

**Automatización de informes de disponibilidad de infraestructura de red mediante
plataforma de inteligencia de negocios**

Wilmar Andrés Jaramillo González

Asesor y Director

Rafael Gaitán Ospina

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas Tecnología e Ingeniería ECBTI

Especialización en Ciencia de Datos y Analítica

2025

Resumen

Este proyecto se centra en crear una plataforma automatizada que facilita la extracción, transformación y generación de informes sobre la disponibilidad de la infraestructura de red. El objetivo es optimizar los procesos de monitoreo y mejorar la toma de decisiones operativas dentro de la organización. La solución integra datos de PRTG Network Monitor a través de un flujo ETL desarrollado en Python, presentando la información en paneles interactivos mediante Power BI. La metodología utilizada combina el Team Data Science Process (TDSP) para un tratamiento estructurado de los datos y las mejores prácticas de ITIL v4 para la gestión del servicio.

Los resultados muestran una significativa reducción del 87.5 % en el tiempo necesario para generar los informes, pasando de un proceso manual que tomaba alrededor de 10 horas a uno automatizado que solo requiere 1.2 horas para un total de 989 dispositivos. Además, se ha mejorado la precisión, consistencia y disponibilidad de la información, lo que reduce el riesgo de errores humanos y fortalece la continuidad operativa. En resumen, el proyecto resalta la efectividad de la automatización como una estrategia clave para mejorar la eficiencia, calidad y puntualidad de los informes de red en la gestión tecnológica de la empresa.

Palabras clave: Python, Inteligencia de Negocios, TDSP, ITIL v4, Monitoreo de Red, Automatización ETL

Abstract

This project presents the development of an automated platform for extracting, transforming, and generating network infrastructure availability reports, aimed at optimizing monitoring processes and operational decision-making within the organization. The solution integrates data from PRTG Network Monitor through an ETL workflow implemented in Python and delivers the results through interactive dashboards in Power BI. The methodology combines the Team Data Science Process (TDSP) for structured data handling and ITIL v4 best practices for service management.

The results demonstrate an 87.5% reduction in the time required to generate reports, decreasing from a manual process of approximately 10 hours to an automated process of 1.2 hours for a total of 989 devices. Additionally, the platform increases data accuracy, consistency, and availability, minimizing human error and strengthening operational continuity. Overall, the project validates automation as an effective strategy to enhance efficiency, quality, and timeliness in network reporting within the organization's technological management processes.

Keywords: Python, Business Intelligence, TDSP, ITIL v4, Network Monitoring, ETL Automation

Tabla de Contenido

Introducción	8
Justificación	9
Objetivos.....	10
Objetivo General.....	10
Objetivos Específicos	10
Marco de Referencia	11
Estado del Arte	11
Marco Contextual	12
Marco Teórico	13
Marco Conceptual.....	15
Marco Normativo	17
Normativa Nacional.....	17
Normas Internacionales y Estándares Técnicos.....	18
Metodología	20
Implementación del Sistema ETL y Automatización de Datos	24
Arquitectura de Integración de Datos	24
Estrategia de Extracción de Datos (Ingesta).....	26
Configuración de la Petición	26
Algoritmos de Transformación y Limpieza (Data Wrangling)	27
Modelo de Datos y Carga (Loading)	28
Desarrollo de Inteligencia de Negocios	30
Modelado de Datos (Esquema en Estrella).....	30

Ingeniería de Cálculos (Lenguaje DAX)	31
Cálculo de Disponibilidad General.....	31
Semáforos y KPIs Condicionales	32
Diseño de Interfaz y Visualización de Datos.....	32
Interactividad y Filtros Dinámicos	34
Evaluación de Resultados	35
Definición de Métricas de Evaluación.....	35
Análisis Comparativo de Eficiencia Operativa	35
Análisis de los Resultados	36
Evaluación de Escalabilidad.....	37
Validación de Precisión y Calidad del Dato	38
Conclusiones	39
Referencias Bibliográficas	40

Lista de Figuras

Figura 1 <i>Topología Proyecto</i>	23
Figura 2 <i>Arquitectura ETL</i>	25
Figura 3 <i>Algoritmo de Conexión a la API de PRTG</i>	27
Figura 4 <i>Modelo de Datos Relacional en Power BI</i>	31
Figura 5 <i>Función Promedio Disponibilidad</i>	32
Figura 6 <i>Funcion Estado SLA</i>	32
Figura 7 <i>Dashboard Power BI</i>	33
Figura 8 <i>Impacto de la Automatización en Tiempos de Respuesta</i>	37

Lista de Tablas

Tabla 1 <i>Resumen Flujo ETL</i>	25
Tabla 2 <i>Estructura Tabla DevicesAvailability</i>	28
Tabla 3 <i>Comparación del Tiempo de Procesamiento</i>	36

Introducción

En el mundo empresarial de hoy, tener una infraestructura de red estable y disponible es clave para asegurar que las operaciones sigan funcionando sin problemas y que los servicios tecnológicos sean eficientes. Este reto se vuelve aún más importante en organizaciones que manejan un gran número de dispositivos y diferentes entornos de clientes, donde los métodos manuales de monitoreo y generación de informes pueden ser lentos, propensos a errores y difíciles de mantener.

En el caso de la organización que estamos analizando —una empresa que se dedica a gestionar y supervisar la infraestructura de red para varios clientes corporativos— estas tareas requieren una gran cantidad de tiempo y recursos, lo que complica la obtención de información actualizada sobre el estado real de los dispositivos que se están monitoreando. Frente a esta situación, se hace evidente la necesidad de implementar mecanismos automatizados que optimicen la recolección, transformación y análisis de datos operativos. Esto permitirá contar con informes precisos y actualizados que apoyen la toma de decisiones. En respuesta a esta necesidad, este proyecto propone desarrollar e implementar una plataforma de automatización que mejore el proceso de generación de informes sobre la disponibilidad de la infraestructura de red, integrando técnicas de ciencia de datos y herramientas de inteligencia de negocios.

La creación de esta solución se basa en metodologías como el Team Data Science Process (TDSP), que ofrece un marco estructurado para el manejo de datos, y ITIL v4, que orienta la gestión eficiente de los servicios de TI dentro de la organización. Con esto, el proyecto busca fortalecer las capacidades operativas y analíticas de los equipos técnicos, reducir significativamente los tiempos de procesamiento de la información y avanzar hacia una cultura organizacional que valore la automatización, la precisión y el uso estratégico de los datos

Justificación

En el mundo tecnológico de hoy, la disponibilidad y el rendimiento de la infraestructura de red son esenciales para que cualquier organización funcione de manera eficiente. Sin embargo, los métodos manuales para recolectar y consolidar datos de herramientas de monitoreo a menudo causan retrasos, errores humanos y reportes con información desactualizada. Esto complica la detección temprana de incidentes y limita la capacidad de los equipos técnicos para tomar decisiones rápidas y fundamentadas (Paessler AG, 2023).

