

Propuesta
Sistema de Gestión Gerencial del Consejo Nacional de Lucha contra la Deforestación y otros
Crímenes Ambientales Asociados – CONALDEF.

ANEXO 4. ESQUEMA Y ENTORNO TECNOLÓGICO DE LA SOLUCIÓN

Enero 2021

CONTENIDO

ESQUEMA FUNCIONAL DE LA SOLUCIÓN	3
1. Esquema de la Solución.....	3
1.1 Fuente de datos.	4
1.2 Extracción, Transformación y Carga (Etl).....	4
1.3 Datawarehouse.	4
1.4 Datamart y Cubos.	4
1.5 Explotación de la información.....	5
1.6 Datos, información, conocimiento.	7
2. Entorno tecnológico de la solución.....	8
2.1 Hardware.....	8
2.2 Software.	11
2.3 Redes.....	17
3. Arquitectura física de la solución.....	18

ESQUEMA FUNCIONAL DE LA SOLUCIÓN

La solución es un sistema integrado orientado a la web donde se contará con diferentes módulos para permitir a los usuarios interactuar con los indicadores a través de análisis, consultas y reportes, así mismo la visualización de los comportamientos a partir de mapas tipo semáforo, mapas temáticos, mapas de tendencias, patrones, distribuciones, entre otros que permitan ver la dinámica de los indicadores para soportar la toma de decisiones.

Los datos que soportan los indicadores y otros son obtenidos a través de procesos de extracción, transformación y carga de las diferentes fuentes identificadas a partir de WebServices, bases de datos, archivos de texto, hojas de cálculos, etc.

Por otra parte, el sistema permitirá contar con el calendario de reuniones y eventos; gestión documental donde se podrá observar las actas, convenios, documentos técnicos, entre otros; el directorio de entidades y contactos; y por último los metadatos de los documentos y mapas:

1. Esquema de la Solución.

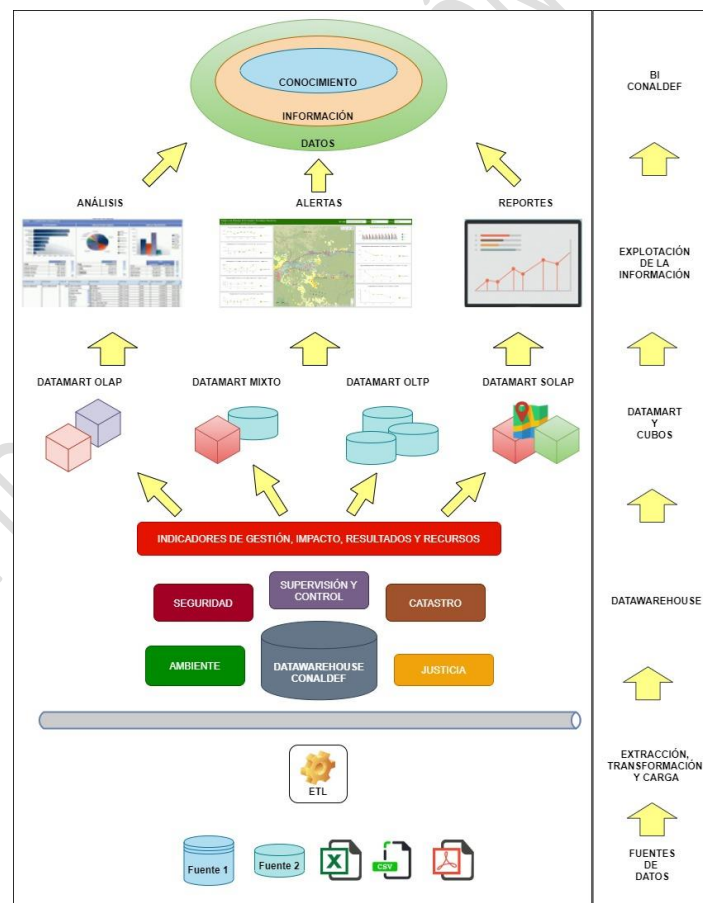


Ilustración 1 Arquitectura solución BI para el CONALDEF.

1.1 Fuente de datos.

La solución de BI que soporta el Sistema de Gestión Gerencial parte de los sistemas de origen de las instituciones que se encuentran para atacar la deforestación en el marco de CONALDEF (bases de datos, hojas de cálculo, archivos de texto...), sobre los que en la mayoría de los casos es necesario aplicar una transformación estructural para optimizar el análisis de la información que se encuentra dispuesta en ellas.

1.2 Extracción, Transformación y Carga (Etl).

En este paso se implementan todos los procesos ETL que permiten extraer los datos de las diferentes fuentes de datos identificadas, realizar su estructuración, estandarización, homologación, clasificación, georreferenciación, entre otros, para luego ser dispuestos en la Datawarehouse de CONALDEF a través de datamarts.

