

**APARTADO 1:**

**DESARROLLO DEL  
SISTEMA DW.**

## **CONTENIDO**

---

1	DATA MART: PRESUPUESTO .....	3
1.1	Requerimientos .....	4
1.2	Análisis .....	11
1.3	Interacción con grupo OLAP .....	33
1.4	Diseño.....	34
1.5	Implementación.....	58
2	DATA MART: BEDELIA .....	59
2.1	Requerimientos .....	59
2.2	Análisis .....	63
2.3	Interacción con grupo OLAP .....	75
2.4	Diseño.....	76
2.5	Implementación.....	89
3	APENDICE .....	90
	Técnica para Identificación de datos requeridos .....	90

**DATA MART:**

**PRESUPUESTO**

## 1.1 REQUERIMIENTOS

---

### ➤ REQUERIMIENTOS TECNOLÓGICOS

---

*Los requerimientos tecnológicos especifican el ámbito de implementación del data mart.*

---

El data mart de Presupuesto se construye sobre una base de datos ORACLE que se encuentra en un servidor UNIX dentro del INCO.

Los datos del data mart podrán ser accedidos utilizando herramientas de consultas a bases de datos relacionales y utilizando herramientas OLAP, en particular Cognos. No se especifica una herramienta o lenguaje de programación especial para implementar el proceso de carga. Sí existen restricciones al respecto, la tecnología debe poder acceder a la base de datos ORACLE, correr sobre UNIX y el INCO debe disponer de la misma.

### ➤ MODELO CONCEPTUAL MULTIDIMENSIONAL

---

#### A. DIMENSIONES

Dimensión/Categoría	Descripción	Datos
Instituto	Institutos de la F.I..	(id_instituto, nombre)
Departamento	Departamentos	(id_depto, nombre)
Secciones	Secciones	(id_seccion, nombre)
Fuente de Financiación	Fuentes de financiación de la F.I.	(id_fuente, nombre)
Cargo	Representa los cargos existentes en la F.I.	(id_cargo)
Cantidad de horas	Cantidad de horas	(id_cant_horas)
Grado	Grado de un docente en un cargo.	(id_grado)
Docente	Representa los docentes de la F.I.	(id_docente, nombre)

Concepto	Conceptos del manejo del presupuesto (*)	(id_concepto, descripcion)
Fecha	Fecha	
Año	Año	(nro_año)
Semestre	Semestre	(nro_semestre[1,2] )
Mes	Mes	(nro_mes [1..12] )

(\*) Descripción de los valores de la dimensión Concepto:

Ejecutado Permanente:	Movimiento fijo del presupuesto dado por los sueldos base de los cargos
Ejecutado Temporal:	Movimiento transitorio del presupuesto, dado por extensiones horarias transitoria y compensaciones temporales sobre los sueldos base
Vacantes:	Movimiento de dinero para financiar cargos que no están siendo ocupados por ningún docente
Reducciones permanentes:	Movimiento del presupuesto dado por variaciones al ejecutado permanente debido a reducciones en el horario base de los cargos
Reducciones temporales:	Movimientos del presupuesto dado por variaciones transitorias al ejecutado permanente
Comprometido con disponibilidad:	Movimiento del presupuesto para financiar cargos que aún no han sido ocupados por docentes (cargos llamados) pero que en un futuro lo serán.
Comprometido sin disponibilidad:	Cargos que no han sido ocupados por docentes (cargos llamados) aún, pero no se dispone del dinero para financiarlos.

## B. MEDIDAS

- Gr1-10hs: Valor equivalente al sueldo base de un grado 1, 10hs.
- Pesos Uruguayos: Valor expresado en pesos uruguayos
- Ahorro (Gr1-10hs): Mide el ahorro en el equivalente gr1-10hs
- Ahorro (Pesos): Mide el ahorro en pesos uruguayos
- Gasto (Gr1-10hs): Mide el gasto en el equivalente gr1-10hs
- Gasto (Pesos): Mide el gasto en pesos uruguayos
- Cantidad horas: Cantidad de horas registradas
- Cantidad docentes: Cantidad de docentes asignados.
- Maximo (Gr1-10hs): Máximo a ejecutar en el equivalente gr1\_10hs
- Maximo (Pesos): Máximo a ejecutar en pesos uruguayos

### C. RELACIONES MULTIDIMENSIONALES ENTRE DIMENSIONES Y MEDIDAS

Dados un instituto, una fuente de financiación, un concepto ejecutado, un docente, un cargo, un grado, una cantidad de horas, podemos medir a lo largo del tiempo el *valor monetario* involucrado, el aporte al *ahorro* y el aporte al *gasto* (tanto en pesos como en el equivalente gr1\_10hs) y la cantidad de docentes. Los conceptos posibles de cruzar son todos menos vacante, comprometido con disponibilidad y comprometido sin disponibilidad que se detallan más adelante.

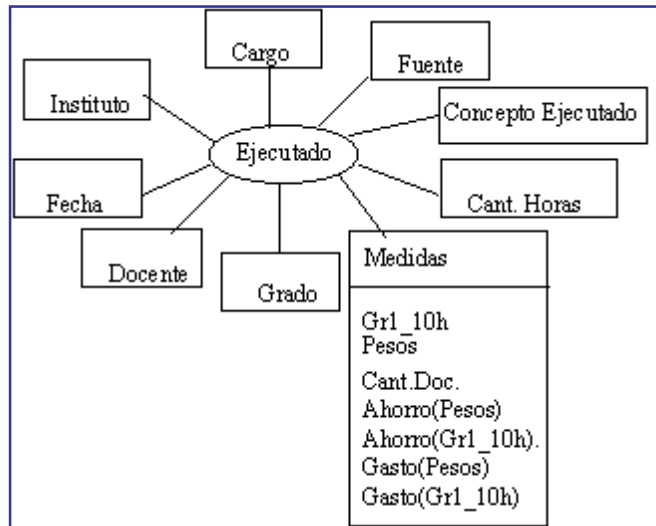


Figura 1: Relación dimensional Ejecutado

Cuando el concepto es vacante, no tiene sentido cruzar la información con la dimensión docente ya que éste tipo de cargo no tiene un docente asignado. De todas formas se pueden medir dados un instituto, un cargo, una fuente de financiación, un grado, una cantidad de horas, el *valor monetario involucrado*, el *ahorro* y el *gasto* (tanto en pesos como en el equivalente gr1\_10hs), a lo largo del tiempo.

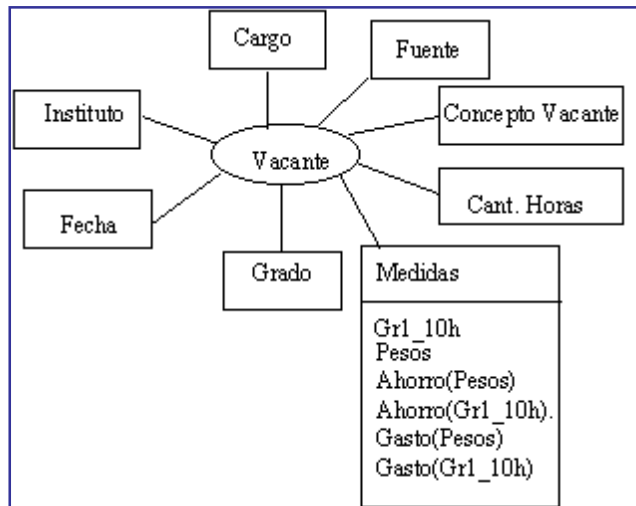


Figura 2: Relación dimensional Vacante

Si el concepto es comprometido con disponibilidad no se lo puede cruzar con información del docente ya que los cargos que son llamados no tienen asignado (formalmente) un docente. De todas formas, se puede medir a lo largo del tiempo, dados: la fuente de financiación, el instituto, el cargo, el grado y la cantidad de horas, el *valor monetario* involucrado, el *ahorro* y el *gasto* (tanto en pesos como en el equivalente gr1\_10hs)

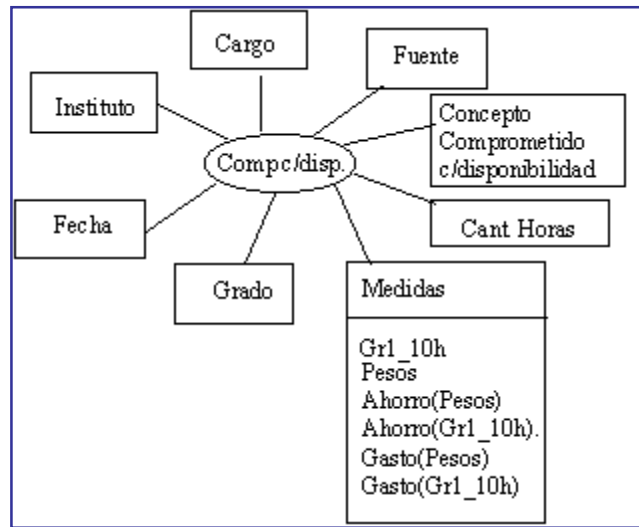


Figura 3: Relación dimensional Comp c/disp

Si el concepto es comprometido sin disponibilidad, no tiene sentido cruzarlo con la información de fuentes de financiación (ya que aún no se dispone de fuente para financiarlo) ni con la información de docentes (por no tener aún un docente asignado). Sí tiene sentido medir a lo largo del tiempo, el *valor monetario* involucrado, el *ahorro* y el *gasto* (tanto en pesos como en el equivalente gr1\_10hs) dados un instituto, un cargo, una cantidad de horas y un grado.

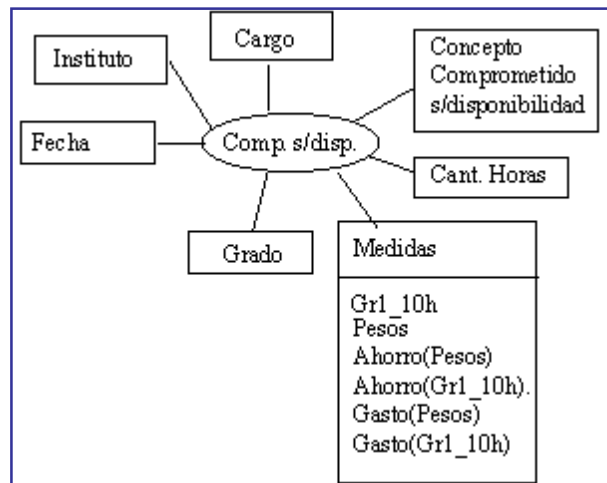


Figura 4: Relación Dimensional Comp s/disp

Interesa saber, dado un instituto, una fuente de financiación y una fecha, el *máximo posible* a ejecutar (gastar) tanto en la medida pesos como en el equivalente gr1\_10hs.

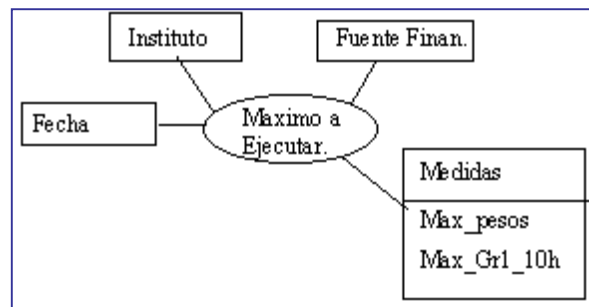


Figura 5: Rel. Dimensional Máximo a Ejecutar

#### D. CRUZAMIENTO INTERESANTE

Interesa medir, dados un concepto, una fuente de financiación, un docente, un grado y una cantidad de horas, el acumulado mensual ahorrado y el acumulado mensual gastado, expresado tanto en pesos como en el equivalente gr1\_10hs. Valen para éste cruzamiento, las restricciones mencionadas anteriormente. No tiene sentido cruzar el concepto comprometido con disponibilidad con información de docente y tampoco cruzar el concepto comprometido sin disponibilidad con fuentes de financiación y/o docentes. Por último, el concepto vacante no puede cruzarse con información de docentes.

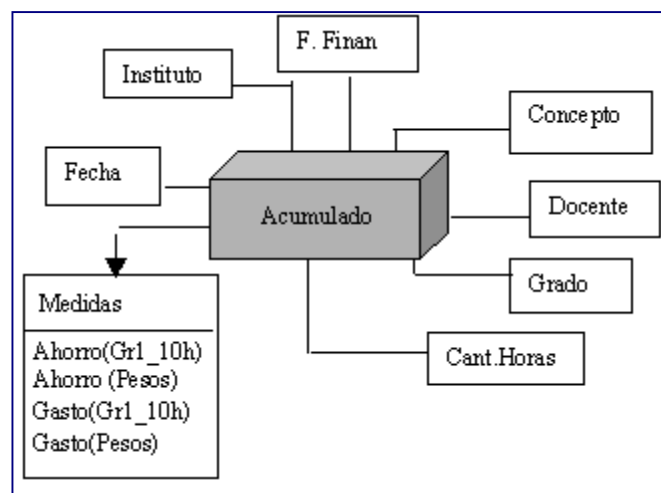


Figura 6: *Cruzamiento Interesante*



## ➤ RELEVAMIENTO DE FUENTES

---

*Se releva información sobre los sistemas operacionales utilizados en la Sección de Personal de la F.I. para llevar el presupuesto de la facultad. Se analiza básicamente la funcionalidad de los sistemas y el uso que se les da a los mismos.*

---

La sección de Personal de la Facultad de Ingeniería trabaja actualmente con dos sistemas operacionales diferentes, los cuales identificamos con el nombre 'Sistema original' y 'Sistema nuevo'. El 'Sistema Original' como lo dice el nombre fue el primer sistema utilizado por la sección de personal y actualmente es manejado por una sola persona dentro de dicha sección la cual es la única que sabe utilizarlo. El 'Sistema Nuevo', es un sistema que fue diseñado para ser utilizado en todas las facultades para llevar el presupuesto. La F.I. tiene algunos puntos de diferencia con el resto de las facultades en cuanto a como organiza el presupuesto y por ende la operativa del 'Sistema Nuevo' no satisface completamente las necesidades de la misma.

### **Sistema Original**

Es un sistema elemental, desarrollado en 4Pro sobre DBase para llevar el presupuesto de la FI. Contiene información actual acerca de los cargos para todos los servicios de la facultad. Maneja información de cargos ejecutados, comprometidos y vacantes, como ser el docente asignado, su grado en el cargo y la cantidad de horas actuales del cargo. También contiene información del sueldo asociado a cada combinación posible de grado/hora.

El sistema no maneja el tema de extensiones y reducciones horarias, ni las compensaciones sobre el sueldo, así como tampoco lleva registro de las diferentes fuentes de financiación de los cargos. Al no adaptarse a los requerimientos específicos del presupuesto de la facultad, es usado de una forma particular, diferente de aquella para la cual fue diseñado.

El sistema, no tiene capacidad de registrar los movimientos efectuados sobre los cargos, los mismos son llevados de forma manual en planillas. El personal de la sección guarda estas planillas mensualmente y se encuentran archivadas desde que se comenzó a usar el sistema.

## Sistema Nuevo

Es un sistema realizado en Clipper sobre DBase para llevar el detalle del presupuesto liquidado, esto es sólo mantiene información sobre cargos ejecutados. Almacena datos actuales de los cargos como ser departamento, funcionario asignado, grado, horas y fuente de financiación. También lleva un registro de todo los movimientos efectuados sobre los cargos desde el momento de creación.

A diferencia del sistema original, fue diseñado para manejar las reducciones y extensiones horarias de los cargos, tanto transitorias como permanentes así como también las compensaciones sobre el sueldo. Maneja mucha información sobre los funcionarios la cual no es de interés para nuestro proyecto.

A pesar de ser un sistema más completo, no sustituye al sistema original porque no maneja los cargos comprometidos y vacantes. La información almacenada en el histórico es muy antigua pero sólo es válida a partir del año 1998, los datos de años anteriores fueron ingresados manualmente y no en forma completa.



Toda la información recabada en esta fase se obtuvo mediante entrevistas con los funcionarios de la sección de Personal. Las entrevistas fueron nuestro primer contacto con los funcionarios y nos brindaron una primera visión general sobre el negocio.

## 1.2 ANALISIS

---

Para realizar el análisis se necesita información sobre los sistemas operativos (OLTP) de la sección Personal, tanto a nivel de base de datos como de operativa, e información sobre las reglas del negocio de dicha sección. Sin embargo, de la operativa del sistema no existe ningún tipo de documentación y las reglas del negocio de la sección tampoco están documentadas. Con lo único que se cuenta para realizar el análisis es el contenido de las bases de datos y el 'knowhow' de los funcionarios.

Durante las tres primeras fases del análisis, se realizaron reuniones esporádicas con funcionarios de la sección de Personal donde se aclararon dudas a diferente nivel de detalle que fueron surgiendo en cada una de dichas fases. Siempre las aclaraciones fueron sobre las reglas del negocio y cómo los funcionarios manejan los sistemas operacionales, nunca acerca del funcionamiento interno de los sistemas.

### ➤ ANÁLISIS GENERAL

---

*Se estudia la información que manejan los sistemas operacionales o sea sus bases de datos. Con qué tablas cuentan, que datos tienen y cómo éstos se relacionan. Esta fase nos brinda información sobre los sistemas operacionales y las reglas básicas del negocio.*

---

Se realiza un análisis general del contenido de las bases de datos de los sistemas fuente. En este punto se plantea un resumen del análisis, especificando el contenido general de cada tabla, el análisis general completo puede verse en el *Apartado2, Metadata del sistema*.

Como no existe documentación sobre los modelos conceptuales de las bases de datos de los sistemas operacionales de la sección de Personal, los mismos se construyen en ésta fase.

## A. ESTUDIO BASES DE DATOS FUENTES

### Sistema Original

Se analiza la estructura de todas las tablas de la B.D. del Sistema Original para obtener información general sobre las tablas que existen y los datos que mantienen.

TABLA/COLUMNA	DESCRIPCION
<i>MAESER</i>	Servicios de la Facultad de Ingeniería.
<i>PERSONAL</i>	Datos sobre cargos comprometidos, vacantes y ejecutados válidos a la fecha.
<i>SUELDOS</i>	Sueldos asociados a cada grado/hora.

### Sistema Nuevo

Se analiza la estructura de todas las tablas de la B.D. del Sistema Nuevo para obtener información general sobre las tablas que existen y los datos que mantienen.

TABLA	DESCRIPCION CONTENIDO
<i>WCGOS</i>	Datos de los cargos ejecutados válidos a la fecha.
<i>WEXTRED</i>	Datos sobre las extensiones/reducciones horarias transitorias de los cargos válidas a la fecha.
<i>WDEDIC</i>	Datos sobre las compensaciones a los funcionarios válidas a la fecha.
<i>WHIST</i>	Historia de los movimientos realizados sobre un cargo desde su creación. Además de registrar los movimientos que generaron datos en las tablas antes descritas (creación del cargo, extensión/reducción transitoria, compensaciones), registra otros movimientos como ext/red permanentes, prórrogas del cargo, prórrogas de las compensaciones, prórrogas de las ext/red, etc. Datos sobre la historia de los cargos.
<i>WINST</i>	Institutos de la Facultad de Ingeniería.
<i>WFINANC</i>	Fuentes de financiación de los cargos.
<i>WPROVIS</i>	Provisión de los cargos.



La información recabada en esta fase se obtiene a partir del estudio de las bases de los sistemas fuentes y algunas entrevistas con los funcionarios de la sección de Personal, en donde se corroboran las deducciones realizadas. Para poder estudiar las bases nos fueron facilitados por la sección, los archivos que contienen los datos de las mismas.

## B. MODELOS CONCEPTUALES BASES DE DATOS FUENTES

Como no existe documentación acerca del modelo conceptual de ninguna de las bases fuentes, éstos se construyen a partir de las estructuras físicas de las bases de datos aplicando las técnicas de ingeniería reversa.

En el *Apartado2, Metadata del Sistema* se documenta el mapeo entre los objetos conceptuales de ambos modelos y las estructuras físicas de las bases de datos.