Este proyecto surge de la necesidad de automatizar la extracción, transformación y análisis de datos sobre la disponibilidad de la infraestructura de red. Para ello, se desarrollará una plataforma en Python que integrará información de diversas fuentes, como PRTG Network Monitor, y la procesará a través de un flujo ETL (Extraer, Transformar, Cargar). Los datos se presentarán mediante una herramienta de inteligencia de negocios (Power BI), que permitirá visualizar indicadores clave y reportes dinámicos actualizados en tiempo real (Kimball & Caserta, 2004; Inmon, 2005).

Desde un enfoque técnico, esta propuesta optimiza la gestión operativa al reducir el tiempo necesario para generar informes, mejorar la precisión de los datos y fortalecer la supervisión de la red. En términos metodológicos, se aplican buenas prácticas del Team Data Science Process (TDSP) para el manejo de datos y de ITIL v4 para la gestión de servicios de TI, asegurando un enfoque estructurado y sostenible.

En el contexto organizacional, la plataforma ayuda a mejorar la eficiencia y agilidad en la toma de decisiones operativas, ya que ofrece información confiable y oportuna sobre el estado de la red. Así, se impulsa la transformación digital, la optimización de recursos tecnológicos y la mejora continua de los procesos de monitoreo y gestión de infraestructura.

Objetivos

Objetivo General

Implementar una plataforma que automatice la extracción de información sobre la disponibilidad de infraestructura de red desarrollada en Python, para mejorar la generación de informes mediante una herramienta de inteligencia de negocios garantizando agilidad en la gestión de decisiones operativas.

Objetivos Específicos

Implementar un sistema automatizado de recolección, limpieza y transformación de datos de infraestructura de red provenientes de herramientas de monitoreo (PRTG) y sistemas de gestión, aplicando técnicas de ETL para garantizar la calidad e integridad de los datos.

Diseñar y desarrollar informes mediante una plataforma de inteligencia de negocios (Power BI) que integre métricas y dashboards interactivos.

Evaluar la efectividad de la plataforma implementada mediante métricas de precisión, efectividad de la aplicación y optimización de tiempo.

Marco de Referencia

Estado del Arte

La automatización de informes en el ámbito de la infraestructura tecnológica ha sido un tema muy explorado desde la óptica de la inteligencia de negocios (BI) y la ciencia de datos. Muchos autores coinciden en que al implementar herramientas de BI, se puede integrar, analizar y visualizar grandes volúmenes de información en tiempo real, lo que facilita la toma de decisiones estratégicas.

Roque Vargas y Cadillo Montesinos (2022) crearon una solución de inteligencia de negocios para una empresa de transporte interprovincial en Perú, con el objetivo de automatizar la gestión de información de ventas utilizando Power BI y Pentaho Data Integration. Los resultados mostraron una notable reducción en los tiempos de generación de reportes, evidenciando el impacto positivo que la automatización tiene en la eficiencia operativa.

De manera similar, Vele Arpi (2020) implementó un sistema web de BI en la Red de Instituciones Financieras de Desarrollo, aplicando las metodologías CRISP-DM y Scrum. Este proyecto permitió automatizar los procesos de extracción, transformación y carga (ETL), mejorando el acceso a la información y optimizando la toma de decisiones.

Por otro lado, Susanibar Bazalar (2020) examinó la relación entre los sistemas de inteligencia de negocios y la automatización de reportes en la empresa Adecco, concluyendo que la disponibilidad y precisión de la información tienen un impacto directo en la calidad de las decisiones empresariales.

Estos estudios resaltan lo importante que son las herramientas de BI para mejorar la gestión de la información en las organizaciones. Sin embargo, se ha notado una falta de atención hacia la automatización de los informes de disponibilidad de red, un campo donde la

combinación de metodologías de ciencia de datos con plataformas de inteligencia de negocios aún tiene mucho potencial para innovar. En este sentido, el proyecto que se presenta propone una solución centrada en la infraestructura de telecomunicaciones, utilizando metodologías TDSP e ITIL para optimizar el monitoreo y la generación automática de reportes.

Marco Contextual

El proyecto se lleva a cabo en el ámbito de la gestión de infraestructura de red en entornos empresariales, donde la disponibilidad y el rendimiento de los servicios tecnológicos son esenciales para mantener la continuidad operativa. En estos escenarios, los equipos de tecnología deben estar en constante vigilancia de parámetros como el ancho de banda, la conectividad, la latencia y la disponibilidad de los dispositivos de red, asegurando que los servicios se mantengan activos y estables.

En la actualidad, la forma en que se recopila y presenta esta información suele ser manual o mediante herramientas aisladas, lo que provoca duplicación de esfuerzos, retrasos en los informes y una capacidad de análisis limitada. Esto dificulta la obtención de indicadores actualizados y una visión completa del estado de la red.

Este proyecto se sitúa en el área de Telecomunicaciones e Infraestructura TI, y su implementación está dirigida a entornos empresariales o institucionales que demandan intensamente servicios digitales, como organizaciones educativas, servicios financieros o industrias tecnológicas. En particular, se toma como referencia la experiencia de la Escuela de Ciencias Básicas, Tecnología e Ingeniería (ECBTI) de la Universidad Nacional Abierta y a Distancia (UNAD), donde la gestión de redes y servicios tecnológicos necesita soluciones ágiles que faciliten la toma de decisiones.

El desarrollo de una plataforma para automatizar informes permitirá centralizar la información sobre el rendimiento de la red, ofreciendo tableros interactivos en Power BI y reportes actualizados en tiempo real. Este avance mejorará la capacidad de supervisión, reducirá los tiempos de respuesta ante incidentes y elevará la calidad del servicio tecnológico en el entorno institucional o empresarial donde se implemente.

Marco Teórico

El desarrollo de soluciones automatizadas para crear informes sobre la infraestructura de red se basa en varios enfoques teóricos que combinan los principios de la ciencia de datos, la inteligencia de negocios (BI) y las metodologías ágiles de gestión tecnológica. Estos enfoques permiten convertir grandes volúmenes de datos en información valiosa para tomar decisiones estratégicas dentro de las organizaciones.

