**Universidad Nacional de Misiones**

**Facultad de Ciencias Exactas Químicas y Naturales**

**Tesis de grado Licenciatura en Sistemas de Información**

**Diseño, desarrollo e implementación de una solución de Inteligencia de Negocios para la evaluación de la fauna íctica en el río Paraná, aguas arriba de la represa Yacyretá.**

**Estudio de caso, Proyecto de Biología Pesquera Regional.**

**Autor: Cinthia Alexandra Cuba Lezcano**

**Tutor: Lic. Jesús Germán Andrés Pautsch**

**Co-tutores:**

**Mgter. Lic. Gladys Graciela Garrido**

**Lic. Danilo Ramón Aichino**

**Año: 2016**

*Dedicatoria*

Resumen

En la actualidad la mayoría de las organizaciones cuentan con sistemas de bases de datos en las cuales se almacena información diariamente. El exceso de dicha información dificulta el proceso de análisis y la extracción de conocimiento de las fuentes de datos creando la necesidad de utilizar una herramienta que facilite dicho proceso.

La inteligencia de negocios cuenta con instrumentos que facilitan la extracción, depuración y automatización de gran parte del trabajo, minimizando la posibilidad de errores y emitiendo de manera rápida y precisa información que da soporte a la toma de decisiones.

El objetivo de la presente tesis se centra en el desarrollo e implementación de una solución de inteligencia de negocios, que contribuya en el proceso de análisis y toma de decisiones sobre el estado y conservación de la fauna íctica del río Paraná aguas arriba de la represa Yacyretá.

Se utilizan ciertas funcionalidades de las metodologías Kimball y Hefesto, que a criterio del autor representan de manera clara y concisa los requisitos del usuario y que son de fácil comprensión para los mismos.

Para lograr la meta propuesta se crea un data Warehouse que permite el análisis multidimensional y a través de técnicas OLAP obtener conocimientospara su posterior análisis y representación mediante tableros de control.xxxxx

Palabras claves:*Inteligencia de negocios, tableros de control, OLAP, datawarehouse, comunidad íctica.*

**Abstract**

Nowadays, most of the companies use databases to store information on a daily basis. The excess of such information makes the process of analysis and knowledge retrieval from the database difficult. Therefore, the creation of a tool to facilitate such process is vital.

Business Intelligence has the tools to facilitate data retrieval, refinement and automation of most of the work, minimising error occurrence and offering information that helps to make decisions ina fast and precise way.

The main objective of this thesis focuses on the development and implementation of a business intelligence solution, that contributes to the analysis process and decision making regarding the state and conservation of the ichtyofauna of the Paraná river upstream the Yacyretá dam.

The functionalities of Kimball’s and Hefesto’s methodologies that are used are, in the author’s opinion, more representative of the users’ requirements and are more easily understood by them.

To accomplish the proposed goal a data Warehouse is created to allow multidimensional analysis and, by means of OLAP techniques, obtain knowledge for its analysis and representation through dashboards**.**

Keywords:*Business intelligence, dashboard, OLAP, datawarehouse*, *ichtyofauna.*

*Reconocimientos*

****Índice****

[CAPITULO 1: Introducción 16](#_Toc444510843)

[1.1 Motivación 16](#_Toc444510844)

[1.2 Objetivos 17](#_Toc444510845)

[1.2.1 Objetivo General 17](#_Toc444510848)

[1.2.2 Objetivos Específicos 17](#_Toc444510849)

[1.3 Estructura del documento 18](#_Toc444510850)

[CAPITULO 2: : Marco Teórico 21](#_Toc444510851)

[2.1 Bases de Datos 21](#_Toc444510852)

[2.1.1 Definición 21](#_Toc444510853)

[2.1.2 Objetivos e Importancia 22](#_Toc444510854)

[2.2 Sistemas OLTP 22](#_Toc444510855)

[2.3 Inteligencia de Negocios 23](#_Toc444510856)

[2.3.1 Breve historia. 23](#_Toc444510857)

[2.3.2 Definición 24](#_Toc444510858)

[2.4 Data Warehouse 25](#_Toc444510859)

[2.4.1 Características 25](#_Toc444510860)

[2.5 Arquitectura de DW 26](#_Toc444510861)

[2.5.1 Orígenes de Datos 26](#_Toc444510862)

[2.5.1.1 Fuentes internas 26](#_Toc444510863)

[2.5.1.2 Fuentes externas 26](#_Toc444510864)

[2.5.2 Subsistema ETL 27](#_Toc444510865)

[ Extracción 27](#_Toc444510866)

[ Transformación 27](#_Toc444510867)

[ Carga 28](#_Toc444510868)

[ Staging Area 29](#_Toc444510869)

[2.5.3 Data Mart 29](#_Toc444510870)

[Estrategia *top-down* 29](#_Toc444510871)

[Estrategia *Bottom-Up* 30](#_Toc444510872)

[2.6 Modelado Dimensional 30](#_Toc444510873)

[2.6.1 Tabla de Dimensiones 30](#_Toc444510874)

[2.6.2 Tabla de Hechos 30](#_Toc444510875)

[2.6.3 Tipos de modelado de un DW 31](#_Toc444510876)

[2.6.3.1 Esquema estrella 31](#_Toc444510877)

[2.6.3.2 Esquema copo de nieve 32](#_Toc444510878)

[2.6.3.3 Esquema constelación 33](#_Toc444510879)

[2.6.4 Operaciones sobre modelos dimensionales 34](#_Toc444510880)

[2.7 OLAP 36](#_Toc444510881)

[2.7.1 Definición 36](#_Toc444510882)

[2.7.2 Tipos de implementaciones OLAP 36](#_Toc444510883)

[2.7.2.1 ROLAP 36](#_Toc444510884)

[2.7.2.2 MOLAP 37](#_Toc444510885)

[2.7.2.3 HOLAP 37](#_Toc444510886)

[2.8 Tableros de control 38](#_Toc444510887)

[2.9 Metodologías 40](#_Toc444510888)

[2.9.1 Metodología propuesta por Kimball 40](#_Toc444510889)

[2.9.2 Metodología propuesta por Inmon 42](#_Toc444510890)

[2.9.3 Metodología propuesta por Bernabeu 43](#_Toc444510891)

[2.9.3.1 Características 43](#_Toc444510892)

[2.9.3.2 Pasos de la metodología Hefesto: 45](#_Toc444510893)

[CAPITULO 3: Descripción del Problema 47](#_Toc444510894)

[3.1 La organización 47](#_Toc444510895)

[3.2 Estructura organizacional 47](#_Toc444510896)

[3.3 Actividades 50](#_Toc444510897)

[3.3.1 Sitios de muestreos 50](#_Toc444510898)

[3.3.2 Metodología de muestreo 51](#_Toc444510899)

[3.4 El problema 55](#_Toc444510900)

[3.5 Objetivo 56](#_Toc444510901)

[CAPITULO 4: Solución Propuesta 59](#_Toc444510902)

[4.1 Materiales y Métodos 59](#_Toc444510903)

[4.2 Herramientas seleccionadas para el desarrollo 60](#_Toc444510904)

[4.2.1 Pentaho 61](#_Toc444510905)

[4.2.2 Motor de Bases de Datos 63](#_Toc444510906)

[4.3 Pasos y aplicación de las metodologías 64](#_Toc444510907)

[4.3.1 Paso 1: Análisis de requerimientos 64](#_Toc444510908)

[4.3.2 Procesos del negocio 66](#_Toc444510909)

[4.3.2.1 Selección de procesos. 66](#_Toc444510910)

[4.3.3 Modelo Conceptual 67](#_Toc444510911)

[Paso 2: Análisis del OLTP 68](#_Toc444510912)

[4.3.3.1 Conformación de Indicadores 69](#_Toc444510913)

[Correspondencias 71](#_Toc444510914)

[4.3.4 Nivel de granularidad 73](#_Toc444510915)

[Modelo conceptual ampliado 76](#_Toc444510916)

[4.4 PASO 3: Modelo lógico del DW 76](#_Toc444510917)

[4.4.1 Tipo de modelo lógico del DW 77](#_Toc444510918)

[Tabla de dimensiones y hechos 77](#_Toc444510919)

[4.4.2 Integración de Datos 78](#_Toc444510920)

[4.4.3 Cubos Multidimensionales 78](#_Toc444510921)

[4.4.4 Tableros de control 80](#_Toc444510922)

[CAPITULO 5: Prueba Experimental 82](#_Toc444510923)

[5.1 Análisis OLAP 82](#_Toc444510924)

[CAPITULO 6: Conclusiones 90](#_Toc444510925)

[6.1 Conclusión 90](#_Toc444510926)

[6.2 Trabajos Futuros 91](#_Toc444510927)

[Anexo 1 92](#_Toc444510928)

[Anexo 2 97](#_Toc444510929)

[Anexo 3 103](#_Toc444510930)

[Bibliografía 110](#_Toc444510931)

**Índice de figuras**

[Figura 1: Arquitectura básica de un DW. 25](#_Toc444494503)

[Figura 2: Proceso ETL. 26](#_Toc444494504)

[Figura 3: Transformación: codificación 27](#_Toc444494505)

[Figura 4: Transformación: medida de atributos 27](#_Toc444494506)

[Figura 5: Transformación: convención de nombramiento. 27](#_Toc444494507)

[Figura 6: Esquema estrella 31](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494508)

[Figura 7: Esquema copo de nieve. 32](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494509)

[Figura 8: Esquema constelación. 33](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494510)

[Figura 9: Metodología Kimball – Bussiness Dimensional Lifecycle. 40](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494511)

[Figura 10: Metodología Hefesto 2.0 44](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494512)

[Figura 11: Estructura organizacional del Proyecto de Biología Pesquera Regional 47](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494513)

[Figura 12: Organigrama del Proyecto Biología Pesquera Regional 48](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494514)

[Figura 13: Sitios de monitoreo. 50](#_Toc444494515)

[Figura 14: Principales medidas biométricas. 52](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494516)

[Figura 15: Ruta de análisis para la obtención de los índices de estructura y rendimientos pesqueros. 54](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494517)

[Figura 16: Empresas internacionales que utilizan Pentaho. 59](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494518)

[Figura 17: Empresas de América Latina que utilizan Pentaho. 59](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494519)

[Figura 18: Modelo conceptual 65](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494520)

[Figura 19: Modelo conceptual - Biología Pesquera Regional 66](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494521)

[Figura 20: Modelo de datos de la organización. 69](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494522)

[Figura 21: Modelo conceptual ampliado - Proyecto Biología Pesquera Regional 74](#_Toc444494523)

[Figura 22: Modelo lógico DM - Proyecto de Biología Pesquera Regional 76](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494524)

[Figura 23: Cubo rendimientos pesqueros 77](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494525)

[Figura 24: Consulta de prueba 1 81](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494526)

[Figura 25: Resultado de la prueba 1. 81](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494527)

[Figura 26: Gráfico de la prueba 1. 81](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494528)

[Figura 27: Consulta de la prueba 2. 82](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494529)

[Figura 28: Resultado de la prueba 2 82](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494530)

[Figura 29: Grafico prueba 2. 82](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494531)

[Figura 30: Consulta prueba 3. 83](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494532)

[Figura 31: Resultado prueba 3. 83](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494533)

[Figura 32: Gráfico prueba 3. 83](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494534)

[Figura 33: Resultado prueba 4. 84](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494535)

[Figura 34: Gráfico prueba 4. 84](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494536)

[Figura 35: Consulta prueba 4. 84](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494537)

[Figura 38: Resultado prueba 5. 85](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494538)

[Figura 38: Consulta prueba 5. 85](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494539)

[Figura 38: Gráfico prueba 5. 85](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494540)

[Figura 39: ETL: trabajo completo 95](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494541)

[Figura 40: ETL: Dimensiones 96](#_Toc444494542)

[Figura 41: Carga tabla de hechos. 97](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494543)

[Figura 42: ETL: Especies 98](#_Toc444494544)

[Figura 43: Transformacion perspectiva tiempo. 98](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494545)

[Figura 44: Transformacion perspectiva lugares. 98](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444494546)

[Figura 45: Transformación perspectiva redes. 99](#_Toc444494547)

[Figura 46: Transformación perspectiva campañas. 99](#_Toc444494548)

[Figura 47: Transformación perspectiva horario. 99](#_Toc444494549)

[Figura 48: Transformación tabla de hechos. 100](#_Toc444494550)

**Índice de tablas**

[Tabla 1: Matriz de Dimensiones (Bus Matrix) 41](#_Toc444495924)

[Tabla 2: Medidas de longitud de redes en metros. 51](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444495925)

[Tabla 3: Superficie de redes empleadas, ordenado por sitio y campaña. 51](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444495926)

[Tabla 4: Arquitectura de Bus 65](#_Toc444495927)

[Tabla 5: Prioridad de procesos 66](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444495928)

[Tabla 6: Correspondencia - perspectiva Campaña 71](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444495929)

[Tabla 7: Correspondencia - perspectiva Especies 71](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444495930)

[Tabla 8: Correspondencia - perspectiva Redes 71](file:///C:\Users\Cinthia\Dropbox\TESIS%20CINTHIA%202015\Tesis%20Cinthia\Aplicacion%20Hefesto\Doc%20FINAL%20EN%20PROCESO.docx#_Toc444495931)

[Tabla 9: Correspondencia - perspectiva Lugares 71](#_Toc444495932)

[Tabla 10: Correspondencia - perspectiva Tiempo 71](#_Toc444495933)

[Tabla 11: Cálculo intermedio CPUE 72](#_Toc444495934)

[Tabla 12: Correspondencia con Indicadores 72](#_Toc444495935)

**Capítulo 1**

*Introducción*

# Introducción

## Motivación

En la actualidad, la influencia de las bases de datos dentro de la vida cotidiana es mayor de lo que se puede percibir a simple vista.Todas las redes sociales, sitios web, y una enorme cantidad de empresas se fundan en el uso de una base de datos que recolecta y almacena información. El exceso de datos se ha convertido en una de las principales problemáticas a enfrentar por todas las organizaciones, con lo cual resulta imprescindible la selección inteligente de la información que responda a las particularidades de cada organización y con ello incida de modo efectivo sobre su desempeño.

El Proyecto de Biología Pesquera Regional (UNaM[[1]](#footnote-2)) a través de convenios con la EBY[[2]](#footnote-3)realiza monitoreos mensuales de la fauna íctica en el río Paranáaguas arriba de la represa Yacyretá, desde el año 1993 a la actualidad, generando una gran cantidad de datos que necesitan ser almacenados y luego analizados.

Actualmente, la información contenida enuna base de datos se utilizapara dar respuestas a los requerimientos operacionales básicos, tomar decisiones y/o emitir recomendaciones en favor del estado y conservación de la fauna íctica.

Además dicha información también es utilizada para análisis y confección de los informes semestrales, anuales y/o históricos, los cuales requieren por su gran volumen de datos un tiempo considerable para el análisis y elaboración, debiendo asignar recursos humanos expertos para tal fin. Esta tarea resulta engorrosa, costosa en cuanto a tiempo y se encuentra sujeta a la visión y posible error del experto debido al enorme esfuerzo que supone la comprensión de tal volumen deinformación.

El objetivo principal de este trabajo consiste en la implementación de una solución de inteligencia de negocios quepueda colaborar en el proceso de análisisy toma de decisiones de manera ágil y dinámica, disminuyendo además la ocurrencia de errores a través del acceso a datos previamente calculados.Por otra parte,xxx presentar la informaciónen interfaces claras para el analista o tomador de decisiones, asegurando una estructura homogénea y consistente en los datos. Para lograr el objetivose utilizarán técnicas de modelos multidimensionales y procesamiento analítico en línea.xxxxx

## Objetivos



### Objetivo General

Diseñar, desarrollar e implementar una solución de Inteligencia de Negocios que permita la extracción de conocimiento a través de técnicas OLAP[[3]](#footnote-4) de la base de datos del Proyecto de Biología Pesquera Regional; y con ello proporcionar una herramientade soporte para la toma de decisiones.

### Objetivos Específicos

* Relevar y analizar xx fuentes bibliográficas.
* Analizar y evaluar xx diferentes metodologías y herramientas para el desarrollo del *DW[[4]](#footnote-5)*.
* Analizar los indicadores claves del negocio para la creación de una solución de inteligencia de negocios.
* Analizar las fuentes de datos existente.
* Diseñar los flujos de datos y control para realizar el proceso de extracción, transformación y carga (ETL)[[5]](#footnote-6).
* Automatizar el proceso ETL. xxxx
* Diseñarel DW.
* Diseñar y desarrollarel DM[[6]](#footnote-7) para el posterior análisis multidimensional.
* Realizar las pruebas y validación de datos.
* Implementar la solución.

## Estructura del documento

La tesis se divide en seis capítulos y tres anexos. A continuación se describe brevemente el contenido de los mismos.

En el capítulo 2 se detalla un marco teórico con los conceptos utilizados a lo largo del presente trabajo.

En el capítulo 3 se realiza una descripción de la organización, el planteo del problema existente y los objetivos que con el presente trabajo se pretenden mitigar.

En el capítulo 4 se presenta la solución propuesta, describiendo los pasos realizados, y las funcionalidades utilizadas de las metodologías Kimball y Hefesto en cada uno de los mencionados pasos.

En el capítulo 5 se describen las pruebas realizadas con los usuarios, los resultados obtenidos y las recomendaciones que surgieron de la prueba experimental.

En el capítulo 6 se plantean las conclusiones de la tesis como así también los trabajos futuros que se pueden implementar.

**Capítulo 2:**

*Marco Teórico*

# : Marco Teórico

El presente capítulo describe los conceptos utilizados a lo largo del documento.

## Bases de Datos

### Definición

Una base de datos es una‘colección’que se compone de una serie de datos organizados y relacionados entre sí.

Según *Date*[1], es una colección de datos que contiene información relevante para una organización y cuya finalidad principal es permitir a los usuarios recuperar, actualizar o eliminar los datos.

A lo largo de las últimas décadas, el uso de estas colecciones creció en todos los sectores y con ello surgieron distintos tipos de bases de datos.Entre ellos se pueden encontrar las bases de datos relacionales, las orientadas a objetos y las multidimensionales.El más popular y utilizado por las organizaciones es el relacional.

Para definir, construir y manipular una base de datos se utilizanSGBD[[7]](#footnote-8)que a través de una agrupación de programas sirven a tal fin.

La mayoría de los SGBD relacionales utilizan un lenguaje universal denominado SQL[[8]](#footnote-9).

“SQL es un estándar definido que se compone de un LDD[[9]](#footnote-10) para especificar el esquema y un LMD[[10]](#footnote-11) para expresar las consultas y modificaciones a realizar”[2].

### Objetivos e Importancia

Los principales objetivos de un sistema de base de datos se focalizan en disminuir la redundancia, inconsistencia y el aislamiento de los datos.

La mayoría de las organizaciones poseenun acceso concurrente a los sistemaspor parte de múltiples usuarios que manipulan de manera simultánea la información.Una base de datos garantiza la integridad y seguridadmediante la limitación del acceso a los usuarios, de tal manera que cierta información podrá ser controlada y accedida solamente por los administradores. En conclusión, aumenta la productividad ya que al tener agrupada toda lainformación en un único lugar se garantizael acceso rápido y eficiente a los datos.

## **Sistemas OLTP**

Los sistemasOLTP[[11]](#footnote-12) representan a toda la información transaccional u operacional que genera la organización en su accionar diario, como así también las fuentes externas con las que pueda llegar a disponer.

Entre las fuentes de datos más habituales que pueden existir en cualquier organización se encuentran:

* Archivos de textos.
* Hipertextos.
* Hojas de cálculo.
* Bases de datos.
* Informes semanales, mensuales y/o anuales, etcétera.

Las aplicaciones de base de datos OLTP son perfectas para administrar datos que cambian con frecuencia. Se diseñan para responder inmediatamente a las solicitudes del usuario, dondecada solicitud se considera una transacción única. Estas solicitudes pueden implicar procesos de adición, recuperación, actualización o eliminación de datos.

