

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA  
Facultad de Contaduría y Administración  
Facultad de Ciencias Químicas e Ingeniería  
Maestría en Tecnologías de la Información y la Comunicación



## **Análisis de información y toma de decisiones, mediante almacén de datos. Caso SIEDCyT.**

TESIS

PARA OBTENER EL GRADO DE  
MAESTRO EN TECNOLOGÍAS DE LA INFORMACIÓN Y LA COMUNICACIÓN

*Presenta:*

**RICARDO JAVIER IBARRA IBARRA**

*Bajo la dirección de:*

M.C.C. NORA DEL CARMEN OSUNA  
MILLAN

*Codirección:*

M.C. MARGARITA RAMIREZ  
RAMIREZ

TIJUANA, BAJA CALIFORNIA, MÉXICO

MAYO 2011

*Quiero dedicar esta tesis a mi esposa  
Nashelly y mi hijo Leonardo por la pa-  
ciencia que me tuvieron durante toda la  
maestría y su incondicional apoyo en el  
logro de mis metas personales. Sin su pa-  
ciencia y comprensión yo no hubiese podi-  
do concluir esta etapa tan importante de  
mi vida.*

*Hoja en blanco para sustituirla por la  
aprobación de impresión de la tesis.*

# Agradecimientos

Quiero agradecer a aquel que me dio la capacidad de realizar todo lo que me propongo, a aquel que siempre me ayuda en los momentos más difíciles y que nunca pierde la fe en mí. Agradezco a Dios por permitirme seguirme superando y que me da fuerzas para seguir adelante. También quiero agradecer a toda mi familia que siempre tiene una palabra de aliento para ayudarme a seguir y no dejar de superarme. Finalmente quiero agradecer a todos mis maestros por su guía, dedicación y apoyo brindado durante mis estudios y la realización de esta tesis.

*Ricardo Javier Ibarra Ibarra*

# Resumen

El estado de Baja California se ha propuesto impulsar el desarrollo tecnológico de la región. Para lograrlo se apoyará del sistema de información y documentación científica y tecnológica (SIEDCyT). Dicho sistema administrará información de investigaciones y proyectos realizados en el estado.

La productividad se dará a conocer por medio de indicadores. Con el propósito de administrar mejor la información y tener una respuesta acertada a las principales preguntas del desempeño científico y tecnológico en Baja California. Los datos serán migrados a un almacén de datos. Posteriormente por medio de herramientas especializadas o BI (por sus siglas en inglés **Business Intelligence**), se podrá visualizar, estudiar y compartir información para la toma de decisiones. El consejo de ciencia y tecnología se beneficiara bastante si cuenta con la capacidad de analizar la información de distintas maneras para determinar de forma asertiva el destino de los apoyos a los proyectos tecnológicos.

# Índice general

<b>Introducción.</b>	<b>i</b>
<b>1. El Problema.</b>	<b>1</b>
1.1. Planteamiento del problema. . . . .	1
1.2. Obejetivos de la investigación. . . . .	2
1.3. Justificación. . . . .	3
1.4. Alcance . . . . .	4
<b>2. Marco Teórico.</b>	<b>5</b>
2.1. Antecedentes Históricos. . . . .	5
2.2. Definición de Términos Básicos . . . . .	6
<b>3. Metodología</b>	<b>18</b>
3.1. Obtener requerimientos de informacion. . . . .	19
3.2. Definición de requerimientos técnicos . . . . .	19
3.3. Diseño de prototipos multidimensionales . . . . .	19
3.4. Diseño Físico . . . . .	20

3.5. Diseño y desarrollo del proceso de carga de datos al modelo . . . . .	20
3.6. Diseño y/o configuración de herramientas para el análisis de información . . .	20
3.7. Metodología aplicada al SIEDCyT . . . . .	21
3.8. Diseño de arquitectura . . . . .	23
3.8.1. Grupo de trabajo de planeación e implementación de arquitectura. . .	23
3.8.2. Requerimientos técnicos basados en las necesidades de los usuarios . .	23
3.8.3. Integración del DW/BI con otros sistemas . . . . .	26
3.9. Equipo e infraestructura . . . . .	26
3.10. Selección del producto . . . . .	27
3.11. Modelado Dimensional . . . . .	29
3.12. Diseño físico . . . . .	52
3.12.1. Estandarización de nomenclatura . . . . .	52
3.12.2. Desarrollo del modelo físico de datos . . . . .	55
<b>4. Extracción transformación y carga de datos</b>	<b>65</b>
<b>5. Aplicaciones de inteligencia de negocios</b>	<b>92</b>
5.1. Plataforma de entrega de reportes . . . . .	95
5.2. Proceso del desarrollo de aplicaciones de inteligencia de negocios . . . . .	96
5.2.1. Estándar y apariencia de la plantilla de los reportes . . . . .	96
5.2.2. Lista de reportes preliminares . . . . .	99
5.2.3. Maqueta y documentación de reportes . . . . .	99
5.2.4. Diseño del marco de navegación . . . . .	100

- 5.2.5. Revisión y aprobación de los usuarios . . . . . 100
  
- 6. Desarrollo de la aplicación de inteligencia de negocios en Microsoft Reporting Services . . . . . 105**

  - 6.1. Portal BI . . . . . 106
    - 6.1.1. Impacto en el diseño . . . . . 107

  
- Líneas futuras . . . . . 115**
  
- Conclusión . . . . . 116**
  
- Glosario . . . . . 117**

# Índice de figuras

2.1. Estructura de un Datawarehouse. . . . .	7
2.2. Diferencias entre metodologías . . . . .	8
2.3. Arquitectura Top-Down . . . . .	9
2.4. Arquitectura Bottom-Up . . . . .	10
2.5. Ejemplo de un cubo multidimensional. . . . .	11
2.6. Ejemplo de la estructura estrella. . . . .	13
2.7. Ejemplo de la estructura copo de nieve. . . . .	13
2.8. Ejemplo estructura constelación de hechos. . . . .	13
2.9. Fases de la Inteligencia de Negocios BI. . . . .	14
3.1. Diagrama del ciclo de vida de Ralph Kimball. . . . .	18
3.2. Grados académicos adquiridos por investigador . . . . .	32
3.3. Patentes registradas . . . . .	33
3.4. Investigaciones realizadas . . . . .	34
3.5. Vinculación entre investigadores y grupos de investigación . . . . .	34
3.6. Artículos generados por investigadores . . . . .	35

3.7. Libros generados por investigadores . . . . .	36
3.8. Bus matrix: Referencia de granularidad y uso de dimensiones por cada hecho	37
3.9. Referencia de granularidad y uso de dimensiones por cada hecho . . . . .	38
3.10. Patentes registradas . . . . .	39
3.11. Investigaciones realizadas . . . . .	39
3.12. Vinculación entre investigadores y grupos de investigación . . . . .	39
3.13. Artículos generados por investigadores . . . . .	40
3.14. Participación de investigadores en la generación de libros . . . . .	40
3.15. Relaciona investigaciones con los financiamientos obtenidos . . . . .	41
3.16. Relaciona instituciones con investigaciones . . . . .	41
3.17. Relaciona investigadores con investigaciones . . . . .	42
3.18. Relaciona investigaciones con sus diferentes líneas de investigación . . . . .	42
3.19. Catalogo de artículos . . . . .	43
3.20. Detalle de grados académicos . . . . .	44
3.21. Grupos de investigación registrados en el siedcyt . . . . .	44
3.22. Grupos de investigación registrados en el siedcyt . . . . .	45
3.23. Detalle de instituciones . . . . .	45
3.24. información de investigaciones . . . . .	46
3.25. Detalle de investigadores . . . . .	47
3.26. Libros generados y capturados en el siedcyt . . . . .	48
3.27. catalogo de líneas de investigación . . . . .	49
3.28. Tipo de participación que tuvo el investigador en los libros . . . . .	49

3.29. Patentes generadas y capturadas en el siedcyt . . . . .	50
3.30. Ubicación geográfica desde ciudad hasta continente . . . . .	51
3.31. Artículos . . . . .	55
3.32. Financiamientos . . . . .	56
3.33. Grados académicos . . . . .	56
3.34. Grupos de investigación . . . . .	56
3.35. Instituciones . . . . .	57
3.36. Investigaciones . . . . .	57
3.37. Investigadores . . . . .	58
3.38. Libros . . . . .	58
3.39. Líneas de investigación . . . . .	59
3.40. Participación en libros . . . . .	59
3.41. Patentes . . . . .	59
3.42. Zona geográfica . . . . .	60
3.43. Relación entre financiamientos e investigación . . . . .	60
3.44. Relación entre instituciones e investigaciones . . . . .	61
3.45. Relación entre investigadores e investigación . . . . .	61
3.46. Relación entre investigaciones y líneas de investigación . . . . .	61
3.47. Grados académicos adquiridos . . . . .	62
3.48. Vinculación de investigadores en grupos de investigación . . . . .	62
3.49. Investigaciones realizadas . . . . .	62
3.50. Participación de investigadores en la creación de libros . . . . .	63

3.51. Participación de investigadores en la creación de artículos . . . . .	63
3.52. Patentes registradas . . . . .	63
4.1. Dimensión áreas de conocimiento . . . . .	67
4.2. Dimensión artículos . . . . .	68
4.3. Dimensión financiamientos . . . . .	69
4.4. Dimensión grados académicos . . . . .	70
4.5. Dimensión grupos de investigación . . . . .	71
4.6. Dimensión Instituciones . . . . .	72
4.7. Dimensión Investigaciones . . . . .	73
4.8. Dimensión Investigador . . . . .	74
4.9. Dimensión libros . . . . .	75
4.10. Dimensión líneas de investigación . . . . .	76
4.11. Dimensión participación en libros . . . . .	77
4.12. Dimensión patentes . . . . .	78
4.13. Dimensión zona geográfica . . . . .	79
4.14. Hechos participación en libros . . . . .	80
4.15. Hechos grados adquiridos . . . . .	81
4.16. Hechos vinculación en grupos de investigación . . . . .	82
4.17. Hechos participación en investigaciones . . . . .	83
4.18. Hechos participación en articulos . . . . .	84
4.19. Hechos patentes registradas . . . . .	85
4.20. Paquete ETL para carga de modelo grados adquiridos. (Vista general) . . . .	87

## ÍNDICE DE FIGURAS

---

4.21. Carga de dimensión zonas geográficas. . . . .	88
4.22. Carga de dimensión instituciones. . . . .	88
4.23. Carga de dimensión de grados académicos. . . . .	89
4.24. Carga de dimensión de investigador. . . . .	90
4.25. Carga de tabla de hechos grados adquiridos. . . . .	91
5.1. Plantilla de reportes. . . . .	98
5.2. Lista preliminar de reportes. . . . .	100
5.3. Maqueta grados académicos adquiridos. . . . .	101
5.4. Maqueta patentes registradas. . . . .	101
5.5. Maqueta investigaciones realizadas. . . . .	102
5.6. Maqueta Grupos de investigación. . . . .	102
5.7. Maqueta participación de investigadores en libros. . . . .	103
5.8. Maqueta financiamientos otorgados. . . . .	103
6.1. Ejemplo de página de inicio de portal BI. Es fácil y accesible encontrar cualquier elemento. . . . .	108
6.2. Ejemplo de Interface de selección de dimensiones y hechos o facts en Microsoft Excel. . . . .	111
6.3. Total y tipo de grados adquiridos por los investigadores por año. . . . .	112
6.4. Rango de ingresos por nivel de SNI. . . . .	113
6.5. Autenticación desde Excel al almacén de datos. . . . .	114

# Introducción.

Con el propósito de impulsar el desarrollo tecnológico de la región., El Consejo Estatal de Ciencia y Tecnología (COCyT) a través del programa de fondos mixtos otorgó a la Universidad Autónoma de Baja California (UABC), la creación del Sistema de Información y Documentación Científica y Tecnológica (SIEDCyT), en el cual ingresará información referente a instituciones, investigadores aliados al Sistema Nacional de Investigadores (SNI), proyectos de investigación, desarrollos tecnológicos, artículos, patentes, libros y desarrollo de software, se hará mención de lo más destacado; el sistema manejará únicamente la información relacionada con ciencia y tecnología del estado de Baja California. El SIEDCyT funcionará en una plataforma web y tendrá incluidos algunos reportes que han sido considerados desde el inicio. La base de datos del sistema almacenará grandes cantidades de información. Aprovecharla al máximo, mediante un análisis profundo se convierte en el reto principal.

Las soluciones basadas en almacenes de datos y herramientas de inteligencia empresarial, buscan proveer al usuario del negocio las herramientas e información necesaria para operar y tomar las decisiones basadas en datos reales y correctos. Para lograrlo es necesario contar con una plataforma enfocada a obtener reportes e información de valor de manera eficiente y oportuna. Los almacenes de datos son repositorios grandes de información y su propósito principal es apoyar a herramientas de inteligencia empresarial en la presentación de información. Otro propósito de estos, es el aumentar la productividad del usuario al no tener

que esperar mucho tiempo para obtener algún reporte. Cuando las bases de datos son muy grandes, el desempeño de los reportes se ve degradado. Un almacén de datos busca evitar este problema y además, permite al usuario crear sus propios reportes mediante las herramientas de inteligencia empresarial. Esto puede lograrse aun sin tener conocimientos técnicos de programación y sin conocer los detalles de las bases de datos. Con la cantidad de información que se almacenara en el SIEDCyT y considerando el crecimiento que puede llegar a tener en un futuro próximo. Se busca recomendar una solución basada en herramientas de inteligencia empresarial para obtener un mejor conocimiento de los datos del sistema de información.

El proyecto abarca el análisis de la fuente de información (base de datos normalizada), el diseño de modelos dimensionales, el diseño de extracción, transformación y carga de datos, el diseño de cubos OLAP, el diseño de reportes clave para COCYT y recomendaciones para desarrollar e implementar el portal de inteligencia empresarial (Portal BI). La información que proporcione el almacén de datos a través de las herramientas de inteligencia empresarial, ayudara a conocer la verdad sobre el desempeño o el grado de actividad en investigación y desarrollo tecnológico en Baja California. A partir de esta información, COCYT podrá tomar mejores decisiones con respecto a que áreas de la investigación apoyar y con mayor seguridad otorgar recursos para el impulso de proyectos que beneficien al estado.

# Capítulo 1

## El Problema.

### 1.1. Planteamiento del problema.

El diseño de la base de datos cumple con todas las reglas de normalización. Esta característica no permite realizar un análisis ágil y profundo de la información. El problema principal radica en que se requiere bastante poder de procesamiento y tiempo para lograr estudiar los datos a lo largo del tiempo.

Un segundo problema se da al momento del diseño de los reportes al no contemplar algún dato que pudiese ser necesario conocer en un futuro. Un programador con conocimientos de SQL podría realizar la tarea de generar nuevos reportes de acuerdo a las necesidades en cuestión, pero aún así el extraer una gran cantidad de registros afecta el desempeño del equipo.

Otros factores se ven afectados cuando la información es analizada desde las fuentes originales, si el servidor donde se hospeda la base de datos realiza otras tareas, es posible que estas no se realicen de manera adecuada. La carga de trabajo en el servidor se vuelve muy intensa al obtener reportes con grandes cantidades de registros y operaciones. Por estos motivos un análisis profundo de la información no se puede realizar de manera óptima sobre una base

de datos normalizada.

Los aspectos técnicos mencionados anteriormente afectan directamente a los usuarios del SIEDCyT en la medida en que no puedan visualizar la información desde la perspectiva más adecuada para la toma de decisiones. Los sistemas normalizados tienen una visión pobre de los datos cuando se trata de conocer el comportamiento de estos. Con una visión cerrada, será difícil para COCYT tomar una decisión respecto a la mejor opción sobre el destino de los apoyos, y a largo plazo se podría llegar a tener pocos argumentos para justificar la inversión en Baja California, entre otros.

Es necesario complementar el SIEDCyT con un sistema adicional para análisis de información y toma de decisiones, que tenga la capacidad de resolver futuras necesidades de información, comúnmente conocido como inteligencia de negocios(**Business Intelligence**).

## 1.2. Objetivos de la investigación.

1. **Objetivo general:** Diseño de almacén de datos del SIEDCyT, para soporte y toma de decisiones mediante herramientas de análisis.

2. **Objetivos específicos:**

- Proveer información a nivel estatal sobre ciencia y tecnología para planear, crear estrategias y tomar decisiones sobre la asignación de recursos en el estado.
- Proporcionar métricos exactos sobre el desempeño y productividad científica y tecnológica del estado.
- Evitar degradar el desempeño del sitio web SIEDCyT al momento de obtener reportes o generar nuevas preguntas
- Crear un ambiente amigable de reportes, en el cual los usuarios puedan generar y mantener sus propios reportes, sin asistencia alguna del personal de tecnologías de información.

- Pronósticos sobre el comportamiento de los datos en el futuro para la toma de decisiones.
- Proveer una fiel y única versión de la información.

### 1.3. Justificación.

El motivo principal de la creación del SIEDCyT es tener información que nos ayude a conocer los proyectos, logros y avances tecnológicos de Baja California. El proyecto original hará una gran labor en recopilar y clasificar información referente a instituciones, investigadores, patentes y la información científica y tecnología que se genere en el estado.

Todos estos datos deben darse a conocer en su momento, sin embargo, no se puede conocer de antemano que reportes específicos se van a requerir. Uno de los objetivos principales de este sistema, es informar sobre la realización de proyectos e innovación tecnológica del estado. Para informar de manera correcta y oportuna; es necesario contar con una estructura de base de datos que permita la fácil extracción de la información aún cuando el desarrollador no piense en un reporte en específico. La base del proyecto es dar una herramienta de información fiel y oportuna a COCyT.

La calidad y tipo de información que nos da como resultado de un almacén de datos es de mucho mayor impacto a nivel de análisis y toma de decisiones. Sin la existencia del almacén, los reportes, indicadores de desempeño, pronósticos y reportes especializados podrían tardar horas en obtenerse, además de afectar en gran manera el desempeño del SIEDCyT debido a la carga y consumo de recursos que estas consultas pueden generar en el equipo. En el almacén de datos todas estas preguntas son previamente calculadas para reducir al mínimo el tiempo de espera y utilización de recursos al realizar este tipo de consultas.

Por todo lo anterior, es necesario crear un almacén de datos que permita lograr los objetivos de manera óptima.

## 1.4. Alcance

Este caso práctico cubrirá el análisis y diseño del almacén de datos. Se podrán apreciar las bondades que brinda el organizar la información orientada a temas y la facilidad de obtener información de valor de una manera amigable, rápida y confiable.

# Capítulo 2

## Marco Teórico.

### 2.1. Antecedentes Históricos.

Dependemos de forma creciente de la ciencia y la tecnología, los procesos de producción, las fuentes de alimentación, la medicina, la educación, la comunicación o el transporte son campos cuyo presente y futuro están fuertemente ligados al desarrollo tecnológico y científico, particularmente a la innovación.

La innovación y el cambio tecnológico no son procesos lineales ni características de una empresa, sino el resultado de un ambiente, un entorno que promueve espacios con capacidad para generar, adoptar, y difundir innovaciones, conocidos como espacios emergentes, regiones que aprenden o entorno innovador.

Un territorio innovador es un sistema productivo vinculado a una o varias actividades, en el que una buena parte de las empresas existentes realizan esfuerzos en el plano de la innovación tecnológica, que incorpora mejoras en los diferentes procesos asociados a su cadena de valor y en los productos que ofrece.

Dos componentes destacan en un contexto de innovación: el grado de desarrollo estructural del sistema, es decir, la cantidad de instituciones que permita asegurar la generación,

adaptación de nuevos conocimientos útiles para la sociedad donde opera, y, las relaciones o condiciones para que los elementos del sistema se comuniquen.[18]

En este marco, el Gobierno de Baja California decide trasladar al COCyT, de la Secretaría de Educación y Bienestar Social (SEBS) a la Secretaría de Desarrollo Económico (SEDECO) a principios del 2008, con el objetivo de fortalecer la formulación e implementación de políticas públicas, la generación de productos de alto valor agregado, la producción de nuevos conocimientos a través de la investigación, la existencia de capital humano con habilidades técnicas y científicas y la habilidad para importar y adaptar el conocimiento producido fuera del país, siendo una herramienta eficaz, eficiente y competitiva.[1]

Algunos estados como Jalisco, Monterrey y Guanajuato, ya cuentan con un sistema de información enfocado a resaltar y dar a conocer todos los proyectos de ciencia y tecnología. Cada estado busca dar a conocer su productividad tecnológica con el fin de atraer nuevas inversiones.

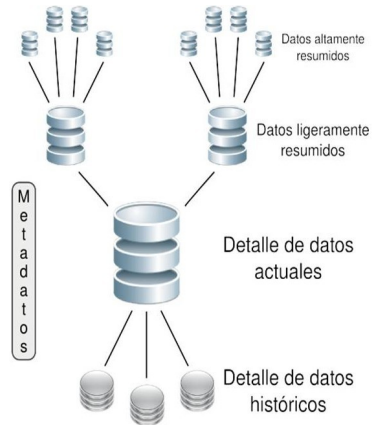
Está claro que entre la ciencia, la tecnología y la sociedad existe una estrecha relación, es decir, en nuestros tiempos la sociedad esta tan ligada con estos dos factores que es imposible separarlos, por todo lo anterior este sistema servirá de base para que el conocimiento tecnológico que se obtenga en la región sea de beneficio para el desarrollo de la entidad.

## 2.2. Definición de Términos Básicos

Para comprender el contexto en que se desarrollará la presente tesis, es importante entender y comprender los conceptos relacionados a los almacenes de datos (**Data Warehouse DW**) y la Inteligencia de Negocios (**Business Intelligence BI**) y todo lo que éstos implican, puesto que dichas definiciones van de la mano con el trabajo a desarrollarse.

1. **Data warehouse:** Es la integración de datos consolidados, almacenados en un dispositivo de memoria no volátil, proveniente de múltiples y posiblemente diferentes fuentes de datos. Con el propósito del análisis y a partir de este tomar decisiones en función de

mejorar la gestión del negocio. Contiene un conjunto de cubos de datos que permiten a través de técnicas de OLAP consolidar, ver y resumir los datos acorde a diferentes dimensiones de estos.[17]



**Figura 2.1:** Estructura de un Datawarehouse.

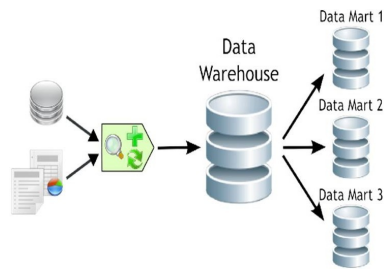
Es importante destacar que existen dos principales metodologías para desarrollar un proyecto de DW/BI, una de ellas es la de *William Inmon* y la segunda es la de *Ralph Kimball*, en la figura 2.2 se ilustran las diferencias entre ambas.

Características	Diferencias	
	Kimball	Inmon
<b>Naturaleza de los requerimientos de la organización para la toma de decisiones</b>	Táctico	Estratégico
<b>Requerimientos de integración de datos</b>	Áreas individuales del negocio	Integración de toda la empresa
<b>Estructura de los datos</b>	métricas de negocio, funcionamiento, medidas y cuadros de mandos	Datos sin métricos y datos a utilizar para satisfacer distintas necesidades de información para el negocio.
<b>Escalabilidad</b>	Necesidad de adaptarse a necesidades altas y volátiles dentro de un alcance limitado.	Alcance indefinido y cambios críticos en los requerimientos.
<b>Persistencia de datos</b>	Los sistemas fuente son relativamente estables.	Alto índice de cambios en los sistemas fuente
<b>Dotación de personal especializado</b>	Pequeños equipos generalistas	Equipos grandes de especialistas
<b>Tiempo para la entrega</b>	La necesidad del primer almacén de datos es urgente.	Los requerimientos de la organización permiten un tiempo más largo de inicio
<b>costo de implementación</b>	Bajo costo de inicio, con cada proyecto subsecuente el costo es aproximadamente lo mismo	Altos costos de inicio, con costos bajos en los proyectos subsecuentes

**Figura 2.2:** Diferencias entre metodologías

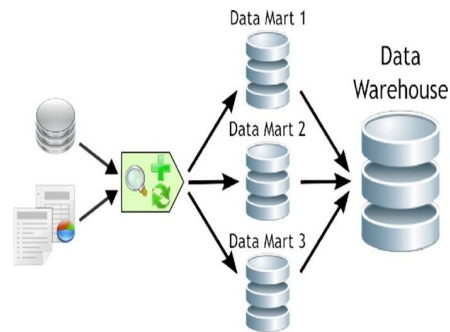
2. **Data Marts DM:** Es un subconjunto del Data Warehouse, usado normalmente para el análisis parcial de los datos. El objetivo de subdividir está dado por la complejidad computacional del análisis global de todas las dimensiones del Data Warehouse y por la necesidad de rapidez. De acuerdo a las operaciones que se deseen o requieran desarrollar, los DM pueden adoptar las siguientes arquitecturas:

a) Top-Down: primero se define el DW y luego se desarrollan, construyen y cargan los DataMart a partir del mismo. En la figura 2.3 se encuentra detallada esta arquitectura:



**Figura 2.3:** Arquitectura Top-Down

- b) Bottom-Up: en esta arquitectura, se definen previamente los DM y luego se integran en un DW centralizado. La siguiente figura presenta esta implementación:



**Figura 2.4:** Arquitectura Bottom-Up

3. **OLTP (Online Transaction Processing):** Se les llama así a las aplicaciones orientadas principalmente a la inserción, actualización y eliminación de datos, diseñada casi siempre usando el modelo Relacional. Estos sistemas están optimizados para realizar estas operaciones en un tiempo corto.
4. **OLAP (Online Analytical Processing):** Son los sistemas que se usan para analizar los datos que las OLTP introducen en la base de datos. Estos casi siempre usan el modelo multidimensional para organizar los datos en la Base de Datos ya que brindan mejores resultados a la hora del análisis.
5. **Base de datos multidimensional:** Es una base de datos en donde su información se almacena en forma multidimensional, es decir, a través de tablas de hechos y tablas de dimensiones. Proveen una estructura que permite, a través de la creación y consulta a una estructura de datos determinada, tener acceso flexible a los datos, para explorar y analizar sus relaciones, y consiguientes resultados.
6. **Tablas de Hechos (Facts):** Son datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones. Los datos presentes en las tablas de hechos constituyen el volumen de la bodega, y pueden estar compuestos por millones de registros dependiendo de su granularidad y

antigüedad de la organización. Los más importantes son los de tipo numérico. El registro del hecho posee una clave primaria que está compuesta por las claves primarias de las tablas de dimensiones relacionadas a este.