El marco teórico de este proyecto se origina en la convergencia entre la ciencia abierta, la gestión efectiva de datos de investigación (Research Data Management, RDM) y la aplicación de metodologías de investigación enfocadas en la automatización de informes. Según Lluís Anglada, la gestión colaborativa del conocimiento y el uso de herramientas digitales son elementos clave para avanzar hacia sistemas automatizados en entornos de red.

García (2018) subraya la importancia de clasificar datos para predecir clases desconocidas y la precisión de esas predicciones, principios que son aplicables a la generación de informes sobre la disponibilidad de la red. Estas técnicas permiten obtener reportes precisos y actualizados, lo que aumenta la confiabilidad de los indicadores de rendimiento de la infraestructura tecnológica.

La inteligencia de negocios abarca un conjunto de estrategias, metodologías y tecnologías diseñadas para analizar datos y ofrecer información útil para la gestión organizacional. Con

herramientas como Power BI, Power Automate o Pentaho Data Integration, se puede integrar diversas fuentes de información, automatizar procesos de análisis y crear visualizaciones interactivas que facilitan la interpretación de los datos.

Diversas investigaciones respaldan el valor de la BI en la automatización de reportes. Roque Vargas y Cadillo Montesinos (2022) desarrollaron una solución de BI para una empresa de transporte, logrando automatizar la gestión de información y reducir significativamente los tiempos de generación de reportes. De igual forma, Vele Arpi (2020) implementó un sistema de BI en la Red de Instituciones Financieras de Desarrollo, empleando las metodologías CRISP-DM y Scrum, obteniendo una mejora sustancial en la accesibilidad y oportunidad de la información.

La automatización de procesos y la minería de datos son temas fascinantes en el mundo actual. La minería de datos, que forma parte de la ciencia de datos, nos ayuda a descubrir patrones, correlaciones y tendencias en grandes volúmenes de información. Según García (2018), la variedad de técnicas disponibles en minería de datos nos permite elegir los métodos más adecuados para abordar problemas específicos. Por ejemplo, en la automatización de informes, estas técnicas son clave para identificar tendencias en el rendimiento de la red y generar alertas tempranas ante posibles degradaciones del servicio.

Este proyecto se basa en la metodología Team Data Science Process (TDSP), propuesta por Microsoft, que guía todo el ciclo de vida de un proyecto de ciencia de datos, desde la comprensión del negocio hasta su implementación. Además, se incorpora el marco de referencia ITIL v4, que se centra en la gestión eficiente de servicios tecnológicos y la mejora continua del desempeño de la infraestructura.

Estas metodologías garantizan un proceso estructurado, permitiendo integrar análisis predictivos, automatización de reportes y gestión de incidentes dentro de una plataforma unificada de inteligencia de negocios.

Susanibar Bazalar (2020) llevó a cabo un análisis sobre la relación entre los sistemas de inteligencia de negocios y la automatización de reportes en la empresa Adecco. Su estudio reveló que la precisión de los resultados y la disponibilidad de la información impactan directamente en la eficacia de la toma de decisiones. Estos hallazgos subrayan la importancia de este proyecto, que busca optimizar la generación de informes sobre la disponibilidad de la red mediante herramientas de BI, asegurando así precisión, rapidez y un valor estratégico en la información proporcionada.

En conjunto, estos aportes teóricos respaldan el diseño e implementación de una plataforma automatizada de informes de red que combine la analítica de datos con la gestión de servicios TI, fortaleciendo la capacidad institucional para la toma de decisiones ágiles y basadas en evidencia.

Marco Conceptual

Este proyecto se basa en una serie de conceptos clave que guían su desarrollo tanto metodológico como técnico, todos ellos relacionados con la inteligencia de negocios, la automatización de procesos, la disponibilidad de red y las metodologías para gestionar datos y servicios.

La inteligencia de negocios es un conjunto de procesos, tecnologías y herramientas que permiten convertir datos en información valiosa para ayudar en la toma de decisiones dentro de las organizaciones. Según Sharda et al. (2018), la BI combina el análisis de datos, la minería de información, la visualización y el uso de tableros interactivos para mostrar métricas relevantes y

facilitar el seguimiento del rendimiento empresarial. En el marco de este proyecto, se implementa la BI a través de herramientas como Microsoft Power BI, que permiten integrar diversas fuentes de datos, automatizar reportes y ofrecer análisis visuales sobre la disponibilidad de la red.

La automatización de informes implica el uso de software y herramientas digitales para generar reportes de forma automática y periódica, lo que reduce la necesidad de intervención manual. Este proceso abarca la recolección, transformación y visualización de datos (ETL: Extract, Transform, Load), asegurando consistencia y rapidez en la entrega de la información. En el sector de las telecomunicaciones, su aplicación permite monitorear el estado de la infraestructura de red en tiempo real, mejorar la eficiencia operativa y anticipar posibles fallos.

La disponibilidad de red se refiere a la capacidad de los sistemas y dispositivos para mantenerse operativos y accesibles durante un tiempo determinado. Generalmente, se mide en porcentaje y depende de la estabilidad de los enlaces, la redundancia de los equipos y la efectividad de los mecanismos de monitoreo. Una alta disponibilidad asegura la continuidad del servicio y reduce el impacto de las fallas técnicas en las operaciones de una organización.

La ciencia de datos es un campo que combina estadística, programación y conocimiento del dominio para extraer información valiosa de los datos. Incluye procesos de recolección, limpieza, análisis y modelado con el objetivo de identificar patrones y tendencias. En este proyecto, se aplica la ciencia de datos para analizar los registros de disponibilidad de red, generando indicadores que apoyen la toma de decisiones técnicas.

El TDSP es una metodología creada por Microsoft para estructurar proyectos de ciencia de datos. Se compone de cinco etapas principales: comprensión del negocio, adquisición y exploración de datos, modelado, implementación y retroalimentación. Su uso asegura un enfoque

organizado para el desarrollo de soluciones analíticas y facilita la colaboración entre los equipos de ingeniería y negocio.

ITIL v4 es un marco de buenas prácticas para la gestión de servicios de TI que busca alinear los servicios tecnológicos con las necesidades del negocio. Define procesos como la gestión de incidentes, la gestión de cambios y la mejora continua. En este proyecto, ITIL guía la implementación del proceso de soporte y mantenimiento de los informes automatizados, garantizando la calidad y continuidad del servicio.

La disponibilidad de la información se refiere a la capacidad de acceder a los datos cuando se necesitan, asegurando que sean íntegros, precisos y oportunos. Este concepto es uno de los pilares fundamentales de la seguridad de la información, junto con la confidencialidad y la integridad, según la norma ISO/IEC 27001. En el marco del proyecto, garantizar la disponibilidad de la información significa que los informes generados estén siempre actualizados y al alcance de quienes gestionan la red.