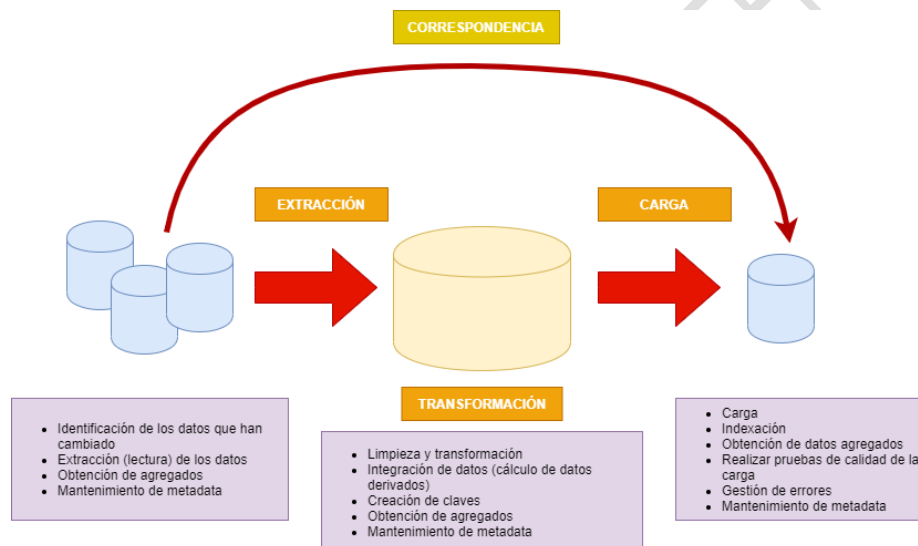


Ilustración 2 Extracción, Transformación y Carga - ETL. Fuente: Herramientas ETL, Carlos Proal. Elaboración propia.

1.3 Datawarehouse.

Los datos alojados en la Datawarehouse o en cada datamart de las diferentes fuentes quedarán en un modelo dimensional que permite por una parte contar datos de manera tal que permite accesos de alto rendimiento.

1.4 Datamart y Cubos.

Los datamart contendrán información de cada una de las temáticas ambiente, seguridad, supervisión y control, y justicia, así mismo la información de la batería de indicadores de impacto, gestión, resultados y recursos seleccionada por las instituciones para realizar los procesos de seguimiento y monitoreo de las acciones contra la deforestación y sus motores, con el fin de tomar decisiones de impacto y fortalecer las políticas públicas.

Los cubos del Sistema de Gestión Gerencial estarán organizados mediante jerarquías y permitirán el análisis rápido de los datos dependiendo de los criterios construidos para su consulta. Cada indicador se puede evaluar en cualquiera de los niveles de las jerarquías. Así, por ejemplo, se pueden obtener las "alertas tempranas" a nivel mensual, o a anual, una vereda, o un departamento...

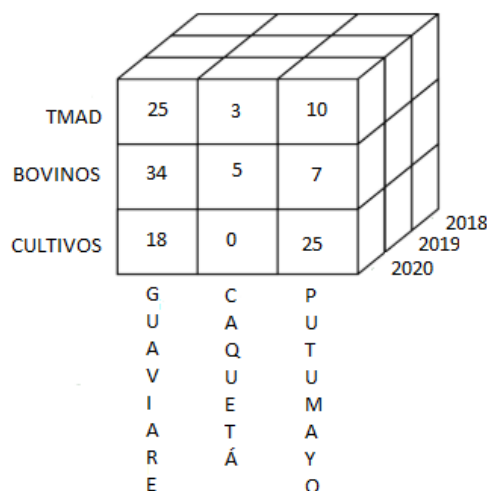


Ilustración 3 Cubos presiones por departamento y período. Elaboración propia.

1.5 Explotación de la información.

La solución BI planteada brinda una gran cantidad de funcionalidades a través de los Cuadros de Mando Integrado y Tableros de Control, los cuales permitirán contar con análisis, alertas y reportes. De esta forma, CONALDEF podrá contar con el estado de la batería de indicadores de gestión, impacto, resultados y recursos integrados en una sola plataforma.

Los Tableros de Control muestran una representación ilustrativa del desempeño de CONALDEF y de cada una de las instituciones. Estos tableros de control cuentan con tablas, gráficos, reglas de medición, cuadrantes y otros indicadores gráficos, como así también formatos condicionales, colores de fondo, entre otros.

La integración de los Cuadros de Mando Integral y Tableros de Control permiten asegurar que cualquier usuario pueda acceder de manera segura a los informes que necesita, en cualquier lugar, en cualquier momento y mediante cualquier dispositivo.

Los Cuadros de Mando Integrado y Tableros de Control brindan información instantánea sobre el desempeño de los indicadores monitoreados. Por lo general se construyen para el nivel estratégico, pero esto no quiere decir que otros niveles como el táctico o el operativo/investigativo también puedan contar con este tipo de herramientas.