7. **Tablas de Dimensión:** definen como están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio. Contienen datos cualitativos. Representan los aspectos de interés, mediante los cuales los usuarios podrán filtrar y manipular la información almacenada en la tabla de hechos.

Cada tabla de dimensión podrá contener los siguientes campos:

- Clave principal o identificador único.
- Clave foráneas.
- Datos de referencia primarios: datos que identifican la dimensión. Por ejemplo: nombre del cliente.
- Datos de referencia secundarios: datos que complementan la descripción de la dimensión. Por ejemplo: e-mail del cliente, fax del cliente, etc.

8. **Cubo multidimensional (hipercubo):** representa o convierte los datos planos que se encuentran en filas y columnas, en una matriz de N dimensiones. Los objetos más importantes que se pueden incluir en un cubo multidimensional, son indicadores, atributos y jerarquías.

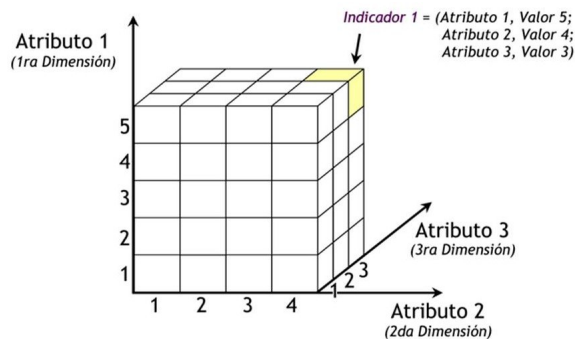


Figura 2.5: Ejemplo de un cubo multidimensional.

9. **Indicadores:** son resúmenes efectuados sobre algún hecho o expresiones, que serán incluidos en algún cubo multidimensional, con el fin de analizar los datos almacenados en el almacén de datos. El valor que estos adopten estará condicionado por los atributos/jerarquías que se utilicen para analizarlos.
10. **Atributos:** constituyen los criterios de análisis que se utilizarán para analizar los indicadores dentro de un cubo multidimensional. Los mismos se basan, en su gran mayoría, en los campos de las tablas de dimensiones y/o expresiones.
11. **Jerarquías:** representan una relación lógica entre dos o más atributos pertenecientes a un cubo multidimensional; la principal ventaja de manejar jerarquías, reside en poder analizar los datos desde su nivel más general al más detallado y viceversa, al desplazarse por los diferentes niveles.

Las jerarquías poseen las siguientes características:

- Pueden existir varias en un mismo cubo.
- Están compuestas por dos o más niveles.
- Se tiene una relación 1-n o padre-hijo entre atributos consecutivos de un nivel superior y uno inferior.

Existen tres arquitecturas del diseño de sistemas de almacenes de datos:[8]

- **Esquema de estrella (Star Schema):** es la arquitectura de almacén de datos más simple. En este diseño del almacén de datos la tabla de variables está rodeada por Dimensiones y juntos forman una estructura que permite implementar mecanismos básicos para poder utilizarla con una herramienta de consultas OLAP. Lo característico de la arquitectura de estrella es que sólo existe una tabla de dimensiones para cada dimensión y esta tabla representa la segunda forma normal.

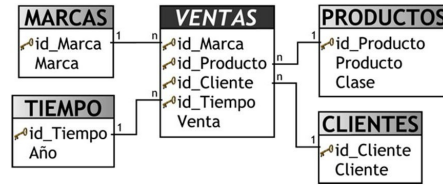


Figura 2.6: Ejemplo de la estructura estrella.

- **Esquema copo de nieve (Snowflaking):** Lo que distingue a la arquitectura en copo de nieve del esquema estrella, es que las tablas de dimensiones en este modelo representan relaciones normalizadas y forman parte de un modelo relacional de base de datos.

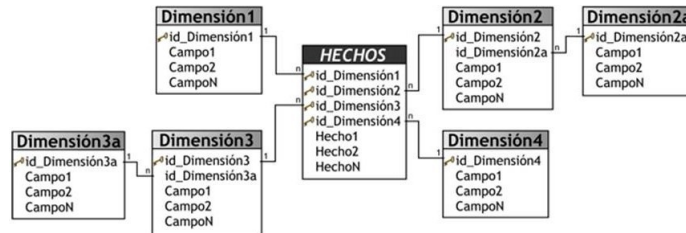


Figura 2.7: Ejemplo de la estructura copo de nieve.

- **Esquema constelación de hechos:** En este esquema las distintas tablas de los hechos están asignadas a las dimensiones relevantes para cada uno de los hechos. Esto puede ser útil cuando los hechos están asignados a un nivel de una dimensión y los otros hechos a otro nivel de detalle de una dimensión.

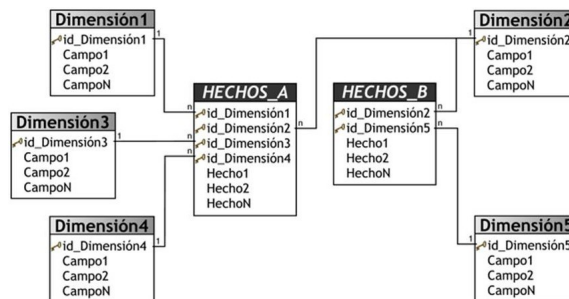


Figura 2.8: Ejemplo estructura constelación de hechos.

- **Ad-hoc querying:** Permite al usuario personalizar una consulta en tiempo real, en vez de estar atado a las consultas prediseñadas para informes. Generalmente las consultas

ad hoc permiten a los usuarios con poca experiencia en SQL tener el mismo acceso a la información de la base de datos. Para esto los sistemas deben poseer una Interface gráfica de usuario (GUI, por sus siglas en inglés Graphical User Interface) para generarlas.

- **Inteligencia de negocios (Business Intelligence BI):** Es el proceso de convertir datos en conocimiento y el conocimiento en acción, para la toma de decisiones. Integra por un lado el almacenamiento y por el otro el procesamiento de grandes cantidades de datos, con el principal objetivo de transformarlos en conocimiento y en decisiones en tiempo real, a través de un sencillo análisis y exploración. [12] La Inteligencia de negocios consta de cinco fases las cuales se explican en la siguiente ilustración:



**Figura 2.9:** Fases de la Inteligencia de Negocios BI.

Entre los beneficios más importantes que BI proporciona a las organizaciones, vale la pena destacar los siguientes:[3]

1. Reduce el tiempo mínimo que se requiere para recoger toda la información relevante de un tema en particular, ya que la misma se encontrará integrada en una fuente única de fácil acceso.
2. Automatiza la asimilación de la información, debido a que la extracción y carga de los datos necesarios se realizará a través de procesos predefinidos.
3. Proporciona herramientas de análisis para establecer comparaciones y tomar decisiones.
4. Cierra el círculo que hace pasar de la decisión a la acción.

5. Permite a los usuarios no depender de reportes o informes programados, porque los mismos serán generados de manera dinámica.
6. Posibilita la formulación y respuesta de preguntas que son claves para el desempeño de la organización.
7. Los usuarios podrán consultar y analizar los datos de manera sencilla e intuitiva.

Los usuarios de los sistemas operacionales son los que hacen que la empresa camine o avance hacia sus objetivos. Ellos realizan ventas, surten órdenes, buscan nuevos clientes, atienden quejas y realizan entregas. Técnicamente la mayoría de las veces estos usuarios trabajan con un solo registro y repetidamente realizan la misma tarea una y otra vez.

Sin embargo, los usuarios de almacenes de datos, observan como camina y avanza la empresa. Por ejemplo, éstos se fijan en el número de órdenes y los comparan con los de la semana pasada, se preguntan el porqué del resultado y se fijan de qué se quejaron los clientes, en teoría los usuarios nunca trabajan con un sólo registro a la vez. Más bien, sus preguntas requieren de la búsqueda de cientos o miles de registros en conjunto para obtener una respuesta. Para complicar aún más las cosas, estos usuarios continuamente cambian sus preguntas.

Comúnmente algunos usuarios conciben un almacén de datos como una réplica de la base de datos del sistema operacional pero almacenado en otro servidor. Si bien es cierto, esto ayuda al desempeño de los sistemas, pero no satisface la necesidad de información de los tomadores de decisiones. Las organizaciones que tienen almacenes de datos de este tipo, no entienden realmente que las necesidades de información de los tomadores de decisiones son totalmente distintas a las necesidades del usuario de sistemas operacionales.

Inicialmente las metas de un almacén de datos están enfocadas a resolver algunos de los siguientes problemas:[9]

1. Se tiene información dispersa, en grandes cantidades y no podemos acceder a ella.
2. No se puede analizar la información de distintas maneras.

3. El poco acceso a la información que existe, no es suficiente o es muy complicado obtenerlo.
4. Se obtiene mucha información pero un porcentaje muy bajo de ella es útil para el usuario.
5. Los usuarios llegan a las juntas con las mismas métricas pero con diferentes resultados.
6. La información que se obtiene no es suficiente para la toma de decisiones.

Para resolver estos problemas es necesario que el almacén de datos cumpla con ciertos requisitos:[8]

- **El almacén de datos debe facilitar el acceso a la información.** Los datos del almacén deben ser entendibles, intuitivos y obvios para el usuario de negocio y no para el usuario de tecnologías de información. Entendible implica legible; los contenidos de los datos deben ser etiquetados con palabras que signifiquen o indiquen algo. El almacén debe permitir separar y/o combinar los datos de múltiples maneras (comúnmente conocido como **slicing and dicing**) para facilitar su análisis. Las herramientas para obtener información del almacén de datos deben ser simples y fáciles de usar. También deben arrojar resultados de las consulta con tiempos de espera mínimos.
- **El almacén de datos debe presentar la información de la organización de manera consistente.** Los datos en el almacén deben ser creíbles. Los datos deben ser cuidadosamente recopilados y organizados a partir de una variedad de fuentes o sistemas en la organización. Los datos solo pueden estar disponibles después de haber sido limpiados y su calidad ha sido validada para el consumo de los usuarios. La información de un proceso de negocio debe empatar con la información en el almacén de datos. Información consistente significa información de calidad. Si dos métricos de desempeño tienen el mismo nombre entonces deben significar lo mismo de lo contrario se les debe

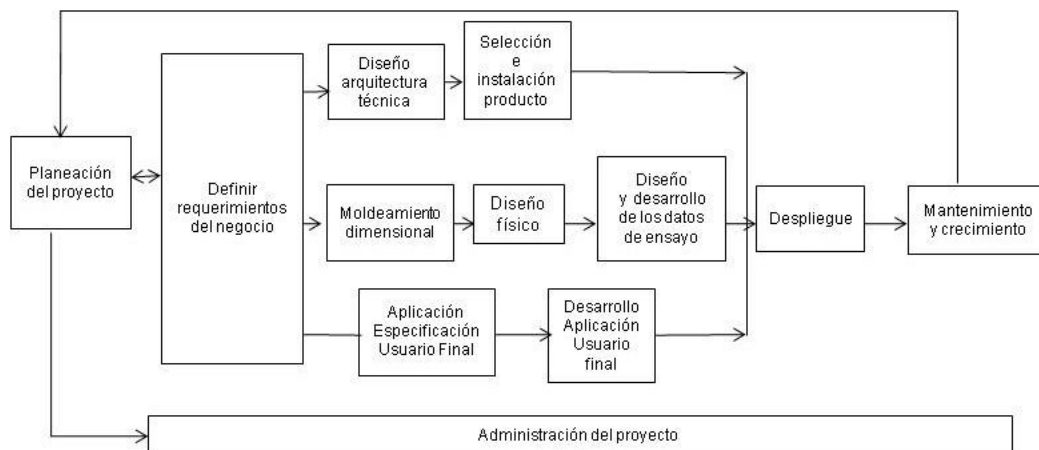
otorgar un nombre distinto. La consistencia implica que las definiciones comunes para el contenido del almacén de datos están disponibles también para los usuarios.

- **El almacén de datos debe ser capaz de adaptarse a cambios.** Los cambios son inevitables. Las necesidades del usuario, las condiciones del negocio, los datos y la tecnología son factores que cambian con el transcurso del tiempo. El almacén de datos debe ser capaz de adaptarse a estos cambios inevitables. Los datos existentes y las aplicaciones no deben ser cambiados o interrumpidos cuando la comunidad empresarial plantea preguntas nuevas o nuevos datos se agrega al almacén.
- **El almacén de datos debe servir como base para la mejor toma de decisiones.** El almacén debe tener la información correcta para apoyar la toma de decisiones. Solo hay una verdadera respuesta del almacén de datos: Es o son las decisiones tomadas a partir de la evidencia que presento el almacén de datos. Finalmente lo que se busca obtener es un sistema de soporte en la toma de decisiones.
- **El almacén de datos debe ser seguro y proteger nuestros activos de información.** Información y datos relevantes en el almacén de datos pueden ser potencialmente dañinos si caen en manos equivocadas. El acceso a la información confidencial de la organización debe ser controlado efectivamente por el almacén de datos.[9]

# Capítulo 3

## Metodología

La metodología a seguir es la del ciclo de vida de almacenes de datos según *Ralph Kimball*. Esta se enfoca principalmente en el diseño de la base de datos que almacenará la información para la toma de decisiones. El diseño se basa en la creación de tablas de hechos (**facts**) que son tablas que contienen la información numérica de los indicadores a analizar, es decir la parte cuantitativa de la información.



**Figura 3.1:** Diagrama del ciclo de vida de Ralph Kimball.

### 3.1. Obtener requerimientos de informacion.

Reconocer los principales casos de uso que involucran a los objetos de estudio. En esta etapa se determinarán de manera muy general los requisitos deseados para la explotación y manejo del almacén de datos. Para lograrlo será necesario llevar a cabo pláticas y entrevistas con el personal, para conocer las maneras en las que actualmente obtienen la información que necesitan, así como, de donde y de qué manera la presentan para su análisis o entendimiento.

### 3.2. Definición de requerimientos técnicos

Determinar características del hardware y software necesarios para llevar a cabo el proyecto. Dichos aspectos o capacidades comprenden poder de procesamiento y capacidad de almacenamiento, donde se cuidará la disponibilidad de los datos mediante redundancia. En el caso del software, se revisarán distintas alternativas existentes en el mercado con el fin de evaluar y seleccionar la mejor opción para el proyecto. Se cuidarán los costos en licencias para garantizar la adquisición de un software de calidad. Es necesario estudiar a fondo toda la información almacenada referente a instituciones, infraestructura, cuerpos académicos, investigadores, investigaciones, líneas de investigación, fuentes de financiamiento, apoyos, reconocimientos, patentes, áreas de conocimiento, producción editorial y de software.

### 3.3. Diseño de prototipos multidimensionales

Se diseñarán los prototipos necesarios cuidando al máximo la similitud al diseño final deseado, el propósito es detectar anomalías que pudiesen impactar la efectividad, desempeño y resultado final del proyecto. El modelo se debe revisar con otros expertos en el área, con el fin de obtener un mejor entendimiento de los alcances del diseño y determinar si es necesario

hacer cambios. Para el diseño se evaluará la mejor opción entre un modelo de tipo estrella y uno copo de nieve. En la reunión con los expertos se tocarán temas como la efectividad de los distintos modelos y su impacto en el proyecto. .

### **3.4. Diseño Físico**

Se generará un ambiente de prueba para codificar el prototipo seleccionado. El modelo dimensional será codificado y hecho realidad en el software que se eligió. Es posible que el diseño original sufra algunos cambios sin embargo, no deben desviarse mucho del diseño original. El diseño físico será evaluado por los expertos para cuidar que se cumpla lo propuesto en el prototipo.

### **3.5. Diseño y desarrollo del proceso de carga de datos al modelo**

Es necesario extraer, transformar y cargar los datos al diseño seleccionado. Para ello se desarrollará un proceso con estas capacidades. Dicho proceso será probado manualmente hasta considerar que esta listo para ejecutarse en modo automático. Una vez listo, el proceso será instalado en el ambiente de prueba para su posterior ejecución calendarizada.

### **3.6. Diseño y/o configuración de herramientas para el análisis de información**

Se presentará la información mediante la selección de algunas herramientas especializadas disponibles en el software, entre ellas, Sql Server, excel y asp.net. Se tomarán en cuenta las

más adecuadas y amigables. Antes de elegir las, se presentarán a los usuarios finales para su evaluación.

### 3.7. Metodología aplicada al SIEDCyT

1. **Definición de requerimientos.** Durante el desarrollo del proyecto SIEDCyT se realizaron diversas reuniones para discutir las necesidades de información del cliente COCyT; en las cuales al inicio el mismo Jorge Ramos director del COCyT y otras personas a su cargo; expresaron de manera muy general cuáles eran sus expectativas referentes al sistema de información, y que reportes esperaban ver en un futuro.

Para reforzar la información de las entrevistas COCyT hizo entrega de un cuestionario, el cual fue utilizado para las entrevistas de los investigadores. Sin embargo como la etapa de análisis ya había concluido, por ello no fue posible incluirlo en su totalidad al SIEDCyT. Pero, la mayoría de las preguntas pueden ser contestadas haciendo consultas directas en la base de datos. Debido a que la información debe estar disponible para las personas que no tienen conocimientos técnicos. Se crearán reportes para que el usuario simplemente los consulte.

Los usuarios con más conocimiento en el manejo de sistemas de información tendrán la capacidad de crear sus propios reportes a partir de otros ya existentes (Ad -hoc).

Las preguntas que tienen respuesta mediante la base de datos son las siguientes:

- **Perfil del investigador:** Es necesario conocer el perfil del investigador, en que empresa está trabajando, sexo, rango de ingresos, nivel de SNI, grados académicos adquiridos, ciudad en que radica, investigaciones, si pertenecen al Programa de Mejoramiento del Profesorado (PROMEP), libros, artículos, patentes y grupo de investigación al que pertenece.

- **Investigaciones:** También es importante conocer datos específicos de las investigaciones que se han llevado a cabo en Baja California. Los datos comprenden: fechas de inicio y fin, equipo e infraestructura utilizada, líneas de investigación, reconocimientos, financiamientos y su tipo (privado, federal) y estatus (aprobado, en proceso, concluido).

La información ya se encuentra en la base de datos, pero no es eficiente ni fácil para los usuarios obtenerla. Por lo mismo, la información se moverá al almacén de datos para su fácil consumo y análisis mediante herramientas BI.

Los datos serán clasificados y conformados debido a que algunos pueden existir en varios reportes aunque signifiquen cosas totalmente distintas. Para clarificar lo anteriormente mencionado, la dimensión zona geográfica; cuando hablamos de investigadores significa lugar de nacimiento. Cuando hablamos de instituciones significa ubicación física de la misma. En el caso de grados académicos, significa lugar donde se obtuvo el grado. Estas perspectivas de ver la información hacen que los datos signifiquen cosas totalmente distintas aunque estemos haciendo referencia a la misma dimensión física.

Las dimensiones creadas en esta fase del proyecto, se harán tomando en cuenta el hecho de que pueden llegar a ser útiles para el desarrollo de una nueva solución, dejando así una plataforma escalable y eficiente para el cliente.

## **3.8. Diseño de arquitectura**

### **3.8.1. Grupo de trabajo de planeación e implementación de arquitectura.**

Para efectos de esta tesis, el equipo de trabajo se compone únicamente por el aspirante de grado. Pero en un ambiente de producción donde el alcance del proyecto es mucho mayor, es necesario involucrar a personas con diferentes roles para diseñar toda la arquitectura de la plataforma DW/BI.

### **3.8.2. Requerimientos técnicos basados en las necesidades de los usuarios**

Tomando en cuenta los requerimientos de información del cliente podemos determinar los siguientes:

1. Necesidad de un portal BI.
2. Herramientas para análisis.
3. Información histórica.
4. Tiempo de respuesta optimo por cada reporte.
5. Integración del DW/BI con otros sistemas.

Los requerimientos anteriores se explican en detalle a continuación:

### **Necesidad de un Portal BI**

El portal BI es una aplicación web, la cual estará disponible para los usuarios desde su navegador de internet. Cumplirá el propósito de entregar reportes preestablecidos, además de mostrar los últimos cambios relevantes en la información. Solo debe ser posible consultarlo a través de la red privada del cliente por cuestiones de seguridad.

### **Herramientas para análisis**

Diferentes usuarios analizarán los datos para elaborar documentos o enviar información seleccionada a la gente que lo requiera. Es necesario que ésta pueda ser fácilmente exportada hacia otras aplicaciones como puede ser Microsoft Excel. Actualmente COCyT maneja google docs, la cual es muy similar a Microsoft office pero con la diferencia de que es en un ambiente web. Para lograr aun más la integración con sus procesos actuales, es necesario proveer una manera eficaz para importar información a las herramientas que actualmente utilizan.

Con las herramientas integradas de Microsoft Excel, es posible analizar de manera más familiar los datos mediante gráficas y personalización del formato. De esta forma el usuario podrá explotar la información sin tener las limitantes que pudiese imponer alguna herramienta preestablecida.

Algunos usuarios necesitarán imprimir reportes para incluirlos en otros documentos y posiblemente los formatos disponibles no se adapten a sus necesidades. Para eliminar este obstáculo es necesario proveerle al usuario una manera de personalizar reportes para imprimir.

### **Información histórica**

Debido a que es importante conocer cómo se va desarrollando la ciencia y tecnología en Baja California a través del tiempo. La información histórica es indispensable. Esto tiene implicaciones técnicas que hay que cuidar. Como se ha comentado anteriormente, en el SIEDCyT

aun no se ha cargado toda la información recopilada por los cuestionarios. Como consecuencia es difícil hacer una estimación exacta sobre las capacidades de almacenamiento necesarias. Otro factor que influye en este problema es el hecho de que nunca se ha desarrollado un sistema para este propósito y con estas características en Baja California.

### Tiempo de respuesta óptimo por cada reporte

Un factor crítico de éxito es el que, la información a analizar se obtenga en no más de cinco segundos por reporte solicitado. Además el desempeño del equipo no debe verse afectado al presentarse solicitudes de reportes. Porque si esto se presentara, no podría haber varios usuarios concurrentes en la plataforma de DW/BI.

Esto presenta algunos retos técnicos, para poder lograrlo es necesario realizar lo que a continuación se presenta:

- **Crear un modelo de base de datos óptimo para análisis de información.** El crear una estructura adecuada para BI, optimiza de manera considerable el desempeño y velocidad de los reportes a entregar al usuario. Se discutirá más a detalle cuando se toque el tema de modelado dimensional.
- **Migrar la información a la base de datos optimizada.** La base de datos del SIEDCyT está en un centro de datos en Estados Unidos. Al estar a una distancia considerable de los usuarios, surgen varios aspectos que pudiesen afectar el tiempo de respuesta. Estos aspectos van desde una conexión lenta hasta la carga saturada de trabajo que pudiese llegar a tener el servidor. Es importante recalcar que el servidor no es dedicado al SIEDCyT. Está en un ambiente compartido con otras aplicaciones, lo cual no nos permite garantizar que los recursos como memoria, CPU y ancho de banda estarán disponibles para cubrir las necesidades del usuario en un tiempo aceptable. Por los problemas mencionados es necesario migrar toda la información a un servidor local en las oficinas de COCyT. No solo se eliminarían los problemas del uso de

recursos si no que también tenemos bajo nuestro control los recursos a emplear para la plataforma. Se describirá más a detalle esta cuestión cuando se toque el tema del proceso ETL.

### 3.8.3. Integración del DW/BI con otros sistemas

Esta característica esta fuera del alcance de este proyecto, pero es necesario mencionarlo porque se debe tomar en consideración que en el futuro podría ser indispensable. Conforme la tecnología avanza y surgen nuevos sistemas de información, es necesario que la plataforma sea capaz de adaptarse a los cambios provocando el menor nivel de problemas posible. Es importante que cumpla con los requerimientos técnicos para integrarse con casi cualquier otro sistema sin tener que recrearse.

## 3.9. Equipo e infraestructura

Una vez identificados todos los requerimientos técnicos para llevar a cabo el proyecto, es necesario revisar que infraestructura se va a utilizar. Siendo el caso de esta tesis, únicamente vamos a proponer la infraestructura pero no vamos a adquirir equipo. Para poner en marcha una plataforma de pruebas no es necesario contar con un equipo tan robusto. Sin embargo se va a proponer el hardware adecuado para el cliente.

Como buena práctica se recomienda separar el sistema DW/BI en diferentes servidores. Esto ayuda a balancear la carga de trabajo además de facilitar el mantenimiento de las aplicaciones.

Para este proyecto se propone adquirir el siguiente equipo:

- 1 Servidor para proceso ETL.
- 1 Servidor para Manejador de base de datos.
- 1 Servidor para herramientas de presentación.

- 1 Dispositivo para almacenaje masivo de red también conocido como SAN por sus siglas en inglés (Storage Area Network).
- 1 Dispositivo de respaldo ya sea de discos o cassetes.

Las características específicas de cada servidor pueden variar. Debido a que no se cuenta con presupuesto para ellos omitiremos las especificaciones detalladas del equipo.