### Sistema Original

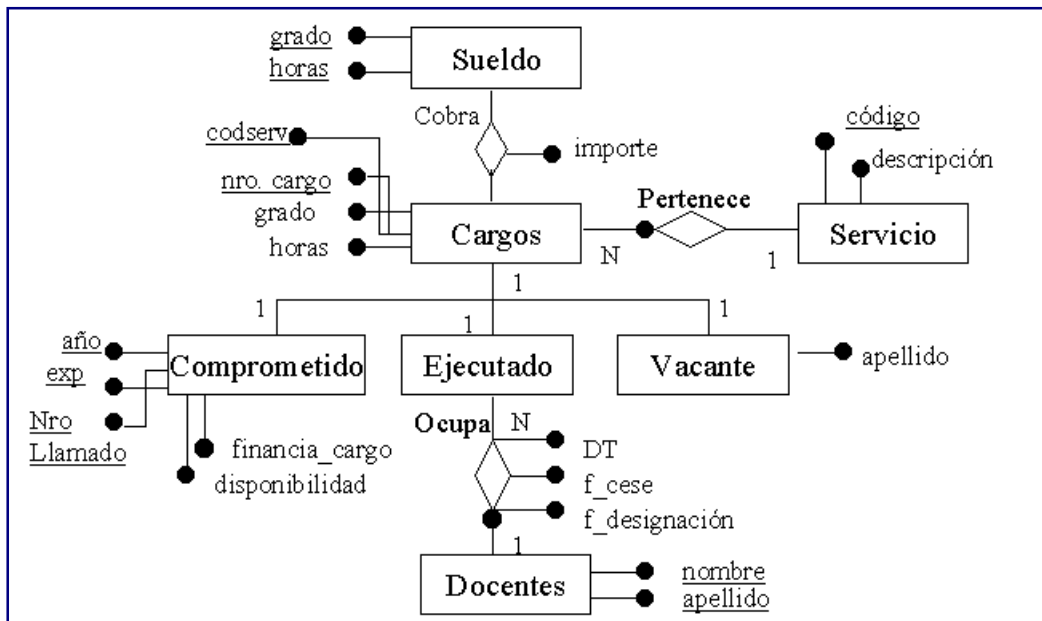


Figura 7: Modelo Conceptual Sistema Original de Presupuesto

Existen tres tipos de cargos: comprometidos, ejecutados y vacantes. Cada cargo se identifica por un número de cargo y tiene grado y horas determinados.

El presupuesto de Facultad se lleva por servicio. De los servicios se conoce el código que lo identifica y su descripción.

Cada cargo pertenece a un solo servicio de la Facultad y no puede existir un cargo que no pertenezca a un servicio. Sin embargo puede haber más de un cargo por servicio.

De los cargos comprometidos se sabe el número de llamado, el expediente y año del llamado, si financia un cargo ejecutado y si hay disponibilidad de dinero.

De los cargos vacantes se conoce el apellido del docente que ocupaba el cargo cuando este quedó vacante.

Los cargos ejecutados tienen docentes asignados a ellos. Sólo puede haber un docente asignado a un cargo.

Los docentes tienen nombre y apellido. Un docente puede ocupar más de un cargo y todos los docentes ocupan por lo menos uno.

De los docentes en un cargo se sabe la fecha de posesión del cargo, la fecha de cese del cargo y si tiene dedicación total al cargo.

Cada combinación grado–hora tiene un importe de sueldo asociado.

### Sistema Nuevo

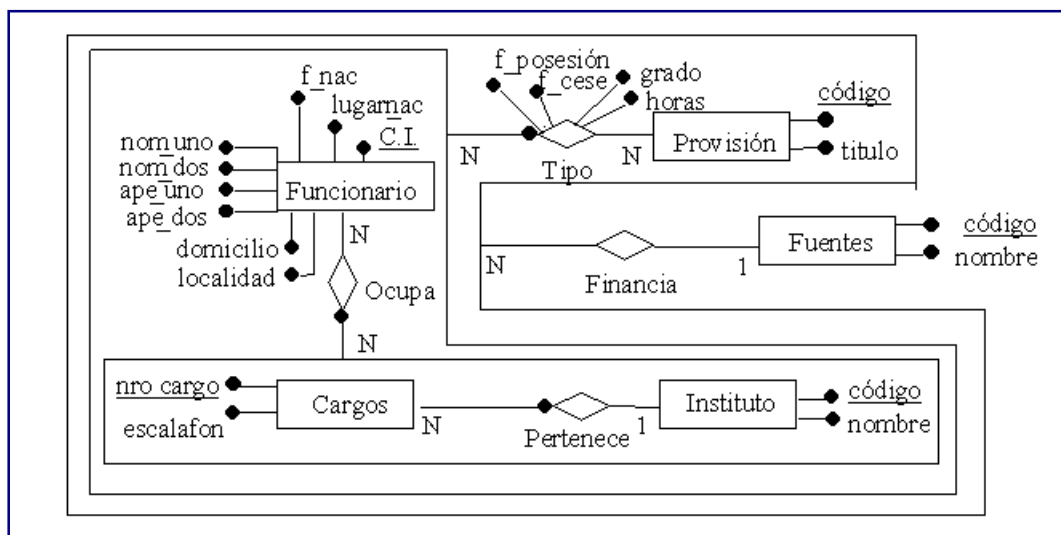


Figura 8: Modelo Conceptual Sistema Nuevo de Presupuesto

El presupuesto de la Facultad es llevado por instituto. De cada instituto se conoce el código que lo identifica y el nombre. Dentro de un instituto existen diferentes cargos.

De los cargos se sabe, el número de cargo que lo identifica y su escalafón. Un cargo debe pertenecer a un solo instituto.

Los funcionarios se identifican por su cédula de identidad, de ellos se conoce su fecha y lugar de nacimiento, nombres, apellidos, domicilio y localidad.

Un funcionario puede ocupar más de un cargo. Todo cargo tiene por lo menos un funcionario asignado.

Existen provisiones de las cuales conocemos el código que las identifica y un título.

Un funcionario ocupando un cargo se asocia con por lo menos una provisión, en un período dado por la fecha de posesión y cese, con un grado y horas determinado.

Una provisión se relaciona con distintos cargos ocupados por funcionarios.

Existen diferentes fuentes de financiación, todas ellas identificadas por un código, se conoce además su nombre. Cada fuente puede financiar las diferentes provisiones de un cargo ocupado por un funcionario.

Todo cargo ocupado por un funcionario con una provisión tiene una fuente que lo financia.



Los modelos conceptuales se realizaron aplicando técnicas de reingeniería de datos. No tuvimos apoyo por parte de los funcionarios de la sección Personal ya que no tenían conocimiento técnico de los sistemas.

### ➤ IDENTIFICACIÓN DE DATOS REQUERIDOS.

---

*Se analizan las dimensiones y medidas documentados en el modelo conceptual multidimensional y se determina si existe información almacenada al respecto, consultando los modelos conceptuales de los sistemas fuente. En caso que exista dicha información, se define cual es el objeto del modelo conceptual que la mantiene.*

---

Dado que existen dos sistemas fuente, Sistema Original y Sistema Nuevo, se realizan dos cuadros para las *Dimensiones* y otros dos para las *Medidas* (uno por cada sistema fuente).

Para no repetir la lista de objetos conceptuales se numeran según el siguiente cuadro.

Sistema Original	Sistema Nuevo
1. Cargos	1. Cargos
2. Comprometido	2. Provision
3. Ejecutado	3. Fuentes
4. Vacante	4. Financia
5. Docentes	5. Tipo
6. Sueldo	6. Pertenece
7. Servicio	7. Instituto
8. Ocupa	8. Funcionario
9. Pertenece	9. Ocupa

**A. DIMENSIONES**

O B J E T O S																			
D I M E N S I O N		Sistema Original									Sistema Nuevo								
		1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
	Instituto						x										x		
	Dpto						x												
	Seccion						x												
	Grado	x													x				
	Fuente Financ			X								x							
	Concepto																		
	Cargos	X								X									
	Docente					X												X	
	Cant Horas	X													X				

Figura 9: Mapeo entre dimensiones y objetos conceptuales

**B. MEDIDAS**

O B J E T O S																			
M E D I D A S		Sistema Original									Sistema Nuevo								
		1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
	Gr110hs	X					X												
	Pesos	X					X												
	Cant. Docente					X												X	
	Gasto (gr110hs)	X		X															
	Gasto (pesos)	X		X															
	Ahorro (g110hs)	X	X	X	X														
	Ahorro (pesos)	X	X	X	X														
	Maximo																		

Figura 10: Mapeo entre medidas y objetos conceptuales



**C. CRUZAMIENTOS**

**Sistema Original**

DIMENSION																
D I M E N S I O N		Instituto	Dpto	Seccion	Grado	Fuente Financ	Ejecutado Permanente	Ejecutado Temporal	Comp s/disp	Comp c/disp	Vacante	Reducción Temporal	Reducción Permanente	Cargos	Docente	Cant. Horas
	Instituto		{7}	{7}	{7,9, 1}	{7,9, 1,3}	{7,9, 1,3}	{7,9, 1,3}	{7,9, 1,2}	{7,9, 1,2}	{7,9, 1,4}	{7,9, 1,3}	{7,9, 1,3}	{7,9, 1}	{7,9, 1,3, 8,5}	{7,9, 1}
	Dpto				{7,9, 1}	{7,9, 1,3}	{7,9, 1,3}	{7,9, 1,3}	{7,9, 1,2}	{7,9, 1,2}	{7,9, 1,4}	{7,9, 1,3}	{7,9, 1,3}	{7,9, 1}	{7,9, 1,3, 8,5}	{7,9, 1}
	Seccion				{7,9, 1}	{7,9, 1,3}	{7,9, 1,3}	{7,9, 1,3}	{7,9, 1,2}	{7,9, 1,2}	{7,9, 1,4}	{7,9, 1,3}	{7,9, 1,3}	{7,9, 1}	{7,9, 1,3, 8,5}	{7,9, 1}
	Grado					{1,3}	{1,3}	{1,3}	{1,2}	{1,2}	{1,4}	{1,3}	{1,3}	{1}	{1,3, 8,5}	{1}
	Fuente Financ						{3}	{3}				{3}	{3}	{1,3}	{3,8, 5}	{1,3}
	Ej. Perm.													{1,3}	{3,8, 5}	{1,3}
	Ej. Temp.													{1,3}	{3,8, 5}	{1,3}
	Comp s/disp													{1,2}		{1,2}
	Comp c/disp													{1,2}		{1,2}
	Vacante													{1,4}		{1,4}
	Red. Temporal													{1,3}	{3,8, 5}	{1,3}
	Red. Perm.													{1,3}	{3,8, 5}	{1,3}
	Cargos														{1,3, 8,5}	{1}
	Docente															{1,3, 8,5}
	Cant Horas															

Figura 11: Mapeo entre cruzamientos y objetos conceptuales S.O

Sistema Nuevo

DIMENSION																	
D		Instituto	Dpto	Seccion	Grado	Fuente Financ	Ejecutado Permanente	Ejecutado Temporal	Comp s/disp	Comp c/disp	Vacante	Reduccion Temporal	Reduccion Permanente	Cargos	Docente	Cant. Horas	
I	Instituto				{7,6 .1,9, 8,5}	{7,6 .1,9, 8,5, 2,4, 3}	{7,6 .1,9, 8,5, 2}	{7,6 .1,9, 8,5, 2}				{7,6 .1,9, 8,5, 2}	{7,6 .1,9, 8,5, 2}	{7,6 .1, }	{7,6 .1,9, 8}	{7,6 .1,9, 8,5}	
M	Dpto																
E	Seccion																
N	Grado					{7,6 .1,9, 8,5, 2,4, 3}	{5,2 }	{5,2 }					{5,2 }	{5,2 }	{1,6 .7,9, 8,5}	{1,6 .7,9, 8,5}	5
S	Fuente Financ						{7,6 .1,9, 8,5, 2,4, 3}	{7,6 .1,9, 8,5, 2,4, 3}				{7,6 .1,9, 8,5, 2,4, 3}	{7,6 .1,9, 8,5, 2,4, 3}	{7,6 .1,9, 8,5, 2,4, 3}	{7,6 .1,9, 8,5, 2,4, 3}	{7,6 .1,9, 8,5, 2,4, 3}	
I	Ej. Perm.													{7,6 .1,9, 8,5, 2}	{7,6 .1,9, 8,5, 2}	{5,2 }	
O	Ej. Temp.													{7,6 .1,9, 8,5, 2}	{7,6 .1,9, 8,5, 2}	{5,2 }	
N	Comp s/disp																
	Comp c/disp																
	Vacante																
	Red. Temporal													{7,6 .1,9, 8,5, 2}	{7,6 .1,9, 8,5, 2}	{5,2 }	
	Red. Perm.													{7,6 .1,9, 8,5, 2}	{7,6 .1,9, 8,5, 2}	{5,2 }	
	Cargos														{1,6 .7,9, 8}	{1,6 .7,9, 8,5}	
	Docente															{1,6 .7,9, 8,5}	
	Cant Horas																

Figura 12: Mapeo entre cruzamientos y objetos conceptuales S.N.

En los cuadros anteriores se identifican los objetos conceptuales que manejan información asociada a cada cruzamiento entre dimensiones. La identificación debe leerse de izquierda a derecha y de arriba hacia abajo. Para no repetir ocurrencias se analizan los cruzamientos correspondientes a la mitad derecha superior a la diagonal.

Los cuadros marcados en gris representan los cruzamientos válidos según el modelo conceptual multidimensional de Presupuesto. Los números dentro de cada cuadro gris corresponden a los objetos del modelo conceptual (ver cuadro de referencia) que mantienen información necesaria para el cruzamiento. Los cuadros grises vacíos representan cruzamientos que tienen sentido según el modelo conceptual multidimensional pero que no son mantenidos en el sistema fuente analizado.

La dimensión Concepto es disgregada en cada uno de sus valores para ser analizada. Se disgrega la dimensión porque de otra manera no se podría visualizar que la información asociada a cada valor se encuentra en diferentes objetos conceptuales.

## ➤ ANÁLISIS DETALLADO

---

*Se realiza un análisis minucioso de los datos en cada tabla, estudiando todos los valores posibles en cada campo y cualquier otro dato que pueda estar oculto detrás de los mismos.*

---

Se puede encontrar la documentación del análisis detallado tanto del Sistema Original como del Sistema Nuevo en el *Apartado 2, Metadatos del sistema.*

El siguiente es un resumen general de la información que se obtuvo de dicho análisis, donde se destacan los problemas de calidad más importantes encontrados en las fuentes de datos analizadas.

Es importante destacar que si bien algunos problemas tienen que ver con valores inválidos en los datos, otros tienen que ver con reglas del negocio ocultas en los mismos. Este tipo de información también es importante a la hora de realizar un resumen general del estado de calidad de los datos para una fuente de datos determinada.

### Sistema Original

El estado de los datos en este sistema no es muy confiable ya que algunos de sus campos se utilizan de manera poco predecible. En particular, para los cargos ejecutados mucha de la información requerida se encuentra perdida dentro de campos de texto.

#### Personal.dbf

Esta tabla almacena información sobre cargos comprometidos, ejecutados y vacantes. Mantiene gran parte de los datos requeridos pero el problema es saber interpretarlos ya que éstos se encuentran dispersos en los distintos campos de la tabla. Los datos con más problemas de calidad son los correspondientes a cargos ejecutados.

Puntos a tener en cuenta:

- El *grado* debe ser un número en el rango 1..5.
- Los cargos que comienzan con '2' son corresponden a cargos comprometidos sin disponibilidad del dinero. Los demás cargos comienzan con '6'.
- Un valor 41 en el campo *horas*, indica un docente con dedicación total.
- Los campos de texto *apellido* y *nombre* son usados a veces para almacenar otra información además del nombre y apellido del docente. Como ser, extensiones horarias, dedicaciones, reducciones horarias, número de llamado, fuente de financiación, y otro tipo de información sin clara utilidad. Los valores en los campos no siguen un estándar definido lo cual dificulta aún más su entendimiento.
- Las fechas de *designación* y *cese*, tienen el año de 2 cifras en vez de 4.
- Los campos nombre y apellido almacenan en ocasiones información acerca de compensaciones, extensiones o reducciones de los cargos. La información no cumple ningún estándar ni se asegura que sea completa y actual.

- Los campos nombre y apellido almacenan en ocasiones información sobre la fuente de financiación. La información no cumple ningún estándar ni se asegura que sea completa y actual.

### **Maeserv.dbf**

Esta tabla almacena información sobre los servicios de la Facultad de Ingeniería. Decodificando el número de servicio según ciertas reglas, se obtiene el instituto y departamento al cual pertenecen. Las reglas de codificación no siguen un estándar.

### **Sueldos.dbf**

Se encuentra en la tabla *sueldos.dbf* información acerca de los sueldos correspondientes a cada pareja grado/hora. La forma en que se encuentra almacenada dicha información es muy particular. Las columnas se corresponden con los grados posibles y las filas con las horas. Por ejemplo en el lugar col:1, fila:20 se encuentra el sueldo asociado a un grado 1 de 20hs semanales.

## **Sistema Nuevo**

### **Wcgos.dbf**

Aquí se encuentran datos correspondientes a cargos ejecutados vigentes al momento. Mantiene información acerca del docente a cargo, escalafón del docente, grado, horas, fuente de financiación del cargo y fechas de posesión y cese del mismo. Si bien se encontraron casos no normales, el estado de calidad de los datos es bueno.

Puntos a tener en cuenta:

- El campo horas registra las horas base del cargo.
- Solo interesan cargos con el campo escalafón en 'G' ya que estos corresponden a cargos de funcionarios docentes.
- El grado para cargos docentes debe estar dentro del rango 1..5.
- La fecha de posesión debe ser menor o igual a la de cese.
- Tiene más importancia la existencia de un cargo en la tabla que el contenido de las fechas de posesión y cese. Si el registro existe, el cargo está vigente.
- El año en las fechas es numérico de 2, debería ser numérico de 4.
- El identificador del cargo es algunas veces un número y otras un string según el tipo del cargo en cuestión.
- Cuando un cargo se prorroga, se lo mantiene en la tabla modificándole la fecha de cese. Hay casos en que se aplica la prórroga antes de recibir la resolución del consejo. Si el consejo no aprueba la prórroga, no se registra en el sistema la pérdida del dinero desembolsado.

### **Wextred.dbf**

Se almacenan en esta tabla la información de las extensiones o reducciones horarias transitorias vigentes al momento. De las mismas, se detalla información sobre el cargo al cual afectan, instituto, fuente de financiación, docente a cargo, escalafón, grado y horas finales del cargo luego de aplicada la extensión o reducción. El estado de calidad de datos es relativamente bueno pero se encontraron ciertas inconsistencias.

Puntos a tener en cuenta:

- El cargo al cual afecta la ext/red debe existir en la tabla Wcgos.dbf.
- El docente registrado en la ext/red debe existir en la tabla Wfuncs.dbf.
- La fecha de posesión debe ser menor o igual a la de cese.
- Tiene más importancia la existencia de la ext/red en la tabla que el contenido de las fechas de posesión y cese. Si el registro existe, la ext/red está vigente.
- El año en las fechas es numérico de 2, debería ser numérico de 4.
- Una reducción es válida si disminuye las horas base del cargo.
- Una extensión es válida si aumenta las horas base del cargo. Si existe otra extensión es válida si aumenta las horas extendidas de la misma.
- Las reducciones se dan sobre la misma fuente de financiación del cargo.
- No puede haber más de una reducción transitoria vigente para un mismo cargo. Si existe más de una reducción sobre un cargo, la de mayor horas es la válida.
- Puede haber más de una extensión horaria sobre el mismo cargo con diferente fuente de financiación. Si existe más de una extensión sobre un cargo para la misma fuente de financiación, la de mayor horas es la válida.
- La cantidad de horas extendidas se pueden calcular a partir de las horas registradas en la extensión menos las horas base del cargo correspondiente. Si existe más de una extensión, la cantidad de horas extendidas se calculan como la diferencia de horas entre ambas extensiones.
- La cantidad de horas reducidas se pueden calcular a partir de las horas base del cargo menos las horas registradas en la reducción.
- En caso que exista una reducción y alguna extensión, la extensión de menor cantidad de horas absorbe la diferencia entre las horas base del cargo y las horas reducidas.
- En caso de recibir extensiones realizadas sobre un cargo que cesó, se crea el cargo nuevamente con la misma información original, se registra la extensión y se da de baja el cargo de forma inmediata. Es por esto que la información queda registrada únicamente en el historial del sistema.