• Análisis

Mediante la funcionalidad OLAP (On-Line Analytical Processing) es posible realizar la forma más sencilla de análisis, permitiendo que cualquier persona pueda ver de manera minuciosa subconjuntos de datos interrelacionados o “cubos”, simplemente con un clic.

Los Análisis permitirán a CONALDEF y a las instituciones tener acceso a la DataWareHouse a través de funcionalidades avanzadas para realizar análisis multivariados a partir de los cubos construidos.

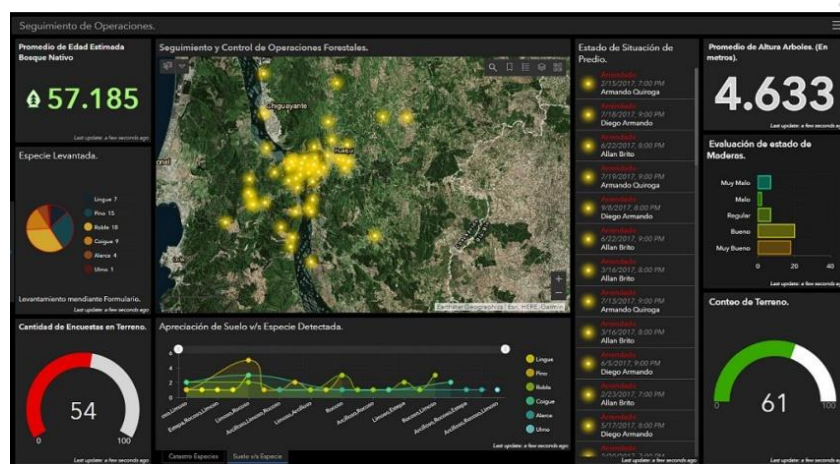


Ilustración 4 Cuadro de Mando Integral. Fuente: https://twitter.com/GIS_Bandit/status/1057361848066928640/photo/1

• Alertas

Los Cuadros de Mando de Alertas evidencian los comportamientos de los indicadores en una ubicación y tiempo determinado y que permiten generar alertas al nivel estratégico, con el fin de tomar decisiones oportunas y focalizadas.

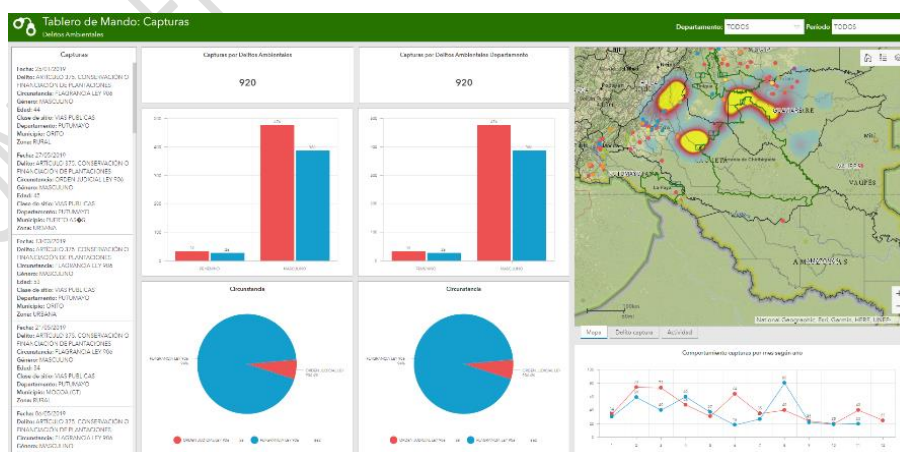


Ilustración 5 Cuadro de Mando Capturas por Delitos Ambientales. Fuente: UNODC.

- **Reportes**

Las funciones de Reportes permiten que el BI llegue a los niveles táctico y operativo/investigativo, información con alto nivel de detalle, lo que impacta fuertemente a los encargados de la toma de decisiones de CONALDEF y de las instituciones.

1.6 Datos, información, conocimiento¹.

En resumen, el Esquema Funcional de la solución nos lleva desde un simple dato a la generación de conocimiento del fenómeno de la deforestación y de sus motores, con el fin de tomar decisiones más acertadas y contundentes sobre los actores que la generan.

Por lo tanto, para entender cada momento a continuación se hace una descripción de los datos, la información y el conocimiento:

- **Datos**

Los datos son la mínima unidad semántica, y se corresponden con elementos primarios de información que por sí solos son irrelevantes como apoyo a la toma de decisiones. También se pueden ver como un conjunto discreto de valores, que no dicen nada sobre el porqué de las cosas y no son orientativos para la acción.

Los datos pueden provenir de fuentes externas o internas, pudiendo ser de carácter objetivo o subjetivo, o de tipo cualitativo o cuantitativo, etc.