La mayoría de las organizaciones, se basan en los sistemas OLTP para guardar una gran cantidad de datos y tener tiempos de respuestas cortos a las numerosas transacciones realizadas cotidianamente.Sin embargo, su eficiencia no se basa en la consulta masiva de grandes cantidades de información y tampoco en el análisis de la misma. Estos sistemas no están diseñados para responder a preguntasglobales sino a pequeños grupos de datos. Las preguntas que involucren consultascomplejas podrían resolverse en un lapso extenso.

## Inteligencia de Negocios

### **Breve historia**.

El términoBI[[12]](#footnote-13)se expuso por primera vez en el año 1958 por Hans Peter Luhnque la define como: “[…] la capacidad de aprender las interrelaciones de los hechos presentados de manera que guíen las acciones hacia una meta deseada*”*[3].

En el año 1989 Howard Dresner propuso un término general para describir Inteligencia de Negocios como *"*[…]los conceptos y métodos para mejorar la toma de decisiones de negocios mediante el uso de sistemas de soporte basados en hechos"[5].

A partir de este momento se multiplicaron las aplicaciones como soporte a la Inteligencia de Negocios, dando lugar a la era de *Business Intelligence 1.0.* Dichasaplicaciones facilitaron el acceso a la información considerablemente pero al existir tantas en el mercado no podía consolidarse en una única solución, y a este problema lo denominaron “la existencia de muchas versiones de la realidad”[6].

No fue hasta el año 2000, la era de *Business Intelligence 2.0*donde las aplicaciones de inteligencia de negocios comenzaron a consolidarse, brindando finalmente diferentes herramientas que dan soluciones completas a las organizaciones para la toma de decisiones.

El año 2012, la era de los *smartphones*,da comienzo a *Business Intelligence 3.0*[7]con un enfoque de colaboración para grupos de trabajos sociales. Se convierte en el catalizador para empresas impulsadas por los resultados del grupo de trabajo colaborativo. El nuevo conjunto de herramientas necesita ser ‘omnipresente’ en todos los dispositivos para que el grupo de trabajo pueda tomar decisiones en tiempo real desde cualquier lugar, sobre cualquier proceso de la organización. Así mismo, esta nueva era necesita más protección de los datos por la complejidad que supone acceder a través de la nube y los sistemas subyacentes utilizados para conseguirlo.

Examinando las distintas definiciones y atravesando cada era, queda claro que el punto de partida son los hechos o datos de una organización, y que la Inteligencia de Negocios no consiste en una metodología o herramienta en particular, sino que es un conjunto de tecnologías, herramientas y métodos que sirven a un objetivo en común,**obtener información útil para la toma de decisiones del negocio.**

### Definición

Inteligencia de Negocios es un conjunto de métodos, tecnologías y herramientas que permiten integrar y homogeneizar los datos de todos los ámbitos de una organización, ya sean sistemas OLTP, fuentes de datos internas y/o externas; para transformarlos en información estructurada para su posteriorexplotación y análisis. De esta manera se obtienen conocimientos que brindan soporte a la toma de decisiones.

## Data Warehouse

El término DW fue definido por W.H. Inmoncomo “una colección de datos orientado a un tema, integrado, no volátil y variable en el tiempo que dan soporte a la toma de decisiones gerenciales”[8].

### Características

A continuación se describe cada característica de la definiciónde DW propuesta por Inmon:

* **Orientado a un tema**:a diferencia de los OLTP que cuentan con información solamente de las áreas para las que fueron implementadas, un DW está orientado a consultar información relevante relacionada a la operatoria general de la organización,
* **Integrado**: se considera la característica más importante del DW. Debido a que una organización puede operar con distintas fuentes de datos, lo más probable es que no haya consistencia entre las aplicaciones respecto de la codificación, convenciones de nombramiento, atributos físicos, cálculos sobre los atributos, etcétera. Por ello, es fundamental establecer una codificación única e independiente del método o la aplicación de origen. De esta manera, se puede asegurar a los usuarios del DW que los datos son coherentes y consistentes,
* **No volátil**: a diferencia de los sistemas transaccionales que en su curso normal de trabajo manipulan los datos, en el DW la información no se actualiza ni elimina. Los datos se cargan masivamente y se mantienen como una ‘foto’, en un formato estático,
* **Variable en el tiempo**: los datos almacenados en el DW corresponden a un lapso de entre 5 y 10 años, lo cual permite analizar a la organización a través del tiempo y contrastar con información histórica.

## Arquitectura de DW

La arquitectura básica del DW integra los siguientes componentes:

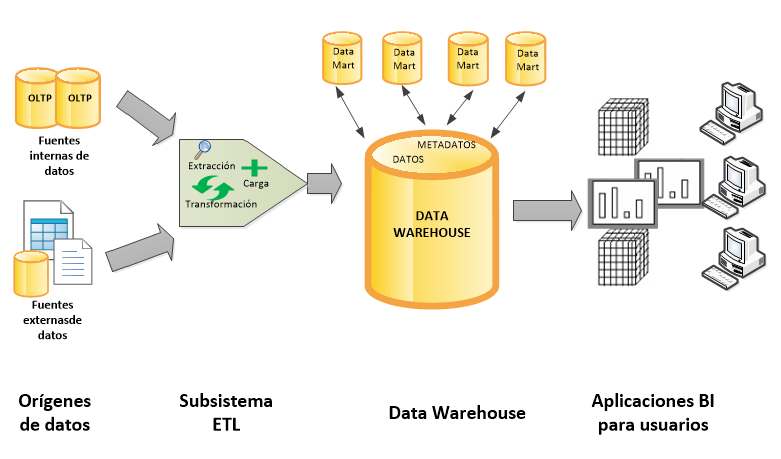


Figura 1:Arquitectura básica de un DW.

* + 1. Orígenes de Datos

Son las distintas fuentes de datos que puede utilizar una organización:

#### **Fuentes internas**

Representa a la/s base/s de datos del/los sistemas de gestión que almacenan las transacciones diarias,

#### Fuentes externas

Representan losdocumentos, bases de datos externas a la organización, hojas de cálculos, etcétera.

### **Subsistema ETL**

Es el proceso que permite fusionar datos desde las distintas fuentes, los reformatea, limpia, sumariza, disgrega y carga al DW para ser analizados. Es una actividad muy importante en el proceso de *data warehousing*, que no es muy visible para los usuarios finales pero que fácilmente ocupa el 70 por ciento del tiempo y esfuerzo en la mayoría de los proyectos de DW[9]. El proceso ETL se puede observar en la Figura 2:

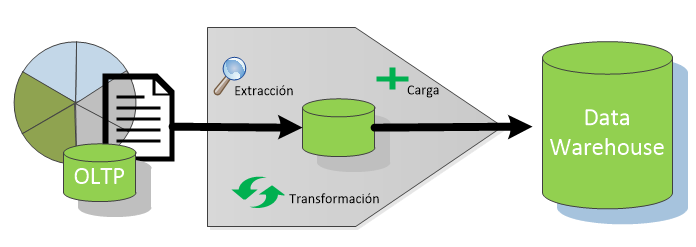


Figura 2: Proceso ETL.

Este proceso se compone de las siguientes etapas:

* + Extracción: basándose en las necesidades y requerimientos de los usuarios, se examinan las distintas fuentes de datos con las que opera la organización y se extrae la información que se considere pertinente para el desarrollo,
  + Transformación: en este pasose convierten los datos incongruentes en un conjunto de datos consistentes para insertarlos en el DW. Aquí se definen los estándares a utilizar respecto de la codificación (Figura 3), medida de los atributos (Figura4), convenciones de nombramientos (Figura 5), limpieza y calidad de datos para integrarlos finalmente al DW,

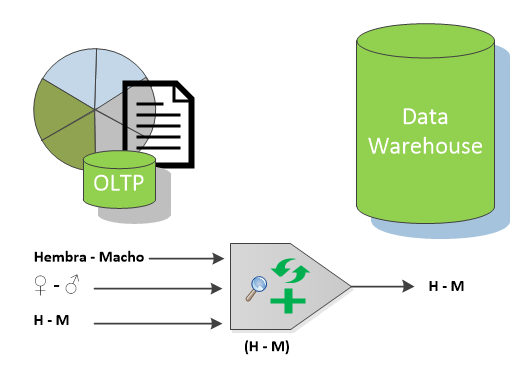


Figura 3:Transformación: codificación

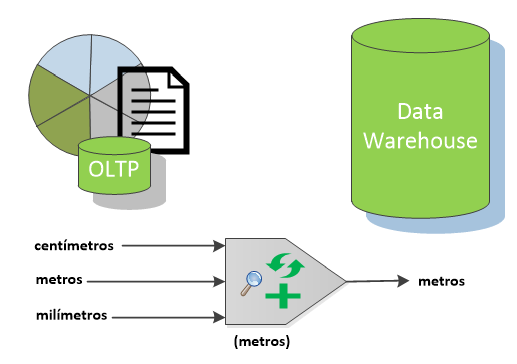


Figura 4: Transformación: medida de atributos

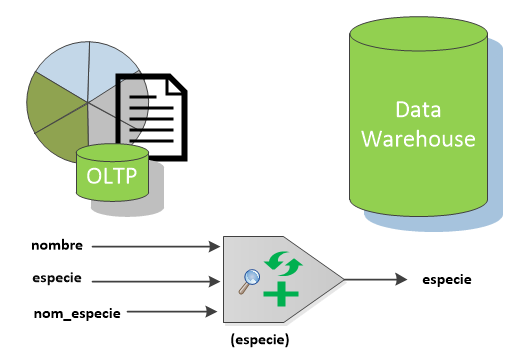


Figura 5: Transformación: convención de nombramiento.

* + Carga: una vez realizada la transformación, se tomanlos datos y se los traslada al DW. Esta función realiza la carga inicial y posteriormente se utiliza para la actualización o mantenimiento del mismo.
  + StagingArea[[13]](#footnote-14):una vez que los datos son extraídosdesde las distintas fuentes, se guardan en un área de almacenamiento intermedio que permite manipular los datos sin interrumpir los OLTP. Al estar en un área intermedia no depende de la disponibilidad de los sistemas transaccionales y tampoco del DW, pudiendo así realizar las manipulaciones de datos necesarias, facilitando la integración de las diversas fuentes internas y externas de datos.

### Data Mart

Un DM es un subconjunto del DW que se encuentra orientado a un proceso o departamento específico de la organización.

Muchos DW comienzan siendo DM, ya que el desarrollo abocado a un solo departamento o sección requiere menos tiempo, minimiza los riesgos y permite realizar una entrega del trabajo correspondiente a BI en tiempos moderados. Es importante realizar una planificación completa para la implementación del DW, ya que incorporar DM sin una correcta planificación puede derivar en redundancia de datos o en DM que se ajuste a la terminología de un solo sector.

Para que la construcción de un DW o DM sea ordenado, se pueden optar entre dos estrategias:

Estrategia *top-down*: consiste en diseñar primero el DW y a partir de éste, desarrollar, construir y cargar los distintos DM. Esta estrategia conlleva un importante tiempo de análisis y es necesario conocer en detalle todos los datos que lo integrarán y sus interrelaciones ya que cualquier cambio impactará en todo el DW.

Estrategia *Bottom-Up*: a diferencia del anterior, esta estrategia define previamente los DM para luego integrarlos en un DW centralizado. La ventaja de ello radica en que un DM puede ser implementado sin que exista el DW, en un corto lapso, a un bajo costo y presentando rápidamente una solución en la que se expongan sus cualidades y beneficios.

## Modelado Dimensional

Es una técnica para el diseño lógico de un DW.Se basa en la generación de tablas de hechos, una colección de tablas de dimensiones y las relaciones entre ellos. A continuación se describe cada uno de estos componentes:

### Tabla de Dimensiones

Las tablas de dimensiones proveen un medio para analizar el contexto del negocio. Almacenan datos cualitativos que describen: quién, dónde, cuándo,cómo y porquéde un evento o hecho.

En resumen, representan los aspectos de interés a través de los cuales los usuarios podrán filtrar y analizar la información almacenada en la tabla de hechos.

### Tabla de Hechos[[14]](#footnote-15)

Estas tablas contienen datos cuantitativos que serán precisamente los indicadores a analizar de la organización. Un hecho define qué analizar del proceso de negocio, es una medida y por tanto siempre un valor numérico como por ejemplo: cantidad, monto total, etcétera.

Existen dos tipos de hechos, los básicos y los derivados:

* + - **Básicos**: son los que se encuentran representados por un campo dentrode la tabla de hechos, por ejemplo ‘precio’ o ‘cantidad’.
    - **Derivados**: son los que surgen de la combinación de uno o más hechos con una operación lógica o matemática de por medio, por ejemplo:

🡪 total = precio \* cantidad

🡪Promedio general = AVG (notas\_alumno)

### Tipos de modelado de un DW

Para elaborar el modelo dimensional del DW es necesario tener en cuenta los diferentes esquemas entre los que se puede optar y que dan lugar al modelo físico de los datos. A continuación se describe cada esquema:

#### Esquema estrella

Este esquema consta de una tabla central de hechos y de varias tablas de dimensiones relacionadas a ella mediante sus respectivas claves foráneas (Figura 6). A diferencia de un modelo relacional, el modelo dimensional debe estar desnormalizado. Con esto se busca evitar uniones entre tablas y mejorar el tiempo de respuesta ante las consultas del usuario. El esquema estrella es el más simple de interpretar, documentar y mantener, sin embargo es el esquema menos robusto para la carga y el más lento de construir. No obstante,posee ventajas interesantes, por ejemplo, su diseño es fácilmente modificable, simplifica el análisis y por sobre todas estas, posee el mejor tiempo de respuesta ante las consultas de usuario,

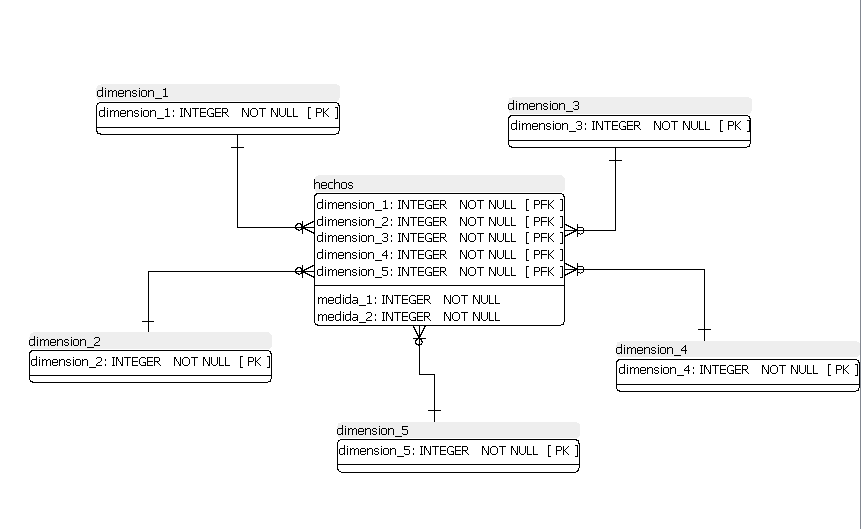


Figura 6: Esquema estrella

#### Esquema copo de nieve

Este esquema es similar al anterior, cuenta con una tabla central de hechos y varias tablas de dimensiones. La diferencia radica en que las dimensiones están parcialmente normalizadas como resultado de las jerarquías establecidas entre ellas (Figura 7). El motivo de utilizar este tipo de modelo reside en la posibilidad de segregar los datos de las tablas de dimensiones y responder a los requerimientos solicitados.

Este esquema posee mayor complejidad en su estructura, y al encontrarse parcialmente normalizada el desempeño puede verse reducido por las uniones entre tablas que deberán realizarse en las consultas.

Se debe tener especial cuidado al crear jerarquías ya que el crecimiento de tablas puede duplicarse rápidamente hasta el punto de ser inmanejables afectando gravemente el desempeño y tiempo de respuesta de los resultados,

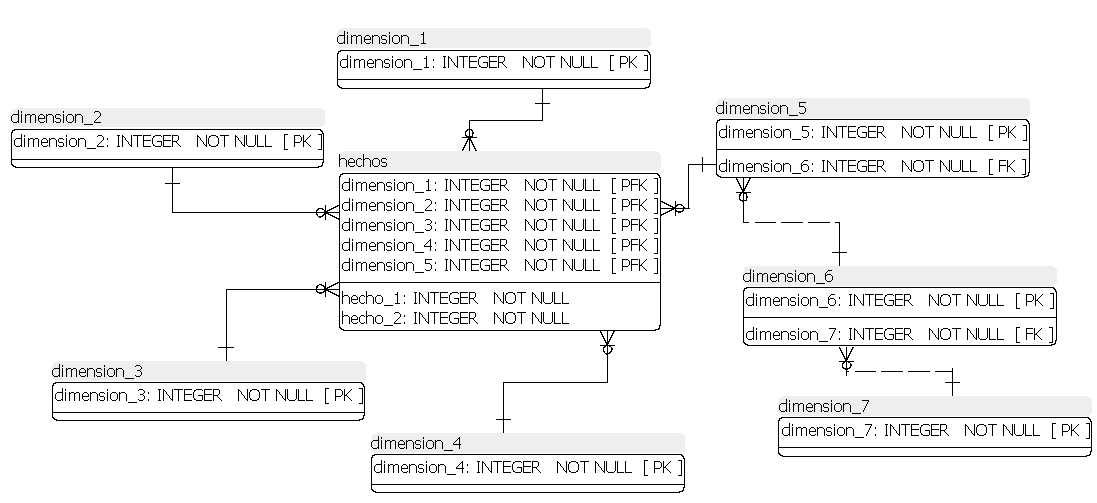


Figura 7: Esquema copo de nieve.

#### Esquema constelación

Este modelo es una variación del esquema estrella. La diferencia se encuentra en que posee más de una tabla de hechos y no necesariamente se relacionan con todas las dimensiones (Figura 8).

Es el esquema menos utilizado por la complejidad que contrae y porque, a diferencia de los otros esquemas, no es soportado por todas las herramientas de consulta y análisis.

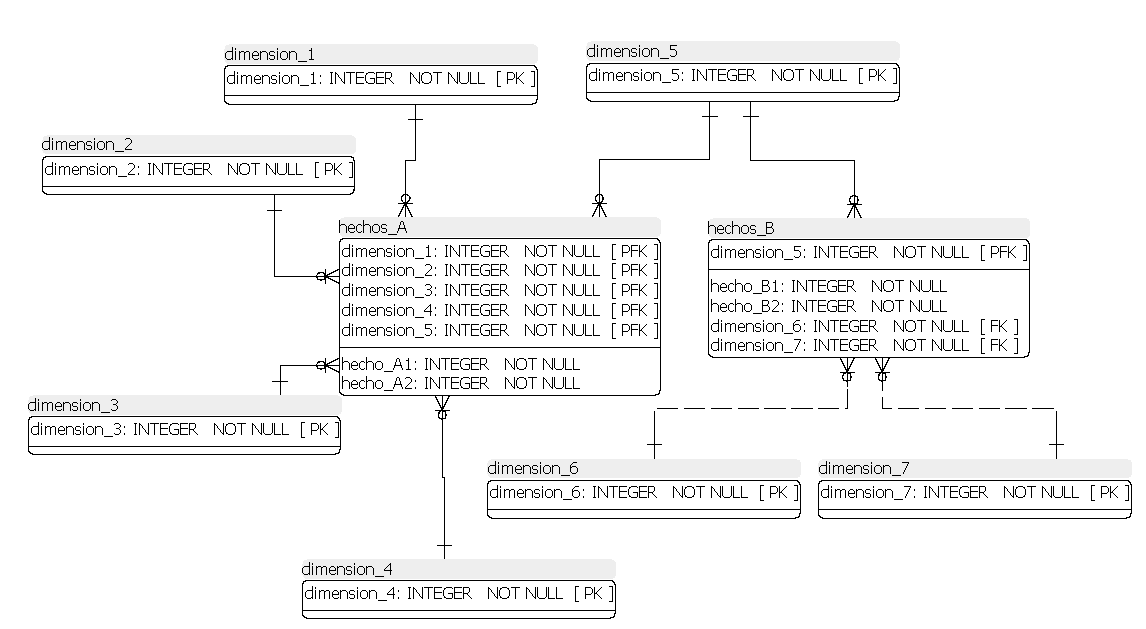


Figura 8: Esquema constelación.