### **Wdedic.dbf**

Esta tabla mantiene información sobre las compensaciones sobre los sueldos de los cargos vigentes al momento. Se encuentran datos sobre el cargo al cual corresponde la compensación, tipo de compensación, docente a cargo, departamento, fuente de financiación, escalafón, grado y fechas de posesión y cese.

Puntos a tener en cuenta:

- El cargo al cual afecta la dedicación debe existir en la tabla Wcgos.dbf
- El docente registrado en la dedicación debe existir en la tabla Wfuncs.dbf.
- La fecha de posesión debe ser menor o igual a la de cese
- El año en las fechas debería ser numérico de 4 y no de 2
- El valor en el campo *horas* de la dedicación, no tienen mucho sentido. El valor correspondiente a la dedicación se calcula como un porcentaje sobre un base de cálculo los cuales varían según el tipo de dedicación (total, parcial, compensada, etc.)
- El valor en el campo porcentaje, debería tener el porcentaje a partir del cual se calcula el valor correspondiente a la dedicación, pero éste no tiene datos correctos.

### **Whist.dbf**

Se encuentran en dicha tabla todos los movimientos efectuados sobre un cargo desde su creación. La misma almacena datos desde el año 1977 pero sólo los registros más recientes tienen cierto nivel de consistencia. De todas formas, la información guardada en esta tabla no es muy confiable ya que se registran en ella tanto los movimientos correctos como los incorrectos. Esto es, la actualización de esta tabla es hecha por el sistema de forma automática y transparente al usuario. El problema es que como el sistema no le provee al usuario de todos los mecanismos necesarios para llevar a cabo el negocio, a veces el usuario debe ‘engañar’ al sistema para poder entrar la información deseada. Este tipo de acción es la que empobrece el nivel de calidad de datos de esta tabla. No se recomienda el uso de la misma.

### **Wfinanc.dbf**

Mantiene las fuentes de financiación de cargos de la Facultad de Ingeniería.

Puntos a tener en cuenta:

- Las descripciones de las fuentes no son del todo claras e informativas.
- Existen fuentes de financiación denominadas ‘honorario’. Todo cargo financiado por dichas fuentes es un cargo honorario.

### **Winst.dbf**

Esta tabla mantiene datos sobre institutos de la F.I. También maneja información de algunos departamentos, secciones, unidades y cátedras de la F.I. No se registra la relación entre estos últimos y los institutos en caso que la hubiera.

Puntos a tener en cuenta:

- Maneja al mismo nivel información correspondiente a institutos, departamentos, secciones, unidades y cátedras.
- Existen institutos denominados ‘becarios’, ‘becario no docente’, ‘docente honorario’, etc. Todo cargo que tenga como instituto dichos códigos es un cargo que no recibe sueldo.

### **Wfuncs.dbf**

Aquí se encuentra información sobre los funcionarios de la facultad, tanto docentes como no docentes.

Puntos a tener en cuenta:

- El nombre se encuentra distribuido en cuatro campos correspondientes al primer y segundo nombre y al primer y segundo apellido.
- Los mismos se identifican por la cédula de identidad.



Según la metodología propuesta para aseguramiento de la calidad de los datos, el análisis detallado brinda un conocimiento total del estado de calidad de los datos fuentes y es la base para definir las rutinas de calidad necesarias. En este caso, además de lo anteriormente dicho, sirve para completar el conocimiento de los sistemas operacionales y reglas de negocio. Permite entender a fondo cómo se manejan los sistemas frente a casos excepcionales no previstos por los mismos, casos no documentados y que difícilmente surjan en una entrevista dado el escaso conocimiento inicial del negocio.



## ➤ SELECCIÓN DE DATOS

*Se seleccionan los objetos a extraer de las bases fuente. Se construye un modelo conceptual que permite visualizar los datos con los que se cuenta para cargar el futuro data mart.*

La selección de datos se realiza en base a la información recabada en las etapas de *Análisis* anteriores, principalmente a partir del *Análisis Detallado* de los datos.

Los datos seleccionados se encuentran tanto en la base de datos del Sistema Original como en la del Sistema Nuevo. A nivel general, la información de los cargos vacantes y comprometidos se toma del S.O. y la del ejecutado, reducciones, extensiones y docentes del S.N. La complejidad de la selección en este caso es que la información se encuentra duplicada en ambos sistemas operacionales, debiéndose hacer una integración.

En los cuadros de las *figuras 13.a y 13.b* se visualizan los sistemas de donde se extraerá información para cargar cada dato requerido (dimensiones, medidas y cruzamientos)

		Sistema Original	Sistema Nuevo
D I M E N S I O N	Instituto	X	X
	Dpto		
	Seccion		
	Grado	X	X
	Fuente Financ		X
	Concepto		
	Cant. Horas	X	X
M E D I D A	Cargos	X	X
	Gr110hs	X	X
	Cant. Docente		X
	Gasto Mes	X	X
Ahorro Mes	X	X	
Máximo			

*Figura 13.a: Mapeo entre objetos multidimensionales y fuentes de datos*

DIMENSION																
D I M E N S I O N		Instituto	Dpto	Seccion	Grado	Fuente Financ	Ejecutado Permanente	Ejecutado Temporal	Comp s/disp	Comp c/disp	Vacante	Reducción Temporal	Reducción Permanente	Cargos	Docente	Cant. Horas
	Instituto				S.N, S.O	S.N	S.N	S.N	S.O	S.O	S.O	S.N	S.N	S.N, S.O.	S.N	S.N, S.O
	Dpto															
	Seccion															
	Grado					S.N.	S.N.	S.N.	S.O.	S.O.	S.O.	S.N.	S.N.	S.N, S.O.	S.N	S.N, S.O.
	Fuente Financ						S.N.	S.N.				S.N.	S.N.	S.N.	S.N.	S.N.
	Ej. Perm.													S.N.	S.N.	S.N.
	Ej. Temp.													S.N.	S.N.	S.N.
	Comp s/disp													S.O.		S.O.
	Comp c/disp													S.O.		S.O.
	Vacante													S.O.		S.O.
	Red. Temporal													S.N.	S.N.	S.N.
	Red. Perm.													S.N.	S.N.	S.N.
	Cargos														S.N.	S.N, S.O.
	Docente															S.N.
	Cant Horas															

Figura 13.b: Mapeo entre objetos multidimensionales y fuentes de datos

### A. MODELO CONCEPTUAL DE FUENTES SELECCIONADAS

El siguiente modelo permite visualizar el conjunto de entidades seleccionadas para cargar el data mart de Presupuesto. El modelo se construye a partir de los modelos conceptuales de los sistemas fuente y la selección de fuentes realizada.

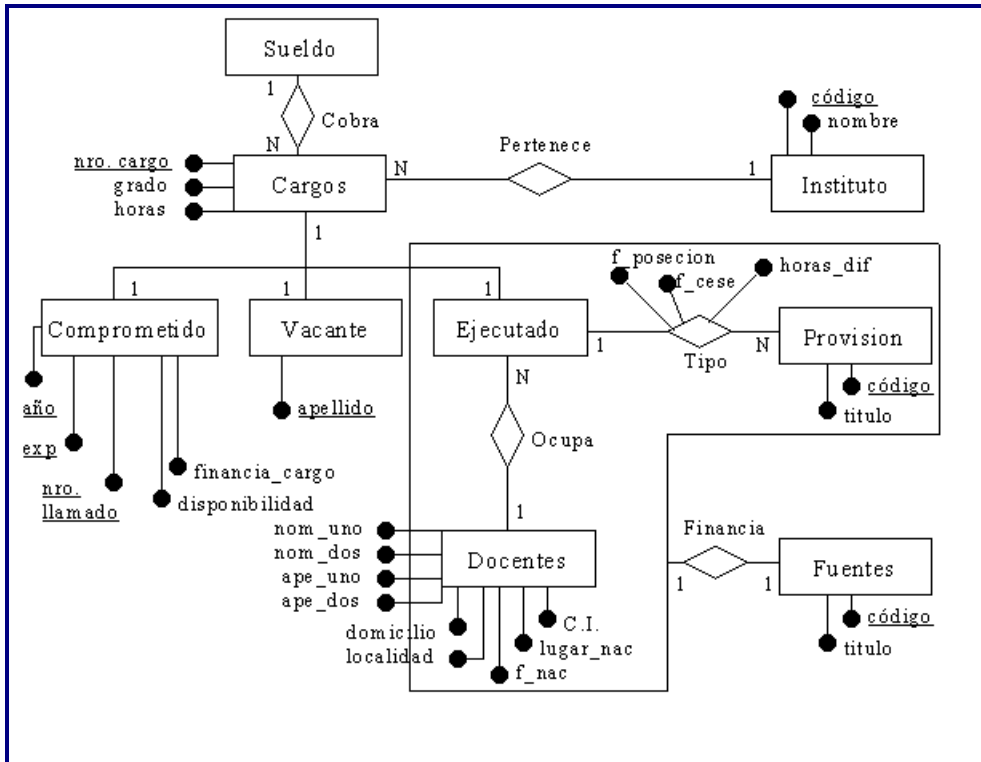


Figura 14: Modelo Conceptual de datos fuente seleccionados

## B. CORRESPONDENCIA DEL MODELO DE DATOS SELECCIONADOS CON MODELOS FUENTE

Se especifica (semiformalmente) las correspondencias entre el Modelo Conceptual de datos seleccionados y los Modelos Conceptuales de los sistemas fuente.

Identificamos los Modelos Conceptuales de la siguiente manera:

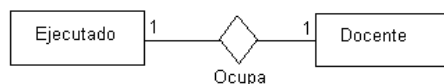
Modelo del Sistema Original	=	S1
Modelo del Sistema Nuevo	=	S2
Modelo Fuentes Seleccionadas	=	S

A cada entidad del Modelo Conceptual de Fuentes Seleccionadas se le asigna la/las entidad/es correspondientes de los Modelos Conceptuales fuente:

S.Sueldos	=	S1.Sueldos
S.Instituto	=	S2.Instituto U S1.Servicios
S.Fuentes	=	S2.Fuentes
S.Provision	=	S2.Provision
S.Docentes	=	S2.Funcionarios
S.Cargos	=	S1.Comprometido U S1.Vacante U S2.Cargos
S.Comprometido	=	S1.Comprometido
S.Vacante	=	S1.Vacante
S.Ejecutado	=	S2.Cargos
S.Tipo	=	S2.Tipo
S.Financia	=	S2.Financia
S.Pertenece	=	S2.Pertenece U S1.Pertenece
S.Ocupa	=	S2.Ocupa

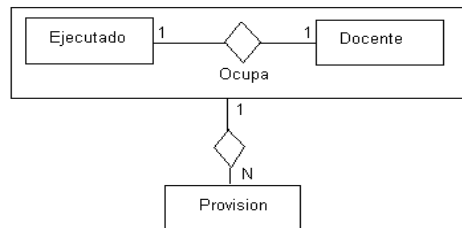
## C. CARDINALIDADES

1)



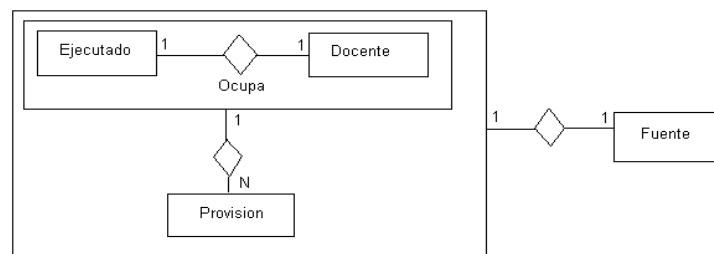
La relación es 1 a 1 porque en un mes dado solo tomamos un docente en un cargo. Los cargos que cambian de docente pasada la mitad del mes, se consideran al mes siguiente.

2)



Un docente asignado a un cargo tiene N provisiones que aportan a los diferentes conceptos del ejecutado (para la fecha dada)

3)



Un docente asignado a un cargo tiene una única fuente de financiación para cada una de sus provisiones.

#### **D. CORRESPONDENCIA DEL MODELO DE DATOS SELECCIONADOS CON TABLAS FUENTE SELECCIONADAS**

En la sección anterior, se especifican las entidades conceptuales de los sistemas fuente que conforman el modelo de datos seleccionados. Por otro lado, se cuenta con mapeos entre los modelos conceptuales de los sistemas fuente y los modelos físicos de los mismos. Falta entonces especificar el mapeo entre el modelo de datos seleccionados y los modelos físicos de los sistemas fuente. El mismo se encuentra en el *Apartado 2, Metadata del Sistema*.

Con este último mapeo, se especifica qué tablas de las bases de datos fuente se utilizarán para extraer la información requerida ya que puede darse el caso de tener dentro de un mismo sistema fuente más de una estructura física que brinde la misma información.

## ➤ ANALISIS DE CARGA

---

*Se describen los problemas que se deberán enfrentar a la hora de cargar el data warehouse.*

---

### 1. Datos que Faltan

- a) Fuentes de Financiación. No se dispone de esta información para los conceptos Comprometidos con disponibilidad y Vacante.
- b) Conceptos. Los sistemas no mantienen la lista de conceptos requerida. Se debe buscar la forma de generarla.
- c) Máximo. La información de máximos por institutos y fuentes de financiación no es llevada por ninguno de los dos sistemas, por lo tanto se debe estudiar la forma de obtenerla o eventualmente generarla.
- d) Dedicaciones especiales. Este tipo de dedicación se calcula como un porcentaje menor al 70% sobre el total percibido por el docente en el cargo. Este porcentaje no es almacenado en los sistemas operacionales.

### 2. Datos a Integrar

- a) Institutos. Ambos sistemas mantienen una lista de institutos. Dado que tomamos cargos de ambos sistemas, se debe integrar esta información.

### 3. Datos a Calcular

- a) Gr.1\_10h. Esta medida no es almacenada en ninguno de los dos sistemas pero se calcula a partir de las horas y grados asociados a un cargo (dato existente en ambos sistemas) y de la equivalencia de sueldos (mantenida por el sistema original).
- b) Disponibilidad. Se deberá obtener este dato para los cargos comprometidos del S. O. Se calcula a partir del número de cargo.
- c) Concepto. Para cruzar los conceptos con los cargos, se deberá calcular esta información. En el S.O. se deriva a partir del tipo de cargo y de la disponibilidad del cargo. En el S.N. se debe asociar a cada provision el concepto correspondiente.
- d) Extensiones/Reducciones. Se deberá calcular las horas correspondientes a cada extensión ó reducción ya que en el S.N. las mismas contienen el total de horas luego de efectuada la extensión/reducción.

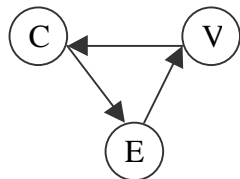
- e) Gasto del Mes, Ahorro del Mes. Esta información no es almacenada en ninguno de los sistemas, pero se calcula a partir del concepto (dato a calcular), del grado y horas correspondientes al cargo (dato existente en ambos sistemas) y de la equivalencia de sueldos (mantenida por el S.O.).
- f) Información en pesos. No se dispone de ésta información de forma directa pero se calcula a partir del grado/hora del cargo (datos existentes en ambos sistemas) y de la equivalencia de sueldos (mantenida en S.O.).
- g) Dedicaciones Compensadas. El equivalente gr1\_10h correspondiente se calcula como el 45% del equivalente gr.1\_10h del cargo base.
- h) Dedicación Total. Se toma el valor correspondiente de la equivalencia de sueldos a partir del grado del cargo y 41 horas. El valor de la dedicación total será la diferencia entre ese valor y el sueldo base del cargo.

#### 4. Datos a Transformar

- a) Docentes. La cédula de identidad del docente no parece ser un identificador primario conveniente. El nombre del docente se encuentra separado en distintos campos de texto.
- b) Sueldo. El esquema de datos de la tabla Sueldos, no es adecuado para los cálculos necesarios a efectuar en la carga del data mart.
- c) Fechas. La información sobre fechas se encuentra en formato ddmmyy, debiéndose pasar el año a yyyy por compatibilidad con el año 2000

#### 5. Datos a Limpiar

- a) Fuentes de Financiación. Las descripciones de las fuentes de financiación del S.N. no están estandarizadas ni son precisas.
- b) Cargos. Un cargo puede ser ejecutado, comprometido o vacante. Para pasar de un estado a otro existen ciertas restricciones. El siguiente esquema muestra los cambios de estados permitidos para un número de cargo. Como los cargos se extraen de más de un sistema fuente, sería interesante controlar los cambios en el estado de un cargo de una carga a la otra.



## **6. Granularidad de Datos**

- a) Depto. y Secciones. Se descarta la información sobre dptos y secciones porque no se puede implementar automáticamente la decodificación del código de servicio.

## **7. Volatilidad de la Información**

- a) La información sobre presupuesto se lleva en forma mensual, alrededor de la primer quincena, se pasa la información a la sección sueldos para que efectúe los pagos. Es en ese momento cuando se deben extraer los datos del mes ya que luego de esa fecha se comienzan a actualizar los datos para el siguiente mes. El sistema nuevo, maneja un histórico con información pero éste no será utilizado por ser su información de muy baja calidad.

## **8. Evolución en el Tiempo**

- a) Se requiere la evolución mensual del presupuesto, no presentándose ningún problema al respecto dada la volatilidad mensual de la información fuente y la facilidad de su cálculo.

## **9. Tecnología Disponible**

- a) No hay tecnología disponible para acceder de forma automática a los sistemas fuente ya que la sección de personal aún no se encuentra conectada a la red de la facultad. La información debe trasladarse entonces utilizando algún medio magnético.
- b) Las bases de datos fuente están compuestas por archivos dbase, por lo tanto sus datos deben ser convertidos para ser ingresados en la base Oracle del data warehouse.

## **10. Volumen de datos**

- a) No se espera tener un volumen muy grande de datos por el momento, no habrá mayores problemas al respecto dado el espacio disponible para el data warehouse. En una carga se espera manejar cantidades no mayores a 5000 registros aproximadamente.



## 11. Accesibilidad a los Datos Fuente

- a) La sección de personal se muestra interesada en el funcionamiento del sistema a construir por lo que brindará periódicamente la información requerida de sus sistemas operacionales. Será responsabilidad del administrador del sistema DW obtener periódicamente la información.

## **1.3 INTERACCION CON GRUPO OLAP**

---

Se acordaron los siguientes puntos:

- Los cubos se cargarán en forma total cada vez, no descartándose la idea de realizar en el futuro cargas incrementales de los mismos.
- Se decide materializar en el data warehouse los cálculos a realizar para obtener el gasto (gr110hs y pesos) y el ahorro (gr110hs y pesos) mensual para mejorar el tiempo de carga de los cubos.
- Se decide que la información del máximo a ejecutar es relevante por lo tanto debe generarse.
- Se decide prescindir de la información acerca de fuentes de financiación asociados a los conceptos Vacante y Comprometido con disponibilidad.
- Se define como formato para la fecha 01MMyyyy. Los cálculos del año y semestre se realizan con las herramientas de usuario final (Power Play de Cognos).
- Se decide cargar la información relacionada a las dedicaciones especiales tomando como valor de porcentaje el 70% aunque no sea un valor exacto.
- El grupo OLAP define una jerarquía grado-docente para implementar un requerimiento específico de usuario. Dado que un docente puede tener más de un grado, por restricción de la herramienta de usuario final (Power Play de Cognos) la jerarquía debe identificarse por duplas grado-docente. Se debe tener en cuenta este punto a la hora de diseñar las tablas del data mart de modo de facilitar la carga de la jerarquía.