- **Información**

La información se puede definir como un conjunto de datos procesados y que tienen un significado (relevancia, propósito y contexto), y que por lo tanto son de utilidad para quién debe tomar decisiones, al disminuir su incertidumbre. Los datos se pueden transformar en información añadiéndoles valor de diferentes formas:

- **Contextualizando:** se sabe en qué contexto y para qué propósito se generaron.
- **Categorizando:** se conocen las unidades de medida que ayudan a interpretarlos.
- **Calculando:** los datos pueden haber sido procesados matemática o estadísticamente.
- **Corrigiendo:** se han eliminado errores e inconsistencias de los datos.
- **Condensando:** los datos se han podido resumir de forma más concisa (agregación).

Por tanto, la información es la comunicación de conocimientos o inteligencia, y es capaz de cambiar la forma en que el receptor percibe algo, impactando sobre sus juicios de valor y sus comportamientos.

¹ https://www.sinnexus.com/business_intelligence/piramide_negocio.aspx

- **Conocimiento**

El conocimiento es una mezcla de experiencia, valores, información y know-how que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción. Se origina y aplica en la mente de los conocedores. En las organizaciones con frecuencia no sólo se encuentra dentro de documentos o almacenes de datos, sino que también está en rutinas organizativas, procesos, prácticas, y normas.

El conocimiento se deriva de la información, así como la información se deriva de los datos. Para que la información se convierta en conocimiento es necesario realizar acciones como:

- Comparación con otros elementos.
- Predicción de consecuencias.
- Búsqueda de conexiones.
- Conversación con otros portadores de conocimiento.

2. Entorno tecnológico de la solución.

2.1 Hardware.

- **Servidor de Base de Datos/Datawarehouse.**

Es el servidor que se encarga de soportar el motor de bases de datos PostgreSQL y su componente espacial PostGIS, dentro de sus tareas están:

- Soportar todo el tráfico y el flujo de información proveniente de las peticiones de datos de los usuarios.
- Dar respuesta a las solicitudes de datos alfanuméricos y geográficos y bodega de datos proveniente de usuarios internos y externos.
- Albergar la capa media del servidor de aplicaciones.
- Proveer la estructura y el almacenamiento de los datos espaciales, cartografía básica y temática.

Por lo cual tiene una alta velocidad de procesamiento, que garantiza una respuesta adecuada a los usuarios del sistema.

Característica	Descripción
Procesamiento	Procesador Intel(R) Xeon(R)
	Dos (2) procesadores.
Memoria RAM	64 GB
Estructura de Almacenamiento	
Sistema Operativo	Linux (Red Hat, Centos, Ubuntu) / Windows
RDBMS Software Libre	PostgreSQL
RDBMS Software Licenciado	Oracle
Motor Espacial Software Libre	PostGIS



Motor Espacial Software Licenciado	Oracle Spatial Nativo
	SDE Gateway ESRI

- **Servidor Web**

El servidor web, es el servidor encargado de disponer el contenido de un sistema orientado a la Web para los usuarios.

La función de este servidor será:

- Recibir del usuario las peticiones.
- Acceder a los requerimientos de la petición realizada.
- Procesar lo solicitado y enviar de acuerdo con el protocolo de transferencia el resultado.
- Exponer los resultados al usuario.

Característica	Descripción
Procesamiento	Procesador Intel(R) Xeon(R)
	Dos (2) procesadores.
Memoria RAM	64 GB.
Estructura de Almacenamiento	
Sistema Operativo	Linux (Red Hat, Centos, Ubuntu) / Windows
IP	Pública

- **Servidor de Aplicaciones**

Servidor de aplicaciones se encarga de publicar los servicios y páginas web que será accedidas por los usuarios internos y externos del Sistema de Gestión Gerencial.

Proporciona soluciones de alta disponibilidad flexibles y automatizadas que aseguran que las aplicaciones desplegadas sobre Apache-Tomcat cumplen con sus objetivos de CONALDEF.

La función de este servidor será servir las aplicaciones; sus tareas son:

- Proveer la entrada de consultas y requerimientos provenientes de Intranet e Internet.
- Dar respuesta a las solicitudes de datos geográficos y alfanuméricos proveniente de los usuarios internos y externos.
- Soportar todo el tráfico y el flujo de información proveniente de las peticiones de datos de los usuarios.
- Ser el servidor y controlador de acceso a las aplicaciones.

El software que soporta el servidor de aplicaciones puede ser: Apache-Tomcat con soporte para PHP, Jboss, o WebLogic.

Característica	Descripción
Procesamiento	Procesador Intel(R) Xeon(R)
	Dos (2) procesadores.
Memoria RAM	64 GB.
Sistema Operativo	Linux (Red Hat, Centos, Ubuntu) / Windows
Servidor de aplicaciones software libre	Apache-Tomcat Jboss
Servidor de aplicaciones software licenciado	IIS WebLogic
Observaciones	Soportar para PHP

- **Servidor BI.**

Aunque puede ser el mismo servidor de aplicaciones se sugiere implementar la solución en un servidor independiente, con el fin de separar los sistemas de información transaccionales de los de inteligencia de negocios.