Marco Normativo

El desarrollo de soluciones tecnológicas en el ámbito institucional y empresarial necesita estar en sintonía con las normas, políticas y estándares que regulan la gestión de tecnologías de la información, la protección de datos personales y la calidad del servicio. En este contexto, el proyecto que presentamos se basa en los siguientes marcos normativos y técnicos:

Normativa Nacional

Ley 1341 de 2009: Define los principios y conceptos sobre la sociedad de la información y la organización de las Tecnologías de la Información y las Comunicaciones (TIC) en Colombia.

Ley 1581 de 2012: Establece disposiciones generales para la protección de datos personales, garantizando la confidencialidad, integridad y seguridad de la información manejada en plataformas tecnológicas.

Decreto 1078 de 2015: Reglamenta parcialmente la Ley 1341 de 2009 y define lineamientos para el uso, seguridad y administración de la infraestructura tecnológica del Estado.

Ley 1266 de 2008: Regula el manejo de información financiera, crediticia, comercial y de servicios, aplicable a proyectos que involucren bases de datos.

Normas Internacionales y Estándares Técnicos

ISO/IEC 27001:2022: Norma internacional para la gestión de la seguridad de la información, que garantiza la protección de los datos procesados en la plataforma de automatización.

ISO/IEC 20000-1:2018: Establece los requisitos para la gestión de servicios de TI, alineándose con los procesos definidos en el marco ITIL utilizado en este proyecto.

ITIL v4: Marco de referencia para la gestión eficiente de servicios tecnológicos, orientado a la mejora continua y la calidad en la entrega de servicios de red.

PMBOK (Project Management Body of Knowledge): Guía de buenas prácticas en gestión de proyectos, que apoya la planificación, ejecución y control de las fases del desarrollo.

Team Data Science Process (TDSP) de Microsoft: Modelo metodológico para el desarrollo estructurado de proyectos de ciencia de datos, aplicado en la automatización y análisis de información.

El cumplimiento de estos marcos normativos asegura que la solución desarrollada se ejecute bajo criterios de seguridad, calidad y trazabilidad, garantizando la validez técnica y ética de los resultados obtenidos.

Metodología

El desarrollo de este proyecto se basa en un enfoque práctico y tecnológico, con el objetivo de diseñar e implementar una plataforma de inteligencia de negocios (BI) que automatice los informes sobre la disponibilidad de la red. Para ello, se utiliza un método mixto - cuantitativo y descriptivo - que combina el análisis de datos de rendimiento con la validación práctica del sistema que se ha implementado.

La metodología emplea los principios del Team Data Science Process (TDSP) de Microsoft, que permite un tratamiento estructurado de los datos, junto con las mejores prácticas de ITIL v4, enfocadas en la gestión de servicios de TI. Esta fusión de enfoques permite abordar el proyecto desde dos ángulos complementarios: la analítica de datos y la gestión operativa del servicio.

Comprensión del negocio

En esta etapa se realiza un levantamiento de requerimientos y un análisis de las necesidades de los usuarios respecto a los informes de disponibilidad de red.

Actividades principales:

- Identificación de indicadores y métricas clave (KPIs) de infraestructura.
- Definición de objetivos de los informes y periodicidad.
- Identificación de usuarios finales y partes interesadas.

Recolección de datos

Se obtendrán los datos desde diversas fuentes de monitoreo de red.

Herramientas empleadas:

- PRTG Network Monitor, para capturar información sobre disponibilidad, latencia, tráfico y ancho de banda.

- APIs REST, para la integración de PRTG con la plataforma diseñada.

Actividades principales:

- Configuración de conectores de datos desde PRTG hacia la base de datos SQL.
- Análisis exploratorio de datos (EDA) utilizando Python y librerías como pandas y matplotlib.

- Identificación de patrones de disponibilidad y tendencias históricas.

Limpieza y análisis de datos

Los datos recolectados se preparan para garantizar su consistencia y calidad

Herramientas empleadas:

- Python (pandas, numpy) para la limpieza y transformación de los datos.
- Power BI para la automatización de procesos ETL (Extract, Transform, Load).

Actividades principales:

- Normalización y estandarización de métricas.
- Creación de variables derivadas para análisis predictivo y correlacional.

Automatización de informes

En esta fase se desarrolla la plataforma de BI que permitirá generar reportes dinámicos e interactivos.

Plataforma:

- Microsoft Power BI Desktop para el diseño de tableros.
- Power BI Service para la publicación y actualización automática de informes.

Componentes desarrollados

- Dashboards interactivos con métricas de disponibilidad, desempeño y alertas.
- Informes programados con actualización periódica desde la fuente de datos.

Gestión de servicios (ITIL v4):

- Se aplican las prácticas de ITIL para garantizar la calidad y sostenibilidad del servicio automatizado.

- Actividades principales:
- Definición de acuerdos de nivel de servicio (SLA) para la disponibilidad de dispositivos.

- Establecimiento de procesos para la entrega y respaldo de los informes.
- Documentación de procedimientos operativos y de soporte.

Herramientas y tecnologías:

- Recolección de datos: PRTG, APIs REST, Python.
- Procesamiento de datos: Python.
- Entorno grafico: Python + Framework Django.
- Automatización: Tareas programadas.
- Almacenamiento: Microsoft SQL Database.

La evaluación de la solución desarrollada se realiza mediante pruebas comparativas y encuestas de satisfacción.

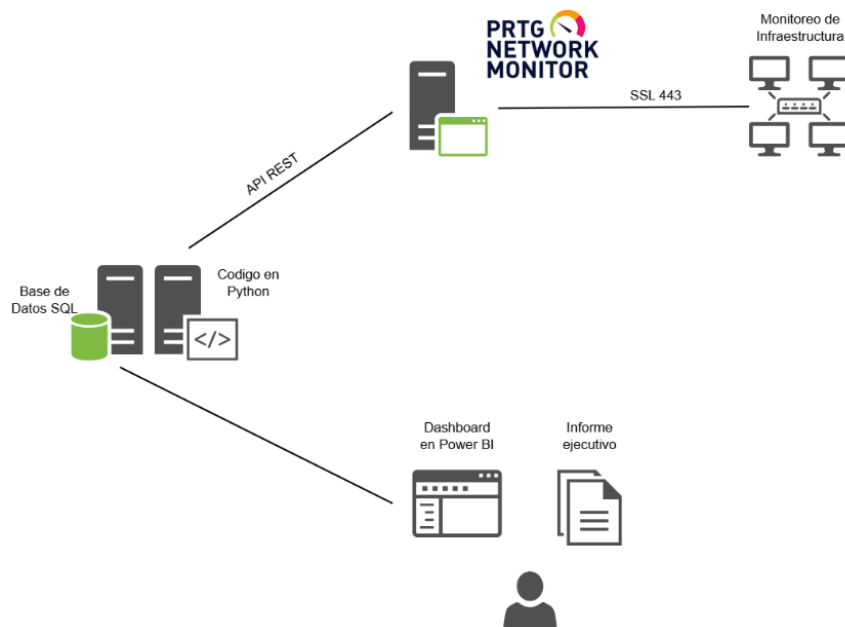
Indicadores de validación:

- Precisión: comparación de los datos reales de disponibilidad frente a los informes automatizados.
- Eficiencia: medición del tiempo requerido antes y después de la automatización.
- Satisfacción del usuario: encuestas dirigidas a los responsables de TI.

