre para las diferentes alarmas y eventos que ocurren en la planta, como se aprecia en la figura 26, existen módulos que siguen en presencia a lo largo de los diferentes periodos de tiempo, lo que hace cuestionar si estas alarmas se pueden lograr disminuir en frecuencia de apariciones.

Top 10 módulos change



Top 10 módulos change



Top 10 módulos change

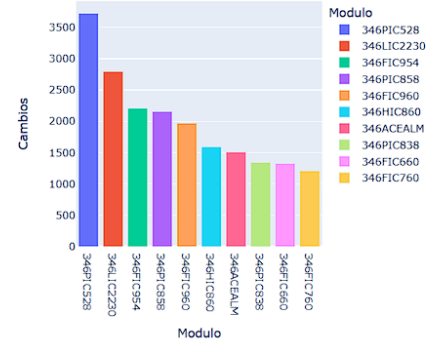
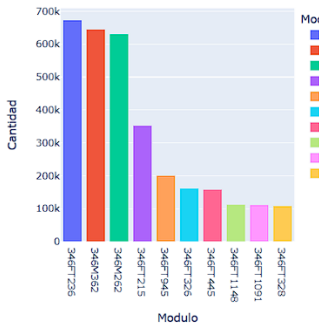
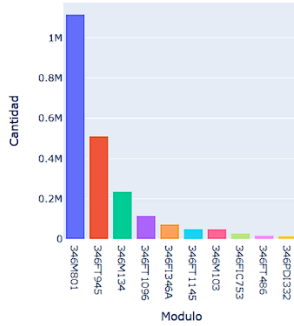


Figura 27: Módulos más frecuentes para los periodos T3-2022 - T4-2022 - T1-2023 de izq. a der. (Elaboración propia)

Top 10 módulos alarm-event



Top 10 módulos alarm-event



Top 10 módulos alarm-event

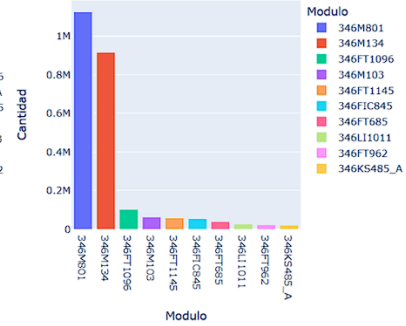
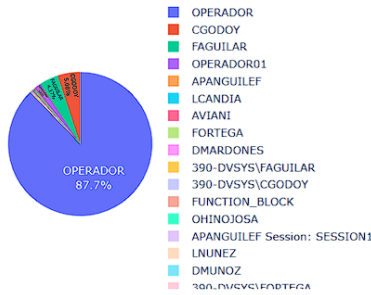


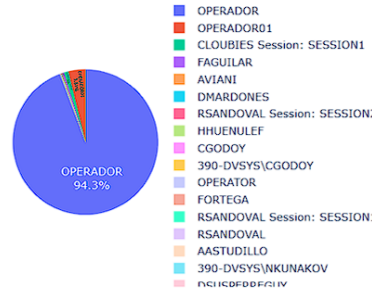
Figura 28: Eventos y alarmas más frecuentes para los periodos T3-2022 - T4-2022 - T1-2023 de izq. a der. (Elaboración propia)

En la figura 28 se aprecia la frecuencia de los usuarios que han intervenido en el área de lavado, con esto el área puede conocer bajo qué usuario se han realizado cambios y la cantidad efectuada. La identificación de los usuarios es clave para determinar a los autores, pues “OPERADOR” y “OPERATOR” son los nombres por los cuales los funcionarios deben operar y no los otros que aparecen en la figura 29. De esta manera, podemos ver un aumento positivo para el usuario “OPERADOR”, pasando de tener una participación del 87.7% al 99.2%, siendo estos resultados gratos para la organización.