La función de este servidor será servir los análisis, reportes y tableros de mando; sus tareas son:

- Proveer la entrada de consultas y requerimientos provenientes de Intranet e Internet.
- Dar respuesta a las solicitudes de SOLAP, OLAP y Reportes de los usuarios internos y externos.
- Soportar los procesos de Extracción, Transformación y Carga, ETL por su sigla en inglés.
- Soportar todo el tráfico y el flujo de información proveniente de las peticiones de datos de los usuarios.

Característica	Descripción
Procesamiento	Procesador Intel(R) Xeon(R)
	Dos (2) procesadores.
Memoria RAM	64 GB.
Sistema Operativo	Linux (Red Hat, Centos, Ubuntu) / Windows
Business Intelligence Software Libre	Pentaho Business Analytics Platform (BA).
	Pentaho Aggregation Designer.
	Pentaho Data Integration - PDI.
	Pentaho Metadata Editor.
	Pentaho Report Designer.
	Pentaho Schema Workbench.
	GeoMondrian
	GeoKettle



Business Intelligence Software Licenciado	Oracle Warehouse Builder
	Oracle Publisher (Reporting and Publishing)
	Oracle Interactive Dashboard
	Oracle Answers (Ad Hoc Analysis)
	Oracle Data Integration

- **Equipos de escritorio/Portátiles**

Son los equipos con los que los usuarios internos o externos se conectan a las aplicaciones web a través de intranet o internet.

Característica	Descripción
Procesador	Intel.
Procesamiento	Un (1) procesador.
Memoria RAM	16 GB.
Sistema Operativo	Microsoft Windows última versión disponible.
Herramientas Ofimáticas	Office Profesional última versión disponible.
Software administración de base de datos para el administrador de la solución	PgAdmin Software Libre Oracle Client Software Licenciado

2.2 Software.

- **Sistema Operativo**

Con respecto al sistema operativo la solución puede correr en un ambiente Linux (Red hat, Centos, Ubuntu) o Windows.

- **Base de datos/Datawarehouse**

Software Libre.

PostgreSQL. PostgreSQL es un sistema de gestión de base de datos relacional de objetos, basado en POSTGRES, desarrollado por la Universidad de California en el Departamento de Ciencias de Computación de Berkeley. POSTGRES fue pionero en muchos conceptos que solo se convirtieron disponibles en algunos sistemas de base de datos comerciales mucho más tarde (PostgreSQL, 2017).

PostgreSQL es el RDBMS que se utilizará como repositorio de datos como solución libre ya sea en la base de datos alfanumérica, la Datawarehouse o la base de datos geográfica, además será el encargado de soportar los requerimientos de información de los diferentes usuarios conectados a través de los Cuadros de Mando Integrado, Tableros de Control, y Reportes.

Para su administración, configuración y afinamiento de la base de datos PostgreSQL se utilizarán las

herramientas propias de PostgreSQL.

Software Licenciado.

Oracle. Es un sistema de gestión de base de datos de tipo objeto-relacional (ORDBMS, por el acrónimo en inglés de Object-Relational Data Base Management System), desarrollado por Oracle Corporation.

Oracle es el RDBMS que se utilizará como repositorio de datos en el caso de que la selección por parte de CONALDEF se a través de software licenciado, ya sea en la base de datos alfanumérica, la Datawarehouse o la base de datos geográfica, además será el encargado de soportar los requerimientos de información de los diferentes usuarios conectados a través de los Cuadros de Mando Integrado, Tableros de Control, y Reportes.

Para su administración, configuración y afinamiento de la base de datos Oracle se utilizarán las herramientas propias de Oracle.

- **Base de datos espacial.**

Software Libre.

PostGIS. PostGIS es un extensor de base de datos espacial para la base de datos relacional de objetos PostgreSQL. Agrega soporte para objetos geográficos permitiendo que las consultas de ubicación se ejecuten en SQL o en caso contrario se utilizará el gateway SDE para el caso que el motor espacial sea SDE sobre PostgreSQL de la casa ESRI.

Es un Gateway utilizado para almacenar sobre un RDBMS PostgreSQL información geográfica y alfanumérica de datos espaciales. PostGIS es la puerta de entrada para el almacenamiento y administración de datos espaciales, soportando los conceptos de: multiusuario, multitareas, multiacceso, multiprocesamiento, entre otros. La información almacenada dentro de un sistema administrador de base de datos (DBMS); permite la edición, consulta y administración en forma multiusuario y provee vistas de transacciones a una Geodatabase. PostGIS juega un papel fundamental en los sistemas SIG corporativos.

Software Licenciado.