### 3.10. Selección del producto

Siendo que el SIEDCyT se desarrollo en la plataforma de Microsoft y si además se considera que el cliente COCYT tiene otros sistemas en esta plataforma. Es recomendable utilizar tecnologías Microsoft para el almacén de datos. De esta manera se facilita la integración y se ahorra tiempo al no tener que realizar tareas adicionales para lograr una integración óptima entre aplicaciones.

El paquete BI de Microsoft contiene todos los componentes necesarios para llevar a cabo cualquier proyecto de DW/BI y los principales son los mencionados a continuación:

- El motor relacional SQL Server (**RDBMS**) para administrar y almacenar la base de datos de almacenamiento de datos dimensional.
- **Integration Services** para crear la extracción, transformación y carga (ETL) del sistema.
- Una base de datos OLAP en **Analysis Services** para admitir consultas de los usuarios, sobre todo para uso ad hoc.
- **Reporting Services** para elaborar informes predefinidos.
- Las herramientas de desarrollo y gestión, especialmente de **SQL Server BI Development Studio** y **SQL Server Management Studio**, para crear y gestionar el sistema

DW/BI.

Este grupo de componentes van enfocados al equipo de desarrollo, pero Microsoft tiene un segundo conjunto importante de herramientas y están diseñadas para el usuario empresarial. Estos incluyen Microsoft Office, Excel especialmente, Office Web Components, un analizador de datos, y servicios de SharePoint.

Como podemos apreciar; Microsoft tiene un conjunto de aplicaciones muy completo. La principal ventaja de esta plataforma, es que las herramientas son muy similares en interface y modo de uso. Esto permite al usuario final familiarizarse más rápidamente con los nuevos sistemas y reducir de manera considerable la curva de aprendizaje.

Otras herramientas ofrecen o intentan ofrecer numerables beneficios en sus productos pero para este proyecto; además de las razones ya mencionadas, se decidió evaluar más a fondo Microsoft SQL Server por los beneficios que ofrece y son los siguientes:

- **Integración:** Desde el sistema operativo, motores de base de datos, entorno de desarrollo y hasta aplicaciones de escritorio. Se puede crear un sistema completo DW/BI utilizando sólo software de Microsoft y tener la garantía de que todos los componentes van a funcionar juntos de manera efectiva.
- **Bajo costo de licenciamiento:** El precio de los productos Microsoft es mucho menor a otros de marcas como Oracle, IBM y SAP que ofrecen bondades similares. Otro factor a considerar es el precio por el soporte el cual es mucho mayor en cualquiera de las marcas ya mencionadas.
- **Compatibilidad con otros sistemas:** Como ya se mencionó, se puede construir un sistema DW/BI completo utilizando únicamente tecnologías Microsoft. Pero no necesariamente es así, se puede tomar algún componente y reemplazarlo por cualquier otro de un proveedor distinto. Al inicio de la definición de los requerimientos se determino

que era necesario que la plataforma seleccionada tuviera la capacidad de integrarse con otros sistemas y Microsoft lo permite.

- **Alto desempeño y escalabilidad:** Debido a la evolución de los sistemas de información, cada vez hay más demanda en volúmenes de información. Microsoft SQL Server tiene la capacidad de adaptarse a cualquier volumen de información sin sacrificar desempeño.
- **La inversión de Microsoft en BI:** Microsoft está altamente comprometido con la creación de herramientas que permitan crear excelentes aplicaciones de inteligencia de negocios. Microsoft asegura que estará en el negocio por mucho más tiempo y seguirá mejorando sus productos con cada nueva versión que salga al mercado.

Tomando en cuenta las necesidades del cliente y las bondades del producto evaluado se tomo la decisión de implementar el sistema DW/BI con la tecnología de Microsoft SQL Server por la integración que promete con los sistemas y aplicaciones existentes.

### 3.11. Modelado Dimensional

En la industria de almacenes de datos e inteligencia de negocios, el modelado dimensional es el más recomendado para presentar la información a los usuarios. Este modelo ayuda a cumplir los siguientes objetivos:

1. Presentar la información a los usuarios de la manera más sencilla posible.
2. Las consultas se resuelven de manera rápida y efectiva.
3. Provee información relevante que describe acertadamente los procesos del negocio.

De acuerdo a los requerimientos se identificaron las siguientes medidas o hechos:

1. **Grados académicos adquiridos por investigador**
2. **Patentes registradas**
3. **Investigaciones realizadas**
4. **Vinculación entre investigadores y grupos de investigación**
5. **Artículos generados por investigadores**
6. **Participación de investigadores en la generación de libros**

Los hechos sin dimensiones no informan o indican mucho. Es necesario apoyarse de las dimensiones para conocer el que, cuando, donde, quien y como. En un almacén de datos las dimensiones pueden ser utilizadas o compartidas por diferentes hechos y para ello no es necesario crear duplicados.

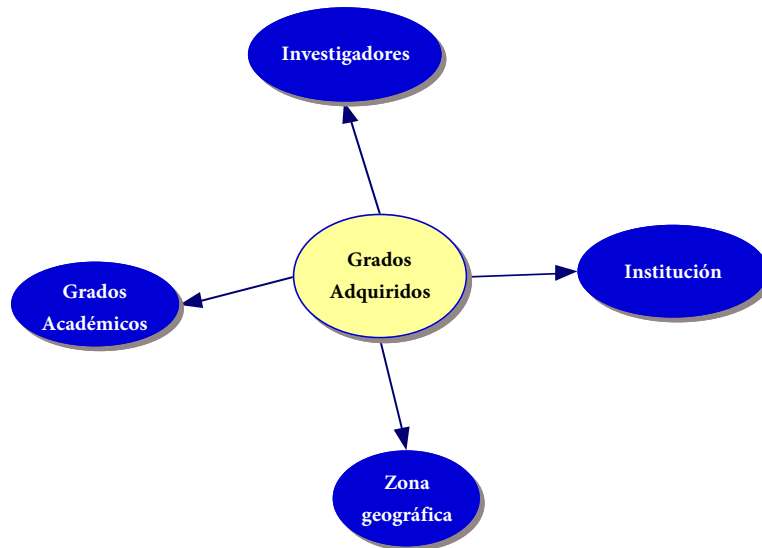
De acuerdo a los hechos identificados, se determino el uso de las siguientes dimensiones:

1. **Investigaciones**
2. **Institución**
3. **Zona geográfica**
4. **Grados Académicos**
5. **Patentes**
6. **Líneas de investigación**
7. **Financiamientos**
8. **Libros**
9. **Grupos de investigación**
10. **Artículos**

11. Participación en libros

12. Investigador

A continuación se presentan los diseños dimensionales de alto nivel y se clarifica el uso de cada dimensión con respecto a los hechos.



**Figura 3.2:** Grados académicos adquiridos por investigador

Este modelo puede resolver un gran número de preguntas pero se destacan las más importantes a continuación:

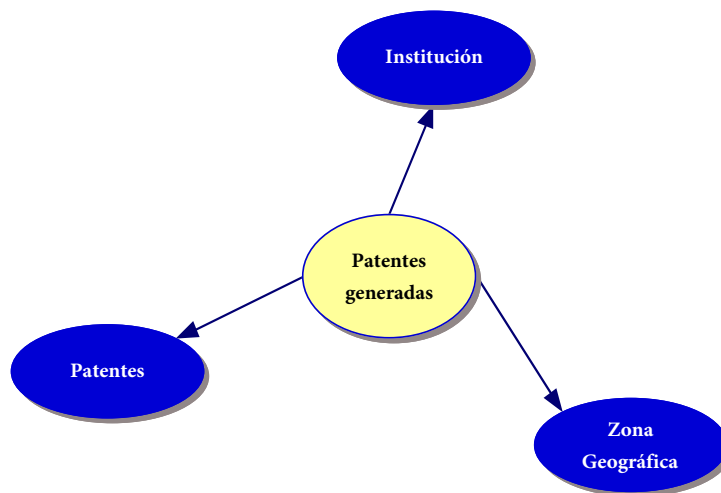
- Cantidad de grados académicos adquiridos en baja california
- Ciudad en donde fueron adquiridos
- Institución que lo emitió
- Investigador que lo adquirió
- Cantidad de licenciaturas, maestrías y doctorados
- Grado académico más y menos adquirido

El modelo de patentes presentado en la figura 3.3 resuelve pero no está limitado a las siguientes preguntas:

- Patentes registradas por institución

- Ciudad en la que fueron registradas
- Año en que se registro
- Áreas de conocimiento que impacto

Las áreas de conocimiento no fueron modeladas como una dimensión independiente, esta se incluyo como parte de los atributos de cada entidad donde se relaciona. Para este caso en específico las áreas y sub áreas de conocimiento forman parte de la dimensión patente.



**Figura 3.3:** Patentes registradas

El modelo de investigaciones realizadas presentado en la figura 3.4 es el más completo. El tener el mayor número de dimensiones nos permite obtener respuestas muy interesantes y no únicamente referente a investigaciones. Las preguntas destacadas que puede contestar este modelo son las siguientes:

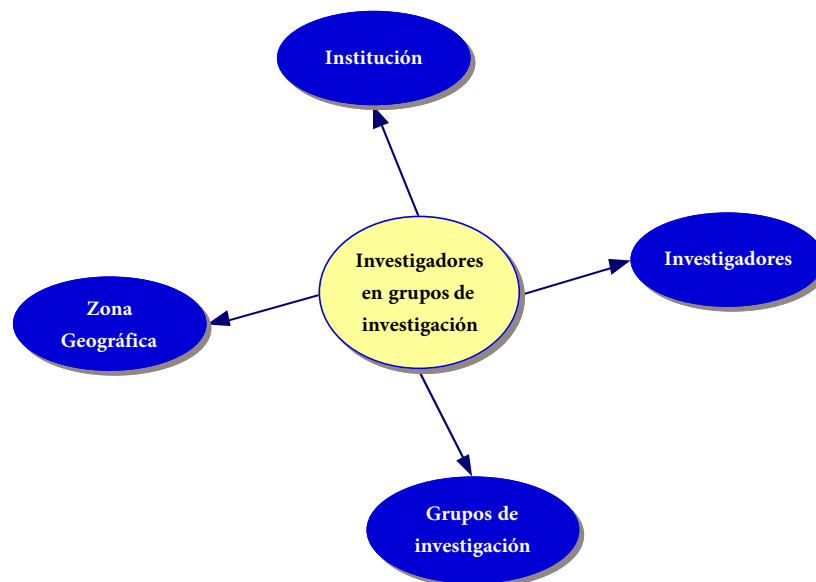
- Qué investigadores y de que ciudades están involucrados en investigaciones específicas
- Cantidad de dinero invertido en investigaciones desglosado por ciudad, líneas de investigación, investigadores y tipo de financiamiento
- En que ciudades hay mas desarrollo de investigaciones

- Cuales líneas de investigación son las más estudiadas y determinar cuales hay que reforzar



**Figura 3.4:** Investigaciones realizadas

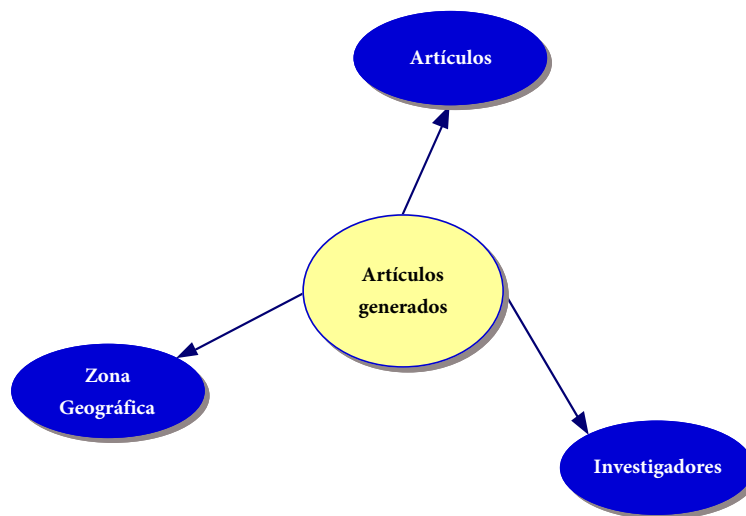
Los hechos de investigadores vinculados a grupos de investigación presentados en la figura 3.5, nos ayudara a conocer de manera muy general que investigadores pertenecen a un grupo de investigación, en que institución se encuentra y en qué ciudad.



**Figura 3.5:** Vinculación entre investigadores y grupos de investigación

Al igual que los modelos anteriores, el hecho de artículos generados presentado en la figura 3.6 cuenta con sus propias dimensiones las cuales nos ayudaran a conocer y sin estar limitado a lo siguiente:

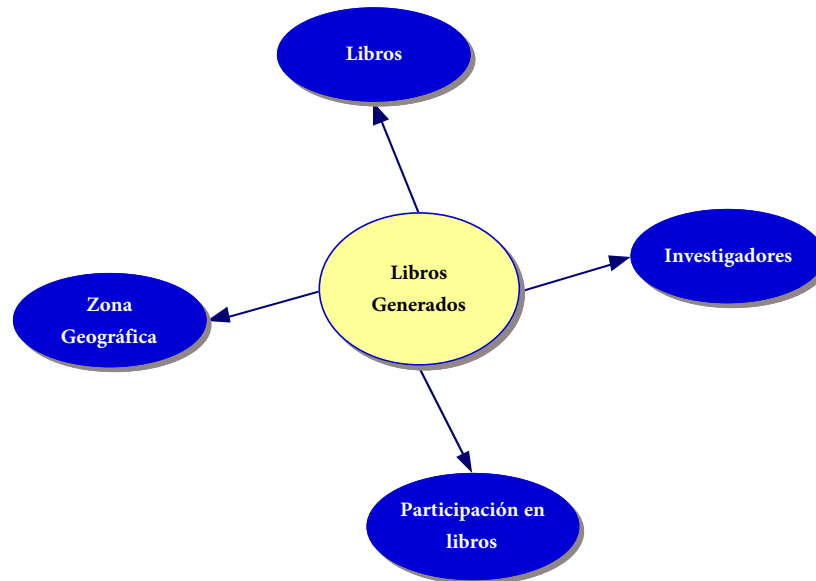
- **Investigadores que producen artículos**
- **Ciudad en donde son publicados**
- **Nivel de SNI que produce más artículos**



**Figura 3.6:** Artículos generados por investigadores

El modelo de libros generados representado en la figura 3.7 nos permitirá conocer a detalle el nivel de involucramiento de los investigadores con referente a la creación de libros. No limitado a ello, también nos ayudara a determinar en qué ciudades se producen más libros.

Como se puede observar estos modelos cumplirán con la demanda de información establecida en los requerimientos. Sin embargo no estará limitado a las preguntas previamente mencionadas. Cada nuevo usuario, con su nueva necesidad de información, podrá hacer nuevos cuestionamientos y el modelo será capaz de responder a la mayoría de ellos, siempre y cuando los planteamientos vayan enfocados al tema o modelo de estudio. Esto es posible debido a que



**Figura 3.7:** Libros generados por investigadores

toda la información es previamente cargada y las herramientas de BI hacen fácil y posible la obtención de esta nueva información de las fuentes establecidas en el almacén de datos.

Para que estos modelos funcionen de acuerdo a los requerimientos establecidos, es necesario definir el nivel de detalle con el que almacenaran la información. Para ellos se define el nivel de granularidad por cada hecho relacionado con sus dimensiones. En la figura 3.8 se establece el nivel de granularidad en función de los requerimientos establecidos.

Una vez establecidas las dimensiones a utilizar con el nivel de granularidad, se puede proceder a crear a detalle cada uno de los hechos y dimensiones que posteriormente se implementaran en el manejador de base de datos. Estos modelos deben seguir los lineamientos establecidos en la figura 3.8 para que puedan cumplir con los requerimientos de información. Esta descripción de entidades deber ir lo más detallada posible para no dejar margen de error. Si no se realizar de tal manera, una vez implementado el modelo puede fracasar al no cumplir con lo establecido en los requerimientos. En el peor de los casos se puede incurrir en re trabajo del sistema ETL y por consecuencia el proyecto se retrasara. La documentación del detalle de cada entidad no va en función del manejador de base de datos que se utilizará para

el proyecto. Únicamente es una guía para el modelado físico del almacén de datos.

Entidades Hechos	Granularidad	Investigadores	Institución	Zona Geográfica	Grados Académicos	Patentes	Investigaciones	Líneas Investigación	Financiamientos	Libros	Grupos de investigación	Artículos	Participación en libros
Grados Adquiridos	1 registro por cada grado adquirido	X	X	X	X								
Patentes generadas	1 registro por cada patente generada		X	X		X							
Investigaciones realizadas	1 registro por cada investigación realizada.	X	X	X			X	X	X				
Investigadores en Grupos de investigación	1 registro por cada investigador vinculado	X	X	X							X		
Artículos generados	1 registro por cada participación en artículo	X		X								X	
Libros Generados	1 registro por cada participación de investigador	X		X						X			X

**Figura 3.8:** Bus matrix: Referencia de granularidad y uso de dimensiones por cada hecho

Este tipo de tabla donde se identifica el uso de dimensiones con los hechos, también es conocido como *bus matrix* en la metodología de Ralph Kimball. Brinda un nivel importante sobre el detalle en que se modelara y almacenara la información. Sin este documento, el modelado físico no tiene una guía exacta que le permita seguir especificaciones para crear entidades, relaciones y restricciones. Dichos elementos son necesarios en el almacén de datos para cumplir con los requerimientos planteados al inicio del proyecto. De igual forma los documentos sobre el uso de las herramientas BI se ven afectados al no tener claro el nivel de detalle de la información. Cada usuario puede suponer o pensar distinto de acuerdo a lo que su criterio le indique. Siendo esto último una falla inaceptable del almacén de datos, debido a que no se cumplirá el objetivo de tener una sola versión de la verdad. Porque cada quien interpretaría los resultados a su manera y esta es una cuestión que se busca eliminar para evitar la toma de decisiones basada en suposiciones y no hechos reales.

A continuación se presenta el detalle de las tablas de hechos. Tanto para tablas de dimensiones y hechos se utilizan los siguientes sufijos en los campos:

- **SK:** Llave subrogada.
- **PK:** Llave primaria.
- **FK:** Llave foránea.

Los elementos restantes están definidos en la descripción de las tablas.

<b>Tabla:</b>	FACT_GRADOS_ADQUIRIDOS
<b>Tipo de tabla:</b>	Hechos (Facts)
<b>Alias:</b>	Grados adquiridos
<b>Descripción:</b>	Cuantos grados académicos hay por investigador y cuando fueron adquiridos

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
<b>GradoSK</b>	Llave primaria subrogada		1,2,3,4
<b>GradoFK</b>	Llave primaria del grado académico	Clave de grado	1,2,3,4
<b>InvestigadorFK</b>	Llave primaria de investigador	Clave de investigador	1,2,3,4
<b>AnioInicio</b>	Fecha de inicio del grado	Año de inicio	1995, 1999, 2000
<b>AnioTermino</b>	Fecha en que adquirió el grado	Año terminación	1995, 1999, 2001

**Figura 3.9:** Referencia de granularidad y uso de dimensiones por cada hecho

<b>Tabla:</b>	FACT_PATENTES_REGISTRADAS
<b>Tipo de tabla:</b>	Hechos (Facts)
<b>Alias:</b>	Patentes registradas
<b>Descripción:</b>	Total de patentes registradas

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
PatenteSK	Llave primaria subrogada		1,2,3,4
PatenteFK	Llave foránea de patente	Identificador de patente	1,2,3,4
InstitucionFK	Llave foránea de Institución	Identificador de institución	1,2,3,4

**Figura 3.10:** Patentes registradas

<b>Tabla:</b>	FACT_INVESTIGACIONES
<b>Tipo de tabla:</b>	Hechos (Facts)
<b>Alias:</b>	Investigaciones generadas
<b>Descripción:</b>	Investigaciones que se llevaron o se están llevando a cabo

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
InvestigacionSK	Llave primaria subrogada		1,2,3,4
InvestigacionFK	Llave foránea de investigación	Identificador de investigación	1,2,3,4
Monto	Monto total del financiamiento recibido	Financiamiento	200,000, 350,000

**Figura 3.11:** Investigaciones realizadas

<b>Tabla:</b>	FACT_GRUPOS_INVESTIGACION_VINCULACION
<b>Tipo de tabla:</b>	Hechos (Facts)
<b>Alias:</b>	Vinculación en grupos de investigación
<b>Descripción:</b>	Vinculación entre investigadores y grupos de investigación

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
GrupoSK	Llave primaria subrogada		1,2,3,4
GrupoFK	Llave foránea de grupo de investigación	Identificador de grupo de investigación	1,2,3,4
InvestigadorFK	Llave primaria de investigador	Clave de investigador	1,2,3,4

**Figura 3.12:** Vinculación entre investigadores y grupos de investigación

<b>Tabla:</b>	FACT_PARTICIPACION_ARTICULOS
<b>Tipo de tabla:</b>	Hechos (Facts)
<b>Alias:</b>	Participaciones en artículos
<b>Descripción:</b>	Total de participaciones en artículos

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
ArticuloSK	Llave primaria subrogada		1,2,3,4
ArticuloFK	Llave foránea de articulo	Identificador de articulo	1,2,3,4
InvestigadorFK	Llave foránea de investigador	Identificador de investigador	1,2,3,4

**Figura 3.13:** Artículos generados por investigadores

<b>Tabla:</b>	FACT_LIBROS_PARTICIPACIONES
<b>Tipo de tabla:</b>	Hechos (Facts)
<b>Alias:</b>	Participaciones en libros
<b>Descripción:</b>	Total y tipos de participación en libros.

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
LibroSK	Llave primaria subrogada		1,2,3,4
LibroFK	Llave foránea del libro	Identificador de libro	1,2,3,4
ParticipacionrFK	Llave foránea del tipo de participación	Identificador de tipo de participación	1,2,3,4
InvestigadorFK	Llave foránea de investigador	Identificador de investigador	1,2,3,4
AnioPublicacion	Año en que se publico el libro	Año publicación	1995, 1999, 2001

**Figura 3.14:** Participación de investigadores en la generación de libros

Existen otro tipo de tablas que ayudan a relacionar los hechos con varios registros de las dimensiones. En el caso de las investigaciones, estas pudiesen llegar a tener dos o más financiamientos y/o dos o más líneas de investigación. Los investigadores pueden tener dos o más grados académicos. Para representar estas situaciones en un modelo dimensional, se utilizan tablas identificadas como tablas puente o de relación. En el modelo de alto nivel no es necesario especificar este tipo de tablas pero en el modelado a detalle es necesario documentarlas como a continuación se presenta.

<b>Tabla:</b>	FACT_REL_FINANCIAMIENTOS_INVESTIGACION
<b>Tipo de tabla:</b>	Puente o relación (Bridge)
<b>Alias:</b>	
<b>Descripción:</b>	Relaciona investigaciones con los financiamientos obtenidos

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
InvestigacionFinanciamientoSK	Llave primaria subrogada		1,2,3,4
InvestigacionFK	Llave foránea de Investigación	Identificador de investigación	1,2,3,4
FinanciamientoFK	Llave foránea de financiamiento	Identificador de financiamiento	1,2,3,4

**Figura 3.15:** Relaciona investigaciones con los financiamientos obtenidos

<b>Tabla:</b>	FACT_REL_INSTITUCIONES_INVESTIGACION
<b>Tipo de tabla:</b>	Puente o relación (Bridge)
<b>Alias:</b>	
<b>Descripción:</b>	Relaciona instituciones con investigaciones

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
InvInstSK	Llave primaria subrogada		1,2,3,4
InvestigacionFK	Llave foránea de Investigación	Identificador de investigación	1,2,3,4
InstitucionFK	Llave foránea de institución	Identificador de institución	1,2,3,4

**Figura 3.16:** Relaciona instituciones con investigaciones

<b>Tabla:</b>	FACT_REL_INVESTIGADORES_INVESTIGACION
<b>Tipo de tabla:</b>	Puente o relación (Bridge)
<b>Alias:</b>	
<b>Descripción:</b>	Relaciona investigadores con investigaciones

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
InvInvSK	Llave primaria subrogada		1,2,3,4
InvestigadorFK	Llave foránea de investigador	identificador de investigador	1,2,3,4
InvestigacionFK	Llave foránea de Investigación	Identificador de investigación	1,2,3,4

Figura 3.17: Relaciona investigadores con investigaciones

<b>Tabla:</b>	FACT_REL_LINEAS_INVESTIGACION
<b>Tipo de tabla:</b>	Puente o relación (Bridge)
<b>Alias:</b>	
<b>Descripción:</b>	Relaciona investigaciones con sus diferentes líneas de investigación

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
InvestigacionLineaSK	Llave primaria subrogada		1,2,3,4
InvestigacionFK	Llave foránea de Investigación	Identificador de investigación	1,2,3,4
LineaInvestigacionFK	Llave foránea de línea de investigación	Identificador de línea de investigación	1,2,3,4

Figura 3.18: Relaciona investigaciones con sus diferentes líneas de investigación

Para concluir con el modelado se presenta a continuación el detalle de las dimensiones a utilizar en el almacén de datos. Aunque la descripción de las tablas indica un significado específico, hay ocasiones donde depende del hecho al que está relacionada de ello obtiene su significado. Tomando como ejemplo el modelo de grados académicos adquiridos; la ciudad puede significar lugar donde se obtuvo el grado, pero tomando la misma tabla desde la perspectiva del investigador, esta significa lugar de origen. Este tipo de características se definen en las herramientas BI.