Frecuencia de usuarios



Frecuencia de usuarios



Frecuencia de usuarios

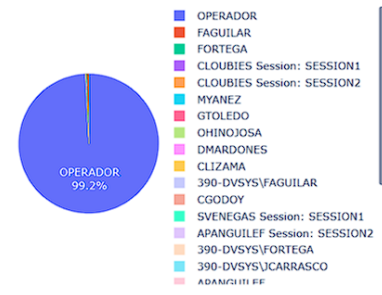


Figura 29: Usuarios más frecuentes para los periodos T3-2022 - T4-2022 - T1-2023 de izq. a der. (Elaboración propia)

Al analizar los resultados en el Dashboard, podemos obtener una visión general y concisa del comportamiento del área en cuestión. Esta representación reducida de información nos permite adquirir una comprensión rápida de la situación, lo que nos lleva a utilizar las secciones diseñadas de cada KPI para examinar con mayor detalle los módulos de usuarios intervenidos y las alarmas generadas.

Estas herramientas, en conjunto, se complementan para brindarnos una visión integral de la información, lo que nos permite llevar a cabo un análisis preciso de los datos más críticos que requieren atención especial. Asimismo, nos proporcionan una visión del rendimiento de las áreas y, por lo tanto, del nivel de automatización, que puede evaluarse mediante los distintos indicadores clave de rendimiento (KPI) obtenidos.

4. Conclusión

El desarrollo e implementación de la herramienta informática propuesta para medir el nivel de automatización en la planta ha logrado cumplir de manera exitosa tanto el objetivo general como los objetivos específicos planteados en esta memoria. La herramienta diseñada y construida permite medir la cantidad y el nivel de intervenciones humanas, al igual que los eventos de alarmas en la Planta de Celulosa Arauco Valdivia. En cuanto a los objetivos específicos, se logró identificar información relevante gracias a la exploración y relaciones entre los datos con el fin de alcanzar los objetivos organizacionales. La construcción de la base de datos ha proporcionado un sustento sólido para el análisis y seguimiento de los indicadores clave de rendimiento (KPIs) y la información crítica identificada. La plataforma web a su vez ha permitido desplegar la información mediante búsquedas accediendo a la base de datos y la implementación del dashboard ha brindado una visualización resumida de los distintos KPIs, facilitando la comprensión y el monitoreo de la información relevante, quedando la plataforma web operativa para los usuarios del área de confiabilidad.

El análisis de datos ha sido fundamental para generar conocimiento significativo que respalde la toma de decisiones informada, esto hace que la herramienta diseñada sea de gran valor por parte de los usuarios, pues permite comprender el comportamiento de los operadores en las distintas áreas y tener la oportunidad de incorporar mejoras en los eventos que son más frecuentemente modificados por los funcionarios de la planta y así lograr la eficiencia y la efectividad de las operaciones en la Planta de Celulosa Arauco Valdivia. Además, le permitirá a los usuarios tomar valores referenciales de otros períodos históricos para poder generar comparaciones y medir su rendimiento en el área.

Para finalizar, a lo largo de este desarrollo en el que se ha examinado y evaluado diferentes opciones para la implementación, se presentan oportunidades para mejorar tanto la ejecución como el análisis de los datos a través de la exploración de otras herramientas. Aunque las herramientas utilizadas para obtener un producto mínimo viable demostraron ser efectivas para la recolección y visualización de datos, considero que la incorporación de herramientas más avanzadas y especializadas podría haber permitido una mayor profundidad en la interpretación y en la generación de análisis de los datos, pues el tiempo destinado para el desarrollo y construcción de la plataforma se hubiese utilizado para estos fines.