## 1.4 DISEÑO

---

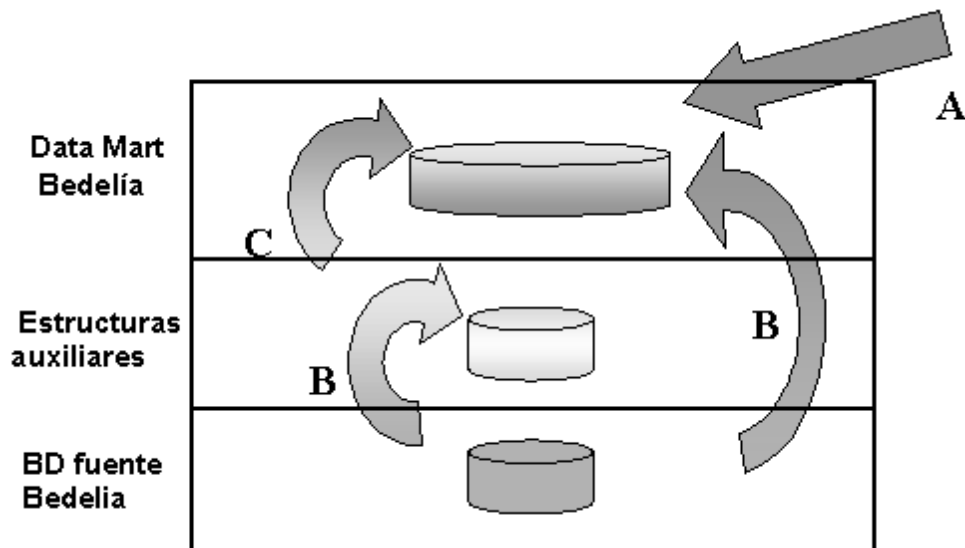
### ➤ ARQUITECTURA DEL DATA MART

---

*Se define en forma general los componentes del data mart, tanto procesos como estructuras de almacenamiento.*

---

Como punto inicial del diseño del data mart de Presupuesto se establece la arquitectura del mismo. En la *Figura 15* se distinguen claramente los distintos componentes de dicho data mart.



*Figura 15: Arquitectura Data mart Presupuesto*

Actualmente las bases de datos de los sistemas fuente ( Sistema Original y Sistema Nuevo) se encuentran en dos máquinas stand alone en la oficina de la Sección de Personal de la Facultad de Ingeniería. Las mismas, se componen por un conjunto de archivos dBase. Dichos archivos pueden ser fácilmente convertidos a un formato que el comando **LOAD** de Oracle pueda entender. En un principio se considera viable este mecanismo, donde los datos fuente se cargan en estructuras auxiliares en la base de datos del data warehouse mediante un LOAD. En un futuro, si el servidor del data warehouse se conectase a las bases de los sistemas fuente mediante una red intra-facultad, sería interesante automatizar el proceso de extracción de datos de las bases fuente.

Según el modelo multidimensional de presupuesto, y el análisis realizado, vemos que no todos los datos que se necesita almacenar en el data mart se encuentran en las bases de datos fuente. Es por esto que se decide diseñar un conjunto de **tablas base** dónde almacenar los datos faltantes y desarrollar dos procesos de carga diferentes, uno para cargar los datos que faltan, y otro para cargar los datos fuente.

El **proceso A**, es el encargado de ingresar a las tablas del data mart los datos que no se encuentran en las bases fuente.

El nivel de calidad en que se encuentran los datos fuente no es el más fortuito. Los problemas detectados son variados, pero aún es posible cargar el data mart a partir de dichos datos, realizando un conjunto importante de validaciones y transformaciones sobre los mismos antes de que éstos ingresen al data mart.

Las primitivas utilizadas para el diseño presentadas en la sección 2.4 *Diseño de un Data warehouse, Informe Principal* [P22], definen las consultas necesarias para convertir los datos, pero no brindan soluciones a los problemas de calidad. Queda entonces descartada la idea de que la carga sea un conjunto de queries a los datos fuente cuyo resultado se ingrese al data mart.

Se toma como solución desarrollar un proceso de carga que, seleccione los datos fuente, realice las validaciones y transformaciones necesarias e ingrese los datos a un conjunto de tablas auxiliares dentro de la base de datos del data mart. El proceso continúa, tomando los datos limpios de las tablas auxiliares y mediante consultas, ingresa los datos a las tablas del data mart. En esta segunda etapa sí pueden utilizarse las consultas definidas en las primitivas.

Esta solución tiene como ventaja principal que el proceso de carga queda más fácil de implementar y de mantener. El conjunto de tablas auxiliares, llamémosle **tablas limpias**, puede utilizarse también como fuente de consulta a los datos de los sistemas fuente.

El **proceso B** es el que carga los datos fuente. Se compone de dos subprocesos que realizan las validaciones y chequeos necesarios con la diferencia que, uno ingresa datos limpios en tablas limpias y el otro lo hace directamente al data mart. El **proceso C**, toma los datos limpios de las tablas limpias y los ingresa al data mart.

## ➤ DISEÑO DE TABLAS

---

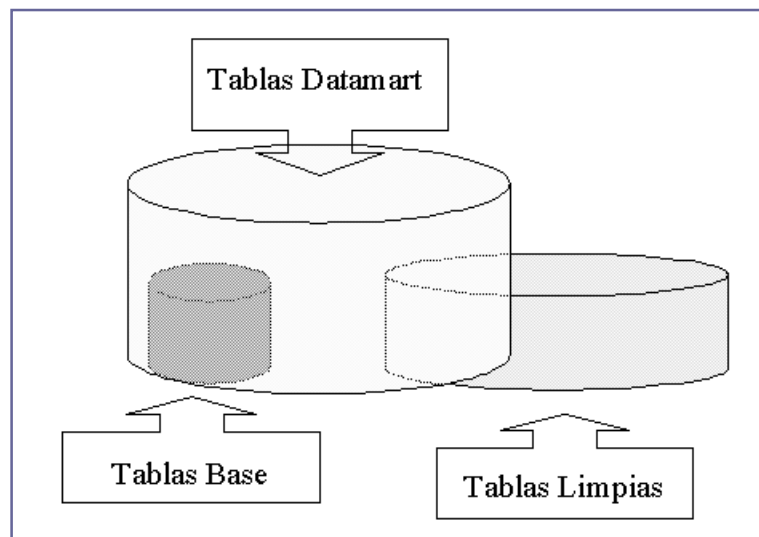
*Se realiza el diseño de las tablas relacionales del data mart y de cualquier otra estructura auxiliar necesaria tanto para facilitar la carga como para la limpieza y transformación de datos.*

---

Dentro del conjunto de tablas diseñadas pueden distinguirse tres conjuntos diferentes (no disjuntos) de tablas: **tablas base**, **tablas limpias** y **tablas del data mart**. Las tablas base almacenan los datos generados para el data mart que no se encuentran en las bases fuentes. Las tablas limpias almacenan datos provenientes de los sistemas fuente, luego de haber sido limpiados. Las tablas del data mart almacenan propiamente los datos del data mart.

El diseño de las **tablas base** se realiza de forma que su estructura sea apropiada para almacenar los datos faltantes especificados como necesarios en el modelo conceptual multidimensional. Las **tablas limpias** surgen como apoyo a la integración y transformación de datos y se diseñan en su mayoría a partir del modelo conceptual de datos seleccionados. Para diseñar las **tablas del data mart**, se aplican las primitivas a las tablas que componen los dos conjuntos antes descritos. En algunos casos, las tablas ya se encuentran en la forma adecuada para el data mart y no se les aplica ninguna primitiva. En otros casos, las tablas son transformadas mediante la aplicación sucesiva de primitivas hasta llegar a obtener una estructura adecuada para el data mart.

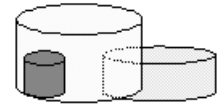
La *Figura 16* resalta los tres conjuntos de tablas y como éstos se relacionan.



*Figura 16: Conjuntos de tablas Presupuesto*

Estándar de nomenclatura utilizado: las tablas que componen el data mart propiamente dicho se identifican por comenzar con 'PR\_'.

## Diseño Tablas Base



Se describe a continuación las tablas base definidas. Las mismas, surgen como respuesta a los problemas planteados en el *Análisis de Carga*. Por más información a cerca del diseño de dichas tablas, referirse al *Apartado 2, Metadata del Sistema*.

### PR\_INSTITUTO.

Almacena la lista de institutos de la Facultad de Ingeniería. Si bien los sistemas fuentes mantienen esta información, ninguno lo hace correctamente y se prefiere definir una lista de institutos para el data mart. *Ver punto 2.a del Análisis de Carga*.

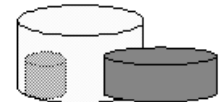
### PR\_CONCEPTO.

Contiene los conceptos del ejecutado especificados en el modelo multidimensional. Los conceptos no se deben actualizar, dado que todo el sistema gira alrededor de la definición de los mismos. *Ver punto 1.b del Análisis de Carga*.

### PR\_MAXIMO.

Mantiene por mes, para cada instituto de la F.I. y cada fuente de financiación el importe máximo disponible en pesos uruguayos y en su equivalente grado 1-10 horas. *Ver punto 1.c del Análisis de Carga*.

## Diseño Tablas Limpias



Se describen las tablas limpias definidas. Cada tabla hace referencia al punto dentro del Análisis de Carga que provocó su diseño. Por más información a cerca de las tablas, referirse al *Apartado 2, Metadata del Sistema*.

### PR\_MAPEO\_INSTITUTO.

Integra los códigos originales de institutos de ambos sistemas y los asocia a los códigos de la tabla PR\_INSTITUTO (institutos definidos para el data mart). Esta tabla permite filtrar datos de institutos fuente que no son de interés para el data mart (no asociándolos a ningún instituto definido) o consolidar datos de más de un instituto fuente (asociándolos a un mismo instituto definido). Se mantiene como clave el campo clave original porque no es un dato actualizable en el sistema fuente. La descripción original en caso que cambie, debe ser actualizada. Se toma esa postura porque la descripción difícilmente cambie y si lo hace será válida ya que tendrá el mismo significado. *Ver punto 2.a de Análisis de Carga*.

#### **PR\_MAPEO\_PROVISION.**

Mapea los códigos de provisión del Sistema Nuevo con los conceptos definidos para el data mart en la tabla PR\_CONCEPTO. Esta tabla permite filtrar datos de provisiones que no se corresponden a ninguno de los conceptos definidos en el data mart. De esta forma, todo código original que aporte a algún concepto, tendrá un código de concepto asociado, al resto se le asociará un valor nulo. Se mantiene como clave el campo clave original por no ser un dato actualizable en el sistema fuente. La descripción original en caso que cambie, debe ser actualizada. Se toma esa postura porque la descripción difícilmente cambie y si lo hace será válida ya que tendrá el mismo significado. *Ver punto 3.c del Análisis de Carga.*

#### **PR\_FUENTE.**

Mantiene la lista de fuentes de financiación de los cargos de la Facultad de Ingeniería provenientes del Sistema Nuevo. Como las descripciones originales no son lo suficientemente claras y precisas, se agrega un campo que tiene en un principio la descripción original de la fuente y puede ser actualizado por el usuario del data warehouse con una descripción de mejor calidad de ser necesario. Se mantiene como clave el campo clave original ya que éste no es un dato actualizable en el sistema fuente. *Ver punto 5.a del Análisis de Carga.*

#### **PR\_DOCENTE.**

Almacena la lista de docentes de la Facultad de Ingeniería, provenientes del Sistema Nuevo. Se define como identificador un campo numérico secuencial, diferente del identificador original (cédula del docente). La cédula queda como una clave alternativa. La lista de nombres originales se integra en un único campo que contiene el nombre completo. *Ver punto 4.a del Análisis de Carga.*

#### **SUELDO.**

Almacena para cada combinación de grado y horas el sueldo correspondiente en pesos uruguayos y en su equivalente medida grado 1-10 horas. Los datos provienen del Sistema Original. *Ver punto 4.b del Análisis de Carga.*

#### **CARGOS.**

Mantiene la lista completa de cargos, facilitando la integración de los cargos provenientes de los sistemas Nuevo y Original. Se agrega un campo que describe el estado actual del cargo (Comprometido, Ejecutado o Vacante) y otro que indica si el cargo goza o no de sueldo. El número de cargo se mantiene como clave ya que no existen dos cargos en el mismo período de tiempo con el mismo número identificador. La jerarquía de Cargos, Vacante, Comprometido y Ejecutado, indicada en el modelo conceptual de datos seleccionados, es del tipo total y excluyente. Los datos que conforman dicha jerarquía son extraídos de ambos sistemas fuentes. El diseño se realiza para facilitar la detección de cambios incorrectos en el estado de un cargo. Por lo que los datos de la tabla deben

mantenerse entre carga y carga y actualizarse en caso de ser necesario. *Ver punto 5.b del Análisis de Carga.*

#### VACANTE.

Contiene los datos de cargos vacantes correspondientes al mes de la carga provenientes del Sistema Original. Se adecua el esquema de la tabla para que quede consistente con los anteriores cambios realizados en las demás tablas fuente y se agregan algunos datos calculados como se indica en el *punto 3 del Análisis de Carga.*

#### COMPROMETIDO.

Contiene los datos de los cargos comprometidos correspondientes al mes de la carga provenientes del Sistema Original. Se adecua el esquema de la tabla para que quede consistente con los anteriores cambios realizados en las demás tablas fuente y se agregan algunos datos calculados como se indica en el *punto 3 del Análisis de Carga.*

#### EJECUTADO.

Contiene los datos de los cargos activos correspondientes al mes de la carga y sus movimientos asociados (extensiones/reducciones transitorias, reducciones permanentes, dedicaciones especiales, etc). Dicha información proviene del Sistema Nuevo. Se agregan los campos necesarios, correspondientes a modificaciones en tablas asociadas y se calculan algunos datos como se indica en el *punto 3 del Análisis de Carga.*

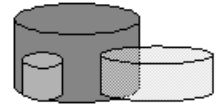
#### EJECUTADO\_ANT.

Contiene los datos de los cargos activos correspondientes al mes anterior de la carga y sus movimientos asociados (extensiones/reducciones transitorias, reducciones permanentes, dedicaciones especiales, etc), información proveniente del Sistema Nuevo. Esta tabla se define para poder verificar movimientos de los cargos ejecutados, específicamente es de utilidad para determinar las reducciones horarias permanentes sin necesidad de consultar el histórico de movimientos almacenado en el S.N. Se definen los mismos campos que en EJECUTADO para poder identificar si es el mismo cargo, el número de cargo es identificatorio en un mes de carga pero puede representar otro cargo al mes siguiente.

#### LOG\_ERRORES.

Tabla para el manejo de errores. En dicha tabla se ingresa información de registros que se descartan, con una descripción de porqué se desechan e información de registros que se ingresan pero por seguridad se informa un warning.

## Diseño Tablas Data mart



Para diseñar las tablas del data mart de presupuesto, se utilizan el conjunto de primitivas presentadas en la sección *2.4 Diseño de un Data Warehouse, Informe Principal*. Las primitivas se aplican a los conjuntos de tablas base, tablas limpias y tablas resultado de aplicar primitivas anteriores a éstas mismas tablas. Es un proceso cíclico que finaliza cuando se obtiene un diseño adecuado al modelo conceptual multidimensional del data mart de presupuesto requerido por el grupo 'Sistema DW: OLAP'.

Las primitivas se aplican en función de optimizar las consultas sobre las tablas del data mart sin convertir el proceso de carga de datos en algo demasiado complejo. Si bien lo óptimo sería tener una gran fact table que almacene todos los cruzamientos de dimensiones y medidas, desfavorecería en gran medida la actualización de los datos en el data mart y los controles de calidad.

El diseño intenta también que las tablas del data mart puedan ser actualizadas mediante inserts únicamente, optimizando así la carga y permitiendo así realizar carga incremental a las herramientas OLAP.

Algunas tablas dimensionales encontradas en los conjuntos de tablas base y limpias no fueron transformadas mediante las primitivas por ser su estructura adecuada para el data mart.



## A. APLICACIÓN PRIMITIVAS – DATA MART PRESUPUESTO

El diagrama muestra la secuencia de aplicación de primitivas hasta llegar al diseño final del data mart relacional.

Las tablas que no necesitan la aplicación de ninguna primitiva, se representan con una flecha directa.

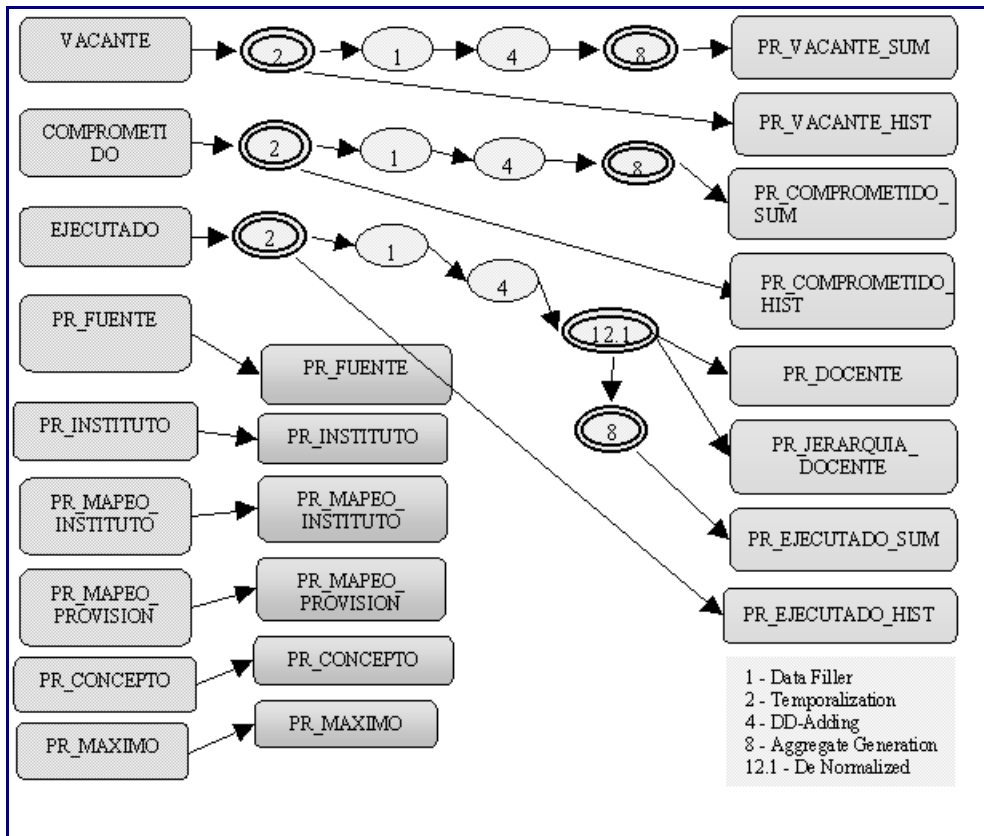


Figura 17: Aplicación de primitivas Data mart Presupuesto

## B. FORMALIZACION DE APLICACIÓN DE PRIMITIVAS

Se formaliza la aplicación de las primitivas, tomando como referencia la tabla del data mart que se obtiene como resultado de su aplicación. Por más información ver sección 2.4 *Diseño del Data Warehouse, Informe Principal*.

### **PR\_VACANTE\_HIST**

#### 2 TEMPORALIZATION

Entrada:

- Vacante ( ca\_nrocargo, mi\_codorig, mi\_indsist, in\_codinst, cp\_codconcepto, va\_grado, va\_horas, va\_gr1\_10hs, va\_plata)
- va\_fecha / DOM(va\_fecha) = DATE
- TRUE

Resultado:

- Vacante' ( ca\_nrocargo, va\_fecha, mi\_codorig, mi\_indsist, in\_codinst, cp\_codconcepto, va\_grado, va\_horas, va\_gr1\_10hs, va\_plata)

#### 1 DATA FILTER

Entrada:

- Vacante' (ca\_nrocargo, va\_fecha, mi\_codorig, mi\_indsist, in\_codinst, cp\_codconcepto, va\_grado, va\_horas, va\_gr1\_10hs, va\_plata)
- { mi\_codorig, mi\_indsist }

Resultado:

- Vacante" (ca\_nrocargo, va\_fecha, in\_codinst, cp\_codconcepto, va\_grado, va\_horas, va\_gr1\_10hs, va\_plata)

#### 4.1 DD ADDING

Entrada:

- Vacante" ( ca\_nrocargo, va\_fecha, in\_codinst, cp\_codconcepto, va\_grado, va\_horas, va\_gr1\_10hs, va\_plata )
- AhorroGasto\_gr1\_10hs(cp\_codconcepto)  
AhorroGasto\_gr1\_10hs: define según el código de concepto si el movimiento aporta al ahorro o al gasto

Resultado:

- Vacante"" (ca\_nrocargo, va\_fecha, in\_codinst, cp\_codconcepto, va\_grado, va\_horas, va\_gr1\_10hs, va\_plata, va\_ahorro\_gasto )

La salida de la primitiva anterior se materializa en la tabla *PR\_VACANTE\_HIST*. Tener en cuenta que el número de cargo (*ca\_nrocargo*) deja de ser clave foránea a la tabla limpia *CARGOS*.