- Eficacia operativa: reducción en el tiempo de generación de reportes y detección temprana de incidentes.

Figura 1

Topología Proyecto



Nota. Arquitectura de integración que conecta una base de datos SQL y procesamiento en Python con un sistema de monitoreo (PRTG) mediante API REST y conexión segura SSL, complementado con visualización en Power BI e informes ejecutivos para la toma de decisiones.

Para dar cumplimiento a esta metodología, el desarrollo del proyecto se estructura en los siguientes capítulos técnicos:

- Implementación del Sistema ETL.
- Desarrollo de Inteligencia de Negocios.
- Evaluación de Resultados.

Implementación del Sistema ETL y Automatización de Datos

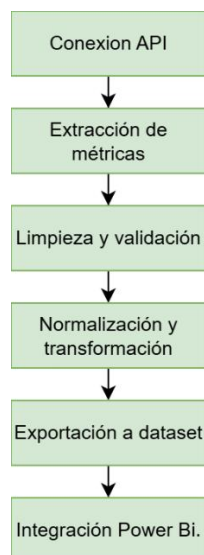
Se detalla la arquitectura de la solución, los algoritmos de extracción desde la API de PRTG y las reglas de negocio aplicadas durante la transformación de datos utilizando Python.

Arquitectura de Integración de Datos

Para garantizar la escalabilidad y la integridad de la información, se diseñó una arquitectura de tipo ETL (Extract, Transform, Load) modular. A diferencia de las conexiones directas, esta arquitectura desacopla la fuente de datos (PRTG) de la capa de visualización (Power BI), permitiendo un pre-procesamiento avanzado y almacenamiento histórico.

La solución se compone de tres capas lógicas:

- Capa de ingesta (Extract): Responsable de la comunicación con el servidor PRTG mediante protocolo HTTPS y autenticación segura.
- Capa de procesamiento (Transform): Script desarrollado en Python 3.x utilizando la librería *Pandas*, encargado de la normalización, validación de tipos y cálculo de métricas de disponibilidad.
- Capa de persistencia (Load): Base de datos Microsoft SQL Server que actúa como *Data Warehouse* intermedio, estructurando los datos para su consumo eficiente.

Figura 2*Arquitectura ETL*

Nota. Flujo de procesamiento de datos que incluye conexión vía API, extracción de métricas, limpieza, transformación y carga final en Power BI para su análisis.

Tabla 1*Resumen Flujo ETL*

Etapa	Descripción	Herramientas
Extracción	Consumo de datos desde la API REST de PRTG	Python, requests
Transformación	Limpieza, depuración, normalización y cálculo de métricas	Python (pandas, numpy)
Carga	Envío de dataset procesado a Power BI Service	Python + API Power BI
Visualización	Dashboards interactivos	Power BI Desktop / Service

Estrategia de Extracción de Datos (Ingesta)

La extracción de datos se realiza mediante el consumo de la API RESTful (o basada en XML/JSON) provista por PRTG Network Monitor. Se configuró un cliente HTTP en Python utilizando la librería `requests` para gestionar las peticiones al servidor.

Configuración de la Petición

Se definieron los siguientes parámetros para la consulta de datos históricos

("historicdata"):

- Endpoint: `https://prtg03.e-global.com.co/api/`
- Parámetros: Se filtraron los sensores por `id`, `sdate` (fecha inicio) y `edate` (fecha fin)

para optimizar el ancho de banda.

- Autenticación: Se utilizó un *passhash* seguro para evitar exponer credenciales en texto plano.

El siguiente fragmento de código ilustra la función principal de conexión y manejo de la respuesta:

Figura 3

Algoritmo de Conexión a la API de PRTG

```

if not self._stop_event.is_set():
    print(f'#####DEVICE AVAILABILITY {client.dbProd} {sdate}')
    params["id"] = r.ID_Sensor
    params["sdate"] = sdate+intervalhour
    params["edate"] = edate+intervalhour
    try:
        response = APIRequest.get_data(client.prtgURL.prtgbaseURL, 'historicdataupt.xml', params)
        if response:
            XTemp = ET.fromstring(response)
            cdata_values = [elem.text.strip() for elem in XTemp.iter() if elem.text and elem.text.strip()]
            if "?" not in cdata_values[1]:
                writedb.SP_SaveDeviceAvailability(r.ID_Sensor, r.Name, r.ID_Device, cdata_values[1], cdata_values[2], sdate, client.dbProd)
                #####
            response = APIRequest.get_Details(client.prtgURL.prtgbaseURL, 'historicdata.xml', params)
            #result = response.content.decode('utf-8')
            root_node = ET.fromstring(response)
            date_list = []
            if root_node is not None:
                for element_item in root_node.findall("./item"):
                    print(element_item)
                    status_node = element_item.find("status")
                    status_value = status_node.text if status_node is not None else ""
                    if "Down" in status_value:
                        date_time_value = element_item.find("datetime").text
                        date_list.append(date_time_value)
                        # Guardar las fechas de tiempo de inactividad en la base de datos
                        writedb.SP_SaveDeviceAvailabilityDetails(r.ID_Sensor, date_time_value, sdate, client.dbProd)
                    else:
                        self.write_HELog(current_date, client.dbProd, r.ID_Sensor, r.ID_Device, r.Name, 'Root Node empty', task_id)
            else:
                self.write_HELog(current_date, client.dbProd, r.ID_Sensor, r.ID_Device, r.Name, 'Device Availability NO SENSOR DATA', task_id)
        except Exception as Err:
            try:
                self.write_HELog(current_date, client.dbProd, r.ID_Sensor, r.ID_Device, r.Name, f'Error exception in Device Availability: {Err}')
            except Exception as Err:
                self.writeTXTlog(f'No controlado Model_HELog {current_date}. Err: {str(Err)}')
                continue
            continue
    else:
        self.write_HELog(current_date, client.dbProd, 0, 0, 0, 'Stop event set', task_id)
        self.updateTask('stop', f'tarea detenida por {requestUser} .', task_id, True)
        return

```

Nota. Script que consulta métricas de disponibilidad vía API, procesa datos XML, identifica periodos de inactividad y registra resultados y eventos en la base de datos.