### Operaciones sobre modelos dimensionales

Las operaciones que se pueden realizar sobre modelos dimensionales son las que en efecto, permitirán a los usuarios explorar e investigar los datos.

Existen cuatro tipos básicos de operaciones en OLAP para el análisis de datos. Por un lado se encuentran las operaciones que definen el nivel de granularidad, entre las que se encuentran*drill-down* y *roll-up* y por otro lado, las que permiten navegar entre las dimensiones, siendo*slice*y *dice*[10].

Relacionado a lo anterior, Hefesto [11]agrega operaciones a las ya mencionadas para el análisis de datos. A continuación se describen cada una de ellas:

* *Slice and dice*: separa y combina los datos de maneras diversas,
* *Swap*: permuta dos dimensiones de análisis, es decir que altera filas por columnas,
* *Down*: baja el nivel de visualización de las filas, a una jerarquía inferior,
* *Drill-down*: permite apreciar los datos en un mayor nivel de detalle. Esto se logra a través de las jerarquías definidas en el cubo. *Drill- down* es ir de lo general a lo específico,
* *Drill-up*: es opuesto a *drill-down*, permite apreciar los datos en menor nivel de detalle, subiendo por las jerarquías definidas en el cubo. Va de lo específico a lo general,
* *Drill-across*: funciona de forma similar a *drill-down* con la diferencia que *drill-across* no se realiza sobre una jerarquía, sino que su forma de ir de lo general a lo específico es agregar un atributo o dimensión a la consulta como nuevo criterio de análisis,
* *Roll-across*: funciona de forma similar a *drill-up* con la diferencia que *roll-across* no se realiza sobre una jerarquía, sino que la forma de ir de lo específico a lo general se logra quitando un atributo de la consulta, eliminado de esta manera un criterio de análisis,
* *Pivot*: permite seleccionar el orden de visualización de los atributos e indicadores, con el objetivo de analizar la información desde diferentes perspectivas o dimensiones,
* *Drill-through*: permite apreciar los datos en su máximo nivel de detalle. Esto brinda la posibilidad de analizar cuáles son los datos relacionados al valor de un indicador que se ha sumarizado dentro del cubo multidimensional.

## OLAP

### Definición

El termino OLAP, propuesto por Codd, E.[12]en el año 1993, promueve la idea de permitir a los analistas, gerentes y/o ejecutivos de una organización, consultar de manera rápida, consistente e interactiva grandes volúmenes de datos.

OLAP es un tipo de procesamiento que se caracteriza, entre otras cosas, por permitir el análisis multidimensional.Los datos provienen generalmente de un DW, y por ello, la estructura que permite este tipo de análisis consta de una tabla de hechos cuyos atributos describen el objeto a analizar (por ejemplo ventas) y de tablas de dimensiones, que describen el contexto del hecho analizado (por ejemplo, productos o clientes que intervinieron en las ventas). A esta estructura multidimensional se denomina “**cubo**” y su implementación puede ser abordada de distintas maneras.

### Tipos de implementaciones OLAP

Existen diferentes tipos de implementaciones de sistemas OLAP. A continuación se describe cada una de ellas:

#### ROLAP[[15]](#footnote-16)

Son siglas de OLAP Relacional.Este tipo de organización física se implementa sobre tecnología relacional y cuenta con los beneficios de un SGBD relacional. ROLAP es una arquitectura flexible y funciona de manera óptima para pequeños volúmenes de datos. Generalmente, los datos del DW son almacenados y gestionados mediante SGBD Relaciones, que aporta la ventaja de realizar consultas directamente por medio de lenguaje SQL.

#### MOLAP[[16]](#footnote-17)

Son siglas de OLAP Multidimensional. Los datos son almacenados en una base multidimensional y para ello requiere una estructura específica de almacenamiento. El sistema MOLAP utiliza una arquitectura de dos niveles, la base de datos multidimensional y el motor analítico. La base de datos multidimensional es la encargada del manejo, acceso y obtención de los datos. El nivel de aplicación es el responsable de la ejecución de las consultas OLAP.

Generalmente los datos se encuentran precalculados, para ello utiliza técnicas de compactación de datos que luego favorecen el desempeño del DW.

A diferencia del tipo de implementación anterior, MOLAP para la generación de cubos multidimensionales necesita de un software específico que administre y gestione los datos de manera multidimensional.

#### HOLAP[[17]](#footnote-18)

Constituye un sistema híbrido entre MOLAP y ROLAP. Consiste en almacenar parte de los datos en un motor relacional y otros en una base de datos multidimensional.

Los datos precalculados y agregados forman parte del almacenamiento multidimensional y los de menor nivel de detalle se almacenan en estructuras relacionales. Básicamente, se utiliza ROLAP para navegar y explorar datos y MOLAP para la realización de los tableros de control.

## Tableros de control

Un tablero de control, es una herramienta cuya característica básica es visualizar información útil para diagnosticaruna situacióny contribuir en la toma de decisiones[13].

El tablero de control permite comprender la situación del negocio a través de la visualización de los indicadores claves de desempeño (KPIs[[18]](#footnote-19)).

El desarrollo de estos tableros se focaliza en responder a preguntas esenciales sobre el dominio de la organización, alerta al usuario sobre fluctuaciones que puedan afectar al negocio y brinda soporte a la toma de decisiones.

Según el propósito de su implementación y de las necesidades de las empresas se pueden desarrollar cuatro tipos genéricos de tableros:

* **Tablero de Control Operativo**: ofrece información sobre los procesos básicos de la organización. Responde preguntas esenciales de un sector, área o informe en particular. El usuario principal de este tipo de tableros es generalmente el responsable de área,
* **Tablero de Control Directivo**: este tipo de tablero posibilitael monitoreode los resultados de la empresa en su conjunto y particularmente de los sectores que fueran definidos como áreas claves de análisis. El conjunto de indicadores está orientado al resultado interno de la organización,
* **Tablero de Control Estratégico**: presenta tanto la información interna como externa a la organización para compararsi la estrategia actual de la organización es rentable respecto del entorno, que a menudo es cambiante. De esta manera permite realizar y coordinar un seguimiento estratégico,
* **Tablero de Control Integral**: este tipo de tablero incluye información tanto operativa, estratégica como gerencial. Como se mencionó anteriormente, un tablero de control cumple la función de diagnosticar e informar. El tablero de control integral cumple la función de alinear la organización a través de un modelo de negocio.[14]

Para el desarrollo de tableros de control se recomiendan seguir los siguientes pasos[15]:

* + - * 1. Selección de los datos a mostrar basado en los requerimientos del usuario.
        2. Elección del formato de presentación. A través de un boceto se presentan las secciones y elementos que contendrá el tablero.
        3. Planificación de la interacción del usuario. Se refiere a la dinámica que tendrá el tablero.
        4. Implementación del tablero.

## Metodologías

Hoy en día existen diversas metodologías sobre diseño y construcción de DW. Cada fabricante de software referente a Inteligencia de Negocios ha buscado imponer una metodología asociada a sus productos. Sin embargo, entre las pioneras y más predominantes se encuentran las metodologías propuestas porKimball[16] e Inmon[17]. En el año 2007 surge la primera edición de Hefesto, una metodología desarrollada por Bernabeu, R.[11] cuya propuesta se encuentra fundamentada en una amplia investigación y como valor agregado aporta las experiencias de quienes utilizaron la metodología en diversos países y con distintos fines.

A continuación se realiza una breve descripción de cada una de estas metodologías.

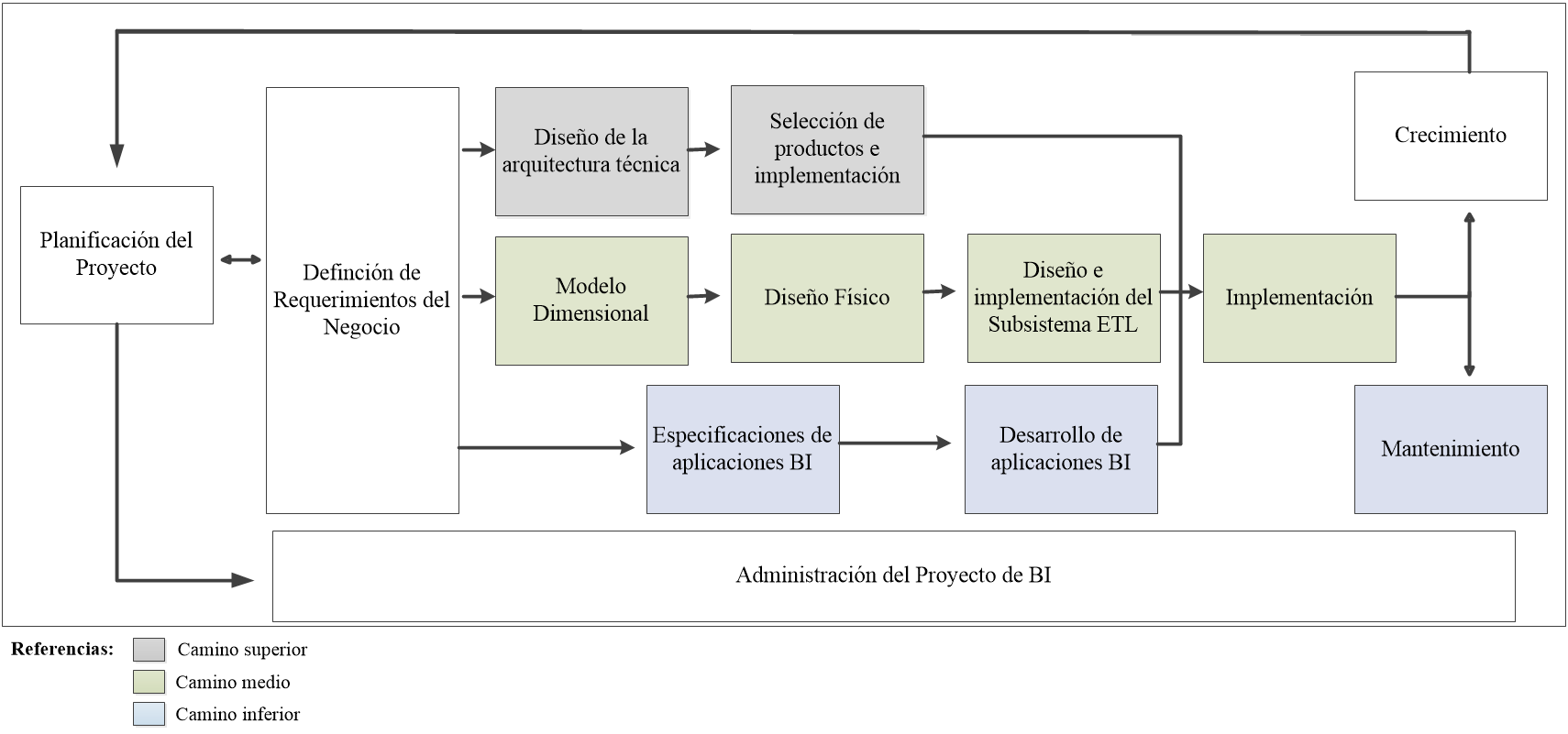
### Metodología propuesta por Kimball

La metodología se basa en lo que Kimball denomina “Ciclo de vida dimensional del negocio”[[19]](#footnote-20). Las tareas que incluyen esta metodología se pueden observar en la (Figura 9).

Todo el proceso se basa en cuatro principios básicos:

* Centrarse en el negocio.
* Construir una infraestructura de información adecuada.
* Realizar entregas en incrementos significativos.
* Ofrecer la solución completa.

Figura 9: Metodología Kimball – Bussiness Dimensional Lifecycle.



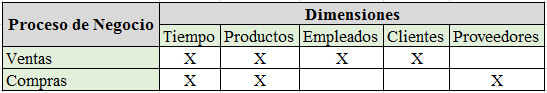
Kimball propone una arquitectura ascendente (bottom-up), partiendo de la idea de que un DW es la unión de todos los DM de una organización. Con ello apunta a realizar una solución que se puede implementar en poco tiempo y mediante el cual se obtienen resultados rápidos para demostrar el valor e importancia de la solución a los involucrados.

Del ciclo de vida planteado, es preciso resaltar el rol central de la tarea de *definición de requerimientos* ya que partiendo del mismo se podrán ejecutar las tareas referentes a las siguientes tres áreas:

* + **Camino superior – Tecnología**: implica tareas relacionadas a la elección de la arquitectura a utilizar y la selección de herramientas o software específico para el desarrollo,
  + **Camino medio – Datos:** en esta sección se diseña e implementa el modelo dimensional, además del subsistema ETL para transportar los datos desde las distintas fuentes al DW,
  + **Camino inferior – Aplicaciones de Inteligencia de Negocios:** aquí se diseñan, desarrollan e implementan las aplicaciones para los usuarios finales.

Una vez que se hayan determinado los procesos de negocio, se utiliza una herramienta denominada ‘matriz de dimensiones’[[20]](#footnote-21). En la matriz se exponen los procesos y dimensiones identificadas durante el relevamiento de requisitos. Los procesos son ubicados en las filas y las dimensiones formaran parte de las columnas.Cada dimensión representa un criterio por medio del cual se podránsumarizar los datos numéricos que se van a analizar. Estos datos numéricos representan las medidas, tales como cantidad, monto total, etcétera. Se puede observar un ejemplo en la Tabla 1, en donde cada X en la intersección de filas y columnas significa que el proceso de negocio se identifica o puede ser analizado desde dicha dimensión.

Tabla 1: Matriz de Dimensiones (Bus Matrix)



Finalmente se busca priorizar los requerimientos o procesos asignando una valoración, entre las que se puede optar por baja, media o alta. La opción “alta” implica a un proceso prioritario que debe ser seleccionado para el desarrollo.

### Metodología propuesta por Inmon

A diferencia del anterior, Inmon plantea una arquitectura descendente (top-down), partiendo de la premisa de que los DM se deben diseñar y desarrollar después de haber construido el DW completo de la organización.

Esta propuesta es recomendable cuando resulta difícil representar el modelo a través de dimensiones y la complejidad de la solución es muy alta. De lo contrario se corre el riesgo de invertir mucho tiempo, esfuerzo y dinero sin poder presentar un resultado para validación al usuario.

### Metodología propuesta por Bernabeu

El autor la denominó **Hefesto**. Es una metodología propia de Bernabeu, que se fundamenta en una amplia investigación, en la comparación de metodologías existentes y en experiencias propias y de colegas en procesos de confección de DW.

[“*La idea principal es comprender cada paso que se realizará, para no caer en el tedio de tener que seguir un método al pie de la letra sin saber exactamente qué se está haciendo, ni por qué.*”] (Bernabeu, R.)[11]

El objetivo principal de la metodología es que de manera metódica y sencilla se comprendae interprete cada paso que se ejecuta durante el proceso y no caer en el error de seguir un método sin entenderqué se está haciendo.

Por otra parte, Hefesto busca entregar una primera implementación que satisfaga parte de las necesidades de la organización, demostrando así las ventajas del DW y que sirva de motivación a los usuarios.

#### Características

A continuación se describen las principales características de la metodología:

* + - Se basa en los requerimientos del usuario, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez a los cambios del negocio,
    - Reduce la resistencia al cambio ya que involucra a los usuarios finales a participar en cada paso del proceso.Los usuarios finales van guiando la implementación, tomando decisiones respecto del comportamiento y funciones del DW,
    - Utiliza modelos conceptuales y lógicos que son fáciles de comprender, interpretar y analizar,
    - Es independiente del ciclo de vida que se utilice por lo tanto la metodología pueden embeberse fácilmente en el ciclo de vida elegido para el desarrollo,
    - Por otra parte es independiente de las estructuras físicas que contengan al DW;y también del tipo de modelado seleccionado, ya que se aplica tanto para DW como para DM,
    - Los objetivos y resultados esperados en cada etapa se distinguen fácilmente y son sencillos de comprender, lo que favorece al proceso general.Cada resultado es el punto de partida para llevar a cabo los siguientes pasos de la metodología.

#### **Pasos de la metodología Hefesto**:

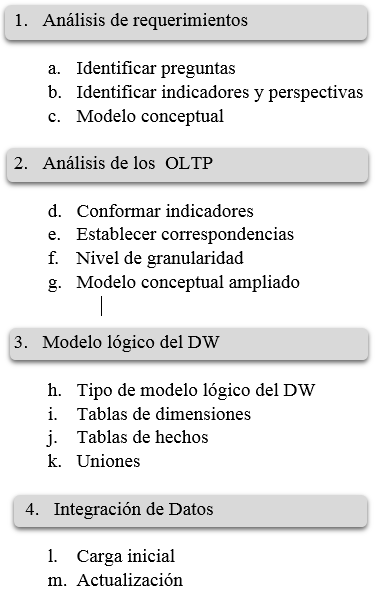


Figura 10: Metodología Hefesto 2.0

**Capítulo 3**

*Descripción del problema*

# Descripcióndel Problema

En este capítulo se describe a la organización, para este caso el Proyecto de Biología Pesquera Regional (sección 3.1), su estructura organizacional (sección 3.2) y las actividades que realiza (sección 3.3).

## La organización

El Proyecto de Biología Pesquera Regional de la Universidad Nacional de Misiones realiza investigación básica y presta servicios a terceros.A través de convenios con la Entidad Binacional Yacyretá,se realizan monitoreos de la fauna íctica en el río Paraná aguas arriba de la represa Yacyretáen forma continua desde el año 1993 a la actualidad.El objetivo principal de los monitoreos consiste enobtener información sobre la estructura de la comunidad íctica,la composición específica de las capturas y los rendimientos pesquerosen diferentes sitios a lo largo del tiempo*.* Conjuntamente se analizandiversos parámetros pesqueros y características de la comunidad de estudio.

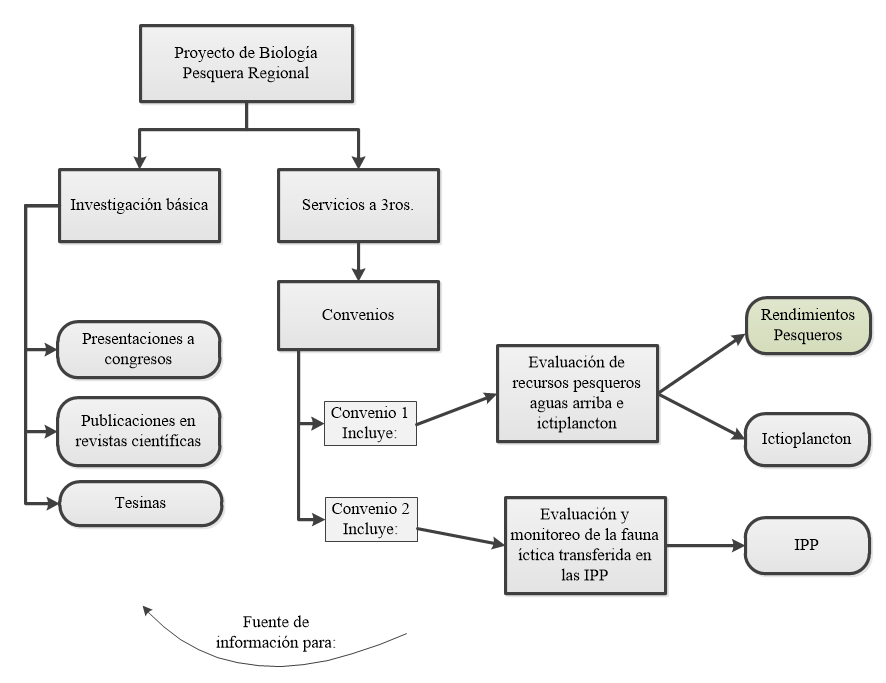
## Estructura organizacional

Con la finalidad de comprender el funcionamiento de la organización se describe su estructura organizacional ylos roles de cada actor dentro del proyecto.

En la Figura 11 se puede observar la estructura general de la organización:

xxx

Figura 11: Estructura organizacional del Proyecto de Biología Pesquera Regional



La organización actualmente tiene dos convenios activos como se puede observar en la Figura 11. Para el desarrollo del presente trabajo se analiza el convenio “Evaluación de los recursos pesqueros aguas arriba e ictioplancton”, y de este, entre todos los objetivos que tiene, el referido a “**Recursos Pesqueros**”. El monitoreo del río Paranárepresenta la parte práctica del proceso. Los datos obtenidos son fuente de información para realizar distintos análisis que retroalimentan al segundo objetivo del proyecto, la investigación básica.Como resultado de dicha investigación se producen publicaciones en revistas y enreuniones científicas.Por otra parte,los estudiantes universitarios avanzados que forman parte del proyecto, profundizan el estudio de alguna temática particular y realizan sus tesinas de graduación.