## ***PR\_VACANTE\_HIST***

### 8 AGGREGATE GENERATION

Entrada:

- Vacante''' (ca\_nrocargo, va\_fecha, in\_codinst, cp\_codconcepto, va\_ahorro\_gasto, va\_grado, va\_horas, va\_gr1\_10hs, va\_plata)
- { va\_fecha, in\_codinst, cp\_codconcepto, va\_ahorro\_gasto }
- { va\_horas, va\_gr1\_10hs, va\_plata }
- { SUM(va\_horas), SUM(va\_gr1\_10hs), SUM(va\_plata) }
- { ca\_nrocargo, va\_grado }

Resultado:

- Vacante''' ( va\_fecha, in\_codinst, cp\_codconcepto, va\_ahorro\_gasto, va\_horas, va\_gr1\_10hs, va\_plata )

El resultado de la primitiva anterior se materializa en la tabla *PR\_VACANTE\_SUM*, que tiene como clave los campos *va\_fecha*, *in\_codinst* y *cp\_codconcepto*.

## ***PR\_COMPROMETIDO\_HIST***

### 2 TEMPORALIZATION

Entrada:

- Comprometido ( ca\_nrocargo, mi\_codorig, mi\_indsist, in\_codinst, cp\_codconcepto, co\_disponibilidad, co\_fin\_cargo, co\_grado, co\_horas, co\_gr1\_10hs, co\_plata )
- co\_fecha / DOM(co\_fecha) = DATE
- TRUE

Resultado:

- Comprometido' (ca\_nrocargo, co\_fecha, mi\_codorig, mi\_indsist, in\_codinst, cp\_codconcepto, co\_disponibilidad, co\_fin\_cargo, co\_grado, co\_horas, co\_gr1\_10hs, co\_plata)

### 1 DATA FILTER

Entrada:

- Comprometido' (ca\_nrocargo, co\_fecha, mi\_codorig, mi\_indsist, in\_codinst, cp\_codconcepto, co\_disponibilidad, co\_fin\_cargo, co\_grado, co\_horas, co\_gr1\_10hs, co\_plata)
- { mi\_codorig, mi\_indsist, co\_disponibilidad, co\_fin\_cargo }

Resultado:

- Comprometido'' (ca\_nrocargo, co\_fecha, in\_codinst, cp\_codconcepto, co\_grado, co\_horas, co\_gr1\_10hs, co\_plata)

### 1 DD ADDING

Entrada:

- Comprometido'' (ca\_nrocargo, co\_fecha, in\_codinst, cp\_codconcepto, co\_grado, co\_horas, co\_gr1\_10hs, co\_plata)
- AhorroGasto (cp\_codconcepto)  
AhorroGasto : define según el código de concepto si el movimiento aporta al ahorro o al gasto

Resultado:

- Comprometido''' (ca\_nrocargo, co\_fecha, in\_codinst, cp\_codconcepto, co\_ahorro\_gasto, co\_grado, co\_horas, co\_gr1\_10hs, co\_plata )

La salida de la primitiva anterior se materializa en la tabla *PR\_COMPROMETIDO\_HIST*. Tener en cuenta que el número de cargo (*ca\_nrocargo*) deja de ser clave foránea a la tabla limpia *CARGOS*.

### ***PR\_COMPROMETIDO\_SUM***

#### 8 AGGREGATE GENERATION

Entrada:

- Comprometido''' (ca\_nrocargo, co\_fecha, in\_codinst, cp\_codconcepto, co\_ahorro\_gasto, co\_grado, co\_horas, co\_gr1\_10hs, co\_plata )
- { co\_fecha, in\_codinst, cp\_codconcepto, co\_ahorro\_gasto }
- { co\_horas, co\_gr1\_10hs, co\_plata }
- { SUM(co\_horas), SUM(co\_gr1\_10hs), SUM(co\_plata) }
- { ca\_nrocargo, co\_grado }

Resultado:

- Comprometido''' ( co\_fecha, in\_codinst, cp\_codconcepto, co\_ahorro\_gasto, co\_horas, co\_gr1\_10hs, co\_plata,)

El resultado de la primitiva anterior se materializa en la tabla *PR\_COMPROMETIDO\_SUM*, que tiene como clave los campos *co\_fecha*, *in\_codinst* y *cp\_codconcepto*.

### ***PR\_EJECUTADO\_HIST***

#### 2 TEMPORALIZATION

Entrada:

- Ejecutado ( ca\_nrocargo, mi\_codorig, mi\_indsist, in\_codinst, do\_coddoc, fu\_codfuente, mp\_codorig, cp\_codconcepto, ej\_fecha\_pos, ej\_fecha\_cese, ej\_grado, ej\_horas, ej\_gr1\_10hs, ej\_plata)
- ej\_fecha / DOM(ej\_fecha) = DATE
- TRUE

Resultado:

- Ejecutado' ( ca\_nrocargo, ej\_fecha, mi\_codorig, mi\_indsist, in\_codinst, do\_coddoc, fu\_codfuente, mp\_codorig, cp\_codconcepto, ej\_fecha\_pos, ej\_fecha\_cese, ej\_grado, ej\_horas, ej\_gr1\_10hs, ej\_plata )

#### 1 DATA FILTER

Entrada:

- Ejecutado' ( ca\_nrocargo, ej\_fecha, mi\_codorig, mi\_indsist, in\_codinst, do\_coddoc, fu\_codfuente, mp\_codorig, cp\_codconcepto, ej\_fecha\_pos, ej\_fecha\_cese, ej\_grado, ej\_horas, ej\_gr1\_10hs, ej\_plata )
- {mi\_codorig, mi\_indsist, mp\_codorig, ej\_fecha\_pos, ej\_fecha\_cese }

Resultado:

- Ejecutado" ( ca\_nrocargo, ej\_fecha, in\_codinst, do\_coddoc, fu\_codfuente, cp\_codconcepto, ej\_grado, ej\_horas, ej\_gr1\_10hs, ej\_plata )

#### 4.1 DD ADDING

Entrada:

- Ejecutado" (ca\_nrocargo, ej\_fecha, in\_codinst, do\_coddoc, fu\_codfuente, cp\_codconcepto, ej\_grado, ej\_horas, ej\_gr1\_10hs, ej\_plata)
- AhorroGasto (cp\_codconcepto)  
AhorroGasto: define según el código de concepto si el movimiento aporta al ahorro o al gasto

Resultado:

- Ejecutado"" (ca\_nrocargo, ej\_fecha, in\_codinst, do\_coddoc, fu\_codfuente, cp\_codconcepto, ej\_ahorro\_gasto, ej\_grado, ej\_horas, ej\_gr1\_10hs, ej\_plata)

La salida de la primitiva anterior se materializa en la tabla *PR\_EJECUTADO\_HIST*. Tener en cuenta que el número de cargo (*ca\_nrocargo*) deja de ser clave foránea a la tabla limpia *CARGOS*.

### ***PR\_DOCENTE\_CARGO***

#### 12.1 DE-NORMALIZED

Entrada:

- Ejecutado"" (ca\_nrocargo, ej\_fecha, in\_codinst, do\_coddoc, fu\_codfuente, cp\_codconcepto, ahorro\_gasto, ej\_grado, ej\_horas, ej\_gr1\_10hs, ej\_plata)  
Docente (do\_coddoc, do\_ci, do\_nombre, do\_nomuno, do\_nomdos, do\_apeuno, do\_apedos)
- { j\_clave, ej\_grado, do\_coddoc }
- { j\_clave }

Resultado:

- Jerarquia\_Docente (j\_clave, ej\_grado, do\_coddoc)
- Ejecutado"" ( ca\_nrocargo, ej\_fecha, in\_codinst, j\_clave, fu\_codfuente, cp\_codconcepto, ahorro\_gasto, ej\_horas, ej\_gr1\_10hs, ej\_plata)

Parte del resultado de la primitiva anterior (Jerarquía Docente) se materializa en la tabla *PR\_DOCENTE\_GRADO*, que tiene como clave un campo secuencial numérico. Como se puede ver en el resultado, la tabla que contenía los campos docente y grado, ahora tiene una clave foránea a la clave de la jerarquía.

***PR\_EJECUTADO\_SUM***

8 AGGREGATE GENERATION

Entrada:

- Ejecutado'''' (ca\_nrocargo, ej\_fecha, in\_codinst, j\_clave, fu\_codfuente, cp\_codconcepto, ej\_ahorro\_gasto, ej\_horas, ej\_gr1\_10hs, ej\_plata)
- { ej\_fecha, in\_codinst, j\_clave, fu\_codfuente, cp\_codconcepto, ej\_ahorro\_gasto }
- { ej\_horas, ej\_gr1\_10hs, ej\_plata }
- { SUM(ej\_horas), SUM(ej\_gr1\_10hs), SUM(ej\_plata) }
- { ca\_nrocargo }

Resultado:

- Ejecutado'''' (ej\_fecha, in\_codinst, j\_clave, fu\_codfuente, cp\_codconcepto, ej\_ahorro\_gasto, ej\_horas, ej\_gr1\_10hs, ej\_plata)

El resultado de la primitiva anterior se materializa en la tabla *PR\_EJECUTADO\_SUM*, que tiene como clave los campos *es\_fecha*, *in\_codinst*, *dg\_codigo* (*clave jerarquía docente grado*), *fu\_codfuente* y *cp\_codconcepto*.

### C. ESQUEMA RELACIONAL DATA MART PRESUPUESTO

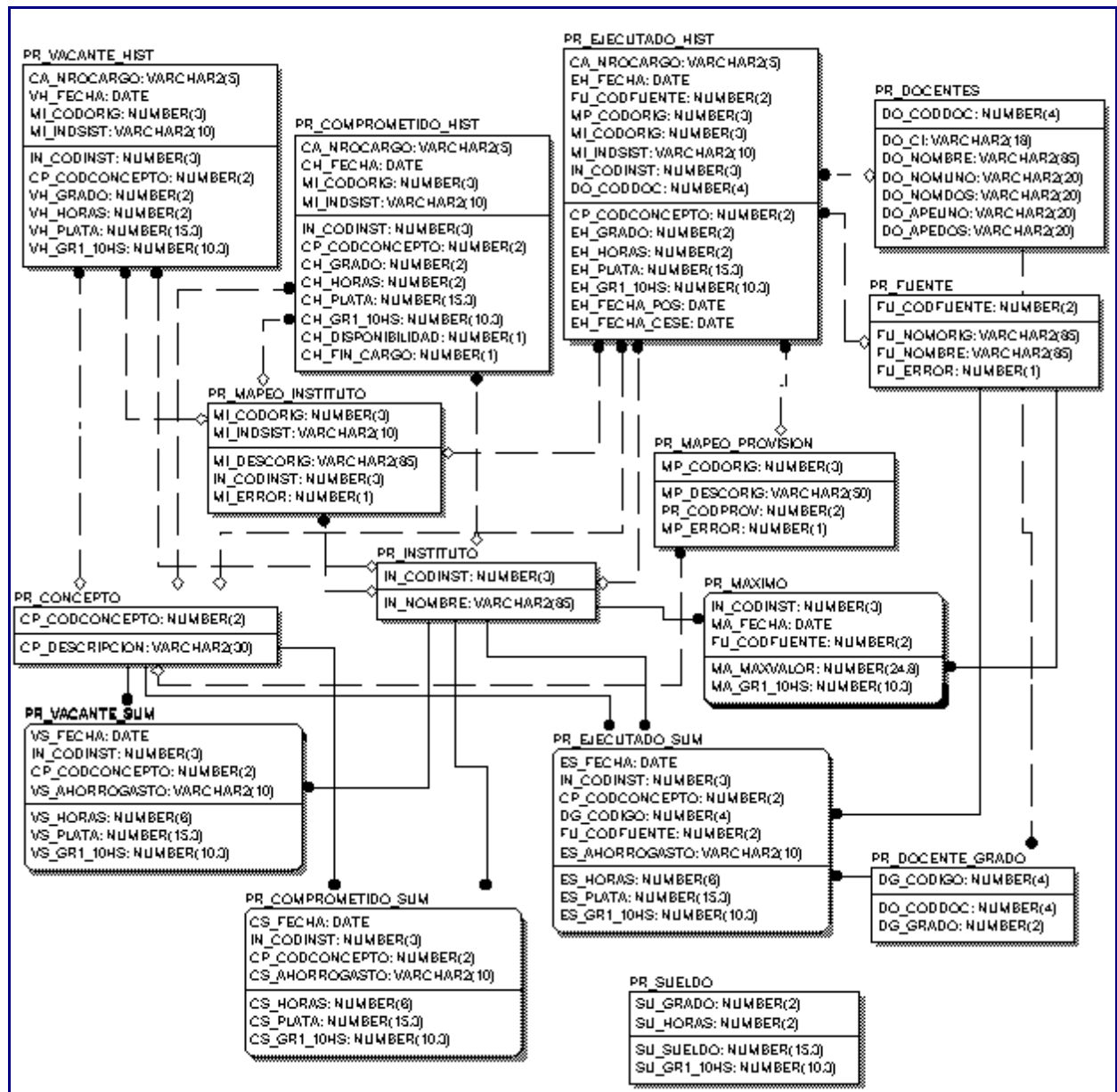


Figura 18: Esquema Tablas Data mart Presupuesto

En la figura 18 se muestra el conjunto de tablas finales que componen el data mart de presupuesto.

Cada tabla tiene una línea horizontal, y los campos por encima de ella representan la clave de la tabla. Las restricciones de integridad entre tablas se representan mediante líneas conectoras entre ellas. Las líneas continuas representan relaciones identificatorias (el hijo tiene la clave foránea como clave) y las líneas discontinuas representan relaciones no identificatorias (el hijo no tiene la clave foránea como clave). El círculo en negro en la punta de los conectores representa la cardinalidad 0 – 1 – N en el hijo. El rombo blanco

representa la cardinalidad 0 – 1 en el padre. Si el conector no tiene una figura al final, representa la cardinalidad 1 en el padre.

## ➤ DISEÑO DE LA CARGA

---

*Se realiza el diseño de las distintas rutinas de extracción y carga de datos al data mart.*

---

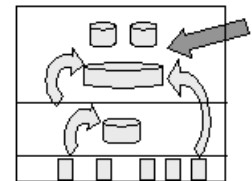
El proceso de carga del data mart de Presupuesto consta de varias partes como ya se especificó en la fase de arquitectura. Básicamente dentro del proceso se distinguen dos etapas, una etapa única donde se cargan los datos inexistentes en las bases fuente y otra etapa periódica que se encarga de cargar los datos que sí existen en las bases fuentes al data mart. Esto es, se define un conjunto de procesos que permiten ingresar al data mart datos inexistentes en las bases fuentes y que son necesarios en el data mart, y se define otro gran proceso que extrae los datos fuentes, los valida, transforma, integra, limpia y los carga al data mart. Llamémosle a estos procesos, Carga Datos Nuevos y Carga Datos Fuente respectivamente.

A continuación veremos un resumen de los procesos de carga y la definición de la periodicidad de la carga. La especificación detallada de todo el proceso de carga se encuentra en el *Apartado 2, Metadata del Sistema*, donde se documentan los pseudocódigos de los procesos diseñados.

### A. PROCESO DE CARGA

#### A.1) Carga Datos Nuevos

Dentro de la etapa de análisis, más precisamente en el análisis de carga, se detectó que cierta información necesaria en el data mart de Presupuesto, no se encontraba en los sistemas fuente o sí se encontraba pero no en la forma adecuada. También se encontró que algunos datos, no tenían el nivel de calidad deseado para el data mart. En particular nos referimos a información del importe máximo disponible por instituto de la F.I., la lista de institutos y los conceptos del ejecutado.



Como solución a estos puntos se deciden ingresar los datos antes mencionados de forma independiente de las bases fuente. El administrador del data warehouse debe encargarse de ejecutar estos procesos e ingresar los datos.

Describimos a grandes rasgos cada uno de los procesos.



### **CargaInstitutos( Nombre )**

Recibe el nombre del instituto y lo ingresa en la tabla *PR\_INSTITUTO* con código identificador generado automáticamente.

Los nombres de institutos ingresados deben corresponderse con los Institutos de la Facultad de Ingeniería que se deseen considerar en el data mart de presupuesto. Este proceso se ejecutaría en la puesta en marcha del sistema y en caso de necesitarse ingresar un nuevo instituto.

### **CargaConceptos( Descripcion )**

Recibe la descripción del concepto y lo ingresa en la tabla *PR\_CONCEPTO* con código identificador generado automáticamente.

Las descripciones de conceptos ingresadas deben corresponderse a los con los conceptos definidos en el modelo multidimensional. El ingreso de estos datos en forma correcta es fundamental para lograr un buen resultado del sistema.

### **CargaMaximoInstituto( Instituto, FuenteFinanciacion, Fecha , ValorMaximo )**

Para el código de instituto, fuente de financiación y fecha dados, ingresa en la tabla *PR\_MAXIMO* el valor máximo indicado y el correspondiente grado 1-10 horas calculado automáticamente.

Este proceso se debe ejecutar mensualmente para ingresar el máximo importe ( pesos uruguayos ) del Presupuesto de la F.I. que dispone cada instituto en el mes correspondiente.

## **A.2) Carga Datos Fuente**

El proceso de carga de datos fuente está diseñado para que tome los datos fuente, los valide, transforme, limpie e integre y por último los ingrese al data mart de Presupuesto.

Para facilitar la transformación y limpieza de datos, se definen un conjunto de tablas auxiliares, llamadas tablas limpias, que almacenan ciertos datos fuente, algunos en forma temporal. Los datos se ingresan en las tablas limpias luego de haber sido previamente validados. En el punto Diseño de Tablas de este capítulo se especifica con detalle estas tablas.

La carga de datos fuente puede verse como dos subprocessos contiguos, un primer subprocesso que toma los datos fuente, los valida, transforma, limpia e integra,

ingresándolos en tablas limpias o directamente en tablas del data mart, y un segundo subproceso que toma los datos limpios de las tablas limpias y los transforma a la forma adecuada para el data mart. En general este segundo proceso está diseñado utilizando las reglas de transformación de datos propuestas en el trabajo 'A transformations based approach for designing the data warehouse' [P22].

Los datos correspondientes a dimensiones en el modelo multidimensional en general se cargan en un paso a las tablas del data mart. Los datos correspondientes a medidas o cruzamientos necesitan de chequeos y transformaciones antes de ingresarse al data mart y son los que se cargan primero a tablas limpias y luego al data mart.

El diseño de este proceso se basa en que los datos fuente ya se encuentran cargados provisoriamente en la base de datos del data mart en estructuras auxiliares (mediante LOAD).

Describimos cada uno de los procedimientos principales que componen ambos subprocesos de carga.

### CargaMapeoInstituto

**Funcionalidad:** Mapea los códigos originales de institutos utilizados en ambos sistemas fuente con los códigos de institutos definidos en el data mart (tabla *PR\_INSTITUTO*), ingresando la correspondencia en la tabla *PR\_MAPEO\_INSTITUTO*.

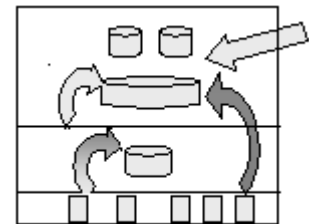
Al finalizar la lectura, en caso de haberse ingresado registros nuevos, dichos registros se presentan en pantalla, donde el administrador encargado de la carga, debe asociar a cada instituto presentado, un instituto del data mart (de la tabla *PR\_INSTITUTO*) o nada si no se desea considerar.