<b>Tabla:</b>	DIM_ARTICULOS
<b>Tipo de tabla:</b>	Dimensión
<b>Alias:</b>	Artículos
<b>Descripción:</b>	Catalogo de artículos

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
ArticuloSK	Llave primaria subrogada		1,2,3,4
ArticuloPK	Llave primaria de articulo	Identificador de articulo	1,2,3,4
Titulo	Nombre del articulo	Titulo de articulo	La ciencia mejora la calidad de vida humana
Url	Dirección de internet del articulo	Enlace web	<a href="http://www.cocytbc.mx/articulo-543.html">http://www.cocytbc.mx/articulo-543.html</a>
CiudadFK	Ciudad donde se genero	Identificador de la ciudad	1,2,3,4

**Figura 3.19:** Catalogo de artículos

<b>Tabla:</b>	DIM_GRADOS_ACADEMICOS
<b>Tipo de tabla:</b>	Dimensión
<b>Alias:</b>	Grados académicos
<b>Descripción:</b>	Detalle de grados académicos

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
<b>GradoSK</b>	Llave primaria subrogada		1,2,3,4
<b>GradoPK</b>	Llave primaria de grado	Grado identificador	1,2,3,4
<b>GradoNombre</b>	Nombre del grado académico	Grado	Informática
<b>GradoNivelPK</b>	Llave primaria del nivel	Identificador de nivel	1,2,3,4
<b>GradoNivel</b>	Descripción del nivel	Nivel académico	Licenciatura, Maestría, doctorado
<b>GradoInstitucionFK</b>	Llave foránea de institución que otorga el grado	Clave de institución	1,2,3,4

Figura 3.20: Detalle de grados académicos

<b>Tabla:</b>	DIM_GRUPOS_INVESTIGACION
<b>Tipo de tabla:</b>	Dimensión
<b>Alias:</b>	Grupos de investigación
<b>Descripción:</b>	Grupos de investigación registrados en el siedcyt

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
<b>GrupoSK</b>	Llave primaria subrogada		1,2,3,4
<b>GrupoPK</b>	Llave primaria del grupo	Identificador del grupo de investigación	1,2,3,4
<b>GrupoNombre</b>	Nombre del grupo de investigación	Nombre	Complejidad y computación
<b>InstitucionFK</b>	Llave foránea de institución donde esta registrado el grupo	Identificador de institución	1,2,3,4

Figura 3.21: Grupos de investigación registrados en el siedcyt

<b>Tabla:</b>	DIM_GRUPOS_INVESTIGACION
<b>Tipo de tabla:</b>	Dimensión
<b>Alias:</b>	Grupos de investigación
<b>Descripción:</b>	Grupos de investigación registrados en el siedcyt

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
<b>GrupoSK</b>	Llave primaria subrogada		1,2,3,4
<b>GrupoPK</b>	Llave primaria del grupo	Identificador del grupo de investigación	1,2,3,4
<b>GrupoNombre</b>	Nombre del grupo de investigación	Nombre	Complejidad y computación
<b>InstitucionFK</b>	Llave foránea de institución donde esta registrado el grupo	Identificador de institución	1,2,3,4

**Figura 3.22:** Grupos de investigación registrados en el siedcyt

<b>Tabla:</b>	DIM_INSTITUCION
<b>Tipo de tabla:</b>	Dimensión
<b>Alias:</b>	Instituciones
<b>Descripción:</b>	Detalle de instituciones

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
<b>InstitucionSK</b>	Llave primaria subrogada		1,2,3,4
<b>InstitucionPK</b>	Llave primaria de institución	Identificador de institución	1,2,3,4
<b>TipoID</b>	Clave de tipo	Identificador tipo	1,2,3,4
<b>TipoDescripcion</b>	Descripción del tipo de institución	Tipo de institución	Gubernamental
<b>SectorID</b>	Clave del sector	Identificador de sector	1,2,3,4
<b>SectorDescripcion</b>	Descripción del sector al que pertenece	Sector nombre	Privado, publico
<b>InstitucionNombre</b>	Nombre completo de institución	Nombre de institución	UABC campus Tijuana
<b>CiudadFK</b>	Llave foránea de ciudad	Clave de ciudad	1,2,3,6

**Figura 3.23:** Detalle de instituciones

<b>Tabla:</b>	DIM_INVESTIGACIONES
<b>Tipo de tabla:</b>	Dimensión
<b>Alias:</b>	Investigaciones
<b>Descripción:</b>	información de investigaciones

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
<b>InvestigacionSK</b>	Llave primaria subrogada		1,2,3,4
<b>InvestigacionPK</b>	Llave primaria de investigación	Identificador de investigación	1,2,3,4
<b>InvestigacionNombre</b>	Nombre o título de investigación	Título de investigación	Análisis Genético Funcional de una PDI
<b>CiudadFK</b>	Llave foránea de ciudad donde se realiza la investigación	identificador de ciudad	1,2,3,4
<b>AnioInicio</b>	Año en que se inicio la investigación	Fecha de Inicio	2000, 2001
<b>AnioTermino</b>	Año en que se termino la investigación	Fecha de terminación	2003, 2004

**Figura 3.24:** información de investigaciones

<b>Tabla:</b>	DIM_INVESTIGADOR
<b>Tipo de tabla:</b>	Dimensión
<b>Alias:</b>	Investigadores
<b>Descripción:</b>	Detalle de investigadores

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
<b>PersonalDataSK</b>	Llave primaria subrogada		1,2,3,4
<b>InvestigadorPK</b>	Llave primaria de investigador	Identificador de investigador	1,2,3,4
<b>PersonalPK</b>	Llave primaria de persona	Identificador de datos personales	1,2,3,4
<b>Nombre</b>	Nombre completo del investigador	Nombre completo	Ricardo Ibarra
<b>AnioNacimiento</b>	Año de nacimiento	Fecha de nacimiento	1979, 1985
<b>MesNacimiento</b>	Mes de nacimiento	Fecha de nacimiento	1, 2, 3, 4 .. 12
<b>DiaNacimiento</b>	Día de nacimiento	Fecha de nacimiento	1,2,3,4,...30,31
<b>Sexo</b>	Sexo del investigador		Masculino, Femenino
<b>SNIPK</b>	Identificador S N I	Clave S N I	1,2,3,7
<b>SNINivel</b>	Descripción del nivel S N I	Nivel S N I	Candidato, Nivel I
<b>PromepPK</b>	Identificador de promep	Clave promep	1,2,3,9
<b>PromepDescripcion</b>	Descripción de promep	Promep	No pertenece
<b>CiudadPK</b>	Llave foránea de ciudad	Clave de ciudad	1,2,3,4
<b>RangolIngresosPK</b>	Identificador de rango de ingresos	Clave rango de ingresos	1,2,3,4
<b>RangolIngresosDescripcion</b>	Descripción de rango de ingresos	Rango de ingresos	10,000 a 20,000

**Figura 3.25:** Detalle de investigadores

<b>Tabla:</b>	DIM_LIBROS
<b>Tipo de tabla:</b>	Dimensión
<b>Alias:</b>	Libros
<b>Descripción:</b>	Libros generados y capturados en el siedcyt

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
<b>LibroSK</b>	Llave primaria subrogada		1,2,3,4
<b>LibroPK</b>	Llave primaria del libro	Identificador del libro	1,2,3,4
<b>Titulo</b>	Nombre o titulo del libro	Nombre del libro	Empezando a programar
<b>SubAreaConocimientoPK</b>	Identificador del sub área de conocimiento	Sub área conocimiento identificador	1,2,3,4
<b>SubAreaConocimientoDescripcion</b>	Nombre de sub área de conocimiento	Sub área de conocimiento	Informática matemática
<b>AreaConocimientoPK</b>	identificador del área de conocimiento	Área de conocimiento identificador	1,2,3,4
<b>AreaConocimientoDescripcion</b>	Nombre del área de conocimiento	Area de conocimiento	Ciencias de la tecnología
<b>Editorial</b>	Nombre de la editorial del libro		Mc Graw Hill
<b>Edición</b>	Edición del libro		segunda edición
<b>Capitulos</b>	Cantidad de capítulos que contiene el libro	Numero de capítulos	12, 13, 14
<b>ISBNLocal</b>	ISBN del libro	ISBN	978-0-596-52735-8
<b>ISBNExtranjero</b>	ISBN del extranjero	ISBN Extranjero	0-596-52735-8_12

Figura 3.26: Libros generados y capturados en el siedcyt

<b>Tabla:</b>	DIM_LINEAS_INVESTIGACION
<b>Tipo de tabla:</b>	Dimensión
<b>Alias:</b>	Líneas de investigación
<b>Descripción:</b>	catalogo de líneas de investigación

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
<b>LineaSK</b>	Llave primaria subrogada		1,2,3,4
<b>LineaPK</b>	Llave primaria de línea de investigación	Identificador de línea de investigación	1,2,3,4
<b>LineaDelInvestigacion</b>	Nombre o titulo de línea de investigación	Línea de investigación	Ciencias computacionales

Figura 3.27: catalogo de líneas de investigación

<b>Tabla:</b>	DIM_PARTICIPACION_LIBROS
<b>Tipo de tabla:</b>	Dimensión
<b>Alias:</b>	Participación en libros
<b>Descripción:</b>	Tipo de participación que tuvo el investigador en los libros

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
<b>ParticipacionSK</b>	Llave primaria subrogada		1,2,3,4
<b>ParticipacionPK</b>	Llave primaria de tipo de participación	Identificador de participación	1,2,3,4
<b>ParticipacionDescripcion</b>	Descripción de participación	Participación	Autor, Coautor, Editor

Figura 3.28: Tipo de participación que tuvo el investigador en los libros

<b>Tabla:</b>	DIM_PATENTES
<b>Tipo de tabla:</b>	Dimensión
<b>Alias:</b>	Patentes
<b>Descripción:</b>	Patentes generadas y capturadas en el siedcyt.

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
<b>Patentes</b>	Llave primaria subrogada		1,2,3,4
<b>PatentePK</b>	Llave primaria de patente	Identificador de patente	1,2,3,4
<b>SubAreaConocimientoPK</b>	Identificador del sub área de conocimiento	Sub área conocimiento identificador	1,2,3,4
<b>SubAreaConocimientoDescripcion</b>	Nombre de sub área de conocimiento	Sub área de conocimiento	Informática matemática
<b>AreaConocimientoPK</b>	identificador del área de conocimiento	Área de conocimiento identificador	1,2,3,4
<b>AreaConocimientoDescripcion</b>	Nombre del área de conocimiento	Área de conocimiento	Ciencias de la tecnología
<b>Patente</b>	Nombre de patente		Sistema de Asistencia Microempresaria
<b>No_Registro</b>	Numero de registro de la patente	Numero de registro	544345555
<b>CiudadFK</b>	Ciudad en donde fue registrada	Identificador de Ciudad de registro	1,2,3,4

**Figura 3.29:** Patentes generadas y capturadas en el siedcyt

<b>Tabla:</b>	DIM_ZONA_GEOGRAFICA
<b>Tipo de tabla:</b>	Dimensión
<b>Alias:</b>	Zona Geográfica
<b>Descripción:</b>	Ubicación geográfica desde ciudad hasta continente

Nombre del Atributo	Descripción	Nombre Alternativo	Valores ejemplo
<b>ZonaPK</b>	Llave primaria subrogada		1,2,3,4
<b>CiudadPK</b>	Llave primaria de ciudad	Ciudad identificador	1,2,3,4
<b>CiudadDescripcion</b>	Nombre de la ciudad	Nombre de la ciudad	Tijuana, Ensenada
<b>EstadoPK</b>	Llave primaria de estado	Estado identificador	1,2,3,4
<b>EstadoDescripcion</b>	Nombre del estado	Estado nombre	Baja California
<b>PaisPK</b>	Llave primaria del país	País identificador	1,2,3,4
<b>PaisDescripcion</b>	Nombre del país	País nombre	México
<b>ContinentePK</b>	Llave primaria de continente	Continente identificador	1,2,3,4
<b>ContinenteDescripcion</b>	Nombre del continente	Continente nombre	América

**Figura 3.30:** Ubicación geográfica desde ciudad hasta continente

Todas las tablas anteriormente presentadas serán la base para el diseño físico. Como se menciono anteriormente la herramienta para el desarrollo del proyecto es SQL Server 2008 Enterprise Edition. Si bien es cierto el diseño físico debe apegarse totalmente al diseño conceptual, es posible que al crear el modelo en la plataforma seleccionada esta cambie un poco. Pero en funcionalidad debe proveer lo mismo que se definió en esta fase.

## 3.12. Diseño físico

En esta etapa es donde hacemos realidad el diseño creado en la fase anterior. Sin embargo, no se trata de únicamente transcribir las tablas al manejador de base de datos. Antes es necesario desarrollar estándares para los nombres de las entidades y atributos. Siendo que el modelo lógico presentado en la fase anterior ya cuenta con un estándar, se tratara de seguir con la misma línea. Esto ayudara a evitar cambios que pudiesen crear confusión. De lo contrario habría que documentar nombres lógicos y nombres físicos haciendo referencia a los documentos de la fase anterior para cada entidad y atributo.

### 3.12.1. Estandarización de nomenclatura

#### Tablas

Cada tabla creada tendrá un prefijo dependiendo del tipo que sea. A continuación se especifica cuáles y como se deben utilizar los prefijos:

- **DIM:** Tablas que son dimensión. *Ejemplo: DIM\_TABLA*
- **FACT:** Tablas que son hechos. *Ejemplo: FACT\_TABLA*
- **FACT\_REL:** Tablas que relacionan a unas con otras. *Ejemplo: FACT\_REL\_TABLA*
- **TMP:** Tablas temporales. *Ejemplo: TMP\_TABLA*

Para ninguna tabla está permitido el uso de espacios en blanco. Este carácter será reemplazado por un guion bajo. *Ejemplo: DIM\_LIBROS, FACT\_GRADOS\_ADQUIRIDOS*. Todos los nombres de las tablas serán en mayúsculas.

### Atributos

Todos los atributos deben iniciar cada palabra con mayúsculas y no deben tener espacio ni guiones. Debido a que el manejador de base de datos seleccionado no permite la letra ñ esta se sustituirá por la letra n y una vocal que haga sonar la palabra lo más parecido a como si se hubiese utilizado la ñ. Ejemplo para AñoInicio se debe utilizar AnioInicio. Ciertos sufijos serán utilizados para identificar rápidamente a los atributos, estos se definen a continuación:

- **SK:** Llave sub rogada
- **PK:** Llave primaria
- **FK:** Llave foránea

Todas las llaves sub rogadas serán de tipo numérico y auto incrementables.

### Valores nulos

En el almacén de datos no habrá campos nulos debido a que estos pueden crear conflicto al generar reportes. Tomando esto en cuenta, por cada valor nulo que exista en la base de datos, este se reemplaza por Ño especificado: En el caso de valores numéricos estos tomaran el valor de 0. Pero es necesario que todos los campos tengan por lo menos un valor y no dejarlos en blanco o nulos.

### Estándares de ubicación de archivos

Los archivos que se utilizan en el almacén de datos ya sea archivos de texto con bloques de código, bases de datos, documentación, etc. Deben contar con una ubicación estandarizada y no estar dispersos en el servidor. La estructura de archivos se creara como se dispone a continuación:

- **Bases de datos.** Ubicación default que utilice el manejador de base de datos para este tipo de archivos, la mayoría de las veces es:

C:\Program Files \Microsoft SQL Server \MSSQL10.INSTANCIA \MSSQL \DATA\

- **Archivos de bloques de código.** Se creara una carpeta en la raíz de una participación del disco duro que sea accesible para los usuarios que estén trabajando con el almacén de datos. Si bien es cierto la letra del dispositivo puede cambiar se optara por nombrar a la carpeta con el nombre de **scripts**.
- **Documentación y archivos de apoyo.** Se utilizará el mismo dispositivo y ubicación que se haya definido para los archivos de bloques de código. Con la única diferencia del nombre de la carpeta, esta tomara el nombre de **Documentación de apoyo SIEDCYTDW**.
- **Otros Archivos.** Pudiese ser necesario guardar otro tipo de archivos relacionados con el proyecto en el servidor. Todos aquellos que no puedan ser clasificados dentro de las categorías previamente mencionadas. Para ellos se creara otra carpeta en la misma raíz de los estándares anteriores pero el nombre de la carpeta a utilizar será: **Archivos SIEDCYTDW**. Cada archivo debe estar acompañado de un archivo de texto plano donde se indique el propósito de la existencia del archivo y si aplica una breve descripción de su funcionalidad.

Con los estándares previamente definidos se puede proceder a crear el modelo físico de datos.

### 3.12.2. Desarrollo del modelo físico de datos

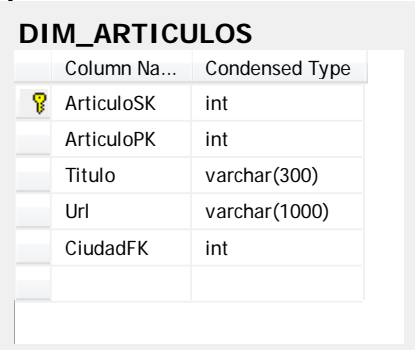
#### Creación de base de datos

Se creara la base de datos DWSiedcyt para alojar las tablas necesarias para el proyecto. Siendo que por el momento se está haciendo el desarrollo en un servidor de pruebas, no se modificaran los ajustes que por defecto aplica SQL Server 2008.

#### Creación de tablas

Para crear las tablas se utilizará la herramienta Microsoft SQL Server Managment Studio. Esta herramienta permite crear tablas por medio de comandos de SQL o por medio de una interface gráfica. Siguiendo los estándares establecidos al inicio de esta fase se procederá a crear cada una de las tablas y su definición es la siguiente:

#### Dimensiones



Column Na...	Condensed Type
ArtículoSK	int
ArtículoPK	int
Titulo	varchar(300)
Url	varchar(1000)
CiudadFK	int

**Figura 3.31:** Artículos

DIM_FINANCIAMIENTOS		
	Column Name	Condensed Type
🔑	FinanciamientoSK	int
	FinanciamientoPK	int
	InstitucionOtorganteFK	int
	TipoID	int
	TipoDescripcion	varchar(100)
	Monto	money


Figura 3.32: Financiamientos

DIM_GRADOS_ACADEMICOS		
	Column Name	Condensed Type
🔑	GradoSK	int
	GradoPK	int
	GradoNombre	varchar(200)
	GradoNivelPK	int
	GradoNivel	varchar(200)
	GradoInstitucion...	int


Figura 3.33: Grados académicos

DIM_GRUPOS_INVESTIGACION		
	Column Name	Condensed Type
🔑	GrupoSK	int
	GrupoPK	int
	GrupoNombre	varchar(100)
	InstitucionFK	int


Figura 3.34: Grupos de investigación

DIM_INSTITUCION		
	Column Name	Condensed Type
	InstitucionSK	int
	InstitucionPK	int
	TipoID	int
	TipoDescripcion	varchar(100)
	SectorID	int
	SectorDescripcion	varchar(100)
	InstitucionNombre	varchar(200)
	CiudadFK	int


**Figura 3.35:** Instituciones

DIM_INVESTIGACIONES		
	Column Name	Condensed Type
	InvestigacionSK	int
	InvestigacionPK	int
	InvestigacionNombre	varchar(100)
	CiudadFK	int
	AnioInicio	int
	AnioTermino	int


**Figura 3.36:** Investigaciones

DIM_INVESTIGADOR		
	Column Name	Condensed Type
	PersonalDataSK	int
	InvestigadorPK	int
	PersonalPK	int
	Nombre	varchar(300)
	AnioNacimiento	int
	MesNacimiento	int
	DiaNacimiento	int
	Sexo	varchar(10)
	SNIPK	int
	SNINivel	varchar(50)
	PromepPK	int
	PromepDescripcion	varchar(50)
	CiudadPK	int
	RangoIngresosPK	int
	RangoIngresosDescripcion	varchar(50)


**Figura 3.37:** Investigadores

DIM_LIBROS		
	Column Name	Condensed Type
	LibroSK	int
	LibroPK	int
	Titulo	varchar(200)
	SubAreaConocimientoPK	int
	SubAreaConocimientoDescripc...	varchar(200)
	AreaConocimientoPK	int
	AreaConocimientoDescripcion	varchar(200)
	Editorial	varchar(100)
	Edicion	varchar(30)
	Capitulos	int
	ISBNLocal	varchar(50)
	ISBNExtranjero	varchar(50)


**Figura 3.38:** Libros

DIM_LINEAS_INVESTIGACION		
	Column Name	Condensed Type
	LineaSK	int
	LineaPK	int
	LineaDeInvestigacion	varchar(100)

**Figura 3.39:** Líneas de investigación

DIM_PARTICIPACION_LIBROS		
	Column Name	Condensed Type
	ParticipacionSK	int
	ParticipacionPK	int
	ParticipacionDescripcion	varchar(100)

**Figura 3.40:** Participación en libros

DIM_PATENTES		
	Column Name	Condensed Type
	PatenteSK	int
	PatentePK	int
	SubAreaConocimientoPK	int
	SubAreaConocimientoDescrip...	varchar(100)
	AreaConocimientoPK	int
	AreaConocimientoDescripcion	varchar(100)
	Patente	varchar(200)
	No_Registro	varchar(50)
	CiudadFK	int

**Figura 3.41:** Patentes

DIM_ZONA_GEOGRAFICA		
	Column Name	Condensed Type
🔑	ZonaSK	int
	CiudadPK	int
	CiudadNombre	varchar(100)
	EstadoPK	int
	EstadoNombre	varchar(100)
	PaisPK	int
	PaisNombre	varchar(100)
	ContinentePK	int
	ContienteNombre	varchar(100)

Figura 3.42: Zona geográfica

### Tablas de relación (bridge)

FACT_REL_FINANCIAMIENTOS_INVESTIGACION		
	Column Name	Condensed Type
🔑	InvestigacionFinanciamient...	int
	InvestigacionFK	int
	FinanciamientoFK	int

Figura 3.43: Relación entre financiamientos e investigación

FACT_REL_INSTITUCIONES_INVESTIGACION		
	Column Name	Condensed Type
🔑	InvInsSK	int
	InvestigacionFK	int
	InstitucionFK	int

Figura 3.44: Relación entre instituciones e investigaciones

FACT_REL_INVESTIGADORES_INVESTIGACION		
	Column Name	Condensed Type
🔑	InvInvSK	int
	InvestigadorFK	int
	InvestigacionFK	int

Figura 3.45: Relación entre investigadores e investigación

FACT_REL_LINEAS_INVESTIGACION		
	Column Name	Condensed Type
🔑	InvestigacionLineaSK	int
	InvestigacionFK	int
	LineaInvestigacionFK	int

Figura 3.46: Relación entre investigaciones y líneas de investigación

## Tablas de hechos (facts)

FACT_GRADOS_ADQUIRIDOS		
	Column Name	Condensed Type
🔑	GradoSK	int
	GradoFK	int
	InvestigadorFK	int
	AnioInicio	int
	AnioTermino	int

Figura 3.47: Grados académicos adquiridos

FACT_GRUPOS_INVESTIGACION_VINCULACION		
	Column Name	Condensed Type
🔑	GrupoSK	int
	GrupoFK	int
	InvestigadorFK	int

Figura 3.48: Vinculación de investigadores en grupos de investigación

FACT_INVESTIGACIONES		
	Column Name	Condensed Type
🔑	InvestigacionSK	int
	InvestigacionFK	int
	MontoFinanciamiento	money

Figura 3.49: Investigaciones realizadas


FACT_LIBROS_PARTICIPACIONES	
Column Name	Condensed Type
 LibroSK	int
LibroFK	int
ParticipacionFK	int
InvestigadorFK	int
AnioPublicacion	int

Figura 3.50: Participación de investigadores en la creación de libros


FACT_PARTICIPACION_ARTICULOS	
Column Name	Condensed Type
 ArticuloSK	int
ArticuloFK	int
InvestigadorFK	int

Figura 3.51: Participación de investigadores en la creación de artículos


FACT_PATENTES_REGISTRADAS	
Column Name	Condensed Type
 PatenteSK	int
PatenteFK	int
InstitucionFK	int

Figura 3.52: Patentes registradas

Cada una de las tablas anteriormente expuestas tiene un papel diferente y muy importante en el almacén de datos. Aunque algunas de ellas juegan más de un rol a la vez dependiendo de la perspectiva de la información. Para optimizar y hacer la experiencia del usuario más amigable, se utilizará una base de datos OLAP. Esta trae grandes beneficios en el análisis de información debido a que permite definir jerarquías, crear planes de partición de datos, estrategias de agregación, análisis de información a distintos niveles de granularidad y un tiempo de respuesta óptimo. Las tablas ya definidas serán el fundamento de la base de datos OLAP. A partir de los hechos identificados y sus dimensiones se crearan cubos OLAP los cuales servirán a las herramientas BI como base para aplicar inteligencia de negocios.

Estos cubos también nos servirán para obtener reportes de manera tradicional, la ventaja es que el usuario tendrá la opción de tomar la información como fue definida o bien únicamente seleccionar lo que desea extraer de acuerdo a sus necesidades. Estas optimizaciones y las anteriores son posibles gracias a los modelos que se definieron y que darán toda la flexibilidad para la obtención de información.

Creado las tablas en el servidor de base de datos y creando los cubos OLAP siguiendo los lineamientos de la tabla Bus matrix presentada anteriormente, se da por terminado el diseño físico.

# Capítulo 4

## Extracción transformación y carga de datos

El proceso de extracción, transformación y carga de los datos también conocido como ETL por sus siglas en ingles extract, transform and load, es la actividad que más tiempo consume en un proyecto de DW/BI. Este proceso se crea en dos vertientes:

- **Carga de la información histórica.** En esta parte del proyecto se toma toda la información desde los inicios de la operación hasta la fecha actual y se carga en el almacén de datos. Este proceso únicamente se ejecuta una sola vez y tiene implicaciones muy específicas. Dependiendo de la cantidad de la información, la extracción, transformación y carga de los datos puede tomar desde algunas horas hasta semanas sin contar el tiempo posterior para la verificación de la carga exitosa.
- **Carga incremental de información (Posterior a la carga histórica).** Con el tiempo los sistemas de información incorporan nuevos datos o actualizan los existentes. Debido a esto, es necesaria la implementación de un proceso de carga incremental, el cual consiste en identificar los registros nuevos o actualizados desde la última sincronización. Dependiendo del tipo de actualización es como se tratan los datos. En el

caso de nuevos registros, estos simplemente se cargan al almacén, en el caso de actualizaciones; es necesario verificar si se trata de un cambio correctivo o de una dimensión que cambia lentamente.