A continuación se presenta el organigrama de la organización (Figura12):

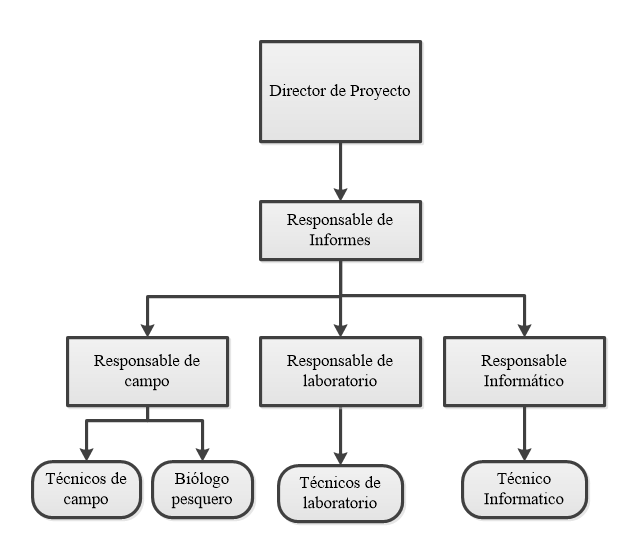


Figura 12: Organigrama del Proyecto Biología Pesquera Regional

A partir del organigrama, se describen brevemente los roles de cada actor:

* + - * **Director de proyecto**:se ocupa de planificar,coordinar y ejecutar con el

recurso humano a cargo, las actividades relacionadas con los objetivos del proyecto,

* + - * **Responsable de Informes**:se encarga de procesar, analizar la información obtenida del trabajo de campo y elaborar los informes en tiempo y forma,
      * **Responsable de campo**: es el encargado de coordinar y ejecutar la logística del trabajo de campo,
  + **Técnicosde campo**: aplican la metodología de muestreo para la obtención de los datos,
  + **Biólogo pesquero**: es el profesional experto en la temática de taxonomía e identificación de peces,
    - * **Responsable de laboratorio**: coordina el uso de mesada y del equipamiento óptico, como lupas y microscopios; mantiene el stock de los insumos utilizados para las diferentes actividades de conservación de las muestras como: soluciones, envases, colorantes, etc.,
      * **Responsable informático:**
  + **Técnico informático**: es responsable de la digitalización de los datos obtenidos en los monitoreos,actualización de la base de datos y de brindar información a los usuarios que la soliciten.

## Actividades

Los servicios prestados a terceros, a través de los convenios, estipulan monitoreos mensuales que se denominan Campañas y se llevan a cabo en un tramo del río que abarca casi 200km,en cinco sitios o estaciones (denominado Lugar para referirse a los sitios o estaciones, indistintamente).

Dichos monitoreos son realizados sobre el eje principal del río Paraná con diferentes frecuencias:muestreos mensuales y muestreos estacionales (se colectan datos cada tres meses).

### Sitios de muestreos

Los sitios se ubican en un tramo del río Alto Paraná entre las localidades de Ituzaingó(provincia de Corrientes) y Corpus (provincia de Misiones), en el área de influencia de la Central Hidroeléctrica Yacyretá.

Dichos sitios, se ubican en tres secciones del cauce principal del río: una región lotica o fluvial (Puerto Maní, SPO); una región de embalse que comprende a Garapé (GPE) y Toma de Agua Eriday (TAE). Además se incluye el análisis de dos cauces secundarios, el Arroyo Yabebiry (YBY)y el arroyo Garupá (GPA) **(**Figura 14**).**

****

Figura 13: Sitios de monitoreo.

Referencias: **1)** SPO, **2)** YBY, **3)** GPA, **4)** GPE y **5)** TAE.

### Metodología de muestreo

Para los muestreos se utilizanredes de espera monofilamento tipo agalleras. Las medidas de las redes (tomadas entre nudos opuestos) son de: 20 cm, 16 cm, 14cm, 12 cm, 8 cm, 7 cm, 6 cm, 5 cm y 4 cm. En la se describen las longitudes y alturas utilizadas teniendo en cuenta las medidas comerciales, de las cuales se considera el largo total como aquel delimitado por la relinga de plomo; y el altorepresenta la longitud del paño estirado.

Porcada monitoreo se registra tanto el alto como el largo de red utilizada, identificandola campaña y el lugar ya que las redes pueden ser reemplazadas por otras con diferentes medidas variando lasuperficie que representan.Un ejemplo de esto se puede observar en la , campaña 2 de TAE.

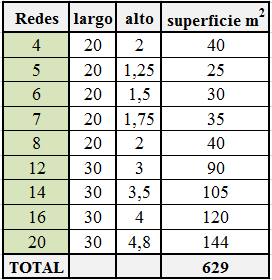


Tabla 2: Medidas de longitud de redes en metros.

xxxx

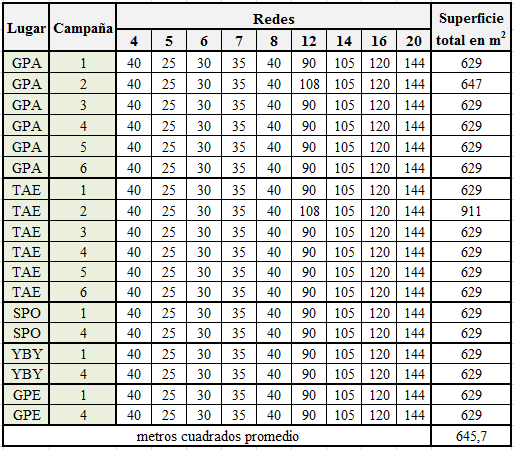
****

Tabla 3: Superficie de redes empleadas, ordenado por sitio y campaña.

A fin de obtener una lista más precisa de la composición específica del área de estudio, se realizan muestreos complementarios con otras artes de pesca como ser: tarrafas, red de arrastre, copos y líneas de mano (reel) en los sitios monitoreados.

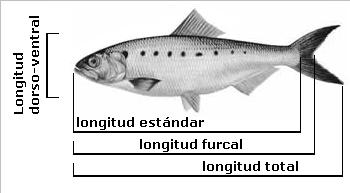
Los datos obtenidos con las redes de espera son utilizados para calcular los demás índices biológicos (diversidad, abundancia y biomasa relativa a CPUE).

En cada estación se instala el conjunto de redes y las mismas son recorridas cada 8hs.

El registro de las colectas se efectiviza en una planillaque se completa en forma manual, con los siguientes datos:

* Numero de campaña.
* Lugar.
* Fecha y Hora de recorrida de las redes.
* Tamaño de malla de la red.
* Nombre científico delejemplar colectado y relacionado a cada uno:
  + Largo estándar, en (mm)[[21]](#footnote-22).
  + Largo Furca, en (mm).
  + Largo total: en (mm).

En la Figura 8 se puede observar cómo se miden las diferentes longitudes del ejemplar.



xxx

Figura 14: Principales medidas biométricas.

* + Peso total: en (g)[[22]](#footnote-23).
  + Altura corporal: en (mm).
  + Sexo: identificación de macho (♂) / hembra (♀).
  + Estadio gonadal:
    - I = virginal,
    - II: reposo,
    - III: inicio maduración,
    - IV: maduración,
    - V: desove,
    - VI: post-desove.
  + Peso Gónadas: en (g).
  + Contenido Estomacal: se indica el volumen según cantidad de alimento ingerido:
    - V: vacío,
    - SV: semi vacío,
    - SL: semi lleno,
    - L: lleno,
  + Contenido Intestinal: se indica el volumen según cantidad de alimento ingerido:
    - V: vacío,
    - SV: semi vacío,
    - SL: semi lleno,
    - L: lleno.
  + Grasa: grado de cobertura en el ejemplar:
    - 0: sin grasa.
    - 1: presencia de grasa.
    - 2: poca grasa.
    - 3: mucha grasa (cubierto totalmente de grasa).
  + Alguna observación extra.

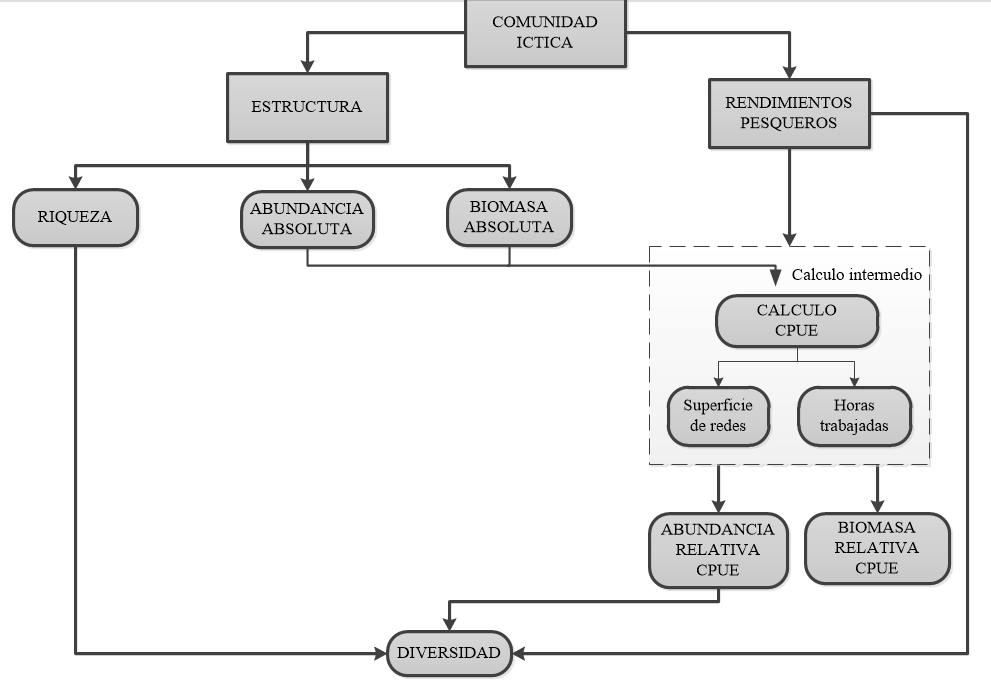
A partir de los datos obtenidos se realizan distintos cálculos que se pueden resumir en la :

Figura 15: Ruta de análisis para la obtención de los índices de estructura y rendimientos pesqueros.

Cada cálculo realizado forma parte de un análisis que luego concluye en los distintos informes que el proyecto debe presentar.

## El problema

Actualmente,para obtener el valor de los índices respecto a la estructura de la comunidad y rendimientos pesqueros se utiliza una planilla de cálculo. El experto en principio debe recopilar la información necesaria, ordenar los datos y posteriormente calcular los índices para, recién en este momento,poder analizar la información.Una vez concluidos estos pasos, el experto se encuentra en condiciones xx confeccionar los informes que demandan los convenios establecidos.

Partiendo de esta situación queda claro que si desde la dirección del proyecto se solicita información precisa o una comparativa histórica paratomar decisiones sobre el recurso íctico, no será posible brindarla de inmediato, debido al tiempo que demora la elaboraciónde los mencionados análisis.

## Objetivo

Como se indica en la sección 1.2, el objetivo principal de este trabajo de tesis consiste en la implementación de una solución de inteligencia de negocios que pueda colaborar en el proceso de análisis,minimizando la ocurrencia de errores cometidos en los cálculos realizados manualmente. Y por otra parte, se pretende brindar la posibilidad de explotar la información mediante tableros de control y análisis OLAP facilitando xx la toma decisiones.

**Capítulo 4**

*Solución propuesta*

# Solución Propuesta

## Materiales y Métodos

La elección de la metodología a utilizar en el presente trabajo, se basa en el estudio y comparación entre el Ciclo de vida de Kimball (sección 2.9), el enfoque de Inmon (sección 2.9) y la metodología Hefesto (sección 2.9).

El enfoque de Inmon propone el desarrollo de un DW que se basa en una arquitectura descendente, top-down (sección 2.9.2), lo que hace que este enfoque no se adecue a las necesidades planteadas, ya que el objetivo del presente trabajo consiste en brindar una solución al usuario final en corto tiempo.

Del análisis de las metodologías Kimball y Hefesto, se observaron características comunes, tales como: el enfoque general de análisis sobre el negocio y en base a ello, la posibilidad de desarrollar y entregar resultadossignificativos en un corto tiempo.

Teniendo en cuenta el objetivo del trabajo, y considerando la facilidad de interpretación que aportan xx herramientas particulares de cada metodología, se considera apropiada la utilización de ambas para el desarrollo del DM.

Ambas inician identificando y analizando los requerimientos. Una vez establecidos los procesos del negocio, Kimball propone una matriz denominada arquitectura de bus que permite exponer y diseñar los DM en un alto nivel. Se utilizará esta herramienta para presentar los procesos de la organización.También se utilizará parte del ciclo de vida de Kimball para exponer la prioridad de cada proceso y en base a ello, seleccionar los más relevantes para el desarrollo.

Para las fases siguientes, relacionadas a la elección de la arquitectura, el modelado dimensional (lógico) y el tipo de implementación (física), se utilizará la metodología Hefesto.

## Herramientas seleccionadas para el desarrollo

Para el desarrollo e implementación de la solución de Inteligencia de negocios se utilizarán herramientas de software libre.

Para determinar qué software utilizar, se realizó una investigación sobre las herramientas existentes en el mercado actual. El grupo Gartner realiza anualmente una investigación sobre los proveedores de tecnología de inteligencia de negocios, tanto aquellos con licencias comerciales como las de software libre. El análisis realizado por este grupo es el más destacado del mercado internacional de BI, ya que ofrece un examen independiente de la estrategia y del estado de madurez que presenta una organización; y por sobre todo, el análisis es independiente de los principales proveedores tecnológicos[18].

Gartner pone énfasis en que centrar la atención en el cuadrante de líderes no es siempre el mejor curso de acción. Todo depende de cómo el proveedor esté alineado con los objetivos del negocio. Concluye en que la elección de una herramienta de BI debe basarse en las necesidades y características de la organización.

Después de analizar cada componente de las distintas suite de Inteligencia de Negocios[19], se llegó a la conclusión de que *Pentaho BI Suite* se adecua al proyecto a desarrollar.

Partiendo de que el objetivo del presente trabajo es proporcionar una solución de inteligencia de negocios se analizan las cualidades de Pentaho, entre las que se encuentran:

* robustezde cada herramienta que forma parte de la suite,
* larga trayectoria en el mercado,
* amplio soporte para la versión *community*,
* herramientas de software libre y open source,
* y la facilidad de uso de todos sus componentes.

Fueron estas características las que impulsaron la elección de la herramienta “PentahoCommunity”.

### Pentaho

Para la implementación de la solución de Inteligencia de Negocios, se utilizará la plataforma Pentaho en su versión community que se encuentra bajo licencia GNU GPL[20]. La versión utilizada corresponde a la *5.4.*

Pentaho es una plataforma open source formada por un conjunto de herramientas que dan soporte a los componentes de Inteligencia de Negocios.Ofrece soluciones para la gestión, análisis y explotación de información. Es una de las plataformas open source más utilizadas y con mayor crecimiento en los últimos años. Catalogada en el año 2016 dentro del cuadrante de herramientas “Visionarias” por Gartner[18].

Algunas de las empresas líderes en el mercado internacional que utilizan Pentaho se pueden observar en laFigura 16, y las que lo utilizan en América latina se observan en la Figura 17. Las imágenes fueron tomadas de la sección de “Customers” de la página principal de Pentaho[21].

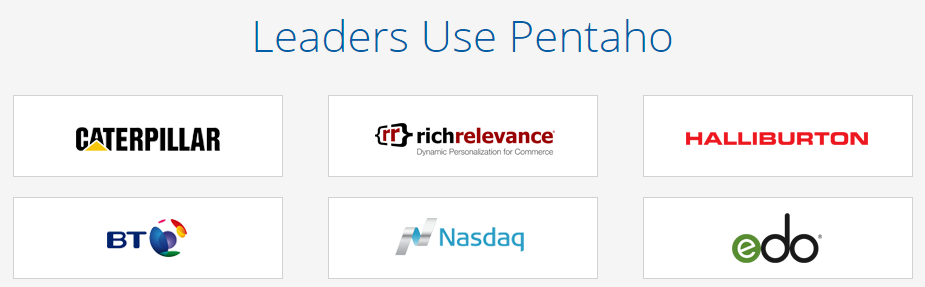
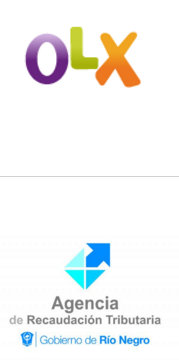


Figura 16: Empresas de América Latina que utilizan Pentaho.

Figura 17: Empresas internacionales que utilizan Pentaho.

La suite integra las siguientes aplicaciones:

* **Pentaho Data Integration (Spoon)**: ofrece potentes funciones para el proceso de extracción, transformación y carga (ETL). Esta aplicación también se utiliza para diseñar visualmente las transformaciones y trabajos encargados de extraer los datos y facilitar la presentación de los mismos para informes y análisis,
* **ReportDesigner**: es una herramienta gráfica que genera informes de los datos a través del motor de Data Integration, sin la necesidad de ninguna fase intermedia. Puede emitir sus informes en varios formatos, incluyendo PDF, Excel, HTML, texto enriquecido, archivo XML y CSV,
* **Mondrian**, ahora rebautizado como **PentahoAnalysisServices**, es una de las aplicaciones más importantes de la plataforma Pentaho BI. Es un servidor OLAP open source que gestiona la comunicación entre una aplicación OLAP (escrita en Java) y la base de datos con los datos fuente. Es decir, que Mondrian actúa como “JDBC para OLAP”[22],
* **SchemaWorkbench**:es una interfaz que permite crear y probar esquemas de los cubos OLAP. A través de esta herramienta se pueden presentar los datos multi-dimensionalmente y permitir a los usuarios seleccionar qué dimensiones y medidas explorar,
* **SaikuAnalytics**:Saiku es una herramienta OLAP destinada a usuarios finales de Pentaho, que permite visualizar y realizar análisis de datos de forma fácil e intuitiva. Reemplazóa Jpivot, que es una herramienta más compleja de entender para los usuarios.Una de las mejorasrespecto de Jpivot comprende la interfaz que se visualiza mediante una página web, mediante la cual sepueden construir vistas personalizadas de datos a través de drag and drop(arrastrando y soltando campos),
* **CDE** (CommunityDashboard Editor): la Comunidad Dashboard Editor es el resultado de las necesidades del mundo real: nació para simplificar la creación, edición y procesamiento de dashboard. CDE y las tecnologías que incluye (CDF, CDA y CCC) permite desarrollar e implementar dashboards en la plataforma Pentaho en forma rápida y efectiva.Es una herramienta muy potente y completa, combinando front-endcon fuentes de datos y los componentes personalizados de una manera fluida,
  + **CDA (Data Access)**: es un plugin diseñado para acceder a las fuentes de datos con gran flexibilidad,
  + **CDF (Dashboard Framework)**: es un marco de código abierto que permite la creación de dashboardsaltamente personalizable. Está basado en estándares de desarrollo web como CSS, HTML5 y JavaScript (aprovechando frameworks como jQuery o Bootstrap). Es una solución eficaz para combinar los datos con una atractiva capa de visualización,
  + **CCC (Chart Components):**plugin de la comunidad componentes Chart, incluye una biblioteca de gráficos que está construido según un top de Protovis.Es muy potente, gratuito y de código abierto. El objetivo de CCC es proporcionar a los desarrolladores los tipos de gráficos básicos, sin perder el principio fundamental: la extensibilidad,
  + **CGG (GraphicsGenerator):**es un plugin que permite al usuario exportar gráficos como imágenes.Permiteademás, la inclusión de gráficos en el interior de CDE.