**Especificación:** Por cada código de instituto original leído, se valida que exista en *MAPEO\_INSTITUTO*. Aquellos códigos originales que no se encuentren en dicha tabla se insertan, los que sí se encuentran son desechados a menos que traigan una descripción diferente, caso en donde se actualiza la misma.

Debe tenerse en cuenta que si a un instituto no se le asocia ningún código de instituto del data mart, todos los cargos correspondientes a el mismo, serán ignorados al momento de cargarlos.

### CargaFuenteFinanciación

**Funcionalidad:** Ingresar nuevas fuentes de financiación a la tabla *PR\_FUENTES*.



Al finalizar la lectura de fuentes, en caso de haberse ingresado registros nuevos, dichos registros se presentan en pantalla, donde el administrador podrá modificar la descripción final de la fuente en caso de que la original no sea clara.

**Especificación:** Se leen los datos de *wfinanc* del S.N.  
Por cada fuente de financiación original, se valida que exista en *PR\_FUENTES*. Aquellos registros que no se encuentren en dicha tabla, serán ingresados a la misma. La descripción final se inicializa con la descripción original del registro.

### **CargaMapeoProvisión**

**Funcionalidad:** Mapea las nuevas provisiones con los conceptos definidos en el data mart ( tabla *PR\_CONCEPTO*), ingresando la correspondencia en la tabla *PR\_MAPEO\_PROVISION*.

Al finalizar la lectura, en caso de haberse ingresado registros nuevos, dichos registros se presentan en pantalla donde el administrador deberá, asociar a cada nueva provisión un código de concepto si dicha provisión se corresponde con alguno de los conceptos definidos en *PR\_CONCEPTOS*. De lo contrario la misma será ignorada y por ende se descartará todo dato que la contenga. Lo mismo podrá hacerse con los registros modificados.

**Especificación:** Los datos se leen de la tabla *wprovis* del S.N.  
Todo código de provisión de *wprovis* que no exista en la tabla *MAPEO\_PROVISION*, es ingresado a la misma sin un concepto asociado. Si el código existe pero con diferente descripción, se actualiza la descripción con el valor nuevo.

### **CargaDocentes**

**Funcionalidad:** Ingresa los docentes nuevos en la tabla *PR\_DOCENTES*.

**Especificación:** Los datos se leen de las tablas *wfunc* y *wegos* del S.N.  
Se toman los registros de *wfunc* que se correspondan a docentes de la facultad, por lo tanto se toman solamente aquellos que estén asignados a un cargo con escalafón 'G'. Esto se hace mediante un join de las tablas *wfunc* y *wegos* por el campo *cedula*.

Todos los docentes seleccionados se ingresan en la tabla *PR\_DOCENTES* si no lo estaban ya en la carga anterior. La manera para identificar un docente, manejando los datos originales, es mirando en primer lugar la cédula y en segundo lugar su nombre. Si dos personas coinciden en el número de cédula pero tienen nombres diferentes, debe indicarse una advertencia al usuario encargado de la carga. Puede ser la misma persona, a la cual se le modificó la manera de escribir su nombre ( por ej.: antes abreviado y ahora no) o puede ser otra persona y por lo tanto ser notificado el error. Luego de definido si el docente existe ya o no, se le asocia automáticamente un identificador al mismo. La cédula pasa a ser un identificador alternativo.

### **Carga Vacantes**

**Funcionalidad:** Carga los cargos vacantes válidos para el mes de carga a la tabla limpia *VACANTE*.

**Especificación:** Borra los registros de la tabla *VACANTE*. Lee los datos de la tabla *personal* del S.O., los valida y transforma. Aquellos que pasan por los distintos chequeos y transformaciones en forma exitosa se ingresan en *VACANTE*.

Si el cargo es nuevo se ingresa en la tabla *PR\_CARGOS* y si no lo es se actualiza su tipo (vacante, comprometido, ejecutado) en caso de haber cambiado.

### **Carga Comprometido**

**Funcionalidad:** Carga los cargos comprometidos válidos para el mes de carga a la tabla limpia *COMPROMETIDO*.

**Especificación:** Borra los registros de la tabla *COMPROMETIDO*. Lee, valida y transforma los datos de la tabla *personal* del S.V.. Aquellos que pasan por los distintos chequeos y transformaciones en forma exitosa se ingresan a la tabla *COMPROMETIDO*.

Si el cargo es nuevo se ingresa en la tabla *PR\_CARGOS* y si no lo es, se actualiza su tipo ( vacante, comprometido, ejecutado) en caso de haber cambiado.

### **Carga Ejecutado**

**Funcionalidad:** La carga de ejecutado se encarga de ingresar los cargos ejecutándose, las extensiones horarias transitorias, las reducciones permanentes y transitorias y las dedicaciones que sean válidas para el mes de la carga.

**Especificación:** Se borra la tabla *EJECUTADO\_ANT*. Los registros *EJECUTADO*, correspondientes a la carga del mes anterior se borran de dicha tabla y se ingresan en *EJECUTADO\_ANT*.

Como primer paso se ingresan los cargo base y las reducciones permanentes. Como segundo paso se ingresan las extensiones/reducciones transitorias. Y por último se ingresan las dedicaciones. Todo se ingresa en la tabla limpia *EJECUTADO*.

Detallamos cada uno de estos pasos por separado para mayor claridad en los puntos siguientes.

### **Carga Permanente**

**Funcionalidad:** Ingresar los cargos base y las reducciones permanentes a la tabla limpia *EJECUTADO*.

**Especificación:** Se leen los datos de la tabla *wcgos* del S.N. se validan y transforman. Aquellos que pasan por los distintos chequeos y transformaciones en forma exitosa se ingresan en la tabla limpia *EJECUTADO*. Por cada cargo leído, si el mismo cargo se encuentra en la tabla limpia *EJECUTADO\_ANT*, se comparan las horas, y en caso que las horas actuales sean menores que las anteriores, se ingresa, además del registro correspondiente a las horas actuales del cargo, un registro a *EJECUTADO* correspondiente a la reducción horaria permanente (con la cantidad de horas reducidas).

### **Carga Ext/Red Transitorias**

**Funcionalidad:** Ingresar las extensiones y reducciones horarias de cargos a la tabla limpia *EJECUTADO*.

**Especificación:** Se leen los datos de la tabla *wextred* del S.N. se validan y transforman. Aquellos que pasan por los distintos chequeos y transformaciones en forma exitosa se ingresan en la tabla limpia *EJECUTADO*.

Se consideran todos los registros independientemente de que la fecha de cese indique que la ext/red cesó, lo anterior no quita que se emita una observación al respecto. Se cargarán aquellas ocurrencias cuya fecha de posesión sea menor o igual a la al día 15 del mes de carga.

Primero se recorren los registros correspondientes a reducciones transitorias y si son válidos se ingresan a *EJECUTADO* con cantidad de horas correspondientes a la reducción. Para cada cargo solamente se toma la reducción de mayor cantidad de horas en caso de que exista más de una.

Luego se recorren los registros correspondientes a extensiones transitorias ordenados en forma ascendente por las horas de la extensión.

La extensión de menor cantidad de horas, en caso de que también exista una reducción transitoria para el cargo en el mes, debe considerarse como horas de la extensión, la diferencia entre las horas de la extensión y las horas base del cargo, más las horas reducidas. La fuente de financiación de dicha extensión financia todas esas horas.

Si hay más de una extensión se deben ir calculando las horas extendidas leyéndolas en el orden de menor a mayor y haciendo la diferencia entre la hora extendida y las horas base más la diferencia extendida hasta ahora, la cual se debe ir acumulando de extensión a extensión.

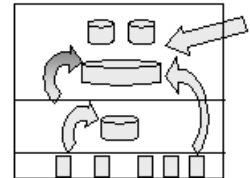
Al ordenar las extensiones de menor a mayor debería cumplirse que también las fechas de posesión estuvieran en ese orden

## Carga Dedicaciones

**Funcionalidad:** Ingresar las deducciones a la tabla limpia *EJECUTADO*.

**Especificación:** Se leen los datos de la tabla *wdedic* del S.N., se validan y transforman. Aquellos que pasan por los distintos chequeos y transformaciones en forma exitosa se ingresan en la tabla limpia *EJECUTADO*.

El importe correspondiente a la deducción, varía según el tipo de la misma. Las *deducciones totales* se calculan como la diferencia entre el sueldo de un DT (se obtiene de la tabla *SUELDO* con el grado del cargo y 41 horas) y el sueldo del cargo base. Las *deducciones especiales* se calculan como un 70 % del total percibido por el docente en el cargo (el porcentaje podría ser menor pero se toma este valor por no tener esta información en el sistema fuente). Las deducciones compensadas se calculan como un 45% sobre el sueldo del cargo base.



## Carga Histórico de Vacantes *VACANTE\_HIST*

**Funcionalidad:** Ingresar los cargos vacantes válidos en el mes de carga a la tabla *PR\_VACANTE\_HIST*.

**Especificación:** Se asocia a cada registro de la tabla limpia *VACANTE*, la fecha de la carga y se ingresan en la tabla *PR\_VACANTE\_HIST*.

## Carga Sumarización Vacantes

**Funcionalidad:** Ingresar la suma de vacantes válidos en el mes de carga para cada instituto y concepto en la tabla *PR\_VACANTE\_SUM*.

**Especificación:** Se asocia a cada registro de la tabla limpia *VACANTE*, la fecha de la carga y un indicador de si el concepto aporta al ahorro o gasto del mes. Se agrupan por fecha de la carga, instituto, concepto e indicador de ahorro/gasto, sumando por horas, equivalencia grado 1-10 horas y equivalencia en pesos uruguayos.

## Carga Histórico Comprometido

**Funcionalidad:** Ingresar los cargos comprometidos válidos en el mes de la carga a la tabla *PR\_COMPROMETIDO*.

**Especificación:** Se asocia a cada registro de la tabla limpia *COMPROMETIDO*, la fecha del mes de carga y se ingresan en *PR\_COMPROMETIDO*.

### **Carga Sumarización Comprometido**

**Funcionalidad:** Ingresar la suma de comprometidos válidos en el mes de carga para cada instituto y concepto en la tabla *PR\_COMPROMETIDO\_SUM*.

**Especificación:** Se asocia a cada registro de la tabla limpia *COMPROMETIDO*, la fecha de la carga y un indicador de si el concepto aporta al ahorro o gasto del mes. Se agrupan por fecha de la carga, instituto, concepto e indicador de ahorro/gasto, sumando por horas, equivalencia grado 1-10 horas y equivalencia en pesos uruguayos.

### **Carga Jerarquía Docente Grado**

**Funcionalidad:** Ingresar las nuevas combinaciones de docente – grado en la jerarquía *PR\_DOCENTE\_GRADO*.

**Especificación:** Se recorren los registros correspondientes a cargos base válidos en el mes de carga de la tabla limpia *EJECUTADO* y si el docente tiene un grado nuevo, por lo tanto no existía la combinación en la jerarquía, se ingresa el código de docente y el grado en *PR\_DOCENTE\_GRADO*, con una clave que se genera secuencialmente.

### **Carga Histórico Ejecutado**

**Funcionalidad:** Ingresar los cargos del ejecutado (cargos base, reducciones permanentes, extensiones/reducciones transitorias, dedicaciones) válidos en el mes de la carga a la tabla *PR\_EJECUTADO\_HIST*.

**Especificación:** Se asocia a cada registro de la tabla limpia *EJECUTADO*, la fecha del mes de carga y se ingresan en *PR\_EJECUTADO*.

### **Carga Sumarización Ejecutado**

**Funcionalidad:** Ingresar la suma de ejecutados válidos en el mes de carga para cada instituto, concepto, combinación docente-grado y fuente de financiación, en la tabla *PR\_EJECUTADO\_SUM*.

**Especificación:** Se asocia a cada registro de la tabla limpia *EJECUTADO*, la fecha de la carga, un indicador de si el concepto aporta al ahorro o gasto del mes y el identificador de la jerarquía que conforman el docente y su grado. Se agrupan por fecha de la carga, instituto, concepto e identificador de jerarquía docente-grado, sumando por horas, equivalencia grado 1-10 horas y equivalencia en pesos uruguayos.

## **B. RUTINAS PARA LA LIMPIEZA DE DATOS**

El conjunto de rutinas fue diseñado de forma de resolver todos los problemas detectados en la etapa de *Análisis*. Algunas intentan corregir errores detectados en el *Análisis Detallado* de los datos, otras resuelven los problemas de integración resultado de la *Selección de Fuentes* realizada, otras transforman datos según lo indicado en el *Análisis de la Carga*. En general, todas ellas apuntan a mejorar el estado de calidad de los datos del data mart y son invocadas todo a lo largo de la carga del mismo. En el *Apartado 2, Metadata del Sistema* puede encontrarse la descripción de todas las rutinas de limpieza diseñadas.



### C. PERIODICIDAD DE LA CARGA

La carga de institutos y conceptos está pensada para realizarse una única vez durante la puesta en marcha del sistema. En caso que se creen nuevos institutos en la F.I. es posible ejecutar el proceso para ingresar dichos datos. El administrador del data mart de presupuesto, persona que debe conocer las reglas de negocio del presupuesto de la F.I. es el encargado de ingresar estos datos.

La carga del importe máximo disponible por los institutos debe realizarse mensualmente antes de que la herramienta OLAP extraiga los datos del data mart de presupuesto.

La carga de datos fuente se debe realizar posteriormente a que la Sección de Personal de la F.I. cierre cada mes. Esto ocurre generalmente durante los primeros 15 días de cadames, por lo que estaría correcto ejecutar la carga alrededor del día 15 de cada mes. El proceso de carga asociará a sus datos la fecha en la que se realiza la carga.

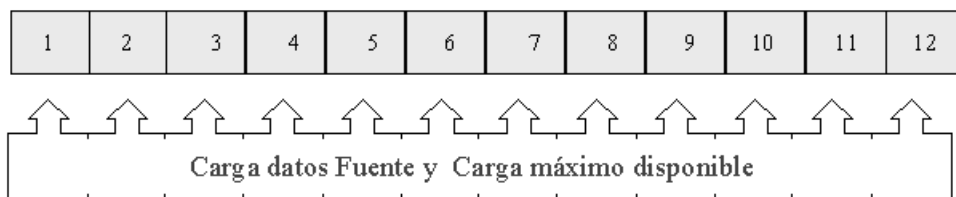


Figura 19: diagrama de periodicidad anual de carga

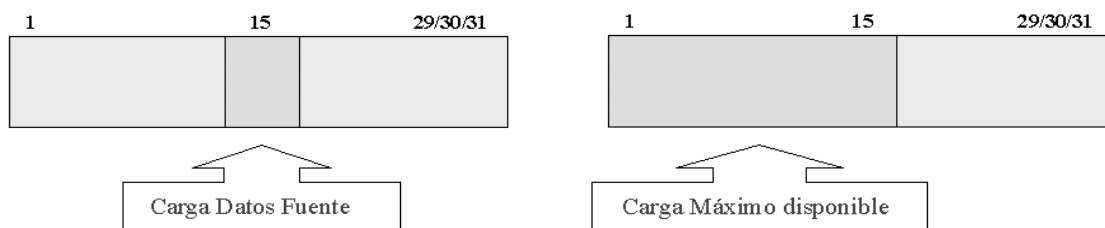


Figura 20: diagramas de carga mensual

## 1.5 IMPLEMENTACIÓN

---

Como resultado del desarrollo del data mart de Presupuesto se obtuvieron los siguientes componentes principales: bd del data mart, procesos de carga y metadata. Se detalla a continuación información de las bd del data mart y los procesos de carga. La metadata se documenta en el *Apartado 2, Metadata del Sistema*.

### Base de datos del data mart.

Nombre: 'bede98'  
DBMS: Oracle 8  
Ubicación: servidor felipe del INCO  
Contenido: información del presupuesto mensual de la F.I.  
Usuario: dw\_bede/dw\_bede

### Procesos de carga.

Proceso de interface:

Nombre: 'Cargas\_Datamarts.exe '  
Ubicación: 'C:\DWTaller5\Vb' del servidor tornasol del INCO  
Leng.Prog.: Visual Basic  
Funcionalidad: Interface central para ejecución de procesos de carga. Permite disparar el proceso de carga de presupuesto. Contiene abm's para la carga de datos a las tablas base. Contiene abm's para realizar los mapeos. Presenta un log de errores y advertencias.

Procesos de Carga:

Nombre: 'Carga de Docentes'; 'Carga de Sueldos'; 'Carga de Fuentes'; 'Carga del Mapeo de Institutos'; 'Carga del Mapeo de Provisiones'; 'Carga de Vacante'; 'Carga de Comprometido'; 'Carga de Ejecutado'; 'Carga de Docentes-Grado'; 'Carga del historial de Ejecutado'; 'Carga de sumarizaciones de Ejecutado'; 'Carga del historial de Comprometido'; 'Carga de sumarizaciones de Comprometido'; 'Carga del historial de Vacante'; 'Carga de sumarizaciones de Vacante'  
Ubicación: base de datos 'bede98'  
Leng.Prog.: PLSQL  
Funcionalidad: Procesos de carga del presupuesto mensual.

## 2 DATA MART: BEDELIA

---

### 2.1 REQUERIMIENTOS

---

#### ➤ REQUERIMIENTOS TECNOLÓGICOS

---

*Los requerimientos tecnológicos especifican el ámbito de implementación del data mart.*

---

El data mart de Bedelía se construye sobre una base de datos ORACLE que se encuentra en un servidor UNIX dentro del INCO. El data mart compartirá la base de datos con el data mart de Presupuesto y juntos formarán el data warehouse del INCO.

Los datos del data mart podrán ser accedidos utilizando herramientas de consultas a bases de datos relacionales y utilizando herramientas OLAP, en particular Cognos.

No se especifica una herramienta o lenguaje de programación especial para implementar el proceso de carga. Sí existen restricciones al respecto, la tecnología debe poder acceder a la base de datos ORACLE, debe correr sobre UNIX y el INCO debe disponer de la misma.

#### ➤ MODELO CONCEPTUAL MULTIDIMENSIONAL

---

##### A. DIMENSIONES

Dimension/Categoría	Descripción	Datos
Carrera	Representa las carreras de la F.I..	
Plan	Plan de la carrera	(Nom_plan [vigente,anterior])
Carrera	Carrera-Ciclo	(Id_carrera, Id_ciclo, nom_carrera, nom_ciclo)
Estudiante	Representa los estudiantes de la F.I..	
Generación	Generación del estudiante	(Id_generación)
Estudiante	Estudiante	(Id_estudiante, nombre)

Desempeño	Desempeño.	(Id_desempeño, nombre [Muy Activo, Activo, Poco Activo, Abandono])
Fecha		
Año	Año	(nro_año)
Semestre	Semestre	(nro_semestre[1,2] )
Período	Período	(nro_período [1..nésimo período] )
Materia	Materias de las carreras de la F.I..	(Id_materia, nombre)
Tipo Actividad	Tipos de actividades.	(Id_actividad, nombre)
Resultado	Resultados de las actividades	
Tipo Aprobación	Tipo de Aprobación.	(Id_taprob [CN-con nota, SN-sin nota, FR-franja resultado])
Tipo Resultado	Tipo de resultado.	(Id_tresult [Aprobado, Reprobado])
Nota	Nota	(nro_nota)

## B. MEDIDAS

- Cantidad de Estudiantes: Cantidad de estudiantes
- Promedio Aprobados: Promedio de notas entre aprobados
- Promedio Real: Promedio de notas real (aprobados y no aprobados)
- Cantidad de Créditos: Cantidad de créditos

### C. RELACIONES DIMENSIONALES ENTRE DIMENSIONES Y MEDIDAS

Dado un estudiante, una carrera, y un desempeño, podemos medir a lo largo del tiempo: la *cantidad de créditos* del estudiante en la carrera, el *promedio real* y el *promedio entre aprobados*.