El alcance de esta tesis, se limita a la carga de información histórica. Se limito de esta manera debido a que únicamente se tiene acceso a una foto estática de la base de datos. Esta foto representa o contiene información que se recopiló mediante trabajo de campo. Una vez el sistema liberado, los investigadores se harán cargo de actualizar su información. Los nuevos cambios ya están en un servidor de producción de COCYT al cual no se tiene acceso.

Como primer paso es necesario identificar las tablas y los atributos que servirán de fuente de datos para el almacén. De cierta forma esto ya se hizo en la etapa de análisis y viabilidad de los modelos dimensionales. Sin embargo, de este punto en adelante es necesario mapear exactamente cada una de las tablas y sus atributos con los del almacén de datos. La tarea no se limita a únicamente identificar fuentes y destinos, sino también en documentar la transformación por la que deben pasar los datos.

El proceso de extracción, transformación y carga de los datos debe documentarse mediante un mapa de alto nivel, donde se pueda apreciar toda la trayectoria de inicio a fin de cada atributo, desde que sale del sistema de información SIEDCYT hasta que llega al almacén de datos. El propósito de este mapa es documentar todo el proceso ETL además de proporcionar una idea clara de la cantidad de trabajo necesario para terminar esta fase del proyecto.

La siguiente sección está dedicada al detalle de los mapas.

Destino		Fuente			
Base de datos	DataWarehouseSiedcyt	Cocytbc			
Tabla	DIM_AREAS_CONOCIMIENTO	Areas_Conocimiento, SubAreas_conocimiento			
Tipo	Dimension				
Destino		Fuente		Transformación	
Atributo	Tipo	Tabla	Atributo	Tipo	
ConocimientoSK	int			int	llave subrogada
AreaConocimientoPK	int	Areas_Conocimiento	id_area	int	Ninguna
AreaConocimientoDescripcion	varchar(200)	Areas_Conocimiento	Descripcion	varchar(200)	Ninguna
SubAreaConocimientoPK	int	SubAreas_conocimiento	subarea_id	int	Ninguna
SubAreaConocimientoDescripcion	varchar(100)	SubAreas_conocimiento	Descripcion	varchar(100)	Ninguna

**Nota:**  
 Para llenar esta dimension se utilizara una vista que trae toda la informacion requerida para la transformacion.  

```

SELECT  aINNER.id_area, aINNER.Descripcion AS Area, s.subarea_id, s.descripcion AS Subarea
FROM    dbo.Areas_conocimientos AS aINNER INNER JOIN
        dbo.SubAreas_conocimiento AS s ON aINNER.id_area = s.area_id
    
```

Figura 4.1: Dimensión áreas de conocimiento

Destino		Fuente			
Base de datos	DataWarehouseStedoyt	Cocytbc			
Tabla	DIM_ARTICULOS	Articulos, DIM_ZONA_GEOGRAFICA			
Tipo	Dimension				
Destino		Fuente		Transformación	
Atributo	Tipo	SCD	Tabla	Atributo	Tipo
ArticuloSK	int	1			llave subrogada
ArticuloPK	int	1	Articulos	ArticuloID	int
Título	varchar(500)	1	Articulos	Título	varchar(500)
URL	varchar(1000)	1	Articulos	Url	varchar(1000)
CiudadFK	int	1	DIM_ZONA_GEOGRAFICA	ZonaSK	select * from (select * from [dbo].[DIM_ZONA_GEOGRAFICA]) [refTable] where [refTable].[CiudadPK] = ?

**Nota:**  
 Para llenar esta dimensión se utilizara una vista que trae toda la información requerida para la transformación.  
 SELECT ArticuloID, Título, Url, Ciudad from Articulos

Figura 4.2: Dimensión artículos

Destino		Fuente			
Base de datos	DataWarehouseSiedcyt	Cocytbc			
Tabla	DIM_FINANCIAMIENTOS	Financiamientos, FinanciamientosTipos			
Tipo	Dimension				
Destino		Fuente			Transformación
Atributo	Tipo	SCD	Tabla	Atributo	Tipo
FinanciamientoSK	int	1			llave subrogada
FinanciamientoPK	int	1	Financiamientos	id_financiamiento	Ninguna
TipoID	int	1	Financiamientos	tipo_financiamiento	Ninguna
TipoDescripcion	varchar(100)	1	FinanciamientosTipos	Nombre	Ninguna
Monto	money	1	Financiamientos	monto	imoney

**Nota:**  
 Para llenar esta dimensión se utilizara una vista que trae toda la información requerida para la transformación.

```

SELECT r.id_relacion, r.id_objeto_primario AS InvestigacionPK, r.id_objeto_destino AS FinanciamientoPK, f.monto AS MontoFinanciado, f.id_otorgante,
       f.tipo_financiamiento AS Id_Tipo_financiamiento, t.nombre AS financiamientoDescripcion
FROM   Relacion_Elements AS r INNER JOIN
       Financiamientos AS f ON r.id_objeto_destino = f.id_financiamiento INNER JOIN
       FinanciamientosTipos AS t ON f.tipo_financiamiento = t.id
WHERE  (r.id_entidad_primaria = 4) AND (r.id_entidad_destino = 12)
    
```

Figura 4.3: Dimensión financiamientos

Destino			Fuente		
Base de datos	DataWarehouseSiedcvt		Cocytbc		
Tabla	DIM_GRADOS_ACADEMICOS		GradosAcademicos, GradosAcademicosClasificaciones		
Tipo	Dimension				
Destino			Fuente		
Atributo	Tipo	SCD	Tabla	Atributo	Tipo
GradoSK	int	1			llave subrogada
GradoPK	int	1	GradosAcademicos	grado_id	llave primaria
GradoNombre	varchar(200)	1	GradosAcademicos	descripcion	Mayusculas
GradoNivelPK	int	1	GradosAcademicos	Nivel	Ninguna
GradoNivel	varchar(200)	1	GradosAcademicosClasificaciones	nivel_nombre	select nivel_nombre from GradosAcademicosClasificaciones as c, GradosAcademicos as g where g.Nivel = c.nivel_id
GradoInstitucionFK	int	1	GradosAcademicos	Institucion_id	ninguna

**Nota:**  
 Sentencia a utilizar  

```

select distinct g.grado_id, g.descripcion, n.nivel_id, n.nivel_nombre, g.institucion_id from Relacion_Elementos as r
join GradosAcademicos as g on r.id_objeto_Destino = g.grado_id
join GradosAcademicosClasificaciones as n on g.Nivel = n.nivel_id
where r.id_entidad_primaria = 2 and r.id_entidad_destino = 8
    
```

Figura 4.4: Dimensión grados académicos

Destino			Fuente		
Base de datos	DataWarehouseSiedcyt		Cocytbc		
Tabla	DIM_GRUPOS_INVESTIGACION		Cuerpos_Academicos, DIM_INSTITUCION		
Tipo	Dimension				
Destino			Fuente		
Atributo	Tipo	SCD	Tabla	Atributo	Tipo
GrupoSK	int	1			int
GrupoPK	int	1	Cuerpos_Academicos	cuero_academico_id	int
GrupoNombre	varchar(100)	1	Cuerpos_Academicos	nombre	varchar(200)
InstitucionFK	int	1	DIM_INSTITUCION	InstitucionSK	int
					select * from (select * from [dbo].[DIM_INSTITUCION]) [refTable] where [refTable].[CiudadPK] = ?

**Nota:**  
 Para llenar esta dimensión se utilizara una vista que trae toda la información requerida para la transformación.  
 SELECT cuerpo\_academico\_id, nombre, id\_institucion from Cuerpos\_Academicos

Figura 4.5: Dimensión grupos de investigación



Destino		Fuente			
Base de datos	DataWareHouseStedcyt	Cocytbc			
Tabla	DIM_INVESTIGACIONES	investigaciones, DIM_ZONA_GEOGRAFICA			
Tipo	Dimension				
Destino		Fuente		Transformación	
Atributo	Tipo	SCD	Tabla	Atributo	Tipo
InvestigacionSK	int	1			int
InvestigacionPK	int	1	investigaciones	id_investigacion	int
Investigacionnombre	varchar(300)	1	investigaciones	investigacion_nombre	varchar(300)
CiudadFK	int	1	DIM_ZONA_GEOGRAFICA	ZonaSK	int
AntioInicio	varchar(100)	1	investigaciones	fecha_inicio	date
AntioTermino	varchar(100)	1	investigaciones	fecha_fin	date
Descripcion	text	1	investigaciones	Descripcion	text

**Nota:**  
 Para llenar esta dimension se utilizara una vista que trae toda la informacion requerida para la transformacion.  
 select id\_investigacion, investigacion\_nombre, ciudad\_id, fecha\_inicio, fecha\_fin, descripcion from Investigaciones

Figura 4.7: Dimensión Investigaciones

Destino		Fuente				
Base de datos	DataWareHouseSiedcvt	Cocytbc				
Tabla	DIM_INVESTIGADOR	Investigador, PersonalData, Sni, Rangolngresos, promet				
Tipo	Dimension					
Destino		Fuente		Transformación		
Atributo	Tipo	SCD	Tabla	Atributo	Tipo	Transformación
PersonalDataSK	int	1			int	llave subrogada
InvestigadorPK	int	1	Investigador	id_investigador	int	Ninguna
Nombre	varchar(300)	1	PersonalData	Nombre+Paterno+materno	varchar(150)	Mayusculas y separacion entre nombres
AnioNacimiento	int	1	PersonalData	FechaNacimiento	date	YEAR(FechaNacimiento)
MesNacimiento	int	1	PersonalData	FechaNacimiento	date	MONTH(FechaNacimiento)
DiaNacimiento	int	1	PersonalData	FechaNacimiento	date	DAY(FechaNacimiento)
Sexo	varchar(10)	1	PersonalData	Sexo	bit	sexo=1? "Masculino" : "Femenino"
SNIPK	int	1	Investigador	id_sni	int	Ninguna
SNINivel	varchar(200)	1	Sni	Descripcion	varchar(200)	Ninguna
PromepPK	int	1	Promet	id_promet	int	Ninguna
PromepDescripcion	varchar(200)	1	Promet	Descripcion	varchar(200)	Mayusculas
CiudadPK	int	1	PersonalData	ciudad	int	Ninguna
RangolngresosPK	int	1	Rangolngresos	id_ingreso	int	Ninguna
RangolngresosDescripcion	varchar(150)	1	Rangolngresos	Descripcion	varchar(150)	Mayusculas

**Nota:**  
Para llenar esta dimension se utilizara una vista que trae toda la informacion requerida para la transformacion.

```

SELECT p.Descripcion AS promet, s.Descripcion AS SNI, i.id_investigador, i.id_personal_data, pd.Nombre + ' ' + pd.Paterno + ' ' + pd.Materno AS investigador, i.estatus,
        i.id_promep, i.id_sni, pd.ciudad
FROM
        dbo.Investigador AS i INNER JOIN
        dbo.Sni AS s ON s.id_sni = i.id_sni INNER JOIN
        dbo.Promet AS p ON p.id_promet = i.id_promep INNER JOIN
        dbo.PersonalData AS pd ON pd.id_personal_data = i.id_personal_data

```

Figura 4.8: Dimensión Investigador

Destino				Fuente			
Base de datos	DataWarehouseSiedcvt			Cocytbc			
Tabla	DIM_LIBROS			libros, Areas_Conocimientos, SubAreas_Conocimiento,			
Tipo	Dimension						
Destino				Fuente			
Atributo	Tipo	SCD	Tabla	Atributo	Tipo	Transformación	
LibroSK	int	1			int	Llave subrogada	
LibroPK	int	1	Libros	libro_id	int	Ninguna	
Título	varchar(300)	1	Libros	titulo	varchar(300)	Ninguna	
SubAreaConocimientoPK	int	1	SubAreas_Conocimiento	subarea_id	int	Ninguna	
SubAreaConocimientoDescripcion	varchar(200)	1	SubAreas_Conocimiento	descripcion	varchar(100)	Ninguna	
AreaConocimientoPK	int	1	Area_Conocimiento	id_area	int	Ninguna	
AreaConocimientoDescripcion	varchar(200)	1	Area_Conocimiento	Descripcion	varchar(200)	Ninguna	
Editorial	varchar(200)	1	Libros	Editorial	varchar(200)	Ninguna	
Edicion	varchar(200)	1	Libros	Edicion	varchar(200)	Ninguna	
Capitulos	int	1	Libros	numero_capitulos	int	Ninguna	
ISBNLocal	varchar(50)	1	Libros	isbn_local	varchar(50)	Ninguna	
ISBNExtranjero	varchar(50)	1	Libros	isbn_extranjero	varchar(50)	Ninguna	
FechaPublicacion	date	1	Libros	FechaPublicacion	date	Ninguna	

**Nota:**

Para llenar esta dimensión se utilizara una vista que trae toda la información requerida para la transformación.

```

SELECT  libros.libro_id AS LibroPK, libros.titulo AS Titulo, libros.fechaPublicacion, libros.sub_area_conocimiento_id AS SubAreaConocimientoPK,
        SubAreas_conocimiento.descripcion AS SubAreaConocimientoDescripcion, Areas_conocimientos.id_area AS AreaConocimientoPK,
        Areas_conocimientos.descripcion AS AreaConocimientoDescripcion, libros.editorial, libros.edicion, libros.numero_capitulos AS Capitulos,
        libros.isbn_local AS ISBNLocal, libros.isbn_extranjero AS ISBNExtranjero
FROM    libros INNER JOIN
        SubAreas_conocimiento ON libros.sub_area_conocimiento_id = SubAreas_conocimiento.subarea_id INNER JOIN
        Areas_conocimientos ON SubAreas_conocimiento.area_id = Areas_conocimientos.id_area

```

Figura 4.9: Dimensión libros

Destino			Fuente		
Base de datos	DataWarehouseStedcylt		Cocytbc		
Tabla	DIM_LINEAS_INVESTIGACION		Lineas de investigacion		
Tipo	Dimension				
Destino			Fuente		
Atributo	Tipo	SCD	Tabla	Atributo	Tipo
LineaSK	int	1			Llave subrogada
LineaPK	int	1	Lineas de investigacion	id_linea	Ninguna
LineaDeInvestigacion	varchar(100)	1	Lineas de investigacion	Descripcion	varchar(100)

Figura 4.10: Dimensión líneas de investigación

Destino		Fuente		
Base de datos	DataWarehouseStedcylt	Cocytbc		
Tabla	DIM_PARTICIPACION_LIBROS	NivelParticipacionLibros		
Tipo	Dimension			
Destino		Fuente		
Atributo	Tipo	Tabla	Atributo	Tipo
ParticipacionSK	int			int
ParticipacionPK	int	NivelParticipacionLibros	idParticipacion	int
ParticipacionDescripcion	varchar(100)	NivelParticipacionLibros	Participacion	varchar(150)
				Eliminacion de espacios en blanco sobrantes
				Llave subrogada
				Ninguna
				Transformación

Figura 4.11: Dimensión participación en libros

Destino			Fuente		
Base de datos	DataWarehouseSiedcyt		Cocytbc, DataWarehouseSiedcyt		
Tabla	DIM_PATENTES		Patentes, Areas_Conocimientos, SubAreas_Conocimiento, DIM_ZONA_GEOGRAFICA		
Tipo	Dimension				
Destino			Fuente		
Atributo	Tipo	SCD	Tabla	Atributo	Tipo
PatenteSK	int	1			int
PatentePK	int	1		patente_id	int
SubAreaConocimientoPK	int	1		subarea_id	int
SubAreaConocimientoDescripcion	varchar(100)	1		descripcion	varchar(100)
AreaConocimientoPK	int	1		id_area	int
AreaConocimientoDescripcion	varchar(100)	1		Descripcion	varchar(200)
Patente	varchar(200)	1		Patente	varchar(300)
No_Registro	varchar(50)	1		noregistro	varchar(50)
CiudadFK	int	1		ZonaSK	int
					select * from (select * from [dbo].[DIM_IZONA_GEOGRAFICA] [refTable] where [refTable].[Ciudad_id] = ?

**Nota:**  
Para llenar esta dimensión se utilizara una vista que trae toda la información requerida para la transformación.

```

SELECT p.patente_id, s.subarea_id, s.descripcion AS SubArea, a.id_area, a.Descripcion AS Area, p.patente, p.ciudad_id, p.noregistro
FROM dbo.Patentes AS p INNER JOIN
      dbo.SubAreas_conocimiento AS s ON p.sub_area_conocimiento_id = s.subarea_id INNER JOIN
      dbo.Areas_conocimientos AS a ON s.area_id = a.id_area
    
```

Figura 4.12: Dimensión patentes

Destino			Fuente		
Base de datos	DataWarehouseSiedcyt		Cocytbc		
Tabla	DIM_ZONA_GEOGRAFICA		Ciudad, Estados, Pais, Continente		
Tipo	Dimension				
Destino			Fuente		
Atributo	Tipo	SCD	Tabla	Atributo	Transformación
ZonaSK	int	1			Llave subrogada
CiudadPK	int	1	Ciudad	id_ciudad	Ninguna
CiudadNombre	varchar(100)	1	Ciudad	nombre	Mayusculas
EstadoPK	int	1	Estados	id_estado	Ninguna
EstadoNombre	varchar(100)	1	Estados	nombre	Mayusculas
PaisPK	int	1	Pais	id_pais	Ninguna
PaisNombre	varchar(100)	1	Pais	nombre	Mayusculas
ContinentePK	int	1	Continente	id_continente	Ninguna
ContinenteNombre	varchar(100)	1	Continente	nombre	Mayusculas

**Nota:**  
Para llenar esta dimension se utilizara una vista que trae toda la informacion requerida para la transformacion.

```

SELECT  dbo.Ciudad.id_ciudad AS CiudadID, dbo.Ciudad.nombre AS Ciudad, dbo.Estados.id_estado AS EstadoID, dbo.Estados.nombre AS Estado,
        dbo.Pais.id_pais AS PaisID, dbo.Pais.nombre AS Pais, dbo.Continente.id_continente AS ContinenteID, dbo.Continente.nombre AS Continente
FROM      dbo.Estados INNER JOIN
        dbo.Ciudad ON dbo.Estados.id_estado = dbo.Ciudad.id_estado INNER JOIN
        dbo.Pais ON dbo.Estados.id_pais = dbo.Pais.id_pais INNER JOIN
        dbo.Continente ON dbo.Pais.id_continente = dbo.Continente.id_continente

```

Figura 4.13: Dimension zona geografica

Destino			Fuente		
Base de datos	DataWarehouseSiedcyt		Cocytbc, DataWarehouseSiedcyt		
Tabla	FACT_LIBROS_PARTICIPACION		relacion_libros_autores, DIM_INVESTIGADOR, DIM_LIBRO, DIM_PARTICIPACION_LIBROS		
Tipo	Hechos				
Destino			Fuente		
Atributo	Tipo	SCD	Tabla	Atributo	Tipo
LibroSK	int	-			int
LibroFK	int	-	DIM_LIBROS	LibroSK	int
ParticipacionFK	int	-	DIM_participacion_libros	ParticipacionSK	int
InvestigadorFK	int	-	DIM_INVESTIGADOR	PersonalDataSK	int
					Transformación
					llave subrogada
					select * from (select * from [dbo].[DIM_LIBROS] [refTable] where [refTable].[LibroPK] = ?
					select * from (select * from [dbo].[DIM_participacion_libros] [refTable] where [refTable].[ParticipacionPK] = ?
					select * from (select * from [dbo].[DIM_INVESTIGADOR] [refTable] where [refTable].[InvestigadorPK] = ?

**Nota:**  
Para llenar esta tabla de hechos es necesario obtener todos los registros donde se relacione investigadores con grados académicos adquiridos. Para obtenerlo se ejecuta la siguiente consulta  
(2 es el código de investigadores)

```
SELECT id_objeto_principal AS InvestigadorPK, participacion AS ParticipacionPK, libro_id AS LibroPK
FROM relacion_libros_autores
WHERE (id_tipo_objeto = 2)
```

Figura 4.14: Hechos participación en libros

Destino			Fuente		
Base de datos	DataWarehouseSiedcyt		Cocytbc		
Tabla	FACT_GRADOS_ADQUIRIDOS		Relacion_Elementos		
Tipo	Hechos				
Destino			Fuente		
Atributo	Tipo	SCD	Tabla	Atributo	Tipo
GradoSK	int	-			int
GradoFK	int	-	RelacionElementos	id_objeto_destino	int
InvestigadorFK	int	-	RelacionElementos	id_objeto_primario	int
AnioInicio	int	-	RelacionElementos	fecha_inicio	date
AnioTermino	int	-	RelacionElementos	fecha_fin	date
					llave subrogada
					select id_objeto_destino where id_entidad_destino = 8
					select id_objeto_primario where id_entidad_destino = 8 /* 8 es el identificador de grados academicos */
					YEAR(fecha_inicio)
					YEAR(fecha_fin)

**Nota:**  
 Para llenar esta tabla de hechos es necesario obtener todos los registros donde se relacione investigadores con grados académicos adquiridos. Para obtenerlo se ejecuta la siguiente consulta  
 (8 es el identificador de grados académicos)  

```
select * from Relacion_Elementos
where id_entidad_primaria = 2 and id_entidad_destino = 8
```

Figura 4.15: Hechos grados adquiridos

Destino			Fuente		
Base de datos	DataWarehouseSiedcyt		Cocytbc, DataWarehouseSiedcyt		
Tabla	fact_grupos_investigacion_vinculacion		Relacion_Elementos, DIM_INVESTIGADOR, DIM_GRUPOS_INVESTIGACION		
Tipo	Hechos				
Destino			Fuente		
Atributo	Tipo	SCD	Tabla	Atributo	Transformación
GrupoSK	int	-			llave subrogada
GrupoFK	int	-	DIM_GRUPOS_INVESTIGACION	GrupoSK	select * from (select * from [dbo].[DIM_GRUPOS_INVESTIGACION] [refTable] where [refTable].[GrupoID] = ?
InvestigadorFK	int	-	DIM_INVESTIGADOR	PersonalDataSK	select * from (select * from [dbo].[DIM_INVESTIGADOR] [refTable] where [refTable].[InvestigadorID] = ?

**Nota:**  
 Para llenar esta tabla de hechos es necesario obtener todos los registros donde se relacione investigadores con grados académicos adquiridos. Para obtenerlo se ejecuta la siguiente consulta

```

SELECT id_objeto_primario AS GrupoID,
       id_objeto_destino AS InvestigadorID
from   Relacion_Elementos
WHERE  (id_entidad_primaria = 3) AND (id_entidad_destino = 2)
    
```

Figura 4.16: Hechos vinculación en grupos de investigación

Destino				Fuente			
Base de datos	DataWarehouseSiedcyt			Cocytbc			
Tabla	Fact_investigaciones_Participacion			Relacion_Elementos, DIM_INVESTIGACIONES, DIM_INVESTIGADOR, DIM_ZONA_GEOGRAFICA			
Tipo	Hechos						
Destino				Fuente			
Atributo	Tipo	SCD	Tabla	Atributo	Tipo	Transformación	
InvestigacionesSK	int	-			int	llave subrogada	
InvestigacionFK	int	-	DIM_INVESTIGACIONES	InvestigacionSK	int	select * from (select * from [dbo].[DIM_INVESTIGACIONES] [refTable] where [refTable].[InvestigacionID] = ?	
InvestigadorFK	int	-	DIM_INVESTIGADOR	PersonalDataSK	int	[refTable] select * from (select * from [dbo].[DIM_INVESTIGADOR] where [refTable].[InvestigadorID] = ?	
ZonaGeograficaFK	int	-	DIM_ZONA_GEOGRAFICA	ZonaSK	int	select * from (select * from [dbo].[DIM_ZONA_GEOGRAFICA] [refTable] where [refTable].[ciudad_id] = ?	
FechaParticipacionPK	DateTime	-	RelacionElementos	fecha_inicio	date	Ninguna	

**Nota:**  
 Para llenar esta tabla de hechos es necesario obtener todos los registros donde se relacione investigadores con grados académicos adquiridos. Para obtenerlo se ejecuta la siguiente consulta  
 (4 es el identificador de investigaciones y 2 identificador de investigadores)

```

SELECT  r.id_objeto_primario AS InvestigacionID,
        r.id_objeto_Destino AS InvestigadorID,
        r.Fecha_inicio,
        i.ciudad_id
from
Relacion_Elementos as r
join Investigaciones as i on r.id_objeto_primario = i.id_investigacion
WHERE  (id_entidad_primaria = 4) AND (id_entidad_destino = 2)
    
```

Figura 4.17: Hechos participación en investigaciones

Destino			Fuente		
Base de datos	DataWarehouseSiedcyt		Cocytbc, DataWarehouseSiedcyt		
Tabla	FACT_PARTICIPACION_ARTICULOS		Relacion_ Elementos, DIM_ARTICULOS, DIM_INVESTIGADOR		
Tipo	Hechos				
Destino			Fuente		
Atributo	Tipo	SCD	Tabla	Atributo	Tipo
ArticulosSK	int	-			int
ArticuloFK	int	-	DIM_ARTICULOS	ArticuloSK	int
InvestigadorFK	int	-	DIM_INVESTIGADOR	PersonalDataSK	int
					llave subrogada
					select * from (select * from [dbo].[DIM_ARTICULOS] [refTable] where [refTable].[ArticuloPK.] = ?
					select * from (select * from [dbo].[DIM_INVESTIGADOR]) [refTable] where [refTable].[InvestigadorPK.] = ?