Como se observa en el algoritmo, se implementó un manejo de excepciones try-except para garantizar que, en caso de caída de la red, el script no detenga su ejecución abruptamente, sino que registre el error para su posterior auditoría.

Algoritmos de Transformación y Limpieza (Data Wrangling)

Una vez obtenidos los datos crudos (en formato XML/JSON), se procedió a la fase de transformación. Esta etapa es crítica para asegurar la calidad del dato antes de la visualización. Se utilizó la librería Pandas para la manipulación de los *DataFrames*.

Se aplicaron las siguientes reglas de negocio y limpieza:

- Parsing de XML a Estructura Tabular: Dado que la respuesta nativa de PRTG es jerárquica (árbol XML), se desarrolló una rutina de "aplanamiento" (*flattening*) para convertir los nodos en filas y columnas utilizando ElementTree.
- Tratamiento de valores nulos: Se identificaron registros donde el sensor no reportó datos (caída de sonda). Estos valores se etiquetaron como "Sin Datos" o se imputaron según la lógica de negocio, evitando vacíos en los reportes de Power BI.
- Conversión de tipos de datos: Los campos de fecha se estandarizaron al formato datetime de ISO 8601 y los valores de disponibilidad se convirtieron de texto a punto flotante (float) para permitir cálculos matemáticos.

Modelo de Datos y Carga (Loading)

Para la persistencia de los datos, se diseñó un modelo relacional en Microsoft SQL Server. Este enfoque centralizado permite mantener un histórico superior al que PRTG ofrece por defecto y facilita consultas rápidas desde Power BI.

La tabla principal, denominada DevicesAvailability, cuenta con la siguiente estructura:

Tabla 2

Estructura Tabla Devices Availability

Column Name	Data Type
id_sensor	int
name	varchar(50)
id_device	varchar(50)
uptime_percent	varchar(50)
downtime	varchar(50)
date	varchar(50)

El proceso de carga utiliza el método `to_sql` de Pandas optimizado con el motor SQLAlchemy, realizando inserciones por lotes (chunks) para no saturar la base de datos durante cargas masivas de los 989 dispositivos gestionados, allí se utilizan procedimientos almacenados para facilitar la inserccion de los datos.

Desarrollo de Inteligencia de Negocios

Tras garantizar la integridad y almacenamiento de los datos en el capítulo anterior, esta fase aborda la construcción de la capa semántica y visual. Se utilizó Microsoft Power BI para modelar la información procesada y permitir la exploración interactiva de los indicadores de disponibilidad de la red.

Modelado de Datos (Esquema en Estrella)

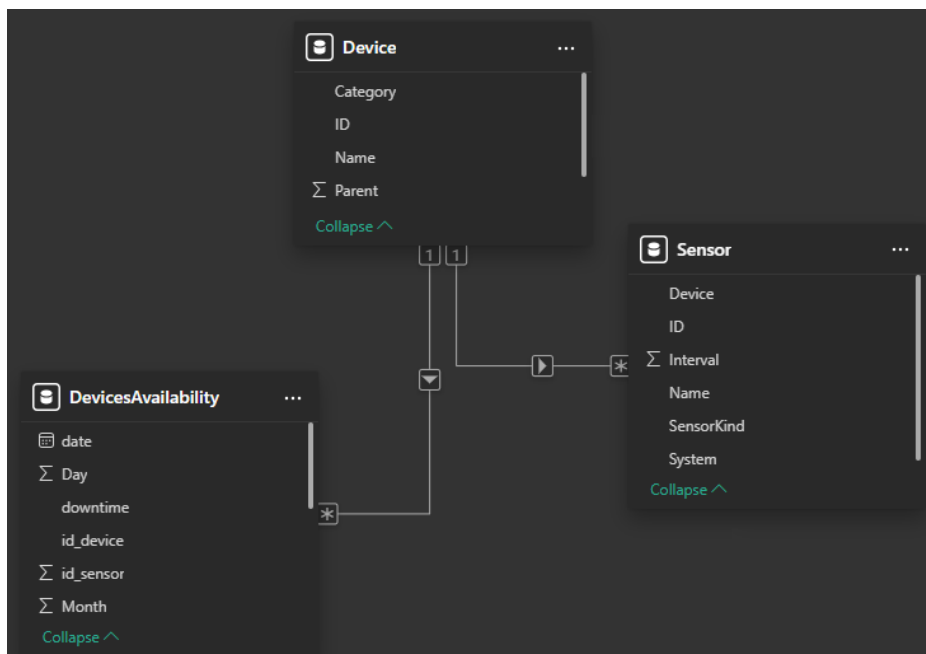
Para asegurar un rendimiento óptimo en las consultas y facilitar el filtrado dinámico, se estructuró el modelo de datos bajo un esquema de Estrella (Star Schema). Este diseño separa los datos transaccionales (hechos) de los atributos descriptivos (dimensiones).

- Tabla de hechos:
- **DevicesAvailability:** Contiene los registros métricos históricos (valor disponibilidad, tiempo de caída) provenientes del proceso ETL.
- Tablas de dimensiones:
- **Devices:** Contiene metadatos (ID, Nombre).
- **Sensors:** Contiene metadatos (ID, Nombre, Device ID).

La relación entre estas tablas se estableció mediante cardinalidad "Uno a Muchos" (1:*), permitiendo que un filtro en una dimensión (ej. seleccionar un Dispositivo) propague el contexto de filtro a millones de registros en la tabla de hechos instantáneamente.

Figura 4

Modelo de Datos Relacional en Power BI



Nota. Modelo de datos que muestra la relación entre dispositivos, sensores y disponibilidad operativa (tiempos de inactividad por fecha)

Ingeniería de Cálculos (Lenguaje DAX)

La inteligencia del reporte reside en las medidas calculadas mediante lenguaje DAX (Data Analysis Expressions). A diferencia de las columnas calculadas, estas medidas son dinámicas y se recalculan según el contexto seleccionado por el usuario.

Cálculo de Disponibilidad General

Para obtener el porcentaje exacto de disponibilidad (Uptime), se implementó una medida ponderada que promedia el valor de uptime extraído desde el PRTG.