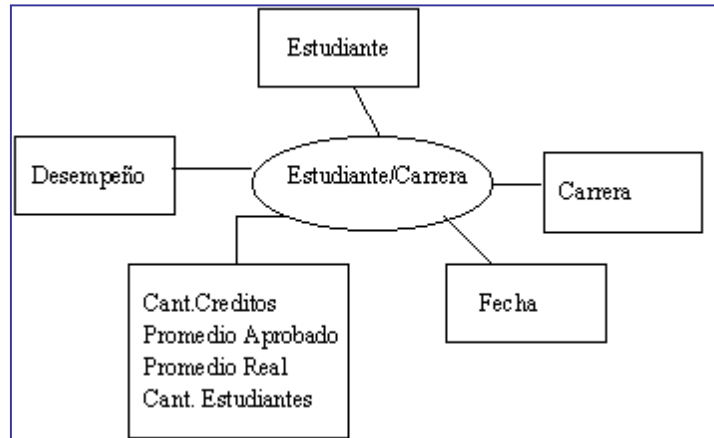


Figura 21: Relación Dimensional Estudiante/Carrera

Podemos medir a lo largo del tiempo, la *cantidad de créditos*, el *promedio real*, el *promedio entre aprobados* y la *cantidad de estudiantes* dado: una carrera, un estudiante, una materia, un tipo de actividad y un resultado.



Figura 22: Relación Dimensional Actividades

## ➤ RELEVAMIENTO DE FUENTES

---

*Se releva información sobre los sistemas operacionales utilizados para llevar la información de Bedelía. Se analiza básicamente la funcionalidad de los sistemas y el uso que se les da a los mismos.*

---

La información de bedelía se encuentra en tablas dentro de la base de datos de data warehouse del INCO. Las tablas mantienen información de estudiantes, carreras, ingresos a carreras, títulos, materias, grupos de materias, actividades de estudiantes y egresos. De aquí en adelante nos referiremos a este conjunto de tablas como la base de datos fuente de bedelía.

Las tablas de bedelía son cargadas periódicamente, con datos provenientes de la bd del sistema operacional de la sección de Bedelía de la F.I. para llevar la información acerca de los estudiantes de ingeniería y su evolución dentro de la carrera. La tarea de definición, implementación y ejecución de la carga de los datos de dicho sistema a las tablas en data warehouse del INCO no compete a este proyecto.



Por motivos ajenos al proyecto, el relevamiento del sistema no se realiza en la forma correcta. Lo adecuado hubiese sido obtener información precisa del sistema operacional de bedelía, accediendo a documentación del mismo y entrevistando usuarios y/o desarrolladores. Dado que esto no fue posible, el relevamiento se realiza sobre las tablas con datos de bedelía, utilizando un documento que describe su contenido. El documento brinda la información suficiente como para realizar el relevamiento del sistema pero no tiene el nivel de detalle necesario para las etapas de análisis posteriores.

## 2.2 ANALISIS

---

### ➤ ANALISIS GENERAL

---

*Se estudia la información que manejan los sistemas operacionales o sea sus bases de datos. Con qué tablas cuentan, que datos tienen y cómo éstos se relacionan. Esto nos brinda una visión global de la información que mantienen los sistemas operacionales.*

---

Estudiamos la base de datos fuente de bedelía, analizando el contenido de las tablas que las componen. En el *Apartado 2, Metadata del Sistema* se documenta el análisis general completo, en esta sección se describe únicamente el contenido general de las tablas.

A partir del modelo físico de la bd se construye un modelo conceptual representando la realidad que maneja la bd fuente. En el *Apartado 2, Metadata del Sistema* se documenta el mapeo entre las estructuras de la base de datos y los objetos del modelo conceptual.

#### A. ESTUDIO BASES DE DATOS FUENTES

TABLA	DESCRIPCION
INGRESOS_ACTIV	Actividades de estudiantes del plan nuevo
ADAP_ACTIV	Actividades de estudiantes adaptados al plan nuevo
VIEJOS_ACTIV	Actividades de estudiantes de planes anteriores
INGRESOS_ICARR	Inscripción a carreras de estudiantes del plan nuevo
ADAP_ICARR	Inscripción a carreras de estudiantes de planes anteriores que se adaptan al plan nuevo
VIEJOS_ICARR	Inscripciones a carreras de estudiantes en planes anteriores
INGRESOS_ESTUD	Datos de estudiantes en plan nuevo
ADAP_ESTUD	Datos de estudiantes adaptados al plan nuevo
VIEJOS_ESTUD	Datos de estudiantes de planes anteriores
MATERIAS	Materias de la F.I.

<b>TABLA</b>	<b>DESCRIPCION</b>
<i>MAT_CARR</i>	Materias para cada carrera ciclo
<i>PREVIATURAS</i>	Previaturas de materias
<i>CARRERAS_CICLOS</i>	Carreras ciclos de la F.I.
<i>LUGARES</i>	Lugares
<i>INSTITUTOS</i>	Institutos
<i>GRUPOS</i>	Grupos
<i>MAT_GRUPO</i>	Materias que pertenecen a grupos
<i>ADAP_EGRESOS</i>	Egresos de estudiantes adaptados al plan nuevo
<i>VIEJOS_EGRESOS</i>	Egresos de estudiantes de planes anteriores
<i>DEF_TITULOS</i>	Títulos de las carreras ciclos

La información referida a estudiantes, inscripciones a carreras y actividades está particionada horizontalmente en tablas, distinguidas por la finalización de su nombre (\*\_ESTUD, \*\_ACTIV y \*\_ICARR) según los datos se refieran a estudiantes que hayan:

- ingresado a Facultad por el plan nuevo, a partir del año 97 (INGRESOS\_\*).
- adaptado al nuevo plan, o sea, optado por el nuevo plan estando inscriptos en alguna carrera de planes anteriores al 97 (ADAP\_\*).
- seguido en las carreras de planes anteriores al 97 (VIEJOS\_\*).

También los datos de egresos están divididos en forma similar, pero sólo en estudiantes adaptados al nuevo plan o los que siguen en los planes anteriores al 97, ya que aún no hay ningún egreso del plan nuevo (tablas ADAP\_EGRESOS y VIEJOS\_EGRESOS).



El análisis de las tablas y sus relaciones fue realizado a partir del documento mencionado en la fase de relevamiento de sistemas.





Las materias forman parte de carreras. Una materia puede formar parte de más de una carrera, donde tiene asociado un nombre y la cantidad de créditos que aporta aprobar esa materia. Las carreras están formadas por muchas materias.

Las materias forman parte de grupos. De los grupos sabemos su código identificador, el nombre y la cantidad de créditos mínima y máxima. Una materia puede formar parte de más de un grupo y dentro de un grupo puede haber más de una materia.

Los estudiantes realizan actividades de materias, y se sabe el código identificador de la actividad, la fecha, el tipo, la clase de actividad, la nota y observaciones.

Los estudiantes egresan obteniendo un título en una carrera, en una fecha con indicación de si se obtuvo por reválida.

Los estudiantes residen en lugares de los que conocemos un código identificador y su nombre. En los lugares se encuentran institutos identificados por un código de los que sabemos además su nombre y su tipo.

## ➤ IDENTIFICACIÓN DE DATOS REQUERIDOS

---

*Se analizan las dimensiones y medidas documentados en el modelo conceptual multidimensional y se determina si existe información almacenada al respecto, consultando los modelos conceptuales de los sistemas fuente. En caso que exista dicha información, se define cual es el objeto del modelo conceptual que la mantiene.*

---

En los cuadros siguientes se encuentra especificado el mapeo entre las dimensiones y las medidas del modelo multidimensional de Bedelía contra los objetos del modelo conceptual de la bd fuente de bedelía. Para no repetir los objetos conceptuales, se numeran en el siguiente cuadro.

Sistema Bedelía
1. Materia
2. Mat_Car
3. Mat_Grup
4. Grupos
5. Previas
6. Carreras
7. Estudiante
8. Activ
9. Icar
10. Obtiene
11. Titulos
12. Egresos
13. Lugares
14. Reside
15. Ubica
16. Institutos

**A. DIMENSIONES**

		<b>O B J E T O S</b>																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
<b>D I M E N S I O N</b>	Estudian tes							X										
	Genera cion							X		X								
	Carrera						X											
	Plan						X											
	TipoActi vidad								X									
	Tipo Aprob.								X									
	Tipo Result.								X									
	Nota								X									
	Fecha																	
	Desemp eño																	
Materias	X																	

Figura 24: Mapeo entre dimensiones y objetos conceptuales

**B. MEDIDAS**

		<b>O B J E T O S</b>															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
<b>M E D I D A</b>	Cantida dEstudi antes							X									
	Promed ioAprob ados								X								
	Prome dio Real								X								
	Cantida d Credito		X														

Figura 25: Mapeo entre medidas y objetos conceptuales

**D. CRUZAMIENTOS**

D I M E N S I O N												
D I M E N S I O N		Estudiante	Generación	Carrera	Plan	Tipo Actividad	Tipo Aprobación	Tipo Resultado	Nota	Desempeño	Materias	
	Estudiante		{7,9 }	{7,9 ,6}	{7,9 ,6}		{7,8 }	{7,8 }	{7,8 }	{7,8 }		{7,8 ,1}
	Generación			{9,6 }	{9,6 }	{9,7 ,8}	{9,7 ,8}	{9,7 ,8}	{9,7 ,8}			{9,7 ,8,1 }
	Carrera				{6}	{6,9 ,7,8}	{6,9 ,7,8}	{6,9 ,7,8}	{6,9 ,7,8}			{6,9 ,7,8, 1}
	Plan					{6,9 ,7,8}	{6,9 ,7,8}	{6,9 ,7,8}	{6,9 ,7,8}			{6,9 ,7,8, 1}
	Tipo Actividad						{8}	{8}	{8}			{8,1 }
	Tipo Aprob.							{8}	{8}			{8,1 }
	Tipo Result.								{8}			{8,1 }
	Nota											{8,1 }
	Desempeño											
	Materias											

*Figura 26: Mapeo entre cruzamientos y objetos conceptuales*

## ➤ ANALISIS DETALLADO

---

*Se realiza un análisis minucioso de los datos en cada tabla, estudiando todos los valores posibles en cada campo y cualquier otro dato que pueda estar oculto detrás de los mismos.*

---

El análisis detallado se realiza sobre las tablas de la base de datos fuente de bedelía, en particular sobre las tablas que corresponden a objetos del modelo conceptual identificados en los cuadros de mapeos vistos en el punto anterior.

El resultado completo del análisis detallado se encuentra en el *Apartado 2, Metadata del Sistema*, se presenta a continuación un resumen del mismo, con resultados a nivel general y de cada una de las tablas analizadas.

La base de datos fuente de bedelía presenta sus datos a un nivel aceptable de calidad. En base a qué afirmamos eso, primero porque los datos que se encuentran en las tablas, son en general, coherentes a la documentación de las mismas, segundo porque la información está normalizada, con un buen uso de codificación, tercero porque los datos se adaptan fácilmente a los datos requeridos y por último, porque si bien encontramos problemas, pensamos que tienen solución.

Sin embargo, a pesar de lo dicho, no se confía plenamente en la información fuente. Dado que no se tiene conocimiento de los procesos que realizan la extracción de los datos del sistema operacional de bedelía, no se puede afirmar que el proceso no modifique la información, ni que tome toda la información y ni que lo haga correctamente.

Dada una situación similar en otro proyecto de construcción de un data mart, esta base de datos de bedelía se descartaría como fuente de datos. En nuestro caso, como no sólo tenemos de objetivo construir el data mart de bedelía sino que también hay fines didácticos y de experimentación, y dado que es la única fuente existente, decidimos considerar como fiables los datos fuente, o sea considerarlos como exactos y completos.

### **Carreras\_ciclos**

En esta tabla se mantiene la lista de carreras y ciclos de la F.I. No tiene la lista completa de todas las carreras que existen en realidad, sí tiene todas las correspondientes al INCO.

### **Ingresos\_Estud, Adap\_Estud y Viejos\_Estud**

Se encuentra en estas tablas información de los estudiantes de la F.I., tanto actuales como egresados. El campo estci es usado como clave del estudiante. Los campos estci, nomest, nacido, sexo y est son adecuados para una identificación completa de la persona y tienen valores correctos. El campo nacido sin embargo no es año 2000 compatible. El resto de los campos no tiene valores estandarizados y a veces vacíos, igualmente no es información importante para el data mart.

### **Ingresos\_icarr, Adap\_icarr y Viejos\_icarr**

Contienen información de las inscripciones de estudiantes a carreras – ciclos y su estado de inscripción. Los campos de fecing y feceg contienen valores de fechas que no son año 2000 compatibles. Los códigos identificatorios de carreras-ciclos se corresponden correctamente a los códigos de la tabla carreras\_ciclos. Los valores en el campo estci se corresponden a los valores en alguna de las tablas ingresos\_estud, adap\_estud y viejos\_estud. Sin embargo no se corresponden según la partición horizontal realizada, o sea, los que están en ingresos\_icarr supuestamente deberían estar en ingresos\_estud y lo mismo para adap\_\* y viejos\_\*.

Los valores en los campos calisnc y porcambio tienen una codificación estándar, no hay valores fuera de esa codificación.

### **Materias**

La tabla materias mantiene la lista de todas las materias dictadas en F.I. El código identificatorio de las materias en casi todos los casos esta formado por letras que representan la carrera correspondiente a la materia.

### **Mat\_carr**

Mantiene la relación entre las materias y las carreras-ciclos, indica qué materias pertenecen a qué carreras\_ciclos. Los códigos identificatorios de carreras-ciclos se corresponden correctamente a los códigos de la tabla carreras\_ciclos. Lo mismo sucede con los códigos de materias y los códigos de la tabla materias.

### **Grupos**

Contiene los grupos de materias. El código identificador de los grupos parece representar una carrera correspondiente al grupo.

### **Mat\_grupo**

Mantiene la relación entre las materias y los grupos, indica qué materias pertenecen a qué grupos. Tanto los códigos de materias como los de grupos se corresponden a los códigos de las tablas materias y grupos respectivamente.

### **Def\_titulos**

Mantiene los títulos que se obtienen en las distintas carreras. Los códigos de carreras, ciclos se corresponden a los ingresados en la tabla carreras\_ciclos.

### **Adap\_egresos, Viejos\_egresos**

Estas tablas contienen los egresos de estudiantes. El campo feceg contiene valores de fechas no compatibles con el año 2000. Tanto códigos de carreras, ciclos, estudiantes como

títulos se corresponden a los códigos ingresados en las tablas carreras\_ciclos, \*\_estudiantes y def\_titulos respectivamente.

## ➤ SELECCIÓN DE DATOS

*Se seleccionan los objetos a extraer de las bases fuente. Se construye un modelo conceptual que permite visualizar los datos con los que se cuenta para cargar el futuro data mart.*

Casi toda la información mantenida en la base de Bedelía es requerida para la construcción del data warehouse como se puede ver en el punto Identificación de datos requeridos.

En este caso tenemos una única fuente de información por lo que no es necesario tomar decisiones respecto de cuál es la fuente más apropiada.

### A. MODELO CONCEPTUAL DATOS SELECCIONADOS

El modelo conceptual de datos seleccionados se mantiene igual al de la base de datos fuente de bedelía a menos de algunas diferencias. Se descartan los objetos *recide*, *lugares*, *ubica* e *institutos*. Estos objetos no son requeridos por el modelo multidimensional y tampoco consideramos que su información sea de relevancia para el data mart.

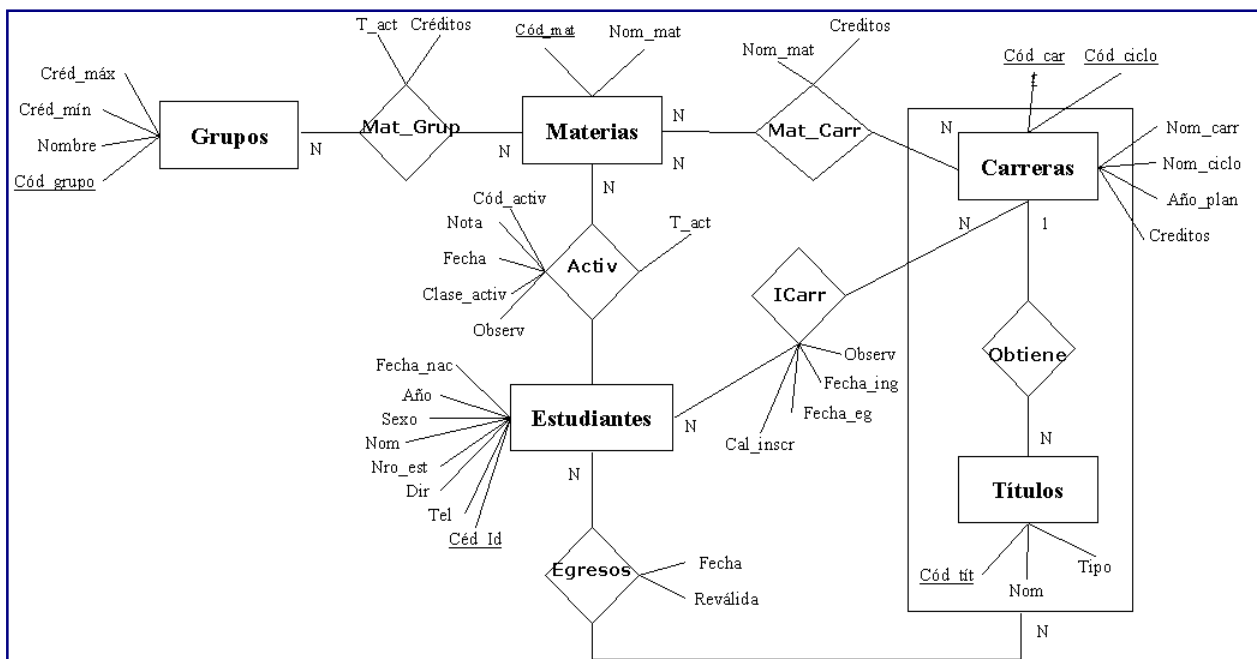


Figura 27: Modelo conceptual bedelía de datos seleccionados

Si bien no hay requerimientos explícitos acerca de estudiantes egresados y títulos obtenidos, se considera que estos datos pueden ser de utilidad para reportes específicos e incluso chequeos de validación por eso se seleccionan los objetos *obtiene*, *título* y *egresos*.

## **B. CORRESPONDENCIA DEL MODELO DE DATOS SELECCIONADOS CON TABLAS FUENTE SELECCIONADAS**

No se considera necesario realizar el mapeo entre el modelo de dato seleccionados y el modelo físico de datos fuente por ser éste muy similar al modelo conceptual de la base de dato fuente a menos de algunas entidades que no fueron seleccionadas. El mapeo del modelo conceptual del sistema fuente con el modelo físico se encuentra documentado en el *Apartado 2, Metadata del Sistema*.

### **➤ ANÁLISIS DE CARGA**

---

*Se describen los problemas que se deberán enfrentar a la hora de cargar el data warehouse.*

---

#### **1. Datos que faltan**

- a) Notas de parciales. El sistema esta diseñado para mantener esta información pero no está disponible al momento de realizar este análisis. Se espera en un futuro disponer de la misma. La ausencia de estos datos afecta los cálculos de promedios.
- b) Desempeño. La lista de desempeños posibles no se encuentra en el sistema fuente y debe encontrarse la forma para generarla.

#### **2. Datos a integrar**

- a) Como la información de Bedelía se encuentra solamente en un sistema, no surgen problemas de integración de datos entre diferentes fuentes. En este caso tampoco tenemos información duplicada dentro del mismo sistema.

#### **3. Datos a calcular**

- a) Promedio aprobado. Esta información no es mantenida como tal por el sistema. Se dispone de información sobre las notas de actividades aprobadas, a partir de la cual puede calcularse el promedio entre aprobados. Se dispone además de información sobre cuales actividades son aprobadas con nota o sin nota, dato que es importante para considerar solamente en el cálculo las actividades con nota.