Oracle Spatial/SDE. Oracle Spatial es un extensor de base de datos espacial nativo para la base de datos relacional de objetos Oracle. Agrega soporte para objetos geográficos permitiendo que las consultas de ubicación se ejecuten en SQL o en caso contrario se utilizará el gateway SDE para el caso que el motor espacial sea SDE sobre Oracle de la casa ESRI.

Es un Gateway utilizado para almacenar sobre un RDBMS Oracle información geográfica y alfanumérica de datos espaciales. Oracle Spatial/SDE es la puerta de entrada para el almacenamiento y administración de datos espaciales, soportando los conceptos de: multiusuario, multitareas, multiacceso, multiprocesamiento, entre otros. La información almacenada dentro de un sistema administrador de base de datos (DBMS);

permite la edición, consulta y administración en forma multiusuario y provee vistas de transacciones a una Geodatabase. Oracle Spatial/SDE juega un papel fundamental en los sistemas SIG corporativos.

- **Business Intelligence**

Plataforma de Pentaho Community Edition.

Pentaho Open Source BI Suite Community Edition (CE), incluye ETL, análisis OLAP, metadata, minería de datos, reportes, dashboards y una plataforma que permite crear soluciones complejas a los problemas del negocio. (Pentaho, 2017)

Pentaho Business Analytics Platform (BA).

Pentaho Business Analytics Platform (BA) es el núcleo del software de Pentaho, una plataforma web que aloja tanto contenido generado en la propia web o ficheros publicados al servidor desde aplicaciones de escritorio.

Pentaho Aggregation Designer.

Pentaho Aggregation Designer simplifica la creación y el despliegue de tablas agregadas que mejoran el rendimiento de sus cubos OLAP de Pentaho Analysis (Mondrian). Pentaho Analysis es un motor OLAP puro y relacional que funciona únicamente con los datos almacenados en su base de datos relacional en lugar de proporcionar su propio modelo de almacenamiento de datos multidimensional. Esto simplifica la implementación y la administración de datos, pero limita el rendimiento cuando se trabaja con conjuntos de datos muy grandes (tablas de hechos con más de 10 millones de registros y / o cubos con una alta cardinalidad de niveles y miembros). Para mejorar el rendimiento en estos escenarios, Pentaho Analysis admite tablas agregadas. Las tablas agregadas coexisten con la tabla de hechos base y contienen medidas agregadas previamente creadas a partir de la tabla de hechos.

Pentaho Aggregation Designer le proporciona una interfaz simple que le permite crear tablas agregadas a partir de niveles dentro de las dimensiones que especifique. En función de estas selecciones, el Diseñador de agregación genera el Lenguaje de definición de datos (DDL) para crear las tablas agregadas, el Lenguaje de manipulación de datos (DML) para llenarlas y un esquema Mondrian actualizado que hace referencia a las nuevas tablas agregadas. Si no está familiarizado con los conceptos de diseño de tabla agregada, el Diseñador de agregación también incluye un asesor inteligente que evalúa la estructura y la cardinalidad de su cubo OLAP y recomienda algunas tablas agregadas iniciales para crear para mejorar el rendimiento.

Pentaho Data Integration – PDI

Pentaho Data Integration (PDI, también llamada Kettle) es el componente de Pentaho responsable de la extracción, transformación y carga (ETL) de procesos. Aunque las herramientas ETL son los más usados en entornos de almacenes de datos, Kettle también se puede utilizar para otros fines:

- Migración de datos entre las aplicaciones o bases de datos.

- Exportación de datos desde bases de datos para archivos planos.
- Carga de datos de forma masiva en bases de datos.
- Limpieza de datos.

Como principales características se tiene:

- Cada proceso es creado con una herramienta gráfica donde se especifica qué se va a hacer sin necesidad de escribir código que indique cómo hacerlo.
- Admite una amplia gama de formatos de entrada y salida, incluyendo archivos de texto, hojas de datos, archivos XML, propiedades de Java y los motores de base de datos software libre y propietarios.
- Arquitectura extensible permite desarrollar conectores y plugins fácilmente.
- Basado en repositorio facilita la reutilización de componentes de transformación, colaboración y administración de modelos, conexiones, etc.
- Depurador integrado.
- Todos los pasos de base de datos existen un botón SQL que generará el código requerido para las tablas y/o índices. También está disponible a nivel de transformación y trabajo.

GeoKettle.

GeoKettle es una versión de Pentaho Data Integration (Kettle) con capacidad de tratamiento de datos espaciales. Es una potente herramienta ETL (siglas en inglés de Extract, Transform and Load: Extracción, Transformación y Carga) orientada al uso de metadatos y con funcionalidades espaciales dedicada a la integración de diversos orígenes de datos para la construcción y/o actualización de bases de datos espaciales y almacenes de datos espaciales.

GeoKettle permite la Extracción desde los orígenes de datos, la Transformación de los datos con el objetivo de corregir errores, realizar limpieza de los mismos, modificar su estructura, hacer que cumplan los estándares y por último la Carga de los datos transformados en un Sistema Gestor de Bases de Datos (DBMS por sus siglas en inglés), un archivo SIG, o un servicio web geoespacial.