Figura 5

Medida DAX para el Cálculo de Promedio de Disponibilidad

```
% Uptime = AVERAGE(DevicesAvailability[Uptime])
```

Nota. Cálculo del promedio del tiempo de funcionamiento (Uptime) de los dispositivos.

Semáforos y KPIs Condicionales

Se desarrollaron medidas de control visual para alertar sobre degradaciones del servicio.

Por ejemplo, la métrica de cumplimiento del SLA (Service Level Agreement):

Figura 6

Función Estado SLA

```
Estado SLA =  
IF(  
    [% Uptime] >= 0.99, "Cumple",  
    IF([% Uptime] >= 0.95, "Riesgo", "Incumplimiento")  
)
```

Nota. Relación entre sensores y dispositivos para el monitoreo del estado operativo.

Diseño de Interfaz y Visualización de Datos

La interfaz gráfica se diseñó siguiendo principios de Usabilidad (UX) y Narrativa de Datos (Data Storytelling), orientada a responder preguntas de negocio de lo general a lo particular (Drill-down).

Figura 7

Dashboard Power BI



Nota. Panel de Power BI que visualiza la disponibilidad de dispositivos, con indicadores de cumplimiento del SLA y análisis por mes y día.

Según se evidencia en el Dashboard final, la estructura visual se compone de tres niveles:

- Nivel estratégico (KPIs de cabecera):
- Se utilizaron tarjetas de indicadores (*Cards*) para mostrar la Disponibilidad

General del periodo seleccionado (ej. 99.99%).

- Justificación: Permite a la gerencia conocer el estado de salud de la red en menos de 5 segundos.

- Nivel táctico (análisis de tendencias):

- *Gráfico de Barras:* "Disponibilidad por Mes". Permite comparar el rendimiento histórico y detectar estacionalidad en fallos.

- *Gráfico de Líneas*: "Disponibilidad por Día". Esencial para identificar caídas puntuales en fechas específicas.
- Nivel operativo (detalle):
- *Tabla Detallada*: Ubicada en la sección inferior derecha, lista los eventos específicos de desconexión por dispositivo.
- *Justificación*: Provee al ingeniero de soporte la información granular (IP, hora exacta, duración) necesaria para realizar análisis de causa raíz (*Root Cause Analysis*).

Interactividad y Filtros Dinámicos

Para potenciar la capacidad de análisis, se implementaron paneles de segmentación (Slicers) en el margen izquierdo del reporte. Esto permite al usuario filtrar la información por:

- *Año / Mes / Día*: Para acotar ventanas de tiempo específicas durante incidentes.
- *Dispositivo / Categoría*: Para evaluar si el problema es generalizado o aislado en un equipo crítico.
- *Sede*: Para comparativas de rendimiento entre diferentes ubicaciones geográficas.

Evaluación de Resultados

Para dar cumplimiento al tercer objetivo específico, se realizó un estudio comparativo entre el proceso manual preexistente y el nuevo flujo automatizado, evaluando dimensiones de eficiencia temporal, escalabilidad y calidad de la información.

Definición de Métricas de Evaluación

Para objetivar el impacto de la solución, se definieron tres indicadores clave de desempeño (KPIs) que permiten medir el éxito del proyecto:

- Eficiencia temporal (E_t): Mide la reducción porcentual del tiempo dedicado a la generación de informes. Se calcula mediante la fórmula:

$$E_t = \frac{T_{manual} - T_{auto}}{T_{manual}} \times 100$$

Donde T_{manual} es el tiempo promedio del proceso humano y T_{auto} es el tiempo de ejecución del script Python.

- Integridad de los datos: Evaluación cualitativa sobre la consistencia de la información (ausencia de errores humanos, duplicidad o vacíos) tras la ingesta ETL.
- Escalabilidad del proceso: Capacidad del sistema para manejar un incremento en el volumen de dispositivos (N) sin degradar el rendimiento de manera lineal.

Análisis Comparativo de Eficiencia Operativa

Se realizaron pruebas de ejecución contrastando los tiempos requeridos para procesar diferentes volúmenes de dispositivos (50, 200, 500 y el total de 989). Los datos obtenidos demuestran una optimización drástica en la carga de trabajo.

La siguiente tabla resume los resultados de las pruebas de estrés:

Tabla 3
Comparación del Tiempo de Procesamiento

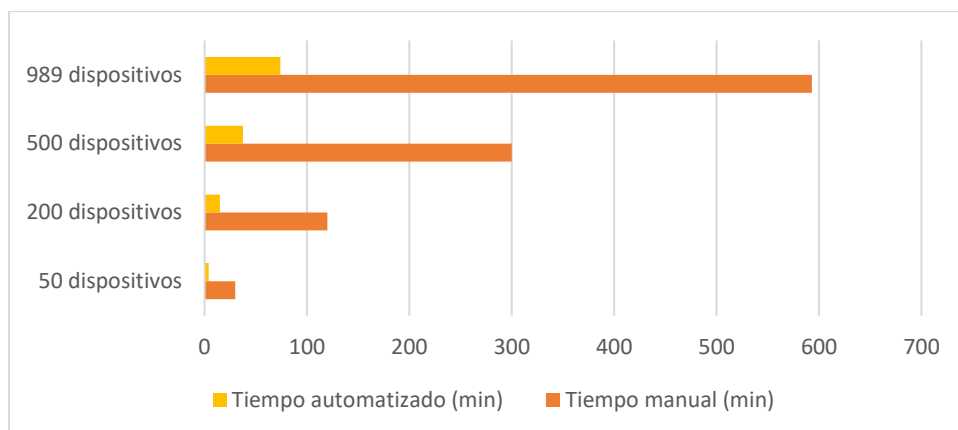
Cantidad de dispositivos	Tiempo manual (min)	Tiempo manual (h)	Tiempo automatizado (min)	Tiempo automatizado (h)	Ahorro (%)
50	30	0.5	4	0.06	86.60%
200	120	2	15	0.25	87.50%
500	300	5	37.5	0.62	87.50%
989	593	9.88	74	1.23	87.50%

Análisis de los Resultados

Como se evidencia en la Tabla 3, para el volumen total de la infraestructura (989 dispositivos), el tiempo de ciclo se redujo de 9.88 horas a solo 1.23 horas. Esto representa una liberación de recursos del 87.5%. Esta reducción implica que el ingeniero encargado, que anteriormente dedicaba más de una jornada laboral completa (aprox. 10 horas) exclusivamente a consolidar datos, ahora puede disponer del reporte actualizado al inicio de su jornada con una intervención mínima.

Figura 8

Impacto de la Automatización en Tiempos de Respuesta



Nota. Comparación del tiempo manual vs automatizado en la gestión de dispositivos, evidenciando mayor eficiencia con automatización.