- c) Promedio real. Esta información no es mantenida por el sistema. Se dispone de información sobre las actividades reprobadas y aprobadas junto con la nota correspondiente, a partir de la cual puede calcularse el promedio real. Se dispone además de información sobre cuales actividades son aprobadas con nota o sin nota, dato que es importante para considerar solamente en el cálculo las actividades con nota.
- d) Cantidad de créditos. Se dispone de la cantidad de créditos que se obtienen al aprobar una materia de una carrera-ciclo para realizar los cálculos correspondientes.
- d) Cantidad de estudiantes. Dado que se tiene información de todos los estudiantes es fácilmente calculable la cantidad de ellos.
- e) Desempeño. Dado que se guardan las actividades aprobadas por un estudiante junto con la fecha de aprobación es posible realizar asociar a cada estudiante en un año el desempeño correspondiente.
- f) Aprobación de Actividades. Una materia puede pertenecer a muchas carreras y un estudiante puede estar inscripto en muchas carreras, por esto, cada vez que el estudiante apruebe una actividad de una materia dada, ésta será válida para todas las carreras en las que esté inscripto el estudiante en las cuales dicha materia esté presente.

#### **4. Datos a transformar**

- a) Generación. Del estudiante se conoce la fecha de inscripción a la carrera, a partir del año de dicha fecha se obtiene la generación del estudiante.
- b) Plan anterior o vigente. De cada carrera se conoce el año del plan de la misma, de donde se deduce si el plan es el vigente (1997) o anterior (<1997).
- c) Tipo Actividad. A partir del código del tipo de actividad en \*\_activ se puede deducir el tipo de actividad que representa (asistencia, examen, curso, parcial1,..., parcial9).
- d) Nota. Los tipos de actividades asistencia y curso no llevan nota. No se puede determinar en el sistema si el curso se aprueba con nota o no.

#### **5. Datos a limpiar**

- a) Las tablas ( \*\_activ, \*\_estud, \*\_icarr) que supuestamente dividen lógicamente la información de estudiantes en el plan nuevo de los que se adaptaron al plan nuevo y de los que continúan en planes anteriores, no lo hacen correctamente. Hay datos repetidos en distintas tablas. Uniendo la información se soluciona el problema.
- b) Actividades. Interesan solamente las actividades cuyo tipo de actividad no sea caducado y aquellas que sean curriculares.

c) Fechas. Todas las fechas no son compatibles al año 2000 y deben transformarse adecuadamente.

## **6. Granularidad de datos**

a) Notas. El sistema maneja actividades con o sin nota, por lo que no todos los tipos de resultados tienen una nota asociada.

## **7. Volatilidad de la información**

En general toda la información mantenida por la bd fuente de bedelía no es volátil, los datos se guardan durante años y se van actualizando.

Las tablas materias, mat\_carr, carreras\_ciclos, grupos, mat\_grupos y titulos contienen información válida al momento de la carga. Los datos no varían mucho en el tiempo, difícilmente sufran actualizaciones pero sí aparecen datos nuevos.

Las tablas \*\_estud mantienen información de todos los estudiantes que ingresaron a F.I. desde hace unos años hasta la fecha. Se actualizan los datos personales en caso que cambien pero los datos identificatorios del estudiante no se modifican. Anualmente, en entre los meses de febrero y marzo se ingresan los estudiantes nuevos. Suponemos que no se borran registros.

Las tablas \*\_icarr mantienen información de los ingresos de todos los estudiantes a F.I. desde hace unos años hasta la fecha. Los datos se actualizan al momento que el estudiante, por algún motivo, finaliza la carrera. Anualmente, entre los meses de febrero y marzo se insertan los ingresos a carreras nuevos. Suponemos que no se borran registros.

Las tablas \*\_activ guardan los datos de los resultados de las actividades de materias realizadas por los estudiantes desde hace unos años hasta la fecha. Se ingresan registros en estas tablas durante todo el año, sobre todo al finalizar los períodos de exámenes. La información difícilmente se borre y tampoco se actualice.

## ➤ INTEGRACIÓN CON OTROS DATA MARTS

---

*Se analizan posibles problemas de integración con la información ya almacenada en el data warehouse relacional.*

---

El análisis de integración con otros data marts se realiza en caso que los data mart constituyan un único data warehouse como en nuestro caso.

En el caso del data mart de bedelía, los posibles conflictos con el data mart de presupuesto podrían haber surgido con la información acerca de institutos. Si bien el nombre es el mismo, el data mart de presupuesto maneja datos de institutos de la F.I pero los institutos de la bd fuente de bedelía corresponden a institutos de educación secundaria de donde provienen los estudiantes que ingresan a la facultad. Esta información no fue considerada para el data mart, por lo tanto no se necesitó resolver el conflicto.

Si en un futuro se decide ingresar al data mart estos institutos manejados por bedelía, deberían ingresarse al data mart identificándose bajo otro nombre, que no diera a confusión.

## **2.3 INTERACCION CON GRUPO OLAP**

---

Se acordaron los siguientes puntos:

- Los cubos se cargarán en forma total cada vez, no descartándose la idea de realizar en el futuro cargas incrementales de los mismos.
- Se definen formatos para los valores de ciertos campos de las tablas del data mart.

## 2.4 DISEÑO

---

### ➤ ARQUITECTURA DEL DATA MART

---

*Se define en forma general los componentes del data mart, tanto procesos como estructuras de almacenamiento.*

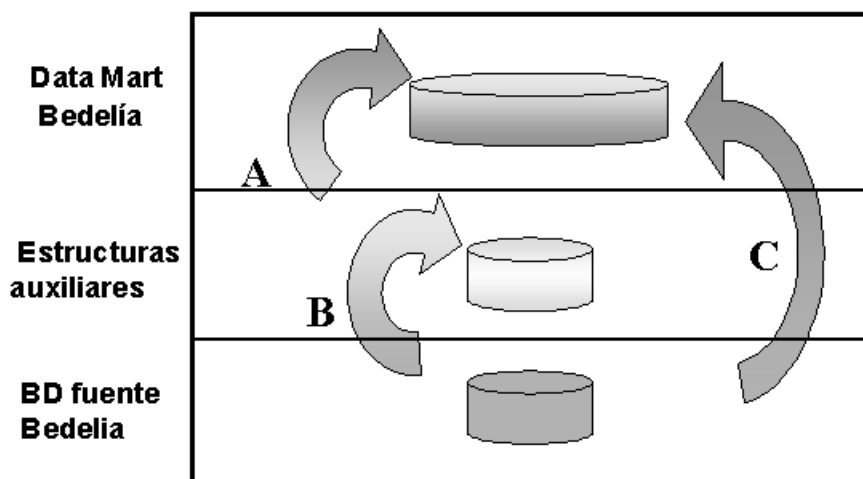
---

La base de datos fuente en este caso, se encuentra dentro de la misma base de datos donde se construirá el data mart. Esto permite acceder fácilmente a los datos fuente.

Las tablas del data mart se diseñan aplicando primitivas a las estructuras de las tablas fuente.

Según el análisis de carga se necesitan realizar algunas transformaciones y cálculos pero en general los datos fuente se pueden convertir fácilmente a una forma adecuada para el data mart, siguiendo las pautas del modelo multidimensional. Se diseña un proceso de carga que extraiga los datos fuente, los transforme, limpie y por último ingrese en las tablas del data mart.

El nivel de granularidad a almacenar los datos es el representado por el modelo multidimensional de bedelía.



*Figura 28: Arquitectura Data mart Bedelía*

En la *figura 28* se presentan los distintos componentes de la arquitectura del data mart de bedelía. Se tiene un conjunto de tablas que forman la bd fuente de bedelía, otro conjunto de tablas auxiliares que solucionan el tema de la mala división de datos según estudiantes en

planes anteriores, actuales y adaptados. Por último están las tablas de data mart que mantienen la información al nivel de mayor granularidad.

El proceso B realiza la unión de los datos que se encuentra particionados horizontalmente ( ver punto limpieza dentro del análisis de carga). Los procesos C y A trasladan, transforman, validan e ingresan los datos a las tablas del data mart.

## ➤ DISEÑO DE TABLAS

---

*Se realiza el diseño de las tablas relacionales del data mart y de cualquier otra estructura auxiliar necesaria tanto para facilitar la carga como para la limpieza y transformación de datos.*

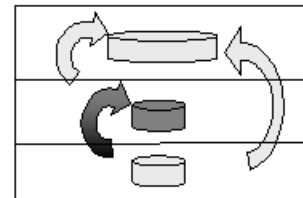
---

Se crean vistas para poder integrar la información clasificada en ingresos, adaptados y viejos de las tablas de estudiantes, inscripciones a carreras, actividades y egresos.

El diseño de las tablas del data mart se realiza utilizando las primitivas presentadas en la sección 2.4 *Diseño de un Data Warehouse, Informe Principal*. Se parte de la estructura de las tablas de la bd de bedelía y de las vistas creadas, se aplican las primitivas reiteradamente hasta llegar al diseño de las tablas del data mart.

Estándar de nomenclatura utilizado: las tablas que componen el data mart propiamente dicho se identifican por comenzar con 'BD\_'.

### Diseño Estructuras auxiliares



#### ESTUDIANTES

```
Create view ESTUDIANTES as
select (estci, nomest, lugar, nacido, sexo, anio, inst, dir, tel, est) from INGRESOS_ESTUD union
select (estci, nomest, lugar, nacido, sexo, anio, inst, dir, tel, est) from ADAP_ESTUD
union
select (estci, nomest, lugar, nacido, sexo, anio, inst, dir, tel, est) from VIEJOS_ESTUD
```

#### INSC\_CARR

```
Create view INSC_CARR as
select (estci, carr, ciclo, fecing, feceg, lugar, obs, calinsc, 'N') from INGRESOS_ICARR
union
select (estci, carr, ciclo, fecing, feceg, lugar, obs, calinsc, porcambio) from ADAP_ICARR
union
select (estci, carr, ciclo, fecing, feceg, lugar, obs, calinsc, porcambio) from VIEJOS_ICARR
```

#### ACTIVIDADES

Create view ACTIVIDADES as

```
select (estci, mat, tact, nota, fecha, curri, tgen, per, tipoper, obs) from INGRESOS_ACTIV
union
select (estci, mat, tact, nota, fecha, curri, tgen, per, tipoper, obs) from ADAP_ACTIV
union
select (estci, mat, tact, nota, fecha, curri, tgen, per, tipoper, obs) from VIEJOS_ACTIV
```

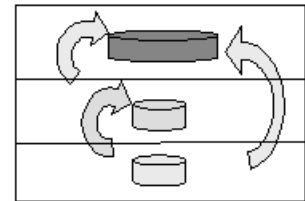
### *EGRESOS*

Create view EGRESOS as

```
select (carr, ciclo, titulo, estci, feceg, revalida) from INGRESOS_EGRESOS
union
select (carr, ciclo, titulo, estci, feceg, revalida) from ADAP_EGRESOS
union
select (carr, ciclo, titulo, estci, feceg, revalida) from VIEJOS_EGRESOS
```

### **Diseño Tablas Data mart**

A partir del modelo físico de la base de datos de Bedelía y las vistas construidas, aplicando las primitivas presentadas, se diseña el modelo físico del data mart relacional de Bedelía.



Se presenta un resumen global de la aplicación de las primitivas y luego el detalle, donde para cada una de las tablas se indican las primitivas utilizadas, junto con los parámetros de entrada y el resultado obtenido.

**A. APLICACIÓN PRIMITIVAS – DATA MART BEDELÍA**

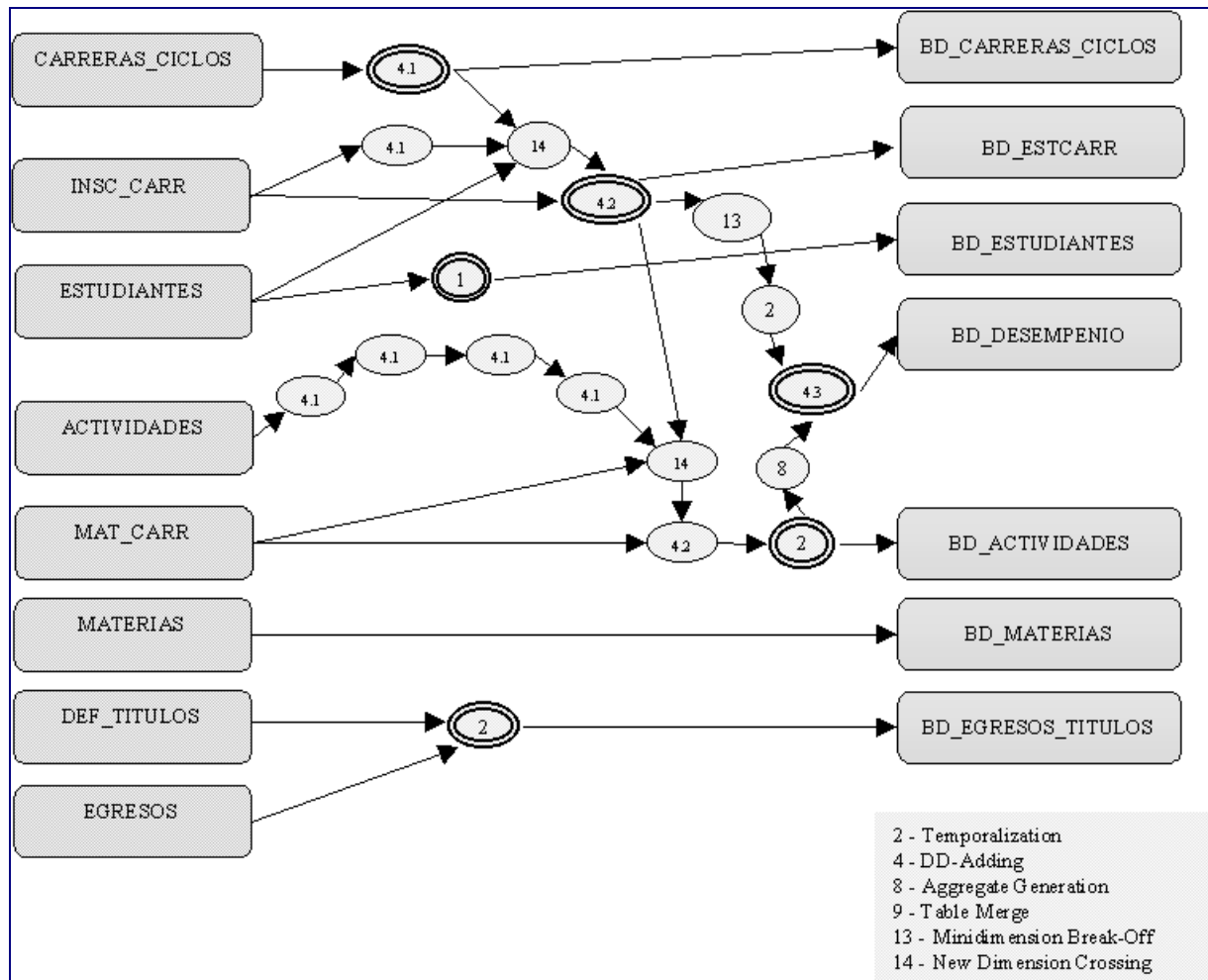


Figura 29: Diagrama de aplicación de primitivas Bedelía

## B. FORMALIZACION DE APLICACIÓN DE PRIMITIVAS

Para cada tabla del data mart, construida a partir de primitivas, se detallan las secuencias de primitivas aplicadas a las estructuras de entrada hasta llegar a la estructura de la tabla del data mart.

### **BD\_CARRERAS\_CICLOS**

#### 4.1 DD\_ADDING 1-1

Entrada:

- Carreras\_ciclos( carr, nomcar, plan, ciclo, nomcic, creditos)
- Tipo\_Plan ( plan )  
Tipo\_Plan: dado una fecha devuelve el valor 'Anterior' si el año es menor a 1997 y el valor 'Vigente' si el año es mayor a 1997.

Resultado:

- Carreras\_ciclos'( carr, nomcar, plan, ciclo, nomcic, tipoplan, creditos )

### **BD\_ESTCARR**

#### 4.1 DD\_ADDING 1-1

Entrada:

- Insc\_carr(estci, carr, ciclo, fecing, feceg, lugar, obs, calinsc)
- Generacion(fecing)  
Generacion: dada una fecha devuelve la generación correspondiente (gen96, gen97, ....)

Resultado:

- Insc\_carr'(estci, carr, ciclo, fecing, feceg, lugar, obs, calinsc, generacion)

#### 14 NEW DIMENSION CROSSING

Entrada:

- Estudiantes(estci, nomest, lugar, nacido, sexo, año, inst, dir, tel, est),  
Insc\_carr'(estci, carr, ciclo, fecing, feceg, lugar, obs, calinsc, generacion),  
Carreras\_ciclos'(carr, nomcar, plan, ciclo, nomcic, tipo\_plan, creditos)
- Y1={nomest, lugar, nacido, sexo, año, inst, dir, tel}, Y2={obs, lugar, feceg},  
Y3={plan, tipoplan, nomcar, nomcic}

Resultado:

- Est\_carr(estci, carr, ciclo, fecing, generacion, calinsc, creditos)

#### 4.2 DD-ADDING N.1

Entrada:

- Est\_carr(estci, carr, ciclo, fecing, generacion, calinsc, creditos),  
Insc\_carr(estci, carr, ciclo, fecing, feceg, lugar, obs, calinsc)
- Fecha\_fin: Devuelve la fecha de finalización de la inscripción dadas la calidad de inscripción en Est\_carr y Insc\_carr.
- {estci, carr, ciclo}
- false

Resultado:

- Est\_carr'(estci, carr, ciclo, fecing, generacion, calinsc, creditos, fecha\_fin)



## **BD\_ESTUDIANTES**

### 1 DATA FILTER

Entrada:

- Estudiantes(estci, nomest, lugar, nacido, sexo, año, inst, dir, tel, est)
- { lugar, año, inst, dir, tel }

Resultado:

- Estudiantes'(estci, nomest, nacido, sexo, est)

## **BD\_ACTIVIDADES**

### 4.1 DATA\_FILTER

Entrada:

- Actividades( estci, mat, tact, nota, fecha, curri, tgen, per, tipoper, obs)
- { tgen, obs }

Salida:

- Actividades'( estci, mat, tact, nota, fecha, curri, per, tipoper )

### 4.1 DD\_ADDING 1-1

Entrada:

- Actividades'( estci, mat, tact, nota, fecha, curri, per, tipoper )
- TipoActividad( tact )  
TipoActividad: dado el código de tipo de actividad, devuelve el nombre del tipo de actividad correspondiente (Asistencia, Examen, Inscripción, Curso, Curso caducado, Curso inválido por curso posterior, Parcial1,..., Parcial9).

Resultado:

- Actividades''( estci, mat, tact, nota, fecha, curri, per, tipoper, tipoactividad )

### 4.1 DD\_ADDING 1-1

Entrada:

- Actividades''( estci, mat, tact, nota, fecha, curri, per, tipoper, tipoactividad )
- TipoAprobacion( tact )  
TipoAprobacion: dado el tipo de actividad, devuelve el tipo de aprobacion ( con nota o sin nota )

Resultado:

- Actividades'''( estci, mat, tact, nota, fecha, curri, per, tipoper, tipoactividad, tipoaprobacion )

### 4.1 DD\_ADDING 1-1

Entrada:

- Actividades'''( estci, mat, tact, nota, fecha, curri, per, tipoper, tipoactividad, tipoaprobacion )
- TipoResultado( tipoaprobacion, nota )  
TipoResultado: devuelve a partir del tipo de aprobación y la nota obtenida, el tipo de resultado ( aprobado, reprobado )

Resultado:

- Actividades''''( estci, mat, tact, nota, fecha, curri, per, tipoper, tipoactividad, tipoaprobacion, tiporesultado )

### 14 NEW DIMENSION CROSSING

Entrada:

- Actividades''''( estci, mat, tact, nota, fecha, curri, per, tipoper, tipoactividad, tipoaprobacion, tiporesultado ), Mat\_carr(carr, ciclo, mat, nommat,tipoexo,créditos), Est\_carr'(estci, carr, ciclo, fecing, generacion, calinsc, creditos, fecha\_fin)
- Y1={