5. Trabajo futuro

- Realizar un proceso de extracción en los archivos para evitar que una persona encargada deba realizar transferencias mediante dispositivos físicos de almacenamiento, esto supondría una integración local entre el equipo que genera constantemente los archivos MDF con el equipo NAS teniendo que desplazar el equipo NAS a la planta de Celulosa Valdivia para ser conectado a la red DCS y generar un enlace de comunicación. Con lo anterior ya realizado se deberá corroborar la integridad y la continuidad de las fechas en los archivos para prevenir problemas.
- La información contenida en la base de datos está compuesta por archivos de la categoría APP-004. En esa base de datos se encuentran registradas algunas áreas de la planta de celulosa de Valdivia. Con el fin de obtener la totalidad de las áreas, es posible integrar los archivos de la categoría APP-001 a una nueva base de datos. Esto permitiría llevar a cabo los procesos de ELT y obtener la visualización de la información de las áreas faltantes, logrando así generar un reporte completo de la planta.
- Integrar nueva funcionalidad en la plataforma, poder realizar la comparación en diferentes periodos de tiempos para módulos, esto con la finalidad de obtener un reporte en dos periodos de tiempos y verificar tendencias en módulos comunes para tener un nuevo indicador de disminución o no de eventos.
- La implementación de un buscador en la plataforma web representa una funcionalidad importante al permitir a los usuarios realizar búsquedas de cualquier tipo en todos los campos disponibles de la tabla de la base de datos, se amplía la capacidad de extracción de información específica y la libertad para explorar y acceder a datos relevantes. Esto permite a los usuarios ingresar consultas claras y obtener resultados relevantes, superando las limitaciones de las búsquedas e Además, la inclusión de opciones avanzadas de filtrado y refinamiento de resultados permite una mayor personalización de las búsquedas, lo que facilita la exploración exhaustiva y la localización precisa de información dentro de la plataforma web.
- Paralelismo utilizando threads para múltiples consultas a la base de datos, esto permitiría poder ejecutar más de una consulta sin tener que realizar una consulta primero y luego esperar a que esta termine para poder realizar otra. Realizar esto implica tener que realizar una mejora en el hardware del equipo de base de datos para poder procesar más rápido las consultas. El realizar múltiples consultas también hace que se envíe más información al equipo host por red, por lo que se debería evaluar el tráfico de red para la cantidad de información enviada.
- Escalamiento del sistema a otras plantas de Celulosa ubicadas en Chile, replicando la metodología de extracción para archivos históricos de eventos efectuados por los operadores y las alarmas para luego tener la visualización de KPIs. Esto permitiría entregar información transversal sobre el rendimiento de las plantas.

6. Bibliografía

[1] Celulosa Arauco y Constitución S.A. Reporte Integrado, 2022. [Online]. Available: arauco.cl/chile/wp-content/uploads/sites/14/2023/04/REPORTE_INTEGRADO_ARAUCO.pdf [Accessed Jun. 07,2023].

[2] Ynzunza Cortés, Carmen Berenice; Izar Landeta, Juan Manuel; Bocarando Chacón, Jacqueline Guadalupe; Aguilar Pereyra, Felipe; Larios Osorio, Martín. El Entorno de la Industria 4.0: Implicaciones y Perspectivas Futuras, 2017. [Online]. Available: <https://www.redalyc.org/journal/944/94454631006/94454631006.pdf> [Accessed Jun. 07, 2023].

[3] Petroc Taylor. Amount of data created, consumed, and stored 2010-2020, with forecasts to 2025 . [Online]. Available: <https://www.statista.com/statistics/871513/worldwide-data-created/> [Accessed Jun.9, 2023].

[4] Kurt Schlegel, Julian Sun, David Pidsley, Anirudh Ganeshan, Fay Fei, Aura Popa, Radu Miclaus, Edgar Macari, Kevin Quinn, Christopher Long. Magic Quadrant for Analytics and Business Intelligence Platforms. [Online]. Available: https://www.gartner.com/doc/reprints?id=1-2955ETOT&ct=220215&st=sb?ocid=lp_pg398450_gdc_comm_az [Accessed Jul. 22, 2023].