Evaluación de Escalabilidad

Uno de los hallazgos más relevantes del estudio es el comportamiento de la curva de esfuerzo.

- En el escenario Manual: La relación es Lineal Directa. Si la empresa adquiere un nuevo cliente y los dispositivos aumentan a 2000, el tiempo manual se duplicaría a ~20 horas, haciendo el proceso insostenible humanamente.
- En el escenario Automatizado: La relación es Marginal. El script de Python procesa registros a velocidad de milisegundos. Duplicar la cantidad de dispositivos solo incrementaría el tiempo de ejecución en minutos (principalmente por latencia de red en la API), manteniendo la eficiencia global.

Esto valida que la solución no solo resuelve el problema actual, sino que está preparada para el crecimiento futuro de la organización sin requerir contratación de personal adicional para tareas operativas.

Validación de Precisión y Calidad del Dato

Más allá del tiempo, se evaluó la efectividad de la aplicación en términos de precisión (Accuracy).

- **Eliminación del error humano:** En el proceso manual, la manipulación de hojas de cálculo (copiar/pegar) inducía riesgos de desalineación de filas o errores de digitación ("fat finger error"). La automatización mediante ETL garantiza que el dato extraído de la fuente (PRTG) es idéntico al dato presentado en el Dashboard (Power BI), asegurando una integridad del 100%.
- **Disponibilidad de la información:** Anteriormente, los reportes tenían una "latencia" de 10 horas (la información mostrada era del día anterior o de la mañana, no del momento). [cite_start]Con la automatización, la ventana de actualización se reduce significativamente, permitiendo una toma de decisiones casi en tiempo real, lo cual es crítico para cumplir con los SLAs de disponibilidad de red.

La implementación de la plataforma ha demostrado ser altamente efectiva. La recuperación de 8.65 horas por ciclo de reporte transforma el rol del equipo técnico: de ser "procesadores de datos" manuales, pasan a ser "analistas de infraestructura", enfocándose en la resolución de incidentes detectados por el sistema y no en la construcción del reporte mismo.

Conclusiones

La automatización del proceso de recolección, transformación y generación de informes sobre la disponibilidad de la infraestructura de red ha sido clave para alcanzar el objetivo general del proyecto, mostrando una mejora notable en la eficiencia operativa. La plataforma que se desarrolló en Python logró reducir los tiempos de consolidación y análisis de datos en un notable 87.5 %, pasando de un proceso manual que tomaba alrededor de 10 horas para 989 dispositivos a un flujo automatizado que solo requiere 1.2 horas. Este resultado subraya la importancia de integrar técnicas de ETL y herramientas de inteligencia de negocios como Power BI en los procesos de monitoreo tecnológico.

La incorporación de metodologías estructuradas como el Team Data Science Process (TDSP) y las directrices de ITIL v4 ha garantizado un desarrollo ordenado, replicable y alineado con las mejores prácticas en la gestión de servicios de TI. Gracias a estos enfoques, el proyecto ha logrado implementar una solución robusta que asegura la calidad e integridad de los datos provenientes de herramientas como PRTG Network Monitor, ofreciendo informes claros, dinámicos y confiables para la toma de decisiones operativas.

La solución diseñada se muestra escalable, sostenible y adaptable a las necesidades reales de la organización, fortaleciendo la capacidad de análisis y la gestión estratégica de la infraestructura de red. La reducción del esfuerzo humano, la disminución del riesgo de errores manuales y la disponibilidad inmediata de información relevante contribuyen a mejorar el desempeño operativo, aumentar la agilidad en la detección de incidentes y consolidar una base tecnológica enfocada en la automatización y la optimización continua de los procesos corporativos.

Referencias Bibliográficas

Anglada, L., & Abadal, E. (2018). ¿Qué es la ciencia abierta? Anuario ThinkEPI, 12, 292–298.

<https://doi.org/10.3145/thinkepi.2018.43>

AXELOS. (2019). ITIL® 4 Foundation: ITIL 4 Edition. The Stationery Office.

García, J., Molina, J. M., Berlanga, A., Patricio, M. A., Bustamante, Á. L., & Padilla, W. R. (s. f.). (2018). Técnicas analíticas y aprendizaje estadístico. Recuperado de

https://d1wqtxts1xzle7.cloudfront.net/64031156/Ciencia_de_datos_2018-libre.pdf

Golfarelli, M., & Rizzi, S. (2009). Data Warehouse Design: Modern Principles and Methodologies. McGraw-Hill.

Hernández-Sampieri, R., Christian, D., & Torres, P. M. (s. f.). Metodología de la investigación: Rutas cuantitativa, cualitativa y mixta. Recuperado de

https://d1wqtxts1xzle7.cloudfront.net/64591365/Metodologia_de_la_investigacion.pdf

IBM. (s. f.). Introducción a la inteligencia de negocios. Recuperado de

<https://www.ibm.com/downloads/cas/6RZMKDN8>

Inmon, W. H. (2005). Building the data warehouse (4th ed.). Wiley.

International Organization for Standardization. (2018). Information technology — Service management — Part 1: Service management system requirements (ISO/IEC 20000-1:2018).

International Organization for Standardization. (2022). Information security, cybersecurity and privacy protection — Information security management systems — Requirements (ISO/IEC 27001:2022).

Kimball, R., & Caserta, J. (2004). The data warehouse ETL toolkit: Practical techniques for extracting, cleaning, conforming, and delivering data. Wiley.

Microsoft. (2023). Power BI Documentation. Recuperado de <https://learn.microsoft.com/power-bi/>

Microsoft. (s. f.). ¿Qué es el Proceso de Ciencia de Datos en Equipo (TDSP)? Recuperado de <https://learn.microsoft.com/es-es/azure/architecture/data-science-process/overview>

Paessler AG. (2023). PRTG Network Monitor User Manual. Recuperado de <https://www.paessler.com/manuals/prtg>

Roque Vargas, E. R., & Cadillo Montesinos, R. B. J. (2022). Solución de inteligencia de negocios que permita automatizar la disponibilidad de la información en una empresa de transporte.

Sharda, R., Delen, D., & Turban, E. (2018). Business intelligence, analytics, and data science: A managerial perspective (4th ed.). Pearson.

Susanibar Bazalar, I. A. (2020). Sistema de inteligencia de negocios y la automatización de reportes en la empresa Adecco.

Vele Arpi, M. G. (2020). Desarrollo de un sistema web de inteligencia de negocios para el manejo de reportes de la Red de Instituciones Financieras de Desarrollo (Tesis de pregrado). Quito, Ecuador.