GeoKettle resulta particularmente útil cuando el usuario pretende automatizar procesos complejos y repetitivos sin tener que generar para ello código de programación, o para convertir entre diversos formatos de datos, o para migrar datos de un DBMS a otro, o cuando quiere alimentar con datos varios DBMS o rellenar almacenes de datos analíticos con finalidad de dar apoyo a la toma de decisiones, etc.

En el ámbito geoespacial GeoKettle puede ser comparado con FME, una herramienta de software privativo para realizar ETL con capacidades espaciales creada por Safe Software. GeoKettle es estable, rápida, cumple con los estándares, tiene cientos de funciones y lee y escribe en diversos formatos de ficheros, servicios y DBMS. GeoKettle está siendo usada por diversas organizaciones en todo el mundo, incluyendo gobiernos, bancos, aseguradoras e integradores de sistemas geoespaciales.

Pentaho Metadata Editor

Pentaho Metadata Editor una herramienta para crear una metadata más avanzada, sin tener que utilizar el motor OLAP y de una forma mucho más sencilla y sin tener conocimientos de los modelos dimensionales.

Pentaho Mondrian

Es el servidor de procesamiento analítico online (OLAP) escrito en Java, permite a los usuarios del negocio analizar grandes y complejos cantidades de datos en tiempo real.

Entre otras características destaca la compatibilidad con el MDX (expresiones multidimensionales) y el lenguaje de consulta XML (XMLA) para análisis y olap4j.

Pentaho Report Designer

Pentaho Report Designer es una solución basada en el proyecto JFreeReport que permite generar informes ágiles y de gran capacidad. Éste permite la distribución de los resultados del análisis en múltiples formatos, todos los informes incluyen la opción de imprimir o exportar a formato PDF, XLS, HTML y texto. Pentaho Reporting permite también la programación de tareas y ejecución automática de informes con una determinada periodicidad. Soporta amplia fuente de datos relacionales, OLAP y XML basados en fuentes de datos.

Pentaho Schema Workbench

Es una interfaz de diseño visual que permite crear y probar los esquemas de cubos OLAP. El diseñador de agregación provee una interfaz simple que permite crear y desplegar tablas de agregado para mejorar el rendimiento de los cubos OLAP de Mondrian.

- **Desarrollo**

Community Dashboard Framework (CDF).

Es un framework de código abierto que permite la creación de cuadros de mando altamente personalizables, están en la parte superior del servidor de Pentaho Business Intelligence. CDF se basa en estándares de desarrollo web como CSS, HTML 5 y JavaScript, aprovechando algunos frameworks comúnmente usados como JQuery o Bootstrap. (Pentaho, 2017)

Principales características de CDF.

- Basado en tecnologías de fuente abierta.
- Usa AJAX.
- Cuenta con un ciclo de vida con componentes que interactúan entre sí.

- Separa la lógica (Javascript) de la presentación (html, css).
- Extensible, el cual da a los usuarios un alto nivel de personalización.
- La biblioteca de componentes es ampliable de forma avanzada.
- Posibilidad de que los usuarios avanzados inserten sus propios fragmentos de código abierto.

Uso de CDF:

- El escenario más común es ser usado como parte de una solución de Pentaho.
- En modo autónomo es una alternativa a la Consola de usuario de Pentaho, PUC por sus siglas en inglés.
- Integrado con Porlets, aplicaciones php, portales de intranet e incluso aplicaciones de escritorio.
- CDF reúne los informes, gráficos y tablas que el servidor BI de Pentaho proporciona en una manera coherente, que el desarrollador puede utilizar para crear visualmente cuadros de mando atractivos y útiles, utilizando html simple y javascript.

Los nuevos componentes se pueden integrar en el framework, y el "MetaLayer" permite a los desarrolladores ahorrar tiempo y evitar errores al copiar y pegar, recopilando la lógica empresarial en unidades bien organizadas que puedan compartirse entre paneles. (Pentaho, 2017).

Indicadores de control de desempeño KPI.

Los indicadores son datos que muestran el estado de una actividad, estos pueden medir cualquier cosa. En el caso de ser indicadores clave, se llaman Indicadores Clave de desempeño o KPI (Key Performance Indicators) por sus siglas en inglés.

Un indicador clave de desempeño o KPI es una métrica de negocio para evaluar los factores que son cruciales para el éxito de una organización. El propósito de usar KPI es enfocar la atención en las tareas y procesos que la gerencia ha determinado como los más importantes para avanzar hacia metas y objetivos declarados. (TechTarget, 2017)

Se puede tomar como ejemplo de indicadores al tiempo promedio que se toma en realizar una actividad en minutos, la eficiencia de un proceso en porcentaje, el consumo de materia prima para realizar una unidad de producto, etc. Para cada uno de estos indicadores se tiene que definir su periodicidad, responsable, entre otros.