[5] Microsoft. Por qué Microsoft Power BI. [Online]. Available: <https://powerbi.microsoft.com/es-es/why-power-bi/> [Accessed Jul 20, 2023]

[6] Tableau Software. ¿Por qué elegir Tableau? [Online]. Available: <https://www.tableau.com/es-es/why-tableau> [Accessed Jul. 20, 2023]

[7] Qlik. ¿Qué es QlikView?. [Online]. Available: https://help.qlik.com/es-ES/qlikview/May2023/Content/QV_HelpSites/QV_AtAGlance.htm [Accessed Jul.24, 2023]

[8] International Business Machines Corporation. ¿Qué es Business Intelligence y cómo funciona? [Online]. Available: <https://www.ibm.com/mx-es/topics/business-intelligence> [Accessed Jul. 19, 2023].

[9] Sarah Laoyan. KPIs: ejemplos de 27 indicadores clave para el éxito de tu proyecto. [Online]. Available: <https://asana.com/es/resources/success-metrics-examples> [Accessed Jul. 19, 2023].

[10] Tableau. Qué debe saber sobre los dashboards de BI. [Online]. Available: <https://www.tableau.com/es-es/learn/articles/business-intelligence/bi-dashboards> [Accessed Jul. 20, 2023].

[11] Emerson Electric Co. Sistema de control distribuido (DCS) DeltaV. [Online]. Available: <https://www.emerson.com/es-cl/automation/control-and-safety-systems/distributed-control-systems-dcs/deltav-distributed-control-system> [Accessed Jul. 18, 2023].

[12] Raunak Jhavar. Extract, transform, and load (ETL). [Online]. Available: <https://learn.microsoft.com/es-es/azure/architecture/data-guide/relational-data/etl> [Accessed Jul. 17, 2023].

[13] Pallets Project. Flask Documentation 2.3.x [Online]. Available: <https://flask.palletsprojects.com/en/2.3.x/>. [Accessed Jul. 17, 2023].

[14] M. Pérez Marqués, *Microsoft SQL Server 2008 R2. Motor de base de datos y administración*. Madrid: RC Libros, 2011.

7. Anexos

7.1 Historias de usuario

1. Como usuario deseo una plataforma web para poder ver la información contenida en una base de datos.

Tarea	Descripción
Diseño de plataforma	Creación de página utilizando framework Flask, asignación de método request get.
Conexión con base de datos	Establecer enlace a base de datos desde Python con el uso de librería pyodbc.
Consultas de prueba en SQL desde Python	Creación de consultas SQL para probar conexión y visualizar en terminal los resultados.
Desarrollar plantilla HTML	Diseño de plantilla con jinja2, integrar navbar y vincular datos obtenidos desde la desde la consulta asignando variables.
Desplegar información de consulta en página web	Mostrar el resultado de la consulta SQL mediante una tabla y gráfica usando librería plotly.

- Como usuario quiero conocer la cantidad y la frecuencia de los módulos que han sido modificados por un operador en un área y tiempo específico para llevar un control de las modificaciones realizadas.

Tarea	Descripción
Creación y prueba de consulta SQL	Diseño de consulta SQL que recupere la información solicitada, luego realizar test para verificar la exactitud de los datos.
Desarrollo de sección en Flask	Asignación de ruta, métodos request (post,get) que capture la fecha y área, integración de consulta, selección de información relevante con dataframe.
Creación de plantilla y añadido de tablas y gráficos.	Diseño de plantilla con jinja2, recepción de información desde la consulta e integración de tabla y gráfico de barras usando plotly.

- Como usuario quiero ver la cantidad y frecuencia de los módulos de eventos de alarmas en un periodo de tiempo y área específico para identificar posibles problemas.

Tarea	Descripción
Creación y prueba de consulta SQL.	Diseño de consulta SQL que recupere la información solicitada, luego realizar pruebas para verificar la exactitud de los datos.
Desarrollo de sección en Flask	Definición de ruta, métodos de request (post, get) para obtener fecha y área, integración de consulta, selección y filtro de datos relevante con dataframe.
Creación de plantilla y añadido de tablas y gráficos.	Diseño de plantilla con jinja2, recepción de información desde dataframe, luego visualización de datos con uso de tabla y gráfico de barras con librería plotly.