**Nota:**  
Para llenar esta tabla de hechos es necesario obtener todos los registros donde se relacione investigadores con grados académicos adquiridos. Para obtenerlo se ejecuta la siguiente consulta.

```

(2 hace referencia a investigadores y 5 refiere articulos)
select
r.id_objeto_primario as InvestigadorPK,
r.id_objeto_Destino as ArticuloPK,
i.fecha_inicio
from Relacion_Elementos as r join Investigaciones as i on r.id_objeto_Destino = i.id_investigacion where r.id_entidad_primaria = 2 and r.id_entidad_destino = 5
    
```

Figura 4.18: Hechos participación en artículos

Destino		Fuente				
Base de datos	DataWarehouseSiedcyt	Cocytbc_DataWarehouseSiedcyt				
Tabla	FACT_PATENTES_REGISTRADAS	Relacion_Elementos, DIM_PATENTES, DIM_INSTITUCION, DIM_ZONA_GEOGRAFICA				
Tipo	Hechos					
Destino		Fuente		Transformación		
Atributo	Tipo	SCD	Tabla		Atributo	Tipo
PatenteSK	int	-			int	llave subrogada
PatenteFK	int	-	DIM_PATENTES	PatenteSK	int	select * from (select * from [dbo].[DIM_PATENTES] [refTable] where [refTable].[PatentePK] = ?
InstitucionFK	int	-	DIM_INSTITUCION	InstitucionSK	int	select * from (select * from [dbo].[DIM_INSTITUCION] [refTable] where [refTable].[InstitucionPK] = ?
CiudadFK	int	-	dim_zona_geografica	ZonaSK	int	select * from (select * from [dbo].[dim_zona_geografica] [refTable] where [refTable].[CiudadRegistroID] = ?

**Nota:**  
Para llenar esta tabla de hechos es necesario obtener todos los registros donde se relacione investigadores con grados académicos adquiridos. Para obtenerlo se ejecuta la siguiente consulta

```

select
  r.id_objeto_primario as InstitucionPK,
  r.id_objeto_Destino as PatentePK,
  c.id_ciudad as CiudadRegistroID
from Relacion_Elementos as r join Patentes as p on p.patente_id = r.id_objeto_Destino
join Ciudad as c on p.ciudad_id = c.id_ciudad
where
  r.id_entidad_primaria = 1 and r.id_entidad_destino = 10
    
```

Figura 4.19: Hechos patentes registradas

Los mapas previamente expuestos son la guía para el desarrollador. Para plasmarlos en programación y hacerlos una realidad, Microsoft provee una herramienta llamada servicios de integración (Integration Services). La cual permite no solamente programar la flujada de los datos si no documentarla de manera gráfica para que otros programadores ajenos al proyecto puedan digerir mucho más rápido todos los pasos involucrados en el proceso. Esto es muy útil cuando el equipo del proyecto inicial entrega el proyecto al cliente, mismo que se encarga de futuros cambios y adecuaciones.

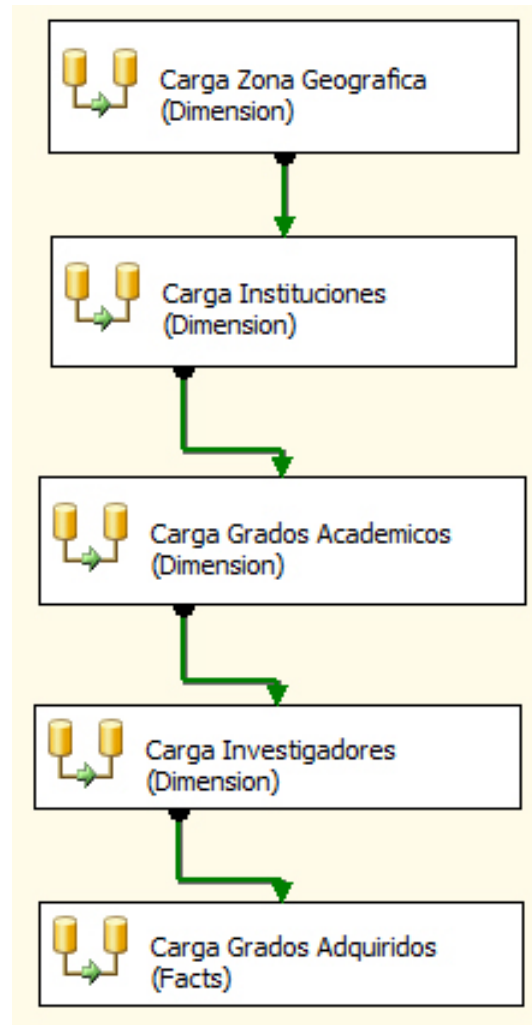
### **Programación de paquetes ETL.**

Por cada modelo presentado se programara un paquete, la programación de este paquete debe basarse y tomar como guía principal los mapas previamente descritos. Existen Dimensiones que se repiten en los distintos modelos. En estos casos únicamente se incluye la programación en el primer paquete donde se presente la dimensión. Para los paquetes posteriores únicamente se incluirán las dimensiones que no fueron previamente cargadas en otro paquete.

Una vez listos los paquetes pasan a ser parte del repositorio de trabajos de SQL Server. Aunque los paquetes realizados para esta tesis únicamente se ejecutaran una sola vez. Pueden quedar almacenados de manera permanente por si se requiere volver a recargar el almacén de datos.

La programación de paquetes con Microsoft integration services es relativamente sencillo. La herramienta permite programar el flujo de los datos de manera gráfica, colocando componentes a los cuales en la mayoría de las veces solo es necesario configurarlos e interconectarlos entre sí. Primeramente se selecciona el componente para extraer los datos de las tablas, se seleccionan los componentes a utilizar para realizar las transformaciones necesarias y finalmente se utiliza un componente especial para guardar los datos en las tablas de dimensión o de hechos.

A continuación se presentan como ejemplo las pantallas de la carga del modelo dimensional de grados adquiridos.



**Figura 4.20:** Paquete ETL para carga de modelo grados adquiridos. (Vista general)

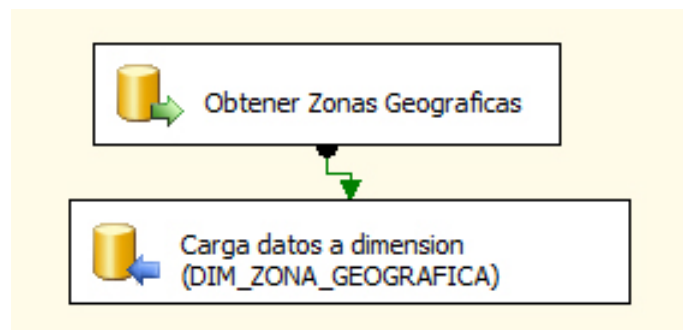


Figura 4.21: Carga de dimensión zonas geográficas.

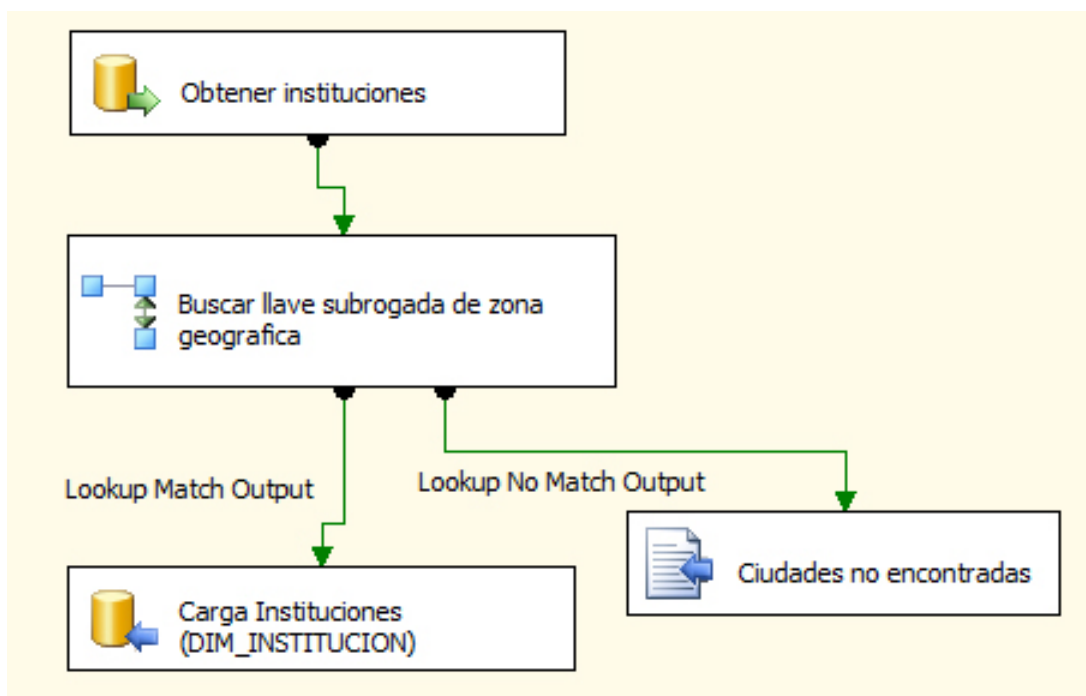


Figura 4.22: Carga de dimensión instituciones.

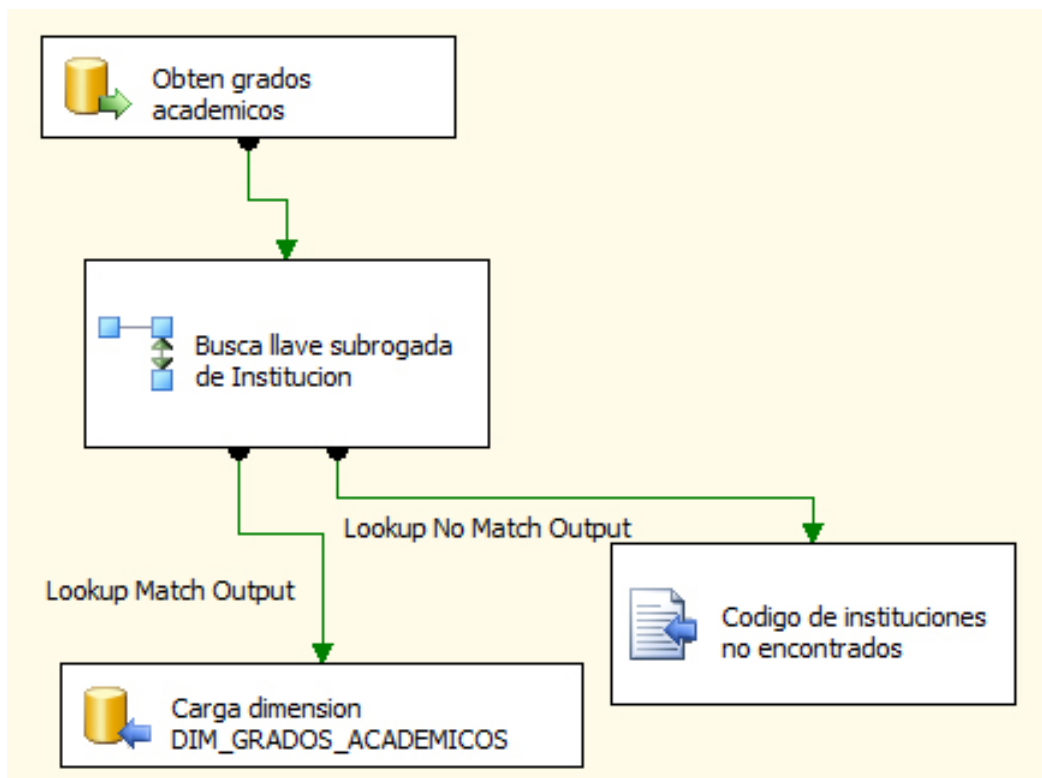


Figura 4.23: Carga de dimensión de grados académicos.

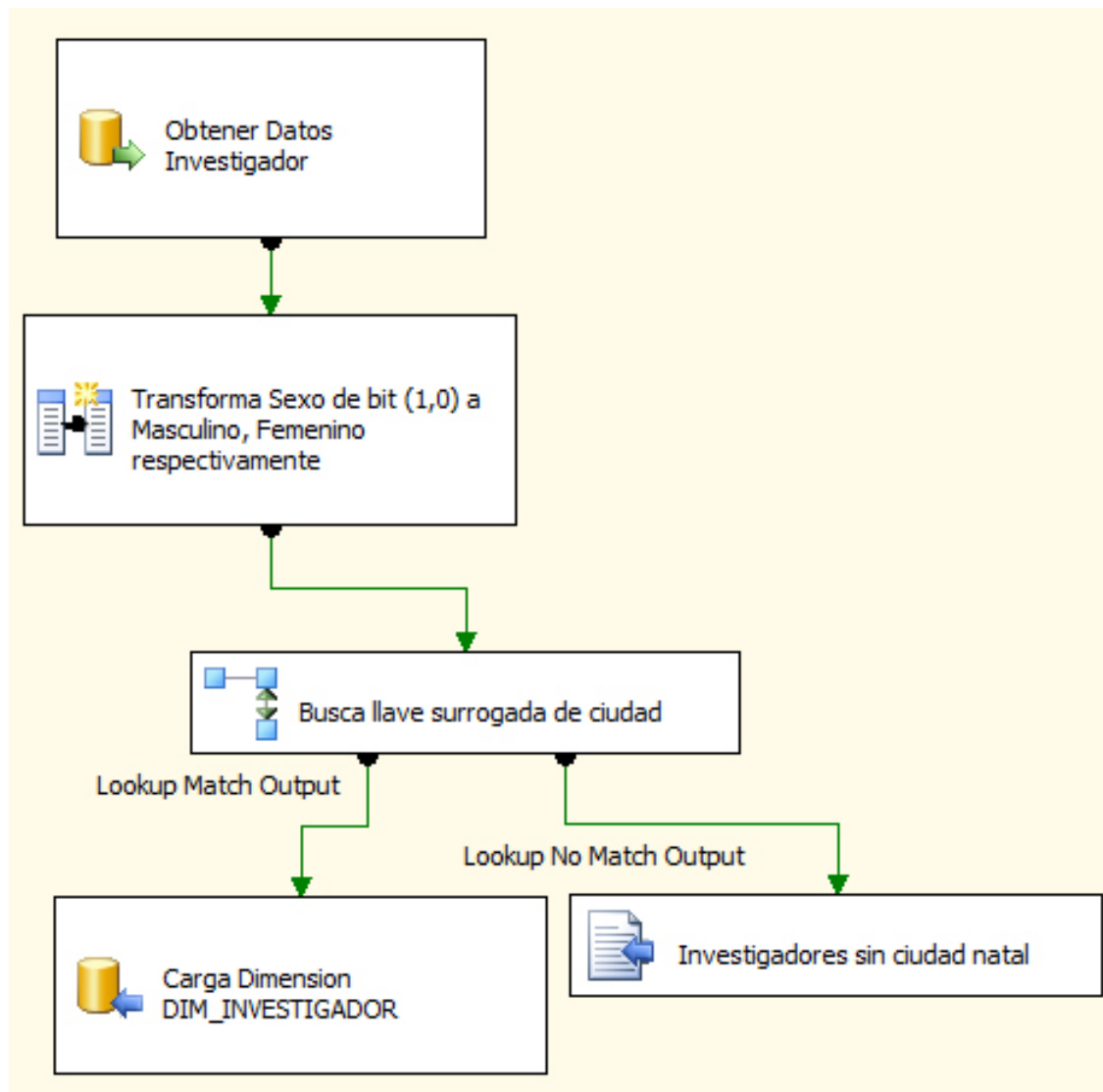


Figura 4.24: Carga de dimensión de investigador.

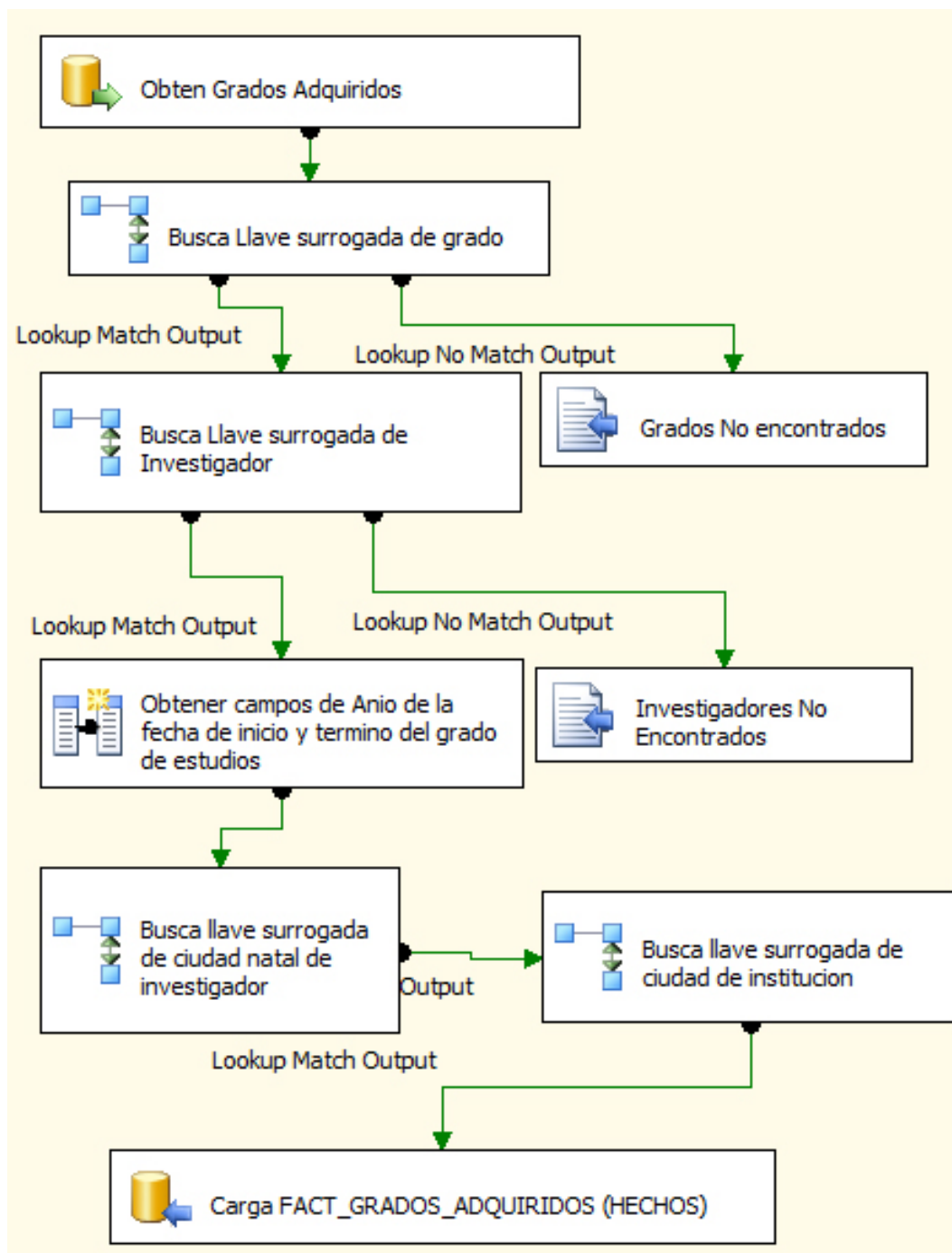


Figura 4.25: Carga de tabla de hechos grados adquiridos.

# Capítulo 5

## Aplicaciones de inteligencia de negocios

Todo lo realizado en los capítulos anteriores es necesario y de suma importancia. Sin embargo; este resulta con poco valor para los usuarios si no se les proporciona una aplicación de inteligencia de negocios. Estas aplicaciones son las que le permitirán al usuario explotar la información del almacén de datos. Por ello a estos proyectos se les llama DW/BI (Datawarehouse/Business intelligence). Es necesario que existan ambos para entregar una solución completa y efectiva. Por medio de estas herramientas el usuario puede visualizar, configurar, imprimir, analizar y enviar por correo electrónico sus reportes.

Las aplicaciones BI incluyen un gran espectro de reportes y datos analíticos. Estos varían y van desde reportes muy sencillos hasta herramientas muy sofisticadas de análisis con algoritmos muy complejos. Debido a lo anterior el espectro se divide en dos ramas principales. Estas ramas son reportes estándar y aplicaciones de analítica.

### **Reportes estándar**

Este tipo de reportes son lo más básico del espectro BI. Usualmente son muy sencillos, cuentan con un formato predefinido, varían resultados de acuerdo a parámetros y en ocasiones la

información que muestran es un pre procesado o simplemente son reportes estáticos. Estos reportes provén al usuario con información básica de lo que está ocurriendo en el negocio. Aplicado al SIEDCyT y por poner un ejemplo, algunos de los reportes que están dentro de este espectro son los siguientes:

1. Investigadores por ciudad y por nivel de SNI.
2. Libros realizados por investigadores de baja california.
3. Investigaciones realizadas por ciudad.
4. Total y clasificación de niveles académicos adquiridos por los investigadores.
5. Financiamientos otorgados por investigación.

Algunos usuarios no desean o simplemente no tienen tiempo de aprender a utilizar nuevas herramientas. Este tipo de reportes elimina el esfuerzo adicional necesario para obtener información.

### **Aplicaciones de analítica**

Las aplicaciones de analítica son más complejas que los reportes estándar y usualmente se enfocan en un proceso específico del negocio y se requiere un amplio conocimiento y dominio del proceso por parte del usuario para poder interpretar los resultados. En ocasiones son tan complejos que incluyen algoritmos creados específicamente para esa aplicación o utilizan minería de datos. Para el trabajo de esta tesis no utilizaremos este tipo de aplicaciones pero se mencionan con el fin de entender las 2 variantes en las que se dividen las herramientas BI.

### **Valor de las herramientas de inteligencia de negocios.**

El desarrollo de aplicaciones BI implica demasiado trabajo y por ello es necesario destacar todos los beneficios que brindan para así poder justificar el tiempo y esfuerzo. Estos son los siguientes:

- **Valor del negocio:** El proceso de identificar y crear aplicaciones BI basados en los requerimientos del negocio casi garantiza que se va a proveer algo de valor significativo para el negocio.
- **Amplio acceso:** Como se mencionaba anteriormente las aplicaciones BI son el medio y facilitan el acceso a la información que se encuentra en del almacén de datos. Sin ellas los usuarios se verían en la necesidad de codificar sus consultas (Querys). Situación principal que buscamos eliminar.
- **Impacto inmediato:** Las herramientas BI demuestran su valor al negocio desde el momento en que se implementan y dan a conocer información de valor sobre los procesos.
- **Validación de información:** las herramientas BI ayudan a validar la información del negocio porque permiten un estudio de la información mediante hechos y dimensiones de una manera antes no hecha.
- **Alto rendimiento en consultas:** igual que el punto anterior gracias a que la información es modelada en hechos y dimensiones. Generar consultas es mucho más fácil y eficiente mediante estas herramientas.
- **Funcionalidad:** Estas herramientas se alinean a las necesidades del negocio.
- **Colaboración entre usuarios:** Al involucrar a los usuarios en el desarrollo de estas herramientas se crea expectativa y motivación para utilizar las nuevas herramientas.
- **Retroalimentación:** Finalmente, al crear una herramienta BI le permite al equipo de desarrollo experimentar y evaluar el impacto de sus decisiones del diseño.

## 5.1. Plataforma de entrega de reportes

Para esta tesis se eligió la plataforma **Reporting Services** para entregar reportes a los usuarios. Se considera como opción viable para el proyecto porque cuenta con las características indispensables para este tipo de proyectos. Las características son las siguientes:

- **Permite definir el contenido, diseño y formato de los reportes.**
- **Provee interacción con los usuarios por medio de navegación, parámetros y además se puede profundizar en la información por medio de sub reportes.**
- **Permite seleccionar distintas maneras de distribución para los reportes. Algunas de las opciones son plataforma web, correo electrónico, archivos físicos en un directorio entre otros.**
- **Se integra con los sistemas de seguridad existentes para restringir o permitir el acceso a reportes a nivel individual.**
- **Conserva los metadatos de cada reporte.**
- **Mantiene una bitácora del acceso y ejecución de cada reporte.**

Microsoft recomienda complementar esta herramienta con algunas otras de colaboración. Sin embargo para la necesidad planteada en esta tesis es suficiente. Si se extendiera el proyecto y fuese necesario personalizar la interface con extenso detalle. Se optaría por integrarlo con **Share Point Services** y **Performance Point Server**. Estos últimos, sirven para crear portales de internet e intranet. Son principalmente herramientas de colaboración útiles para el trabajo en equipo en la organización.

## 5.2. Proceso del desarrollo de aplicaciones de inteligencia de negocios

Esta parte del proyecto está muy apegado a la recolección de requerimientos e incluye las siguientes tareas:

1. Creación del estándar y apariencia de la plantilla de los reportes.
2. Crear una lista de los reportes.
3. Crear una maqueta y documentación para cada reporte.
4. Diseño del marco de navegación.
5. Revisión y aprobación de los usuarios.

### 5.2.1. Estándar y apariencia de la plantilla de los reportes

Este paso es muy importante, porque de ello depende el que los usuarios encuentren lo que necesitan lo más rápido posible. Esto se logra presentando los datos de una manera consistente en cada reporte. Es necesario definir qué y cómo aparecerá la información en ellos. A continuación se presenta algunos de los elementos que es necesario definir.