### Motor de Bases de Datos

Pentaho se pueden integrar tanto con motores de BD open source como aquellos con licencia comercial. En el presente trabajo se utilizará el motor de base de datosPostgreSQL–open source– en su versión 9.2.

Tambien se utiliza **“pgAdmin III”,** versión 1.14.3**,** que es una aplicación gráfica que facilita la gestión y administración de bases de datos PostgreSQL.**pgAdminIII** es una de lasaplicaciones más completas y populares con licencia Open Source. Se encuentra disponible en varios lenguajes y con soporte para varios sistemas operativos, incluyendo Microsoft Windows, Linux, FreeBSD, Mac OSX y Solaris.

## Pasos y aplicación de las metodologías

A continuación se detallan los pasos aplicados de las metodologías elegidas (sección 4.1) para lograr la solución de inteligencia de negocios.

### **Paso 1: Análisis de requerimientos**

Para identificar las necesidades de información y obtener las preguntas claves del negocio se realizaron entrevistas con el director y el responsable de la confección de informes del proyecto de Biología Pesquera Regional.

Lainformación esencial que se desea obtener consiste en:

Determinar la riqueza, abundancia y biomasa absoluta, abundancia y biomasa relativa según CPUE[[23]](#footnote-24) para:

* un convenio específico,
* un conjunto específico de campañas o en una particular,
* todos los sitios o alguno en particular,
* el conjunto de redes utilizadas o una en particular,
* un tiempo y horario determinado,
* y para una o varias especies de interés.

Como se menciona en la sección 3.3, se realizan campañas exploratorias de pesca donde se obtienen y registran datos sobre cada ejemplar colectado.

En este contexto, los procesos requieren obtener:

Estructura de la comunidad íctica.Incluye los siguientes indicadores:

* + **riqueza o composición específica**: indica la cantidad de especies colectadas,
  + **abundancia absoluta**: indica la cantidad de ejemplares colectados,
  + **biomasa absoluta**: indica el peso total de los ejemplares colectados.

Rendimientos pesqueros.Incluye los siguientes indicadores:

* + **captura por unidad de esfuerzo**:es un índice que expresa la captura (cantidad) obtenida a través de un determinado esfuerzo, con un tipo de arte de pesca que posee una determinada superficie y un tiempo de trabajo.

La fórmula que expresa el índice se define de la siguiente manera:

CPUE= cantidad /100 m2/8 hs.

Donde,

* + cantidad = número de ejemplares o peso total de una especie,
  + 100 m2= superficie de la red utilizada en m2,
  + 8 hs= horas de trabajo.
    - **abundancia relativa:**se obtiene a partir de la CPUE y expresa la cantidad de ejemplares que se obtuvieron a través de un determinado esfuerzo (número /100 m2 / 8 hs),
    - **biomasa relativa:** se obtiene a partir de la CPUE yexpresa la cantidad (en kilogramos)de ejemplares que se obtuvieron a través de un determinado esfuerzo (peso total /100 m2/ 8 hs),
    - **diversidad específica**: este índice mide la variabilidad específica en una comunidad (íctica) en un espacio y tiempo determinado.

El índice utilizado para calcularla es el de Shannon y Weaver[23], cuya fórmula es:

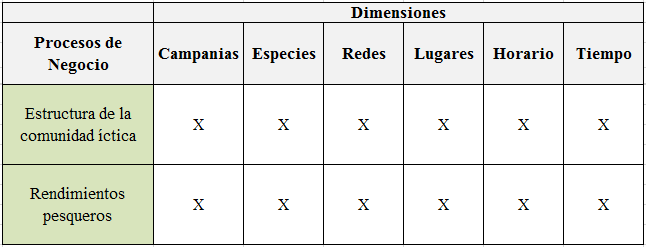


donde ‘n’ es el número de especies presente en la comunidad estudiada [riqueza] y p(i) es la abundancia relativa de la especie (i).

### Procesos del negocio

Identificados los procesos del Proyecto de Biología Pesquera Regional se construye la matriz de procesos/dimensiones que permitirá diseñar los DM de alto nivel. Cada X en la intersección de filas y columnas significa que el proceso de negocio se identifica o puede ser analizado desde dicha dimensión.

Tabla 4: Arquitectura de Bus

****

#### **Selección de procesos**.

A continuación se seleccionan los procesos de negocio a desarrollar.

Para la selección,se estableceuna prioridad a cada proceso. Dicha prioridad se basa en factores tales como,la importancia dentro del proyecto, calidad y disponibilidad de los datos.

A cada factor se le asigna unavaloración, entre las que se puede optar por baja, media o alta. La opción “alta” implica un proceso prioritario que debe ser seleccionado para el desarrollo.

La asignación de prioridades a los distintos procesos se pueden observar en laTabla 5.

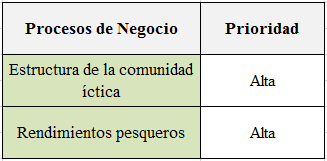


Tabla 5: Prioridad de procesos

A fin de establecer una correspondencia con la terminología utilizada en Hefesto, se aclara que las dimensiones son denominadas‘perspectivas’, las medidas a analizar dentro de cada proceso se denominan ‘indicadores’ y el proceso de negocio en sí mismo, representa la relación entre perspectivas e indicadores.

Esta correspondencia se puede apreciar mediante el modelo conceptual en la Figura18.

### Modelo Conceptual

El modelo conceptual es una representacion gráfica que permite comprender de manera rápida e intuitiva las diferentes perspectivas desde las que puede ser analizado un proceso y cuáles son las medidas o indicadores que involucra cada uno. (Figura 18).

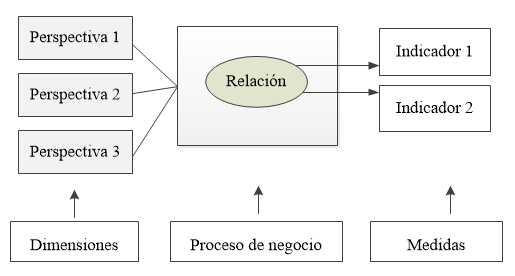


Figura 18: Modelo conceptual

A continuación se presenta el modelo conceptual para cada proceso que será desarrollado (Figura 19).

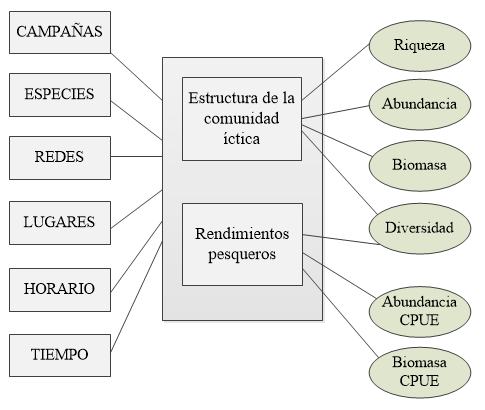


Figura 19: Modelo conceptual - Biología Pesquera Regional

## **Paso 2: Análisis del OLTP**

En esta secciónse analiza la fuente OLTP para determinar el cálculo de los indicadores. Por otra parte se establecen las correspondencias entre el modelo conceptual anteriormente creado y la fuente de datos. Luego se definen los campos que formarán parte de cada perspectiva.Y finalmente se construirá el modelo conceptual ampliado con la información obtenida de los pasos anteriores.

#### Conformación de Indicadores

A continuación se explica cómo se calculan los indicadores, definiendo para ello los hechos que lo componen y la función de sumarización.

1. **Indicador**: **Riqueza**

**Hechos**: cantidad de especies

**Función de sumarización**: COUNT - DISTINCT

**Aclaración**: representa la cantidad de las distintas especies capturadas.

1. **Indicador**: **Abundancia absoluta**

**Hechos**: cantidad de ejemplares por especie.

**Función de sumarización**: COUNT

**Aclaración**: representa la cantidad de ejemplares capturados de cada especie.

1. **Indicador**: **Biomasa absoluta**

**Hechos**: peso total por especie.

**Función de sumarización**: SUM

**Aclaración**: representa el peso total de los ejemplares capturados por especies.

1. **Indicador**: **Abundancia relativa a la CPUE**

**Hechos**: cantidad de ejemplares colectados por especie según la captura por unidad de esfuerzo establecida.

**Función de sumarización**: COUNT

**Aclaración**: representa la cantidad de ejemplares colectados por especie en base al parámetro captura por unidad de esfuerzo.

1. **Indicador**: **Biomasa relativa a la CPUE**

**Hechos**: peso total por especie según la captura por unidad de esfuerzo establecida.

**Función de sumarización**: SUM

**Aclaración**: representa la sumatoria del peso total de cada ejemplar capturado por especie según el parámetro captura por unidad de esfuerzo.

1. **Indicador: Diversidad**

**Hechos:** variabilidad de una comunidad por especies.

**Función de sumarización**: SUM

**Aclaración**: representa la variabilidad de una comunidad por especies colectada.

**Establecer correspondencias**

El objetivo de este paso es analizar el OLTP disponible y establecer las correspondencias entre el modelo conceptualy la fuente de datos.

Para ello se presenta el modelo de datos () utilizado por el Proyecto de Biología Pesquera y luego se asientan las relaciones con el modelo conceptual.

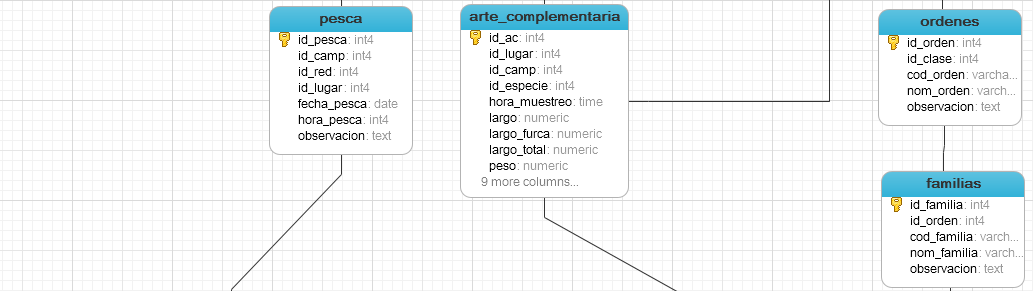
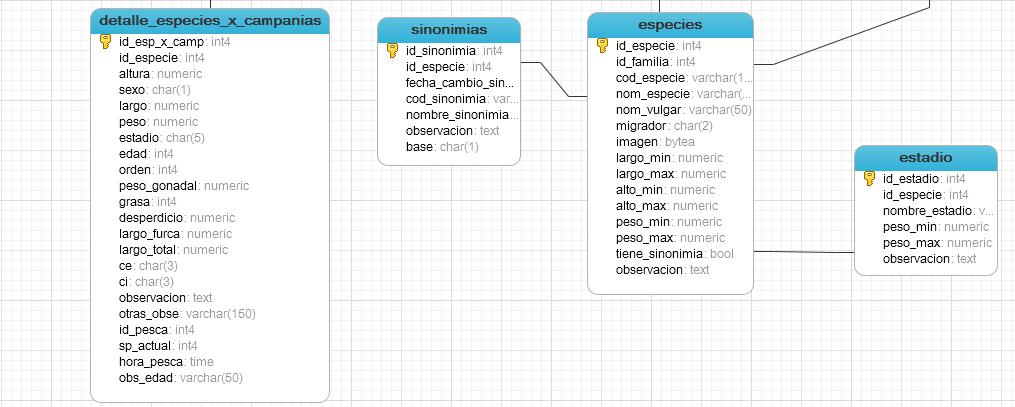
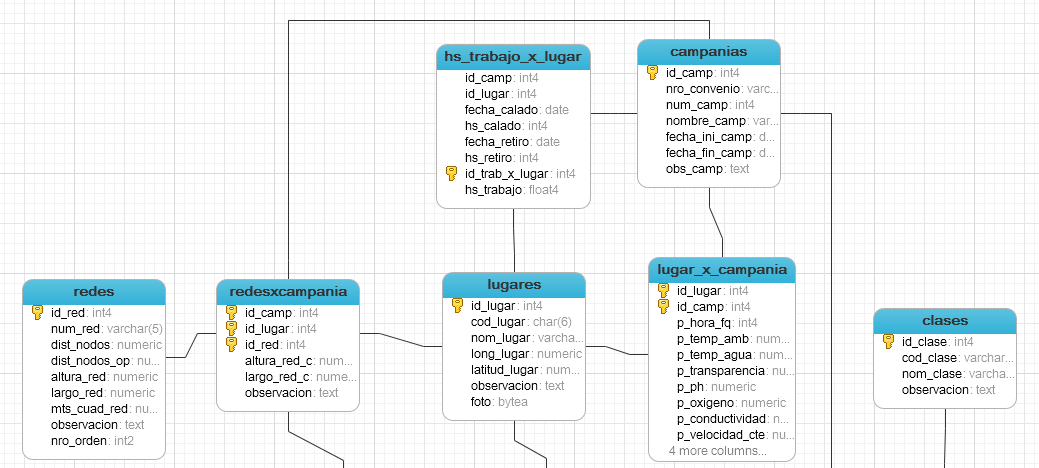


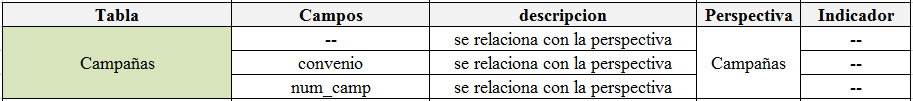
Figura 20: Modelo de datos de la organización.

Correspondencias

Partiendo de cada perspectiva, se establece la correspondencia con la/s tabla/s que involucra del OLTP.

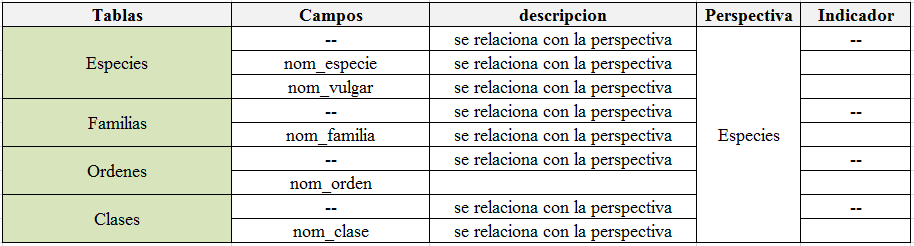
Perspectiva: Campaña

Tabla 6: Correspondencia - perspectiva Campaña



Perspectiva: Especies

Tabla 7: Correspondencia - perspectiva Especies



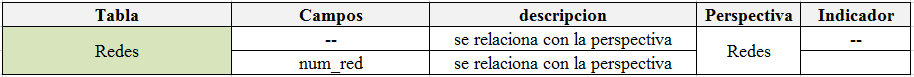
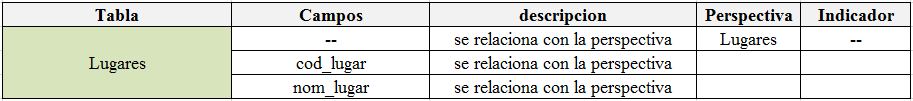
Perspectiva: Redes

Tabla 8: Correspondencia - perspectiva Redes

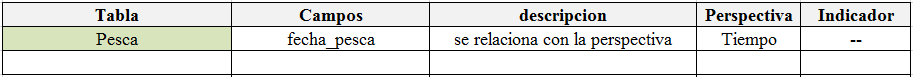
Perspectiva: Lugares

Tabla 9: Correspondencia - perspectiva Lugares



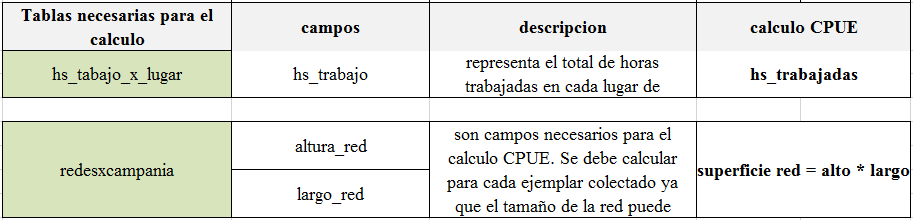
Perspectiva Tiempo:

Tabla 10: Correspondencia - perspectiva Tiempo



Calculo intermedio CPUE:

Tabla 11: Cálculo intermedio CPUE



Correspondencia con Indicadores:

Tabla 12: Correspondencia con Indicadores



### Nivel de granularidad

Una vez establecidas las relaciones con el OLTP se deben seleccionar los campos que contendrá cada perspectiva. Para comprender el significado de los campos residentes en cada tabla, se examina la base de datos y se consulta el diccionario de datos encontrado en el Anexo 1.

A partir de consultas con los usuarios se logra determinar cuáles son los datos que consideran de interés para analizar los indicadores antes expuestos. Los resultados fueron los siguientes:

**Perspectiva Campaña**: los campos que integran la perspectiva son:

* “Id\_camp” de la tabla Campanias, es el identificador del oltp, utilizado para referenciar y ordenar las campañas. No forma parte de la jerarquía.
* “convenio” de la tabla Campanias, representa el nombre y número del convenio en el que se llevan a cabo las campañas.
* “num\_camp” de la tabla Campanias, representa el número de la campaña.

**Perspectiva Especies**: los campos que integraran la perspectiva son:

* “nom\_especie” de la tabla Especies, representa el nombre científico de la especie.
* “nom\_vulgar” de la tabla Especies, representa el nombre vulgar de la especie.
* “nom\_familia” de la tabla Familias, representa el nombre de la familia a la que pertenece la especie.
* “nom\_orden” de la tabla Órdenes, representa el nombre del orden al que pertenece la especie.
* “nom\_clase” de la tabla Clases, representa el nombre de la clase a la que pertenece la especie.
* “interés”: es un campo incorporado durante el proceso para definir algún tipo interés que se tiene sobre la especie.Dado que una especie puede pertenecer a los grupos de interés pesquero, económico o deportivo fue necesario crear un campo para cada tipo de interés.

**Perspectiva Redes**: los campos que integrarán la perspectiva son:

* “id\_red” de la tabla Redes, representa el identificador del oltp utilizado para referenciar y ordenar las redes.No forma parte de la jerarquía.
* “num\_red” de la tabla Redes, representa el numero con el que se identifica una red.

**Perspectiva Lugares**:

* “id\_lugar” de la tabla Lugares, representa el identificador del oltp utilizado para referenciar y ordenar las tuplasde la tabla. No forma parte de la jerarquía.
* “nom\_lugar” de la tabla Lugares, representa el nombre completo de cada lugar de muestreo.
* “cod\_lugar” de la tabla Lugares, representa mediante un código alfanumérico a un lugar de muestreo.
* “nom\_pcia”, es un campo incorporado durante el proceso que asocia un lugar a la provincia a la que pertenece.
* “nom\_pais”, es un campo incorporado durante el proceso que asocia un lugar con al país al que pertenece.

**Perspectiva Horario**: los campos que integraran la perspectiva son:

* “id\_hora”: representa el identificador de la tabla y también el valor de horas en formato hh:mm.
* “franja\_horaria”: representa a través de una etiqueta un agrupamiento de horas. Por ejemplo: de 00:01 a 06: 00hs se representa como ‘madrugada’. De 06:01a 12:00 hs como ‘Mañana’. De 12:01 a 18:00 como ‘Tarde’ y de 18:01 a 00:00hscomo ‘Noche’.
* “franja\_día\_noche”: representa franja del día en dos grupos: día y noche.

**Perspectiva Tiempo**: los campos que integraran la perspectiva son:

* “id\_tiempo”, es un campo de tipo date que además de representar el identificador de la tabla, representa la fecha en formato “aaaa-mm-dd”.
* “año”, es un valor numérico que representa el año.
* “mes”, es un valor numérico que representa el mes.
* “día”, es un valor numérico que representa el día.
* “fecha\_formateada”: es un campo de tipo *date* que representa una fecha en formato “dd-mm-aaaa”.
* Estación\_anio: es un campo que representa a las estaciones del año.