Python. Python es un lenguaje de programación interpretado cuya filosofía hace hincapié en la legibilidad de su código. Se trata de un lenguaje de programación multiparadigma, ya que soporta orientación a objetos, programación imperativa y, en menor medida, programación funcional.

Algunos componentes de los procesos de extracción, transformación y carga serán desarrollados en Python.

- **Minería de datos (Datamining)**

Weka.

Pentaho incorpora la tecnología Weka, es un conjunto integral de herramientas para aprendizaje automático y minería de datos. Su amplio conjunto de clasificación, regresión, reglas de asociación y algoritmos de agrupación se puede utilizar para ayudarle a comprender mejor la empresa y también ser aprovechado para mejorar el rendimiento futuro a través de análisis predictivo. Como principales características tenemos:

- Motor de Data Mining provee de un set de algoritmos de aprendizaje del proyecto Weka incluyendo clustering, segmentación, árboles de decisión, redes neuronales y análisis de componentes principales, entre otros.
 - Integración de Pentaho Data Integration, automatización de procesos de transformación de datos al formato que requiere el motor de Data Mining.
 - Los algoritmos pueden ser aplicados directamente a un set de datos o invocados desde código Java.
 - Provee filtros para normalización, re-muestreo, selección de atributos, transformación y combinación de atributos.
 - Modelos para predicción de variables nominales o numéricas.
 - Herramientas gráficas de diseño para pre-procesamiento de datos, regresión, clustering, reglas de asociación
 -
- **Metadatos.**

Geonetwork². Es un entorno de gestión de información espacial estandarizado y descentralizado, diseñado para permitir acceso a bases de datos georreferenciadas, productos cartográficos y metadatos de varias fuentes, mejorando el intercambio y el compartir entre organizaciones y su audiencia, usando las capacidades de la Internet.

- **Administrador de Contenidos (CMS).**

Wordpress³. Es un sistema de gestión de contenidos, enfocado a la creación de cualquier tipo de página web. Este gestor de contenidos permite la instalación de plugins que permite la gestión documental, manejo de calendarios, directorio de contactos, entre otros.

2.3 Redes.

En lo que tiene que ver con la infraestructura de red, el sistema puede implantarse con la plataforma de comunicaciones actual del IDEAM, que tiene las siguientes características:

² https://es.wikipedia.org/wiki/GeoNetwork_opensource

³ <https://es.wikipedia.org/wiki/WordPress>



Característica	Especificación	Descripción
Medio de Acceso	Red LAN	El acceso al servidor de datos por parte de los usuarios internos usará como medio la Red Local del IDEAM.
Protocolo	TCP/IP	Protocolo de comunicaciones para los niveles de Red y Transporte del Modelo de Interconexión de Sistemas Abiertos (ISO).
Nombre del Servidor	Especificado por el Estándar de Nombres del IDEAM	Se utilizará el estándar asignado para servidores.
Nombre del Dominio	Especificado por el Estándar de dominio del IDEAM	Se utilizará el estándar asignado para el dominio.
Velocidad de Acceso	100 MHz / 1GHz	
Arquitectura de Acceso	LAN	Acceso a consultas y reportes en aplicación web. El ingreso de información de formulación se realizará a través de la web.

- **Características de la red**

La conexión a la base de datos la realiza el servidor de aplicaciones por medio del nombre de servidor, el cual tiene asociado una dirección IP y el nombre de la instancia de la Base de Datos PostgreSQL/Oracle. Físicamente se conectará en la red de servidores del IDEAM.

3. **Arquitectura física de la solución.**

La arquitectura física del Sistema de Gestión Gerencial de CONALDEF parte de la necesidad de la interoperabilidad para acceder a la información vital para medir la batería de indicadores de CONALDEF producida por cada una de las instituciones y fuentes identificadas, ya sea a partir de servicios web alfanuméricos, geográficos o a través de conjuntos de datos estructurados desde las bases de datos principales de cada institución o no estructurados cuando las instituciones no cuenten con sistemas de información que no garanticen un adecuado manejo de la información. Para el caso de los servicios web alfanuméricos, geográficos o a través de conjuntos de datos estructurados a partir de la capa de comunicación serán transferidos al servidor principal (base de datos), utilizando para tal fin, procesos de extracción, transformación y carga automáticos o semiautomáticos cumpliendo con altos estándares de seguridad que garanticen a través de nuevos procesos sean ingresados a la bodega de datos integrando el componente espacial (base datos geográfica) que soportan a partir de datamarts el monitoreo y seguimiento a la batería de indicadores que permitan fortalecer la toma de decisiones a CONALDEF.

A continuación, se puede observar la arquitectura física del Sistema de Gestión Gerencial de CONALDEF.

Sistema de Gestión Gerencial CONALDEF