- **Nombre del reporte:** Nombre claro y descriptivo de cada reporte. Este nombre debe comunicar de una manera muy entendible el contenido del reporte al usuario.
- **Título de reporte:** Es necesario desarrollar un estándar de que y como se mostraran los datos en el título.
- **Cuerpo del reporte:** Diseño de columnas y renglones de los datos incluyendo:
  - **Justificación:** Izquierda para texto y derecho para números.

- **Precisión:** Máximo 2 decimales en los números.
  - **Encabezados o nombres de campos:** Los títulos del detalle se harán en letra fuerte y subrayada.
  - **Fondo y rellenos de color:** Todos los fondos serán blancos y cuando se requiera rellenar de algún color será azul el color del gobierno del estado.
  - **Formato de los totales y subtotales:** Si se trata de dinero se le dará el formato \$XXX,XXX.00 en caso de ser cualquier otro número no sufrirá cambio alguno.
- **Encabezado y pies de página:** Los siguientes elementos deben encontrarse en algún lado del encabezado o del pie de página. Al definirse una sola vez deben respetarse en todos los reportes.
- Nombre del reporte.
  - Parámetros utilizados.
  - Categoría de navegación.
  - Notas importantes.
  - Numero de página.
  - Fecha y hora de ejecución de reporte.
  - Logo del proyecto DW/BI.

Siguiendo las especificaciones anteriores, se definiría la plantilla de reportes como se muestra en la figura 5.1



Figura 5.1: Plantilla de reportes.

Existen otros elementos que es necesario documentar mas sin embargo no es necesario que se muestren como parte del reporte. A continuación se presentan algunos de ellos:

- Variables de usuario necesarias para la interacción con los reportes.
- Metadatos de los reportes, incluyendo descripción, cálculos, autor, fechas de creación, etc.
- Requerimientos de seguridad, incluyendo una lista descriptiva de los grupos que tienen acceso al reporte.
- Ciclo de ejecución, para saber si el reporte se ejecuta automáticamente o de manera periódica.
- Evento disparador de ejecución, en ocasiones algunos reportes se ejecutan cuando ocurre un evento, puede ser la actualización de las fuentes de información del mismo.

- Lista de distribución, usualmente es nombre y correo electrónico.
- Formato estándar de salida como texto, html, pdf, Excel o Word.
- Orientación de la página, tamaño y ajustes de márgenes.

A todo lo anterior se le denomina metadatos del reporte.

### 5.2.2. Lista de reportes preliminares

Al llegar a este punto, es necesario volver a revisar la documentación que se dio como resultado del análisis, tomando los resultados de ello, se puede identificar que reportes se van a desarrollar y quién los va a usar. También es necesario darle una prioridad a cada uno de ellos. Cada reporte es distinto y por ello algunos pueden ser más difíciles de crear que otros. Es importante recordar que el almacén de datos mediante herramientas BI, permite al usuario generar sus propios reportes. Sin embargo existen usuarios que no tienen el tiempo de aprender o simplemente no les interesa armar sus propios reportes, prefieren tenerlo listo y solo consultarlo. Este tipo de usuarios comúnmente son los gerentes generales y altos directivos. Comúnmente no tienen el tiempo de conocer todos los detalles de los sistemas en general y solo se enfocan en conocer los resultados.

De acuerdo a los requerimientos iniciales podemos identificar los reportes presentados en la figura 5.2

Es importante destacar que esta lista es solo para empezar. Una vez que los reportes estén desarrollados y haya retroalimentación de los usuarios se podrán generar muchos otros más.

### 5.2.3. Maqueta y documentación de reportes

Preparar maquetas de cada reporte ayuda en el desarrollo del mismo. Se puede entregar una maqueta al programador y fácilmente entiende lo que se espera del reporte y como

Documento:		Proyecto:	
Lista de reportes candidatos		Siedcyt DW/BI	
#	Nombre Reporte	Descripción breve	Reporte categoría
1	Grados académicos adquiridos	Grados académicos adquiridos por cada investigador Clasificados por tipo.	Preparación académica
2	Patentes registradas	Patentes registradas en Mexico y/o por Institución	Producción
3	Investigaciones realizadas	Investigaciones realizadas en Baja California y por investigadores de ese estado.	Producción
4	Vinculación	Vinculación entre investigadores y grupos de investigación.	Vinculación
5	Artículos por investigador	Generación o participación en artículos	Producción
6	Participación de investigadores en libros	Libros generados o participación en ellos por parte de investigadores de Baja California.	Producción
7	Financiamientos otorgados	Financiamientos y sus tipos otorgados a investigaciones	Financiamientos

**Figura 5.2:** Lista preliminar de reportes.

debe presentar la información. De la figura 5.3 a la 5.8 están las maquetas de los reportes seleccionados.

#### 5.2.4. Diseño del marco de navegación

Una vez ya definidos y clasificados los reportes a realizar, se puede identificar la manera más apropiada de agruparlos para que el usuario encuentre lo que necesita lo más rápido posible. A esta organización se le llama marco de navegación o jerarquía de navegación. Idealmente, esta estructura es auto explicable. De esta forma cualquier usuario que tenga por lo menos una idea de lo que es el SIEDCyT, puede encontrar lo que necesita relativamente rápido.

#### 5.2.5. Revisión y aprobación de los usuarios

En el escenario ideal los usuarios participan en la elaboración del marco de navegación. El equipo de desarrollo muestra el diseño y obtiene retroalimentación. Si es necesario realizar

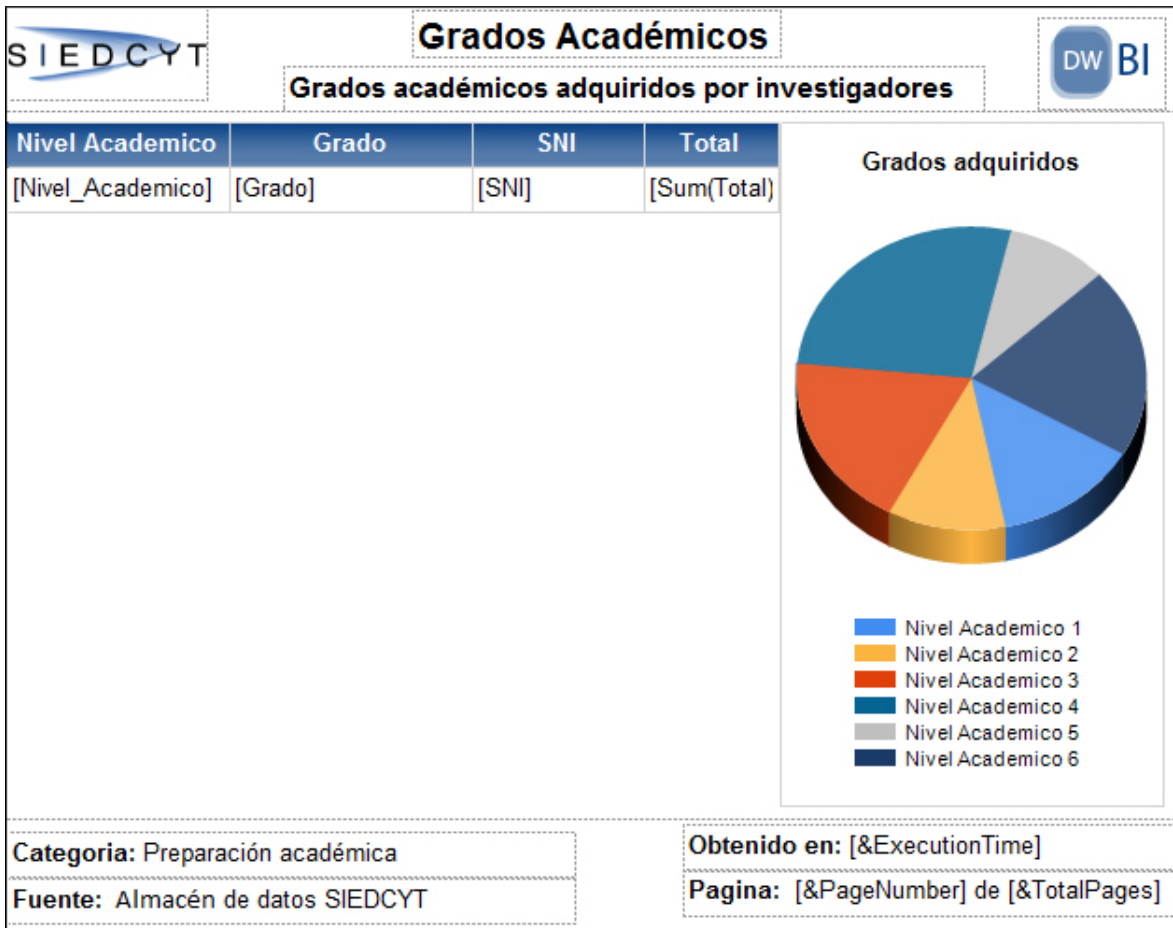


Figura 5.3: Maqueta grados académicos adquiridos.

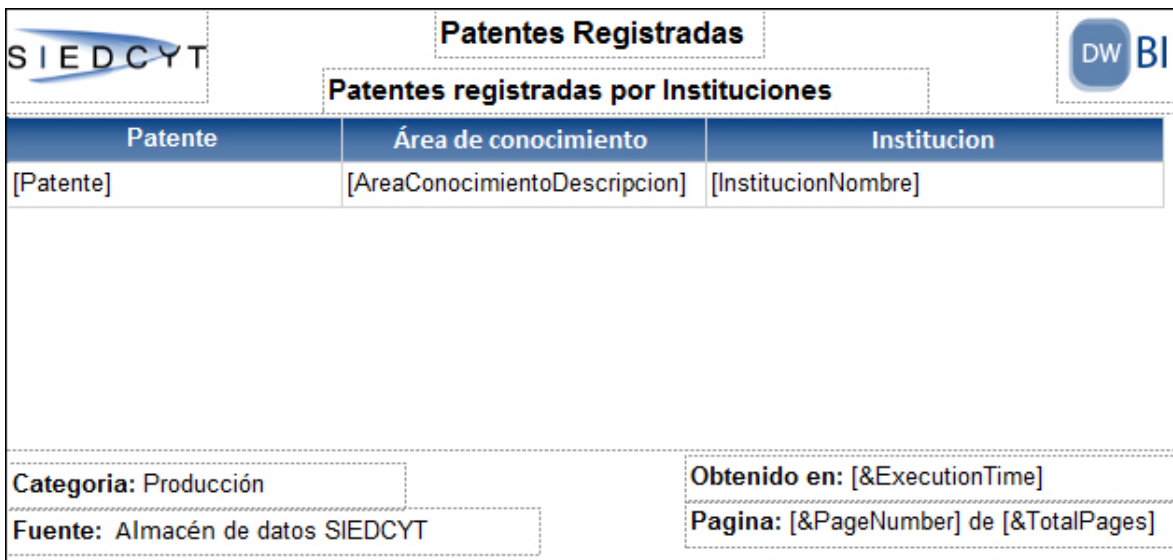


Figura 5.4: Maqueta patentes registradas.

SIEDCYT		Investigaciones Realizadas		DW BI
		Investigacion en Baja California		
Anio	Investigación	Investigador	SNI	
[Anio]	[Investigacion]	[Investigador]	[SNI Nivel]	
Categoria: Producción		Obtenido en: [&ExecutionTime]		
Fuente: Almacén de datos SIEDCYT		Pagina: [&PageNumber] de [&TotalPages]		

Figura 5.5: Maqueta investigaciones realizadas.

SIEDCYT		Grupos de Investigacion		DW BI
		Vinculación entre grupos de investigación e investigadores		
Grupo de investigación	Investigador	SNI		
Categoria: Vinculación		Obtenido en: [&ExecutionTime]		
Fuente: Almacén de datos SIEDCYT		Pagina: [&PageNumber] de [&TotalPages]		

Figura 5.6: Maqueta Grupos de investigación.

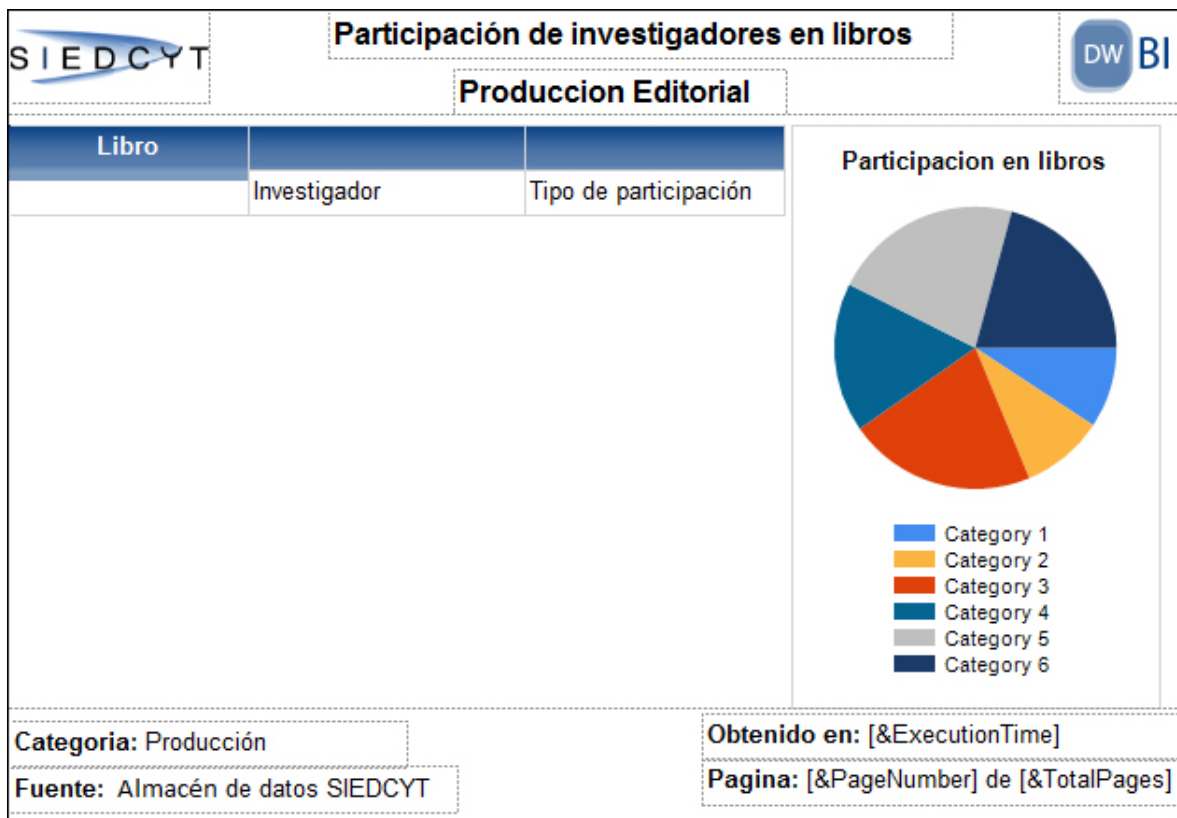


Figura 5.7: Maqueta participación de investigadores en libros.



Figura 5.8: Maqueta financiamientos otorgados.

cambios, estos se evalúan y si es viable hacerlos se hacen. Para efectos de esta tesis no se realizara entrevista alguna. Las entrevistas de esta fase están fuera del alcance considerado al inicio del proyecto.

## Capítulo 6

# Desarrollo de la aplicación de inteligencia de negocios en Microsoft Reporting Services

En esta etapa ya se cuenta con la mayoría de los elementos necesarios para desarrollar los primeros reportes. Microsoft provee una herramienta llamada **SQL Server Business Intelligence Development Studio** que ayuda al desarrollador a crear reportes, probarlos y publicarlos en la plataforma de producción. El primer paso es tomar la lista de reportes preliminares creada en el capítulo anterior y programarlos. Siendo que ya están hechos los modelos dimensionales y la información está cargada. La tarea de crear reportes y publicarlos para su consumo es relativamente sencilla. Gracias a los modelos dimensionales, la creación de las consultas de SQL es más fácil, rápido y óptimo. Lo anterior se da porque ya no es necesario conocer tan a detalle las tablas y tampoco es necesario crear uniones complicadas entre tablas. Todo lo difícil ya fue hecho en la etapa del modelado dimensional. La descripción general para crear un reporte es la siguiente:

1. **Diseño:** Utilizando la maqueta previamente creada se le da forma al reporte.

2. **Selección de fuente:** Dependiendo de los requerimientos del reporte, se selecciona el o los modelos dimensionales a utilizar y se crea la consulta en SQL.
3. **Pruebas:** se prueba el reporte y se hacen las correcciones necesarias.
4. **Publicación:** Se publica o envía el reporte al repositorio en producción.

Una vez que los reportes están en el servidor de producción, pueden ser consumidos de varias formas. El usuario tiene la capacidad de acceder al reporte y verlo en pantalla. Otra opción es calendarizar una entrega por medio de correo electrónico cada vez que este se actualice. Algo muy importante es que esta plataforma tiene la capacidad de entregar los reportes en distintos formatos. Algunos de ellos son XML, PDF, Excel, Word, HTML, etc. **Reporting Services** facilita el trabajo al programador al hacer disponibles varias opciones de entrega y formatos para satisfacer la necesidad de cada usuario a nivel individual. Todos estos beneficios se dan creando una sola versión del reporte, lo cual aumenta grandemente la productividad.

## 6.1. Portal BI

La herramienta Microsoft Reporting Services tiene un portal web por medio del cual hace accesible su administración y visualización de los distintos reportes. Sin embargo el diseño grafico no es muy llamativo y tampoco se apega a los colores y estándares del COCYT. Al hablar de un portal BI se da la impresión de un esfuerzo empresarial de recolectar y categorizar toda la información estructurada y no estructurada para hacerla disponible mediante una rica e inteligente interface gráfica con capacidades de búsqueda y personalización. Por lo anterior es necesario implementar una interface que vaya de acuerdo con los estándares institucionales. Como se menciona es únicamente una interface para personalizar el marco de navegación.

El éxito del sistema DW/BI se determina por el valor que obtiene la organización de él. Para que la organización obtenga algo de valor debe utilizar el sistema. Debido a esto, el portal BI debe proveer la mejor experiencia posible. Para garantizar lo anterior el portal BI debe contar con lo siguiente:

- **Usabilidad:** Los usuarios deben ser capaces de encontrar lo que buscan.
- **Rico en contenido:** Debe incluir mucho más que unos simples reportes. Además de ellos es necesario incluir archivos de ayuda, documentación, ejemplos y si es posible recomendaciones.
- **Limpieza:** Los elementos deben estar bien establecidos para que los usuarios no se confundan o se sientan abrumados con tanta información.
- **Vigente:** Debe ser el trabajo de alguien mantenerlo actualizado. Esto significa: no ligas rotas, elementos con la etiqueta de nuevo cuando ya tienen más de 12 meses, etc.
- **Interactivo:** : Debe incluir funciones que involucre a los usuarios y que los haga regresar al portal.
- **Orientado a proporcionar valor:** Los usuarios que visiten el portal deben sentir que es un recurso de mucho valor y que además le ayuda a realizar mejor su trabajo.

En resumen todos los buenos principios que aplican para un buen sitio web aplican también para el portal BI.

### 6.1.1. Impacto en el diseño

Para lograr los puntos anteriormente mencionados, es recomendable mostrar el máximo de información en los niveles más altos de la jerarquía sin saturar al usuario. Se tiene que evitar que el usuario haga clic muchas veces para llegar a lo que busca. Esto es malo porque

usualmente significa descargar varias páginas inútiles y con ello va un tiempo de espera innecesario. Los contenidos del portal BI deben estar bien definidos y clasificados. En la figura 6.1 se puede apreciar como este portal sería fácil de navegar al utilizar los encabezados para localizar rápidamente el tema de interés y si esto no fuera suficiente inmediatamente después se encuentra un listado de reportes disponibles. Siendo de esta manera, se agiliza la navegación y cualquier persona aun sin haber utilizado antes el portal podrá encontrar fácilmente lo que busca con 1 o 2 clics. La clave de una buena experiencia de navegación es agrupar el contenido por categorías y utilizar el mismo estándar en todas las páginas para que el portal tenga consistencia.

Inicio | Documentos | Crear | Ajustes | Ayuda

DW BI Portal de inteligencia empresarial SIEDCYT

**INICIO** SIEDCYT

<ul style="list-style-type: none"> <li>▪ Documentos</li> <li>▪ Imágenes</li> <li>▪ Publicaciones                             <ul style="list-style-type: none"> <li>-Noticias</li> <li>-Eventos</li> </ul> </li> <li>▪ Discusiones                             <ul style="list-style-type: none"> <li>-Ad-HOC Querys.</li> </ul> </li> <li>▪ Encuestas                             <ul style="list-style-type: none"> <li>-Satisfacción con el portal BI.</li> </ul> </li> </ul>	Reportes Estándar		Información adicional
	<b>Investigaciones</b> <ul style="list-style-type: none"> <li>-Investigaciones por áreas de conocimiento.</li> <li>-Investigaciones por institución.</li> <li>-Investigaciones por ciudad</li> </ul>	<b>Financiamientos</b> <ul style="list-style-type: none"> <li>-Públicos.</li> <li>-Privados</li> <li>-Mixtos.</li> </ul>	<b>Reportes Populares</b> <ul style="list-style-type: none"> <li>-<a href="#">Investigadores por ciudad</a></li> <li>-<a href="#">Infraestructura para investigación</a></li> <li>-<a href="#">Investigaciones por ciudad</a></li> </ul>
	<b>Investigadores</b> <ul style="list-style-type: none"> <li>-Investigadores por ciudad.</li> <li>-Vinculación con instituciones y grupos de investigación.</li> <li>-Investigaciones realizadas.</li> </ul>	<b>Producción Editorial</b> <ul style="list-style-type: none"> <li>-Libros</li> <li>-Artículos</li> </ul>	<b>Estado del Sistema DW/BI</b> <ul style="list-style-type: none"> <li>-El servicio esta arriba y funcionando.</li> <li>-Hay 60 usuarios conectados.</li> <li>-Ultima actualización: hoy 7:00am</li> </ul>
	<b>Instituciones</b> <ul style="list-style-type: none"> <li>-Ubicación geográfica de instituciones.</li> <li>-Patentes registradas.</li> <li>-Investigaciones realizadas.</li> </ul>	<b>Grados Académicos</b> <ul style="list-style-type: none"> <li>-Grados adquiridos por investigadores.</li> <li>-Niveles académicos adquiridos por nivel de SNI</li> </ul>	<b>Documentación del sistema BI</b> <ul style="list-style-type: none"> <li>-Explorador de metadatos</li> <li>-Tutoriales Ad-hoc</li> <li>-Descripción de reportes estándar</li> <li>-Modelos de datos.</li> </ul>
<b>Infraestructura</b> <ul style="list-style-type: none"> <li>-Infraestructura para investigación.</li> </ul>	<b>Equipo Especializado</b> <ul style="list-style-type: none"> <li>-Equipo por institución.</li> <li>-Equipo por Investigador.</li> </ul>	<b>Recursos de apoyo</b> <ul style="list-style-type: none"> <li>-Formulario para pedir ayuda.</li> <li>-Contacte a un experto.</li> <li>-Foros BI de usuario.</li> </ul>	

Gracias por utilizar el portal BI

Figura 6.1: Ejemplo de página de inicio de portal BI. Es fácil y accesible encontrar cualquier elemento.

## Funciones adicionales

La función principal del portal BI SIEDCYT es presentar reportes, sin embargo es necesario que incluya funciones adicionales para proveer una mejor experiencia al usuario. Dichas funcionalidades se describen a continuación:

- **Búsqueda:** La herramienta de búsqueda sirve como alternativa adicional para localizar reportes. Dependiendo del contenido, en ocasiones no será posible incluir toda la información en la página de inicio. En estos casos es muy útil la búsqueda. Conforme el almacén de datos crezca y haya más reportes, más se complicara la inclusión de estos en la portada.
- **Navegador de metadatos:** El usuario puede aprender más del la solución DW/BI si le permitimos conocer los detalles de los datos. El hecho de proveer algunas páginas con la descripción de la base de datos, tablas, atributos, reglas de negocio, utilización de reportes, etc. Incrementa la comprensión de la herramienta. Esta funcionalidad no es para todos los usuarios, únicamente para aquellos realmente interesados.
- **Publicaciones de ayuda:** Es opcional pero de mucha ayuda el incluir un sistema de soporte en línea. El publicar problemas y soluciones de otros usuarios y hacerlos disponibles mediante el portal BI, puede ayudar a otros usuarios a solucionar problemas similares. De esta forma se reduce la carga de trabajo para el equipo de soporte y se va creando una base de conocimiento sobre el sistema DW/BI.
- **Personalización:** El usuario debe ser capaz de guardar sus propios reportes o liga hacia otros reportes en su página personal. Esta personalización es un gran incentivo para ellos y promueve la visita constante al portal BI.
- **Centro de información:** Opcional pero extremadamente útil. El mostrar noticias o actualizaciones ayuda a mantener el interés en la herramienta de BI. Publicar encuestas

sobre la experiencia que se ha tenido con la solución, ofrecer capacitación en alguna área donde la mayoría tenga problemas, noticias sobre futuras mejoras, etc. Ayuda a promover el uso y mantener interés en el portal BI.

Aunque el alcance de esta tesis no cubre el último punto, se menciona para que se tenga en consideración como una línea futura si alguien más decide llevar el proyecto más allá de su alcance inicial.