### Modelo conceptual ampliado

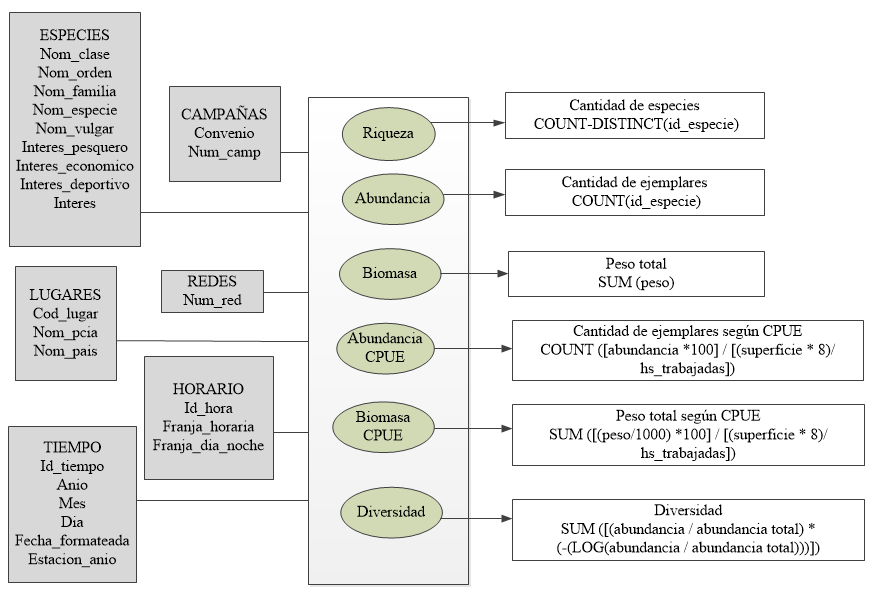
Con el fin de graficar los resultados obtenidos de los pasos anteriores, se amplía el modelo conceptual, colocando bajo cada perspectiva los campos seleccionados por los usuarios y bajo cada indicador su respectiva fórmula de cálculo.

Figura 21: Modelo conceptual ampliado - Proyecto Biología Pesquera Regional

## PASO 3: Modelo lógico del DW

A continuación se confecciona el modelo lógico de la estructura del DW teniendo como base el modelo conceptual anteriormente presentado.

En primer lugar, se define el tipo de modelo a utilizar y a partir de esto se diseñan las tablas de dimensiones y de hecho. Finalmente, se plasman las uniones pertinentes entre estas tablas.

### Tipo de modelo lógico del DW

El tipo de modelo lógico a utilizar esel representado mediante el esquema estrella (sección 2.6.1.1).

Para el caso de estudio se desarrolla el DM unificando los procesos definidos, ya que éstos comparten tanto perspectivas como indicadores.

Tabla de dimensiones y hechos: en este paso se diseñan las tablas de dimensiones y de hechos que formarán parte del DM a desarrollar.

En la Figura 22se puede observar el DM diseñado. La tabla de hechos, denominada ‘*fact\_table*’ es de tipo transaccional,es decir, que cada registro representa una operación que permite el análisis y navegación de datos a un máximo nivel de detalle.

Las tablas de dimensiones se encuentran desnormalizadas y cada una de ellas se designan con el prefijo ‘dim\_’.

La tabla de hechos se encuentra relacionada a las dimensiones a través de claves subrogadas[[24]](#footnote-25), exceptuando a las tablas‘dim\_tiempo’ y ‘dim\_horario’ que cuentan con claves de tipo *date* y *time* respectivamente. Kimball[24] recomienda separar las dimensiones fecha y horario.

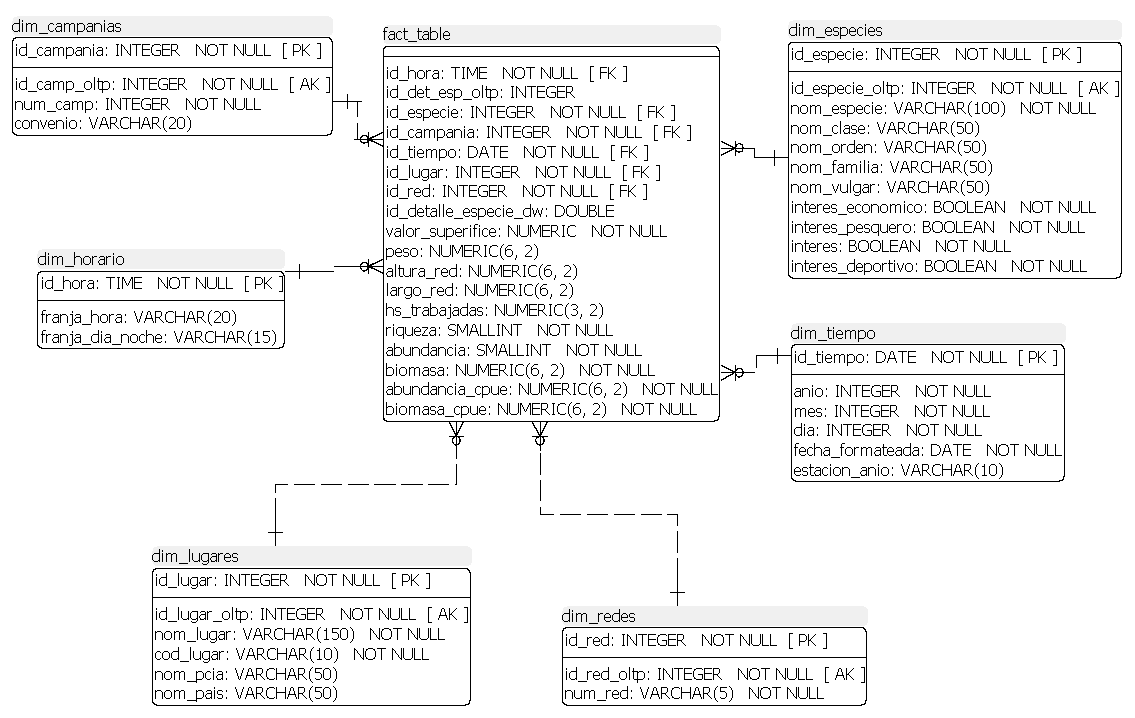
La elección de este tipo de datos se basa en larapidez con los que se obtienen los resultados ya que para la búsqueda compara directamente con el identificador de la tabla. Por otra parte, y basado en el motor de base de datos sobre el que se implementa la solución, PostgreSQL, se utiliza un ejemplo expuesto por la misma empresa para la carga de ambas dimensiones[25].

Figura 22: Modelo lógico DM - Proyecto de Biología Pesquera Regional

### Integración de Datos

Definido el modelo lógico, es necesario poblarlo de datos.Para realizar esta tarea, se utilizan técnicas de limpieza, codificación y unificación de datos a través de procesos ETL.El detalle de cada uno de estos procesos se pueden observar en el Anexo 2.

### Cubos Multidimensionales

Las consultas a través de varias dimensiones son posibles de realizar mediante la definición de cubos multidimensionales que representan al DM. Para el desarrollo del cubo multidimensional se utiliza el tipo de implementación ROLAP, definido en la sección 2.1.1.6.

En la Figura 23se puede visualizar la estructura del cubo definido para rendimientos pesqueros. En el Anexo 3 se encuentra el contenido del archivo XML[[25]](#footnote-26) generado por la herramienta SchemaWorkbench.

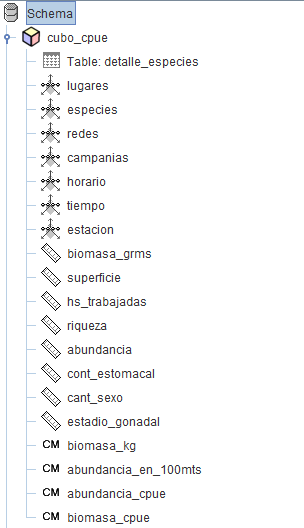


Figura 23: Cubo rendimientos pesqueros – Medidas.

****

Figura 24: Cubo rendimientos pesqueros - Dimensiones (Parte 1)

Figura 25: Cubo rendimientos pesqueros - Dimensiones (Parte 2)

### Tableros de control

Como se menciona en la sección 2.8, un tablero de control es una herramienta de gran utilidad para la gestión y dirección de cualquier tipo de organización.

En el presente trabajo se definirán tableros de control para brindar información resumida y precisa a los responsables del análisis y gestión del proyecto.

A continuación se detallan los tableros a desarrollar, y mediante un boceto se exponen tanto el diseño como los indicadores a incluir.

**Tablero de abundancia absoluta**

A continuación se expone el boceto del tablero de control a desarrollar, en la para el indicador abundancia absoluta.

El tablero permitirá conocer a través de un convenio seleccionado, las campañas que se han llevado a cabo. En el gráfico de torta se aprecia una proporción de la abundancia absoluta por campaña. Al seleccionar una campaña en particular, se actualiza el grafico de barra, indicando de la proporción mencionada, cómo se distribuye en los lugares que fueron monitoreados.

Al seleccionar un lugar, se actualiza el grafico de barra horizontal, describiendo las especies que fueron allí capturadas. Finalmente seleccionado una especie, se puede observar a través de un gráfico de líneas, la evolución que tuvo la especie en relación a la abundancia absoluta durante los últimos 5 años.

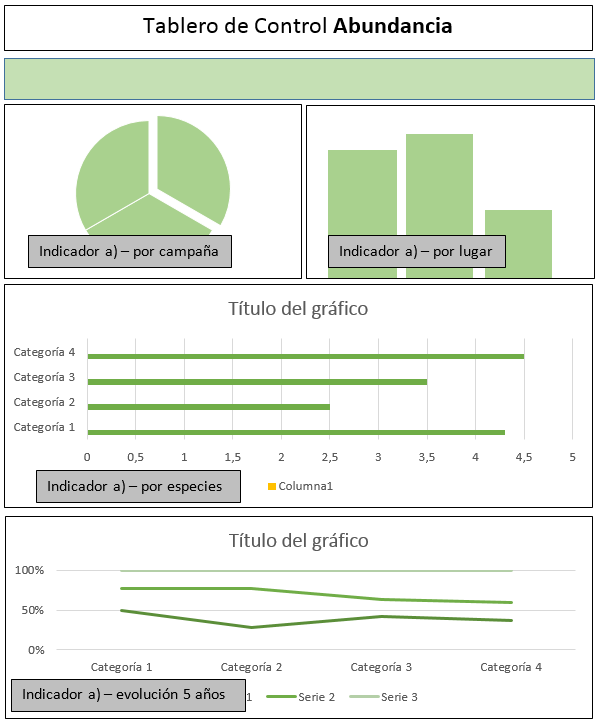


Figura 26: Boceto tablero de control abundancia absoluta.

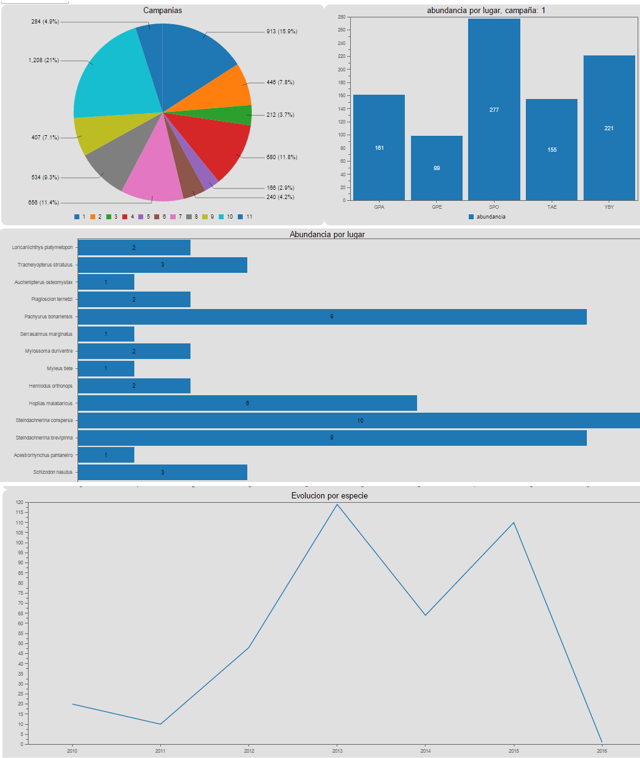
A continuación se muestra una imagen del tablero desarrollado en la *Figura 27*:

Figura 27: Tablero de control Abundancia absoluta.

**Tablero de rendimientos pesqueros**

El siguiente tablero permite conocer la abundancia relativa a la captura por unidad de esfuerzo de los últimos 5 años, determinando de mayor a menor, los lugares con mayor abundancia y las especies más frecuentes.

Permite además conocer la estación del año que en promedio al tiempo mencionado tuvo mayor abundancia relativa. De la misma manera permite conocer el momento del día en el que se obtienen mayor cantidad de ejemplares según la captura por unidad de esfuerzo establecida, en 100 m2 cada 8 hs.

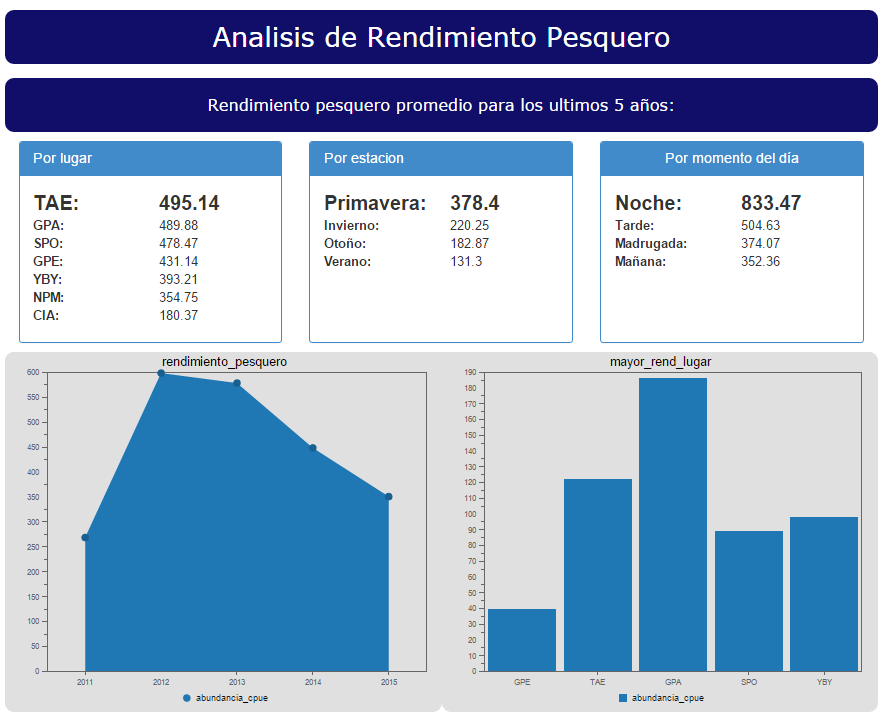


Figura 28: Tablero de control rendimientos pesqueros.

**Capítulo 5**

*Prueba Experimental*

# Prueba Experimental

## Análisis OLAP

A continuación se describen ejemplos de consultas realizadas al cubo de rendimientos pesqueros con el fin de demostrar la utilización de la herramienta y las posibilidades de análisis multidimensional. Para la navegación de los cubos se utiliza la herramienta Saiku de la suite de Pentaho (sección 4) que permite realizar las consultas mediante acciones ‘*drag and drop*’[[26]](#footnote-27).

Las pruebas se realizan junto al usuario, que tras cada resultado expuesto por la herramienta corrobora si son correctos.

En principio se define el indicador al que se va a acceder y las dimensiones por las que se puede filtrar y visualizar la información. Luego se describe el requisito definido por el usuario, la consulta realizada y los resultados obtenidos por la mencionada herramienta.

Dado que los análisis que se solicitan por medio de los convenios siempre requieren información respecto del tiempo y espacio, todas las consultas involucran dos o más dimensiones.

**Riqueza**: este indicador podrá ser analizado a través de las siguientes dimensiones:

* + - * Tiempo
      * Horario
      * Campaña
      * Lugar
      * Red
      * Especies

Prueba 1: determinar la riqueza obtenida en la campaña 1 del año 2015 para todos los lugares de muestreo.

En esta consulta se utiliza la dimensión tiempo con el mayor nivel de jerarquía, año. Por otra parte, para seleccionar el número de campaña (nivel 2 de la jerarquía) es necesario indicar el convenio al que pertenece y luego seleccionar el número de campaña deseado. Finalmente se incorpora la dimensión lugar.

La consulta realizada se puede observar en la Figura 24 y el resultado de la misma en forma tabular en la Figura 25. Por otra parte, se genera un gráfico con la herramienta ‘*Charts Plus*’ en la que se define una proporción de la riqueza respecto de la campaña 1, año 2015 para cada lugar. Se puede observar el gráfico en la Figura 26.

**Observación de la prueba**: ninguna observación, los resultados son correctos.



Figura : Consulta de prueba 1

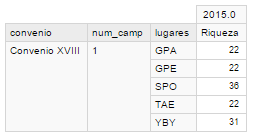


Figura : Resultado de la prueba 1.

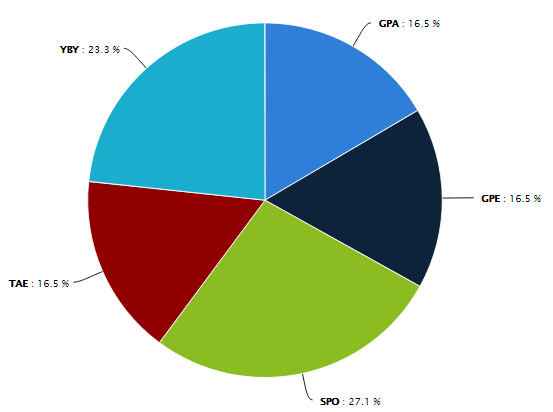


Figura : Gráfico de la prueba 1.

Prueba 2: determinar la riqueza obtenida en la campaña 1 del año 2015, lugar TAE con las redes número 4, 5 y 6.

En esta consulta, al igual que en la anterior se utiliza la dimensión tiempo. Para seleccionar el número de campaña (nivel 2 de la jerarquía) se indica el convenio al que pertenece y luego el número de campaña deseado. Finalmente se incorpora la dimensión redes, de las cuales se seleccionan solamente los números de redes solicitadas.

La consulta realizada se puede observar en la Figura 27y el resultado de la misma en forma tabular en laFigura 28. El gráfico generado por la herramienta ‘*Charts Plus*’ para la mejor comprensión en la Figura 29.

**Observación de prueba 2**: ninguna observación, los resultados son correctos.

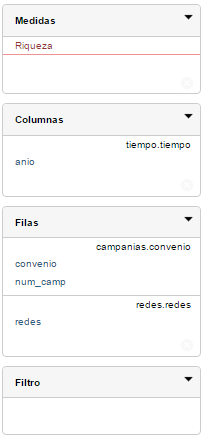


Figura : Consulta de la prueba 2.

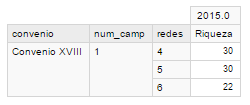


Figura : Resultado de la prueba 2

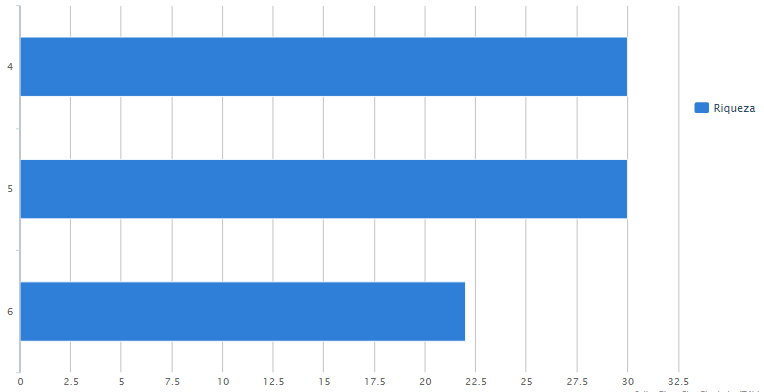


Figura : Grafico prueba 2.

Prueba 3: determinar para el convenio XVIII la franja horaria en que se obtiene mayor riqueza en comparación de todos los lugares muestreados.

En la Figura 30se puede observar la consulta realizada. En las Figuras31y 32se visualizan los resultados, de manera tabular y gráfica, respectivamente.