- Como usuario quiero conocer los nombres de usuario y la cantidad de cambios que han realizado en los módulos para un área y tiempo determinado para conocer qué usuarios están realizando modificaciones.

Tarea	Descripción
Creación y prueba de consulta SQL.	Diseño de consulta SQL que recupere la información solicitada y realizar pruebas para verificar la exactitud de los datos.
Desarrollo de sección en Flask	Definición de ruta e integración de consulta, asignación de métodos (post, get) para obtener fechas y área, filtro y selección de datos con uso de dataframe.
Creación de plantilla y añadido de tablas y gráficos.	Diseño de plantilla con jinja2, recepción de información de dataframe e integración de tabla y gráfico circular con librería plotly.

5. Como usuario quiero ver qué atributos tiene asociado el módulo que ha sufrido cambios o alarmas para identificar que se ha modificado.

Tarea	Descripción
Creación y prueba de consulta SQL	Diseño de consulta SQL y pruebas para corroborar la exactitud de la información.
Desarrollo de sección en Flask	Definición de ruta, asignación de método request (get) e integración de consulta, desarrollo de lógica para agrupar información con dataframe.
Creación de plantilla y añadido de tabla.	Diseño de plantilla con jinja2, recepción de información desde dataframe, luego Integración de tabla.

6. Como usuario quiero poder almacenar datos históricos para poder consultar la información que contiene.

Tarea	Descripción
Creación de base de datos	Integración de archivos a base de datos, selección y orden de entrada para cargar.

Proceso de carga de archivos mediante SGBD	Diseño de procedimiento almacenado para cargar desde el SGBD.
Transformación de los datos	Eliminación de duplicados y campos vacíos, asignación de clave primaria y left join para ingresar datos que no estén contenidos en la base de datos.

7. Como usuario quiero un resumen general con gráficos donde pueda ver los módulos de alarmas e intervenciones más frecuentes y los usuarios que más han realizado cambios dada un área y periodo de tiempo para tener una visión más general de lo que ocurre.

Tarea	Descripción
Creación y prueba de consulta SQL	Diseño de consulta SQL y pruebas para corroborar la exactitud de la información.
Desarrollo de sección en Flask	Definición de ruta e integración de consulta, métodos request (post,get) para la obtención de fechas y áreas ingresadas, desarrollo de lógica para agrupar información.
Creación de plantilla y añadido de gráficos.	Diseño de plantilla con jinja2, recepción de información desde las consultas e integración de los gráficos.

8. Como usuario quiero poder cargar nuevos archivos a la base de datos desde la plataforma web para tener la información actualizada de las áreas de la planta.

Tarea	Descripción
Diseño de procedimiento almacenado	Selección de carpeta para el ingreso de los archivos y ejecución de comando para ver el contenido de la carpeta. Luego creación de procedimiento almacenado que permita adjuntar datos, transformar y cargados a la base de datos origen.
Creación de página	Diseño de plantilla con jinja2 y html, indicador de fecha de los últimos datos cargados y un selector para escoger el archivo a cargar.

9. Como usuario quiero poder descargar un reporte en formato CSV de los módulos más frecuentes y los cambios que hacen los usuarios para poder analizarlos con detalles.

Tarea	Descripción
Modificar plantilla moduleDetails.html	Modificación de dataframe existente para la consulta, cambio en la estructura y tipos de datos. Asignación de botón para ejecutar conversión de dataframe a CSV. Configuración para que archivo sea descargado en el equipo

10. Como usuario quiero poder eliminar información de la base de datos desde la plataforma web para mantener espacio disponible en el disco para futuras cargas de datos.

Tarea	Descripción
Desarrollo de sección en flask	Definición de ruta y métodos request (get y post) para la obtención de fechas, luego conexión con base de datos para ejecutar consulta SQL.
Plantilla HTML	Diseño de las secciones para eliminar información en ambas bases de datos, corroboración de que las fechas sean correctas de inicio a fin, y confirmación para eliminar datos.

7.2 Diagrama de casos de uso