### **Alternativas al portal BI**

Anteriormente se menciona que la plataforma de reportes debe ser capaz de distribuir o presentar la información en diferentes formatos y herramientas. También se menciona la posibilidad de crear reportes ad-hoc. Para proporcionar esta funcionalidad y evitar que el usuario tenga que aprender a utilizar herramientas nuevas, se puede utilizar Microsoft Excel. La mayoría de los usuarios saben utilizar paquetería de office y basta tener un conocimiento básico para poder utilizar el almacén de datos desde Excel. Esto es posible debido a que los reportes se forman de manera gráfica utilizando los modelos dimensionales. En esta parte es donde se da a notar más la utilidad del modelado dimensional para con el usuario. Sin ellos el usuario aun con acceso a la base de datos, se tendría que preocupar por entender el diseño y relación entre atributos para poder crear reportes. Obviamente esto último no es viable. Sin embargo como los datos están modelados por hechos y dimensiones el usuario solamente se tiene que preocupar por seleccionar los atributos y hechos del reporte a crear como se muestra en las figuras 6.2, 6.3, 6.4. De una manera muy sencilla y solamente arrastrando y colocando atributos puede obtener el reporte deseado. Al hacerlo directamente desde Excel, el usuario puede crear graficas y otros elementos visualmente atractivos que le ayuden a analizar mejor la información.

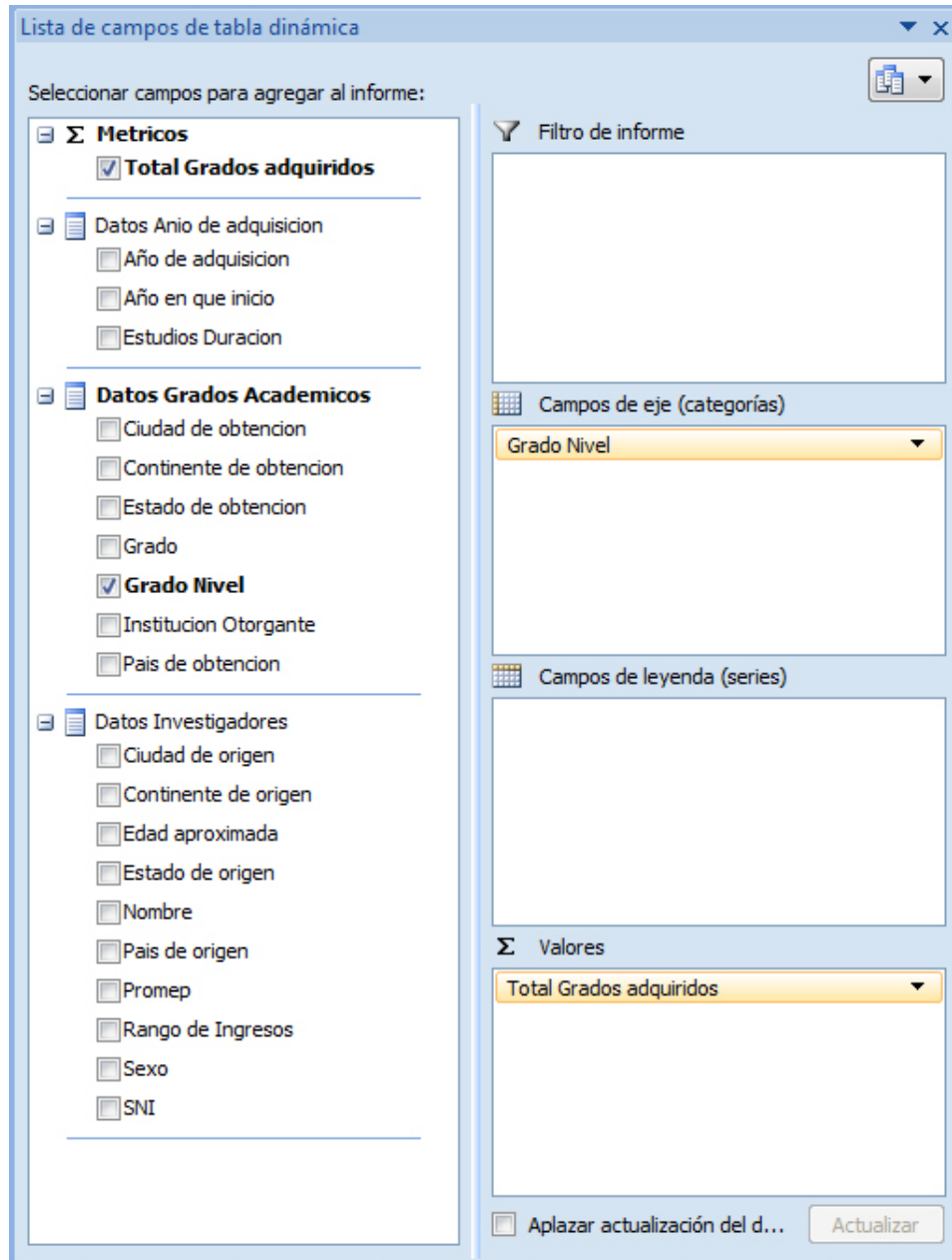


Figura 6.2: Ejemplo de Interface de selección de dimensiones y hechos o facts en Microsoft Excel.

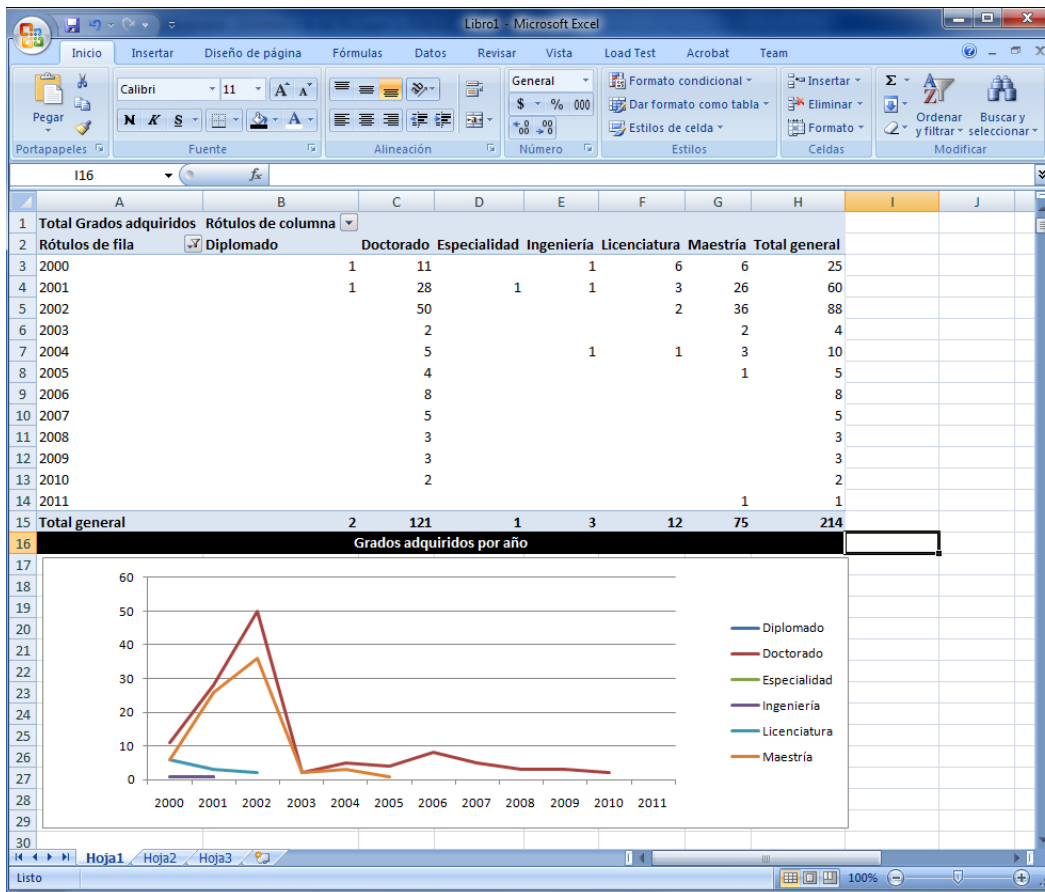


Figura 6.3: Total y tipo de grados adquiridos por los investigadores por año.

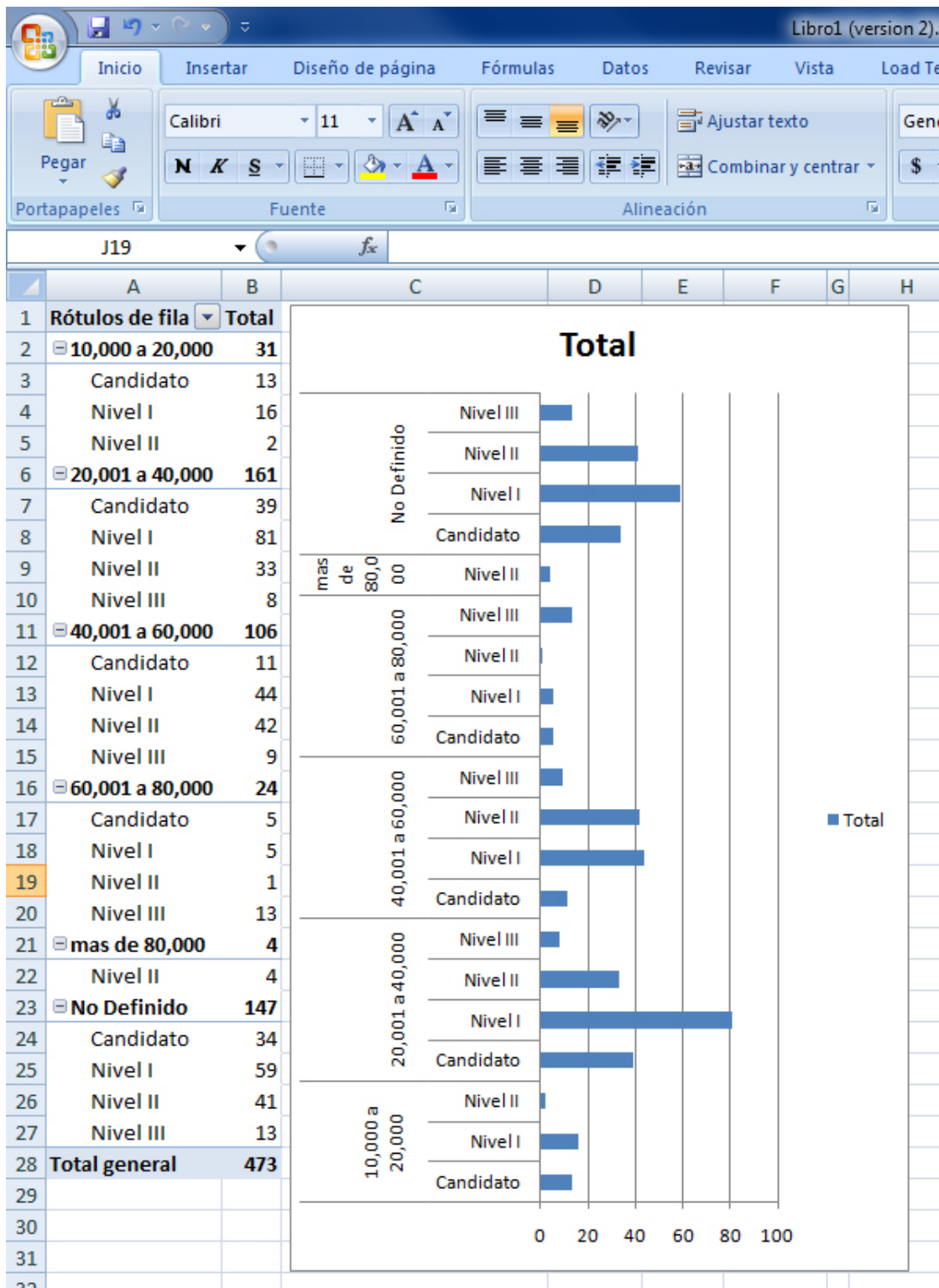


Figura 6.4: Rango de ingresos por nivel de SNI.

## Seguridad

Un elemento muy importante en cualquier sistema es la seguridad y para el almacén de datos no es excepción. Si bien es cierto el acceso a los datos desde otras aplicaciones es muy sencillo y amigable. Esto no quiere decir que es poco seguro. Antes de cualquier interacción con el almacén de datos, es necesario que el usuario o aplicación se autentifique como un usuario válido. Ver figura 6.5. Una vez autenticado, las aplicaciones del sistema DW/BI hacen disponibles los recursos de acuerdo a los privilegios establecidos por el administrador del sistema. De esta forma, no importando desde donde se acceda a los datos, la información siempre está segura. Aun conectado desde Excel, los datos no son almacenados localmente a no ser que el usuario con permisos establecidos, indique que desea guardar una copia local. De lo contrario Excel únicamente funge como un medio por el cual explotar el almacén de datos desde una herramienta que el usuario ya sabe utilizar.

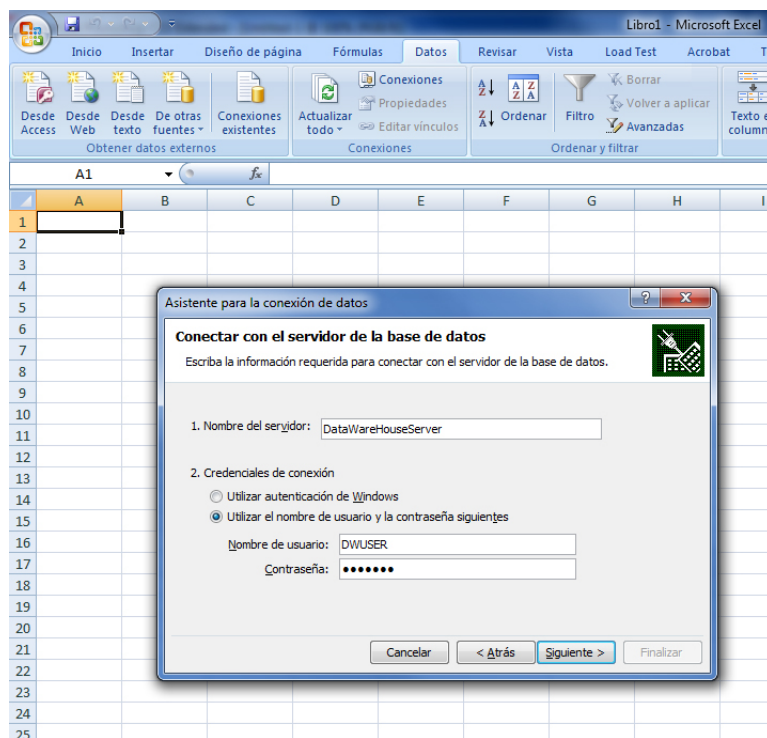


Figura 6.5: Autenticación desde Excel al almacén de datos.

# Líneas futuras

Existen otras formas de proveer aun mayor valor a las herramientas de inteligencia empresarial. Una de ellas es la minería de datos, donde se buscan patrones y/o realizar pronósticos sobre el comportamiento de los datos si cambia una o más variables. Como una mejora a futuro se podría pensar en la posibilidad de implementar estas tecnologías. Para ellos es necesario que exista un experto capaz de interpretar los datos del análisis. Es un proyecto que se tiene que trabajar muy de cercas con COCYT para determinar si realmente aportara un beneficio que justifique su implementación. Aun así vale la pena considéralo como una línea futura.

# Conclusión

Una solución DW/BI es un gran complemento para cualquier sistema de información. Sin embargo no es tarea fácil desarrollarlo e implementarlo. Los beneficios que se logran con el éxito de un proyecto de este tipo son bastante buenos y es lo que hace atractivo invertir en la solución. Todos los sistemas maduros utilizan este tipo de tecnologías para apoyar el crecimiento del negocio y aumentar la productividad de sus usuarios. Al proveer información correcta, oportuna y de una manera fácil de entender, ayuda a que los usuarios realmente apliquen inteligencia empresarial. Sin información de valor, el negocio puede llegar a perder oportunidades, o en el peor de los casos puede perder grandes sumas de dinero tomando malas decisiones debido a incorrectas o inactualizadas bases de información. La razón principal por la que COCYT busco el desarrollo del SIEDCyT es precisamente conocer con el mayor detalle posible la productividad científica y tecnológica del estado. El SIEDCYT como todo sistema de información provee reportes, aun así dista mucho de ser una solución tan completa como un almacén de datos con herramientas de inteligencia empresarial. Otro beneficio adicional no contemplado en esta tesis, es que pudiese integrarse al almacén de datos información de otros procesos o sistemas de información. Esto aumentaría en gran manera el valor de la solución. Como último punto se desea dejar claro que el sistema DW/BI es una plataforma exclusivamente hecha para reportes. Siendo que el SIEDCyT se desarrollo para recopilar información más que como plataforma de reportes, el proyecto DW/BI es el complemento ideal para entregar una solución realmente completa.

# Glosario

- **Ad-hoc query.** Es una consulta formulada al momento de acuerdo a la necesidad de ese instante.
- **Almacén de datos.** En el contexto de la informática, un almacén de datos (del inglés data warehouse) es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.
- **Analysis Services.** Herramienta de Microsoft SQL Server para crear cubos OLAP, indicadores de desempeño y además permite probar modelos de minería de datos.
- **asp.net.** es un framework para aplicaciones web desarrollado y comercializado por Microsoft. Es usado por programadores para construir sitios web dinámicos, aplicaciones web y servicios web XML.
- **BI.** Abreviación de Business Intelligence. Traducido es Inteligencia de negocios. El objetivo principal de esta disciplina es proveer información de valor y a tiempo para la toma de decisiones.
- **Bus matrix.** Tabla que sirve para identificar la unión o relación entre hechos y dimensiones.
- **Business Intelligence.** Ver BI

- **COCyT.** Consejo de ciencia y tecnología, cuyas funciones son: La administración de fondos destinados a la investigación para resolver los problemas estatales y/o regionales y Consolidar un mercado regional de conocimiento
- **Cubo multidimensional.** contienen datos resumidos de grandes Bases de datos o Sistemas Transaccionales (OLTP). Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares.
- **Dashboard.** Herramienta que presenta métricos de desempeño del negocio y también es conocido como tablero de mando.
- **Data Warehouse.** Ver almacén de datos
- **Datamart.** base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica
- **Dimensión .** Perspectiva de los hechos o facts. Si se habla de ventas las dimensiones pueden ser clientes, producto y tiempo.
- **DW/BI.** Se refiere a un sistema integral que utiliza un almacen de datos y herramientas de inteligencia empresarial.
- **ETL.** (Extraer, transformar y cargar en inglés, frecuentemente abreviado a ETL) es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio.
- **Granularidad.** Es la especificidad a la que se define un nivel de detalle en una tabla, es decir, si hablamos de una jerarquía la granularidad empieza por la parte más alta de la jerarquía, siendo la granularidad mínima, el nivel más bajo.
- **Integration Services.** SQL Server Integration Services ('SSIS) es un componente de Microsoft SQL Server utilizado para migración de datos . SSIS es una plataforma

para la integración de datos y sus de flujos de trabajo. Tiene importantes tareas para procesamiento de cubos multidimensional de Analisis Services.

- **KPI.** Indicadores clave de desempeño. En su mayoría se representan mediante semáforos donde rojo s malo, amarillo regular y verde excelente.
- **Llave foránea.** Clave cuyo valor representa la llave primaria de una tabla externa.
- **Llave primaria.** Columna o grupo de columnas cuyos valores identifican únicamente a una fila en la tabla.
- **Llave subrogada.** Llave primaria de una tabla que no tiene significado alguno para el negocio.
- **Normalización.** Aplicación de reglas o formas normales que ayudan a evitar redundancia de datos, problemas de actualización e integridad. Existen 5 formas normales pero se puede decir que una base de datos esta normalizada cuando cumple con las 3 primeras formas normales.
- **Metadatos.** Datos que describen otros datos. En general, un grupo de metadatos se refiere a un grupo de datos, llamado recurso.
- **OLAP.** Acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). Es una solución utilizada en el campo de la llamada Inteligencia empresarial (o Business Intelligence) cuyo objetivo es agilizar la consulta de grandes cantidades de datos.
- **OLTP.** Es la sigla en inglés de Procesamiento de Transacciones En Línea (Online Transacción Processing) es un tipo de sistemas que facilitan y administran aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones (gestor transaccional).

- **Performance Point Server.** Herramienta de inteligencia empresarial que se integra con share point.
- **Portal BI.** Sitio de intranet o internet enfocado en entregar reportes de inteligencia empresarial
- **PROMEP.** Programa del mejoramiento del profesorado.
- **RDBMS.** es un Sistema Administrador de Bases de Datos Relacionales. RDBMS viene del acrónimo en inglés Relational Data Base Management System.
- **Reporting Services.** Se utiliza en aplicaciones de inteligencia empresarial como plataforma de entrega de reportes.
- **SAN.** Ver Storage Área Network
- **SCD.** Ver Slowly changing dimension
- **SEBS.** Ver Secretaria de Educación y Bienestar Social
- **Secretaria de Desarrollo Económico.** Tiene como objetivo desempeñar las materias relativas al desarrollo y regulación de las actividades económicas en los sectores agropecuario, industrial, comercial y de servicios.
- **SEDECO.** Ver Secretaria de Desarrollo Económico
- **SharePoint.** Microsoft SharePoint, también conocido como Microsoft SharePoint Products and Technologies, es una colección de productos y elementos de software que incluye, entre una selección cada vez mayor de componentes, funciones de colaboración, basado en el Explorador web, módulos de administración de proceso, módulos de búsqueda y una plataforma de administración de documento
- **SIEDCyT.** Sistema Estatal de Información, Documentación Científica y Tecnológica de Recursos Humanos, Materiales, de Organización y Financieros del Estado de Baja

California. El objetivo fundamental del SIEDCyT es un Sistema de información administrado y accedido por un sitio Web, el cual permitirá almacenar, procesar y distribuir información relevante de Ciencia y Tecnología, en el estado de Baja California, para fortalecer la toma de decisiones en esta área.

- **Slowly changing dimension.** SCD abreviado en inglés y se utiliza en dimensiones versionadas. Al realizar cambios en una base de datos es posible obtener registros de tiempo atrás, antes de los cambios. Es decir se guardan las diferentes versiones de un registro.
- **SNI.** Sistema Nacional de investigadores. Este fue creado Por Acuerdo Presidencial publicado en el Diario Oficial de la Federación el 26 de julio de 1984, para reconocer la labor de las personas dedicadas a producir conocimiento científico y tecnología.
- **SQL.** Lenguaje de consulta estructurado.
- **SQL Server 2008.** Microsoft SQL Server es un sistema para la gestión de bases de datos producido por Microsoft basado en el modelo relacional. Sus lenguajes para consultas son T-SQL y ANSI SQL.
- **Storage Área Network.** Una red de área de almacenamiento, en inglés SAN (storage area network), es una red concebida para conectar servidores, matrices (arrays) de discos y librerías de soporte.

# Bibliografía

- [1] Consejo Nacional de Ciencia y Tecnología (2008, enero 1), (CONACYT.MX), Disponible: [http://www.conacyt.gob.mx/Fondos/Mixtos/BajaCalifornia/2008-01/FOMIX-BajaCalifornia\\_2008-01\\_Demandas-Especificas.pdf](http://www.conacyt.gob.mx/Fondos/Mixtos/BajaCalifornia/2008-01/FOMIX-BajaCalifornia_2008-01_Demandas-Especificas.pdf) (Consultado: 2010, Febrero 28).
- [2] Cameron, Scott. (2009) Microsoft SQL Server 2008 Analysis Services: Step by Step, Microsoft, USA.
- [3] Claudia Imhoff, Nicholas Galembo, Jonathan G. Geiger.(2003),Mastering Data Warehouse Design, Relational and Dimensional Techniques 2nd ed, WILEY.
- [4] Galleli Chris, T. Silverstein Alex. (2011) Microsoft SQL Server 2008 R2: Unleashed. SAMS, USA.
- [5] Golfarelli Matteo, Rizzi Stefano. (2009) Data Warehouse Design: Modern Principles and Methodologies. McGraw-Hill Osborne, USA.
- [6] Joy Mundy, Warren Thorntwaite. (2006) The Microsoft Data Warehouse Toolkit: With SQL Server 2005 and the Microsoft Business Intelligence Toolset. Wiley, USA.
- [7] Kimball Ralph, Caserta Joe. (2004) The Data Warehouse ETL ToolKit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Wiley Canada.
- [8] Kimball Ralph, Ross Margy. (2008) The Data Warehouse Lifecycle Toolkit 2nd ed, Wiley, Canada.

- 
- [9] Kimball Ralph, Ross Margy. (2002) The Data Warehouse Toolkit The Complete Guide to Dimensional Modeling, 2nd ed, Wiley, Canada.
- [10] Kimball, Ralph. Ross, Margy. (2010) The Kimball Group Reader: Relentlessly Practical tools for Data Warehousing and Business Intelligence. Wiley USA and Canada.
- [11] Knight Brian, Knight Devin, Jorgensen Adam, LeBlanc Patrick, Davis Mike (2010) Microsoft Business Intelligence 24-Hour Trainer. Wrox, Indiana.
- [12] Larson Brian. (2008), Delivering Business Intelligence with Microsoft Sql Server 2008, 2nd ed, McGraw Hill, New York.
- [13] Larson Brian. (2008), Microsoft Sql Server 2008 Reporting Services, McGraw Hill, New York.
- [14] Ponniah Paulraj. (2010), Data Warehousing Fundamentals for IT Professionals. Wiley; 2 edition, Canada.
- [15] Rainardi Vincent. (2008), Building a Data Warehouse: With Examples in SQL Server (Expert's Voice). Apress, USA.
- [16] Roland Bouman, Jos van Dongen.(2009),Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL, 1fst. ed, Wiley, Indiana.
- [17] Robert Wrembel and Christian Koncilia.(2007) DATA WAREHOUSES AND OLAP Concepts, Architectures and Solutions, 1fst. ed, IRM Press, USA.
- [18] Valenti, P. (19 de noviembre de 1999), Políticas para la innovación, (Organización de estados iberoamericanos), Disponible: <http://www.oei.es/salactsi/portoalegre.htm> (Consultado: 2010, Enero 28).