**Observación de prueba 2**: los resultados son correctos. Actualmente cada franja horaria se establece cada 6 horas, se solicita cambiar dichas franjas horarias a 8 hs, y de esta manera obtener las etiquetas: ‘Noche – Madrugada’, ‘Mañana’ y ‘Tarde’.

Figura : Consulta prueba 3.

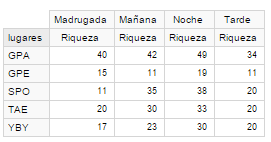


Figura : Resultado prueba 3.

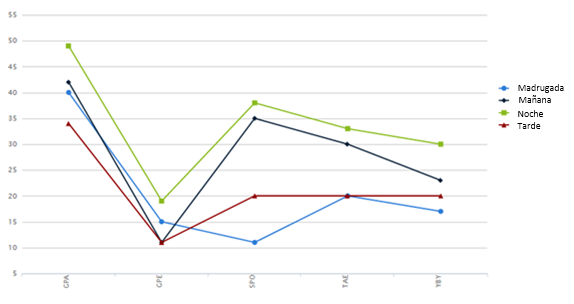


Figura : Gráfico prueba 3.

Prueba 4: abundancia por sitioen los últimos 4 años.

En la Figura 33 se puede observar la consulta realizada. En las Figuras 34 y 35 se visualizan los resultados, de manera gráfica y tabular, respectivamente.

**Observación de prueba 4:** ninguna observación, los resultados son correctos.

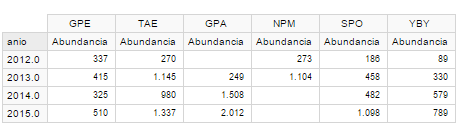


Figura : Resultado prueba 4.

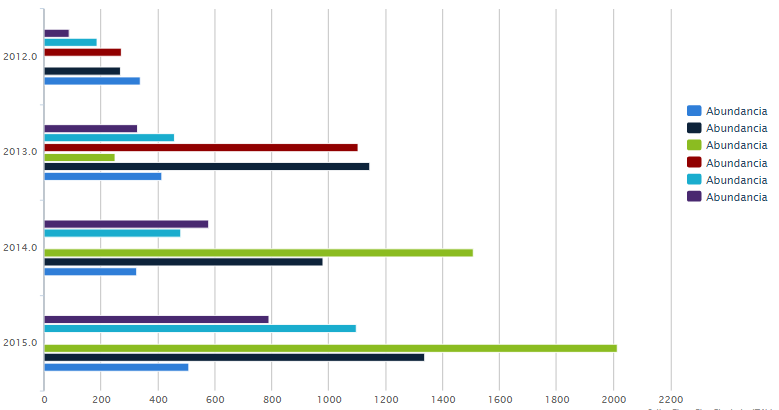


Figura : Gráfico prueba 4.

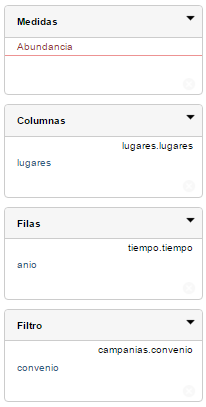


Figura : Consulta prueba 4.

Prueba 5: indicar el horario en que se obtuvo más abundancia en TAE (lugar) en el año 2015.

En la Figura 36 se puede observar la consulta realizada. En las Figuras 37 y 38 se visualizan los resultados, de manera tabular y gráfica, respectivamente.

**Observación de prueba 5:** ninguna observación, los resultados son correctos.

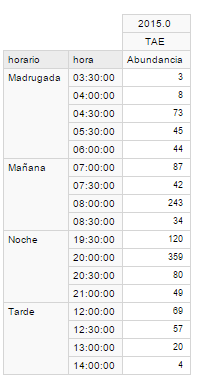


Figura : Resultado prueba 5.



Figura : Consulta prueba 5.

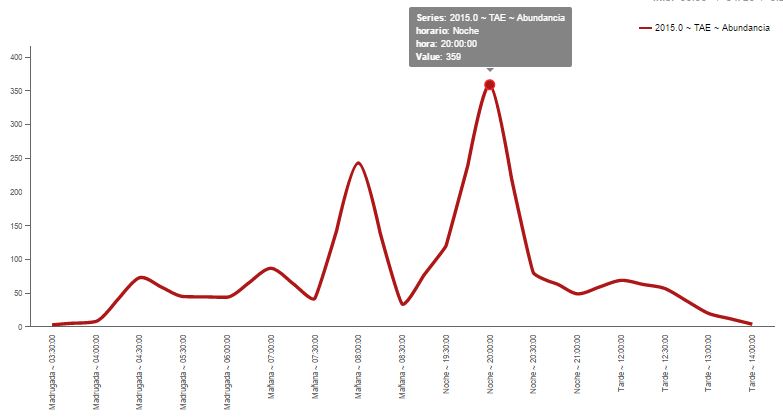


Figura : Gráfico prueba 5.

**Prueba 6**: abundancia cpue, por sitio para las primeras 6 campañas del año 2015.

**Observación de prueba 6:** ninguna, los resultados son correctos.

**Prueba 7**: biomasa cpue, por sitio para las primeras 6 campañas del año 2015.

**Observación de prueba 7:** ninguna, los resultados son correctos.

**Capítulo 6**

*Conclusiones*

# Conclusiones

## Conclusión

El presente trabajo abordó el desarrollo e implementación de una solución de inteligencia de negocios, para la evaluación y toma de decisiones sobre la fauna íctica del río Paraná aguas arriba de la represa Yacyretá.

La principal contribución consistió en la creación de un modelo multidimensional que a través de la automatización de procesos de extracción, transformación y carga permita la población de datos al *datamart* en un solo paso. El proceso ETL fue diseñado de tal manera que pueda ser utilizado por cualquier organización referida al mismo ámbito de estudio. El principal aporte al proceso de análisis, dentro del proyecto, es haber logrado la automatización de los cálculos que determinan el rendimiento pesquero de acuerdo a una captura por unidad de esfuerzo determinada.

Por otra parte, se desarrolló un cubo que a través de técnicas OLAP permite recuperar y explotar información de manera seleccionada, integrada, organizada e historiada.

Para la presentación de la información se desarrollaron tableros de control que mediante gráficos sencillos e intuitivos facilitan la comprensión del estado actual de la comunidad ictica en el área de influencia de la Central Hidroeléctrica Yacyretá.

## Trabajos Futuros

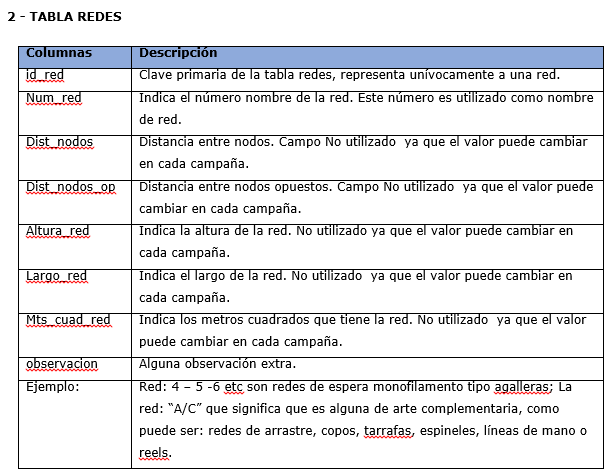
A lo largo del desarrollo del presente trabajo han surgido varias líneas interesantes para ser abordadas, entre ellas:

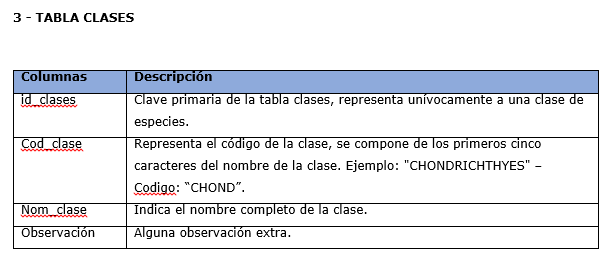
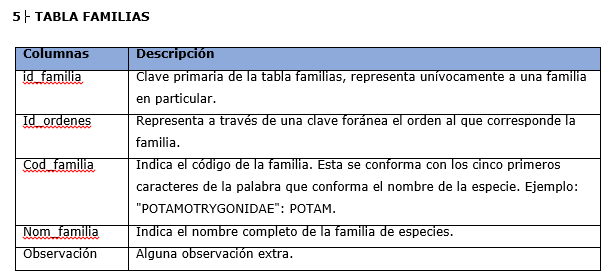
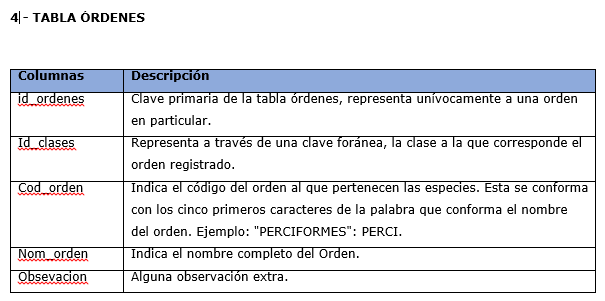
* Mejorar la recolección de datos durante los muestreos realizados con alguna herramienta software, que a través de controles, alerten a los técnicos sobre posibles registraciones erróneas. Un ejemplo de ello sería que determinadas especies no pueden pesar más de cierta cantidad de kilogramos, etcétera.

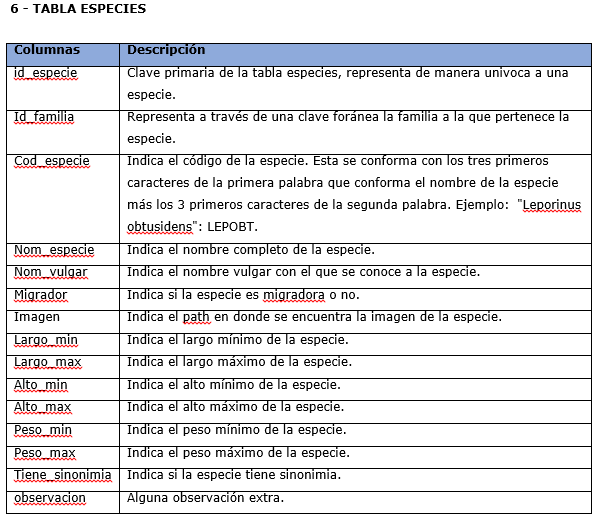
Con esto se lograría que tras cada campaña,los datos se encuentren digitalizados y no requieran luego de exhaustivas correcciones.

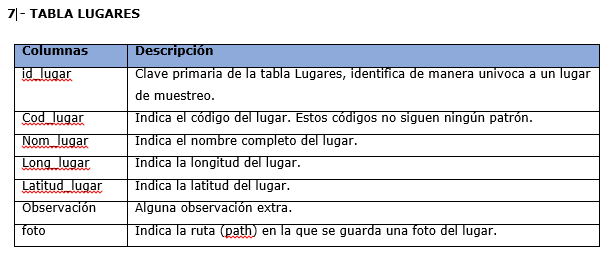
* Por otra parte, se pueden incluir a la base de datos actual y luego al DW, los valores respecto de las variables ambientales que son registradas durante cada monitoreo, a fin de poder analizarlos conjuntamente con los rendimientos pesqueros.
* De la misma manera sería muy interesante incluir al DW, información referida a Ictioplancton y de esta manera poder desarrollar tableros con toda esta información integrada.
* Como último punto, se pueden incluir datos sobre la fauna íctica aguas abajo de la represa Yacyretá y poder comparar la comunidad en toda el área de influencia de la mencionada represa.

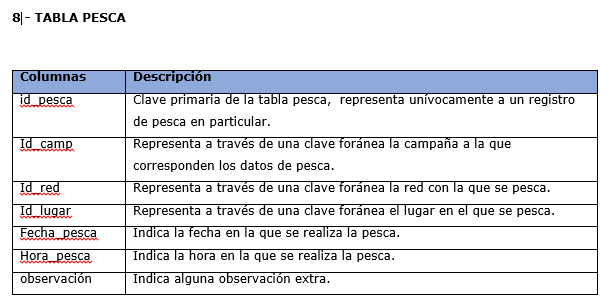
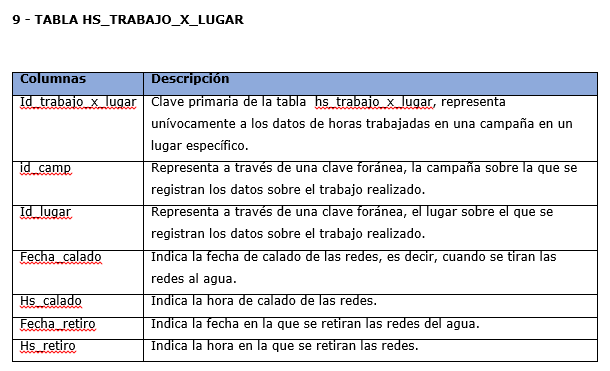
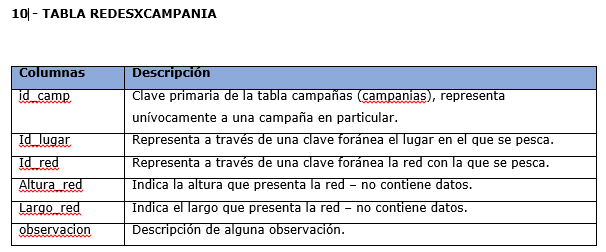
##### Anexo 1











**11 - TABLA DETALLE\_ESPECIES\_X\_CAMPANIAS**

|  |  |
| --- | --- |
| **Columnas** | **Descripción** |
| id\_esp\_x\_camp | Clave primaria de la tabla campañas (campanias), representa unívocamente a una campaña en particular. |
| Id\_especie | Representa a través de una clave foránea la especie capturada. |
| Altura | Indica la altura, en milímetros, del ejemplar. |
| Sexo | Indica el sexo si se ha determinado, o nulo para los ejemplares liberados. |
| Largo | Indica el largo del ejemplar, en milímetros. |
| Peso | Indica el peso, en gramos, del ejemplar. |
| Estadio | Indica el estadio gonadal del ejemplar para determinar su actividad reproductiva. Por ejemplo, I: virginal, II: reposo, III: inicio maduración, IV: maduración, V: desove, VI: post-desove. |
| Edad | Indica la edad del ejemplar. |
| Orden | Indica a través de una numeración secuencial el orden en el que fueron capturados los ejemplares. |
| Peso\_gonadal | Indica el peso de la gónada, en gramos. |
| Grasa | Indica el grado de cobertura de grasa en el ejemplar; por ejemplo: 0: sin grasa, 1: presencia de grasa, 2: poca grasa, 3: mucha grasa. |
| Desperdicio |  |
| Largo\_furca | Indica el largo de Furca del ejemplar. |
| Largo\_total | Indica el largo total del ejemplar. |
| Ce | Indica el contenido estomacal del ejemplar capturado; por ejemplo: vacío (V), semi vacío (SV), lleno (L), semi lleno (SL). |
| Ci | Indica el contenido intestinal del ejemplar capturado; por ejemplo: vacío (V), semi vacío (SV), lleno (L), semi lleno (SL). |
| Observación | Alguna observación extra sobre cada ejemplar en particular. |
| Otras\_obse | Alguna observación extra sobre el muestreo. |
| Id\_pesca | Representa a través de una clave foránea los datos de pesca del ejemplar, en que se detallan: campaña, red, lugar, fecha y hora de la pesca. |
| Sp\_actual | Representa a través de una clave foránea el nombre de la especie actual del ejemplar. Si el id\_especie tiene sinonimia, aquí se indicará el id\_especie de la tabla sinonimia. |
| Hora\_pesca | Indica la hora de pesca. |
| Obs\_edad | Alguna observación sobre la edad del ejemplar. |

Las demas tablas que forman parte del modelo de datos, no seran descriptas ya que no forman parte del presente estudio.

##### Anexo 2

En esta sección se describen los procesos ETL.

Para la carga del DM se diseñó un trabajo llamado “Trab\_completo\_pesca” (Figura 23) que se utiliza como ejecución principal. En principio se ejecuta un trabajo que carga todas las dimensiones; si la carga de cada transformación de dimensión se ejecuta correctamente se procede a la carga de la tabla de hechos. Si todo el proceso se ejecuta correctamente, se muestra un mensaje de Éxito y se envía un e-mail indicando que la carga fue exitosa. De lo contrario, si ocurre un error en alguna instancia de carga, primero se muestra un mensaje en pantalla, luego se aborta el trabajo y finalmente se envía un e-mail indicando en qué punto se produjo el error.

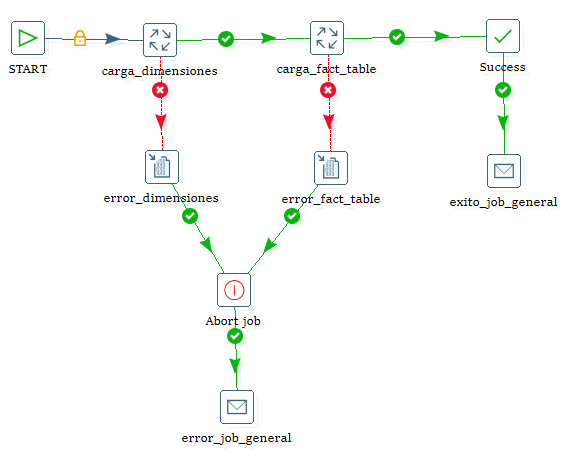


Figura : ETL: trabajo completo

En la Figura 40se puede observar la secuencia completa de ejecución para la carga de las tablas de dimensiones.

Al igual que en el trabajo general, si la ejecución se completa de manera correcta, se muestra un mensaje y envía un e-mail indicado que finalizó correctamente.

Por el contrario, se producirse un error en cualquiera de las dimensiones, se notifica de ello mediante la consola, se aborta el trabajo y envía un email al administrador de sistemas indicando el punto en el que se produjo el error.

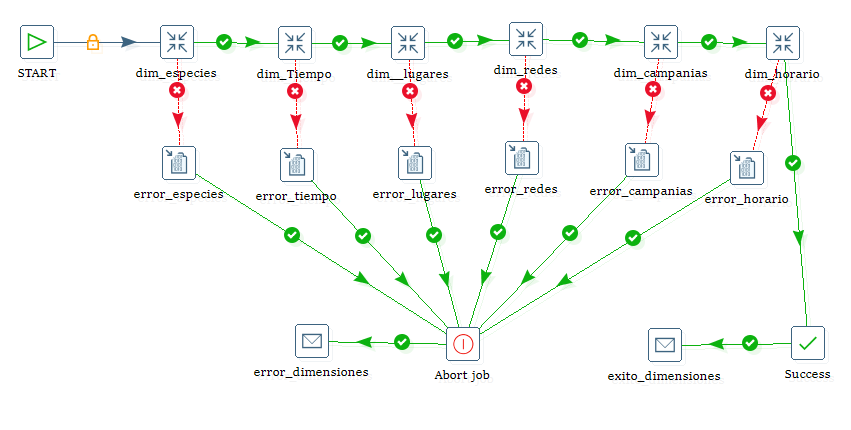


Figura : ETL: Dimensiones

A continuación se presenta el trabajo en el que se muestra la secuencia para la carga de la tabla de hechos, Figura 41. Sigue el mismo proceso descripto anteriormente para indicar el éxito o interrupción de la ejecución.

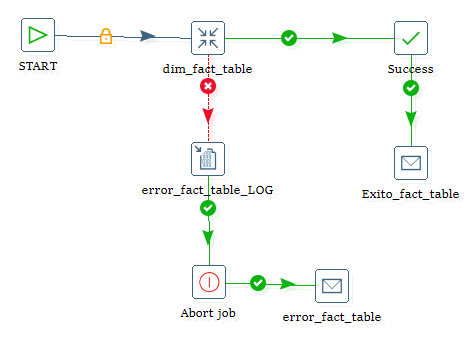


Figura : Carga tabla de hechos.

**ETLDimensiones:**

La mayoría de los procesos de carga de dimensiones siguen una secuencia general. Primero, se leen los datos desde el OTLP de la organización, se realizan las tareas de transformación necesarias y luego se cargan en la tabla correspondiente al DM.

Las dimensiones de tiempo y horario a diferencia de las demás, comienzan la ejecución desde una vista de la base de datos y luego se cargan los datos en las tablas correspondientes del DM.

La última diferencia en la carga de dimensiones se encuentra en la tabla/dimensión de lugares. Dentro OLTP es una tabla muy poco cambiante, pueden trascurrir años hasta que se agregue un nuevo registro y, al menos en este caso, no ha ocurrido que se modifique uno existente. Por dicho motivo y para establecer en el DM un estudio más amplio que el lugar en sí mismo, se agregaron 2 columnas de provincia y país. Para no alterar la tabla de origen pero incluir estos valores al DM, se creó un archivo .xls en donde se establece la relación entre cada lugar y su respectiva provincia y país. Solo para este caso, si se necesita incorporar un nuevo lugar, será necesario modificar el archivo .xls e insertar un nuevo registro.

**Especies**: en la (Figura 42) se pueden observar los pasos de ejecución para la carga de la dimensión Especies.

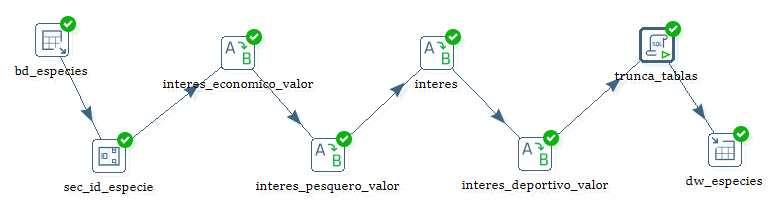


Figura : ETL: Especies

**Tiempo**: en la Figura 43se pueden observar los pasos de ejecución para la carga de la dimensión Tiempo.

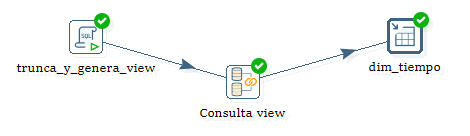


Figura : Transformacion perspectiva tiempo.

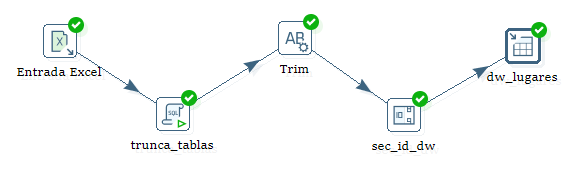
**Lugares**: en la Figura 44se pueden observar los pasos de ejecución para la carga de la dimensión Lugares.

Figura : Transformacion perspectiva lugares.

**Redes**: en la Figura 44se pueden observar los pasos de ejecución para la carga de la dimensión redes.

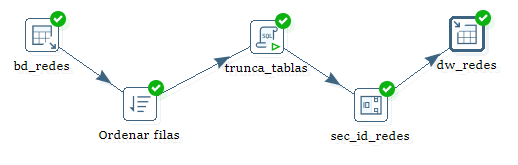


Figura : Transformación perspectiva redes.

**Campañas**: en la Figura 44se pueden observar los pasos de la ejecución para la carga de la dimensión campañas.

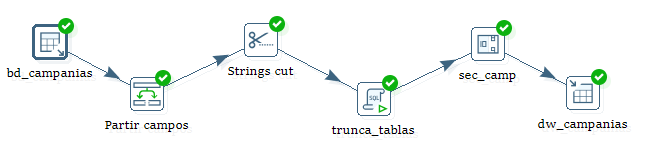


Figura : Transformación perspectiva campañas.

**Horario**: enlaFigura 44 se pueden observar los pasos de la ejecución para la carga de la dimensión horario.

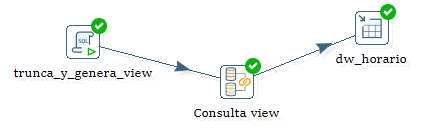


Figura : Transformación perspectiva horario.

**ETLtabladehechos**

En la Figura48 se pueden observar los pasos de la ejecución para la carga de la tabla de hechos.

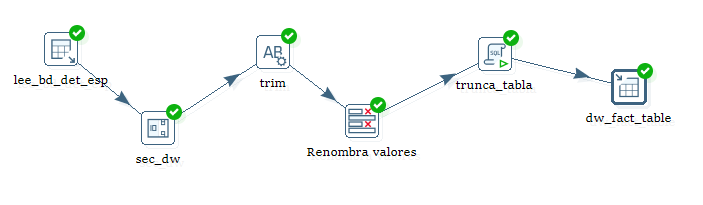


Figura : Transformación tabla de hechos.

##### Anexo 3

A continuación se puede observar el código XML generado por la herramienta Schemaworkbench:

<Schemaname="esquema\_dw\_pesca">

<Dimension type="StandardDimension" visible="true" highCardinality="false" name="dim\_lugar">

<Hierarchy name="lugares" visible="true" hasAll="true" primaryKey="id\_lugar">

<Table name="lugares" schema="public">

</Table>

<Level name="pais" visible="true" column="nom\_pais" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="nom\_pais">

</Level>

<Level name="provincia" visible="true" column="nom\_pcia" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="nom\_pcia">

</Level>

<Level name="lugares" visible="true" table="lugares" column="cod\_lugar" nameColumn="cod\_lugar" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="cod\_lugar">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" highCardinality="false" name="dim\_redes">

<Hierarchy name="redes" visible="true" hasAll="true" primaryKey="id\_red\_oltp">

<Table name="redes" schema="public">

</Table>

<Level name="redes" visible="true" table="redes" column="num\_red" nameColumn="num\_red" ordinalColumn="num\_red" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="num\_red">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="TimeDimension" visible="true" highCardinality="false" name="dim\_horario">

<Hierarchy name="horario" visible="true" hasAll="true" primaryKey="id\_hora">

<Table name="horario" schema="public">

</Table>

<Level name="horario" visible="true" table="horario" column="franja\_hora" nameColumn="franja\_hora" type="String" uniqueMembers="true" levelType="TimeDays" hideMemberIf="Never" captionColumn="franja\_hora">

</Level>

<Level name="hora" visible="true" table="horario" column="id\_hora" type="Time" uniqueMembers="true" levelType="TimeDays" hideMemberIf="Never" captionColumn="id\_hora">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" highCardinality="false" name="dim\_campanias">

<Hierarchy name="convenio" visible="true" hasAll="true" primaryKey="id\_camp\_oltp">

<Table name="campanias" schema="public">

</Table>

<Level name="convenio" visible="true" table="campanias" column="convenio" nameColumn="convenio" ordinalColumn="id\_camp\_oltp" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="convenio">

</Level>

<Level name="num\_camp" visible="true" table="campanias" column="num\_camp" nameColumn="num\_camp" ordinalColumn="id\_camp\_oltp" type="Numeric" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="num\_camp">

</Level>

<Level name="anio\_camp" visible="true" table="campanias" column="anio" nameColumn="anio" type="Numeric" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="anio">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" highCardinality="false" name="dim\_especies">

<Hierarchy name="especies" visible="true" hasAll="true" primaryKey="id\_especie\_oltp">

<Table name="especies" schema="public">

</Table>

<Level name="clases" visible="true" table="especies" column="nom\_clase" nameColumn="nom\_clase" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="nom\_clase">

</Level>

<Level name="ordenes" visible="true" table="especies" column="nom\_orden" nameColumn="nom\_orden" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="nom\_orden">

</Level>

<Level name="familia" visible="true" table="especies" column="nom\_familia" nameColumn="nom\_familia" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="nom\_familia">

</Level>

<Level name="especies" visible="true" table="especies" column="nom\_especie" nameColumn="nom\_especie" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="nom\_especie">

</Level>

<Level name="nom\_vulgar" visible="true" column="nom\_vulgar" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="nom\_vulgar">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="TimeDimension" visible="true" highCardinality="false" name="dim\_tiempo">

<Hierarchy name="tiempo" visible="true" hasAll="true" primaryKey="id\_tiempo">

<Table name="tiempo" schema="public">

</Table>

<Level name="anio" visible="true" column="anio" type="Integer" uniqueMembers="true" levelType="TimeYears" hideMemberIf="Never" captionColumn="anio">

</Level>

<Level name="mes" visible="true" column="mes" type="Integer" uniqueMembers="true" levelType="TimeMonths" hideMemberIf="Never" captionColumn="mes">

</Level>

<Level name="dia" visible="true" column="dia" type="Integer" uniqueMembers="true" levelType="TimeDays" hideMemberIf="Never" captionColumn="dia">

</Level>

</Hierarchy>

<Hierarchy name="fecha" visible="true" hasAll="true" primaryKey="id\_tiempo">

<Table name="tiempo" schema="public">

</Table>

<Level name="fecha" visible="true" column="fecha\_formateada" type="Date" uniqueMembers="true" levelType="TimeHalfYears" hideMemberIf="Never" captionColumn="fecha\_formateada">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" highCardinality="false" name="dim\_estacion">

<Hierarchy name="Estacion" visible="true" hasAll="true" primaryKey="id\_tiempo">

<Table name="tiempo" schema="public">

</Table>

<Level name="estacion" visible="true" column="estacion\_anio" type="String" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="estacion\_anio">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" highCardinality="false" name="especies\_interes">

<Hierarchy name="Interes\_economico" visible="true" hasAll="true" primaryKey="id\_especie">

<Table name="especies" schema="public">

</Table>

<Level name="interes\_pesquero" visible="true" column="interes\_pesquero" type="Boolean" uniqueMembers="true" levelType="Regular" hideMemberIf="Never" captionColumn="interes\_pesquero">

</Level>

</Hierarchy>

</Dimension>

<Cube name="cubo\_rendim\_pesqueros" caption="cubo\_cpue" visible="true" cache="true" enabled="true">

<Table name="detalle\_especies" schema="public">

</Table>

<DimensionUsage source="dim\_lugar" name="lugares" visible="true" foreignKey="id\_lugar" highCardinality="false">

</DimensionUsage>

<DimensionUsage source="dim\_especies" name="especies" visible="true" foreignKey="id\_especie" highCardinality="false">

</DimensionUsage>

<DimensionUsage source="dim\_redes" name="redes" visible="true" foreignKey="id\_red" highCardinality="false">

</DimensionUsage>

<DimensionUsage source="dim\_campanias" name="campanias" visible="true" foreignKey="id\_campania" highCardinality="false">

</DimensionUsage>

<DimensionUsage source="dim\_horario" name="horario" visible="true" foreignKey="id\_hora" highCardinality="false">

</DimensionUsage>

<DimensionUsage source="dim\_tiempo" name="tiempo" visible="true" foreignKey="id\_tiempo" highCardinality="false">

</DimensionUsage>

<DimensionUsage source="dim\_estacion" name="estacion" visible="true" foreignKey="id\_tiempo" highCardinality="false">

</DimensionUsage>

<Measure name="biomasa\_grms" column="peso" aggregator="sum" visible="false">

</Measure>

<Measure name="superficie" column="valor\_superifice" aggregator="max" visible="false">

</Measure>

<Measure name="hs\_trabajadas" column="hs\_trabajadas" aggregator="max" visible="false">

</Measure>

<Measure name="riqueza" column="id\_especie" aggregator="distinct count" caption="Riqueza" visible="true">

</Measure>

<Measure name="abundancia" column="id\_especie" aggregator="count" caption="Abundancia" visible="true">

</Measure>

<CalculatedMember name="biomasa\_kg" caption="biomasa\_kg" formula="[Measures].[biomasa\_grms]/(1000)" dimension="Measures" visible="false">

</CalculatedMember>

<CalculatedMember name="abundancia\_en\_100mts" formula="[Measures].[abundancia]\*100" dimension="Measures" visible="false">

</CalculatedMember>

<CalculatedMember name="abundancia\_cpue" formula="(([Measures].[abundancia\_en\_100mts] / [Measures].[superficie])\*8)/[Measures].[hs\_trabajadas]" dimension="Measures" visible="true">

</CalculatedMember>

<CalculatedMember name="biomasa\_cpue" formula="((([Measures].[biomasa\_kg]\*100) / [Measures].[superficie])\*8)/[Measures].[hs\_trabajadas]" dimension="Measures" visible="true">

</CalculatedMember>

</Cube>

</Schema>

##### Bibliografía

[1] DATE, C. J y RUIZ FAUDON, S. L., *Introducción a los sistemas de base de datos*, 7.a ed. México: Pearson Educación.

[2] SILBERSCHATZ, A., KORTH, H. F., y SUDARSHAN, S., *Fundamentos de Bases de datos*, 4.a ed. Madrid: S.A. MCGRAW-HILL / INTERAMERICANA DE ESPAÑA, 2002.

[3] LUHN, Hans Peter, «A Business IntelligenceSystem.», *IBM J. Res. Dev.*, vol. Vol 2, No 4. (October 1958), pp. 314 - 319.

[4] CODD, E. F., «A relational model of data for large shared data banks», vol. Vol 13 - Issue 6, pp. 377-387, jun-1970.

[5] POWER, D. J., «A Brief History of Decision Support Systems», 2007.

[6] SANSU, G, «Breve historia del Business Intelligence». [En línea]. Disponible en: http://www.businessintelligence.info/definiciones/historia-business-intelligence.html. [Accedido: 30-ago-2015].

[7] GRATTON, Simon, «BI 3.0 The Journey to Business Intelligence. What does it mean? | Blog post», *CapgeminiCapgemini Worldwide*. [En línea]. Disponible en: https://www.capgemini.com/blog/capping-it-off/2012/07/bi-30-the-journey-to-business-intelligence-what-does-it-mean. [Accedido: 30-ene-2016].

[8] INMON, W.H., «Building the Data warehouse». 3rd ed. John Wiley & Sons, Inc., 2002, pp 31.

[9] KIMBALL, R. y CASERTA, J., *The Data Warehouse ETL Toolkit. Practical teschniques for Extracting, Cleaning, Conforming, and Delivering Data*. Wiley, 2004.

[10] CHAUDHURI, Surajit y DAYAL, Umeshwar, «An Overview of Data Warehousing and OLAP Technology», *ACM Sigmod Rec.*, vol. 26, NRO 1, 1997.

[11] BERNABEU, R. D, «Hefesto-v2», 2010. [En línea]. Disponible en: http://www.businessintelligence.info/docs/hefesto-v2.pdf. [Accedido: 15-jul-2015].

[12] CODD, E. F., «Providing OLAP to User-Analysts: An IT Mandate», vol. 32, pp 3-5, 1993.

[13] BALLVÉ, Alberto M., «Creando conocimiento en las organizaciones con el Cuadro de Mando Integral y el Tablero de Control», vol. 3, pp. 13-38, 2006.

[14] KAPLAN, Robert S. y NORTON, David P., «The balanced scorecard - Measures that drive performance.», feb. 1992.

[15] ZAGORECKI, A., RISTVEJ, J., COMFORT, L., y LOVECEK, T., «Executive Dashboard System for Emergency Management», *Komunikacie*, vol. 14, pp. 82-89, 2012.

[16] KIMBALL, R., ROSS, M., THORNTHWAITE, W, MUNDY, J., y BECKER, B., *The Data Warehouse Lifecycle Toolkit*, 2da ed. Canadá: Wiley Publishing, Inc, 2008.

[17] W. H. INMON, *Building The Data Ware House*. Dreamtech, 2003.

[18] SALLAM, Rita, TAPADINHAS, Joao, PARENTEAU, Josh, YUEN, Daniel, y HOSTMANN, Bill, «Magic Quadrant for Business Intelligence and Analytics Platforms», *Tableau Software*, 04-feb-2016. [En línea]. Disponible en: http://get.tableau.com/gartner-magic-quadrant-2016.html. [Accedido: 11-feb-2016].

[19] «Comparativa Open Source BI». [En línea]. Disponible en: http://www.stratebi.es/todobi/jun10/Comparativa\_OSBI.pdf. [Accedido: 21-feb-2016].

[20] «GNU GENERAL PUBLIC LICENSE Version 2, June 1991», *Pentaho*. [En línea]. Disponible en: http://www.pentaho.com/license/gplv2. [Accedido: 11-feb-2016].

[21] Pentaho, «Customer Overview Pentaho», *Pentaho*. [En línea]. Disponible en: http://www.pentaho.com/customers. [Accedido: 21-feb-2016].

[22] «Mondrian | PentahoCommunity». [En línea]. Disponible en: http://community.pentaho.com/projects/mondrian/. [Accedido: 12-feb-2016].

[23] SHANNON, C. E. y W. WEAVER, *The Mathematical Theory of Communication.Univ. Illinois Press, Urbana. 177 p. En MAGURRAN, A. E., 1989. Diversidad Ecológica y su medición. Ed. Vedra, Barcelona. 199 p.* 1963.

[24] KIMBALL, R., ROSS, M., THORNTHWAITE, W, MUNDY, J., y BECKER, B., *The Data Warehouse Lifecycle Toolkit 2ed*, 2da ed. Canadá: Wiley Publishing, Inc, 2008.

[25] «Date and Time dimensions - PostgreSQL wiki». [En línea]. Disponible en: https://wiki.postgresql.org/wiki/Date\_and\_Time\_dimensions. [Accedido: 25-feb-2016].

1. UNaM: acrónimo de Universidad Nacional de Misiones. [↑](#footnote-ref-2)
2. EBY: acrónimo de Entidad Binacional Yacyretá. [↑](#footnote-ref-3)
3. OLAP: acrónimo de procesamiento analítico en línea, del inglés*On Line AnalyticalProcessing*. [↑](#footnote-ref-4)
4. DW: acrónimo de almacén de datos, del inglés *Data Warehouse*. [↑](#footnote-ref-5)
5. ETL: acrónimo de extracción, transformación y carga, del inglés *Extraction, Transformation andLoad.* [↑](#footnote-ref-6)
6. DM: acrónimo de *Data Mart*, es un subconjunto del *data warehouse*. [↑](#footnote-ref-7)
7. SGBD: acrónimo de sistema gestor de base de datos - del inglés *DataBase Management System*. [↑](#footnote-ref-8)
8. SQL: acrónimo de lenguaje estructurado de consulta – del inglés *StructuredQueryLanguage.* [↑](#footnote-ref-9)
9. LDD: acrónimo de lenguaje de definición de datos – del inglés *Data DefinitionLanguage*. [↑](#footnote-ref-10)
10. LMD: acrónimo de lenguaje de manipulación de datos - del inglés *Data ManipulationLanguage*. [↑](#footnote-ref-11)
11. OLTP: acrónimo de procesamiento de transacciones en línea – del inglés*Online TransactionProcessing.* [↑](#footnote-ref-12)
12. BI: acrónimo de inteligencia de negocios, del inglés *Business Intelligence*. [↑](#footnote-ref-13)
13. Staging área: almacenamiento intermedio, del inglés *StagingArea.* [↑](#footnote-ref-14)
14. Tabla de Hechos: del inglés,*FactTable*. [↑](#footnote-ref-15)
15. ROLAP: del inglés Relational On Line Analytic Processing. [↑](#footnote-ref-16)
16. MOLAP: del inglés, Multidimensional On Line Analityc Processing. [↑](#footnote-ref-17)
17. HOLAP: del inglés, Hybrid On Line Analityc Processing. [↑](#footnote-ref-18)
18. KPIs: del inglés Key Performance Indicators. [↑](#footnote-ref-19)
19. Ciclo de vida dimensional del negocio: del inglés, *Business Dimensional Lifecycle.* [↑](#footnote-ref-20)
20. Matriz de dimensiones: del inglés, *Bus Matrix*. [↑](#footnote-ref-21)
21. mm: milímetro es la unidad de longitud según el Sistema Métrico Decimal que corresponde a la milésima parte de un metro. [↑](#footnote-ref-22)
22. **g**: gramos es la unidad utilizada para el peso según el Sistema Métrico Decimal que corresponde a la milésima parte de un kilogramo. [↑](#footnote-ref-23)
23. CPUE: acrónimo de Captura Por Unidad de Esfuerzo, expresa la captura a través de un determinado esfuerzo. [↑](#footnote-ref-24)
24. Claves subrogadas: son claves generadas durante el proceso que no tienen relación con las operaciones del negocio. [↑](#footnote-ref-25)
25. XML: lenguaje de marcas extensible, del inglés*eXtensibleMarkupLanguage.* [↑](#footnote-ref-26)
26. Drag and drop: significa arrastrar y soltar con el cursor del ratón los objetos de una ventana a otra o entre partes de una misma ventana. [↑](#footnote-ref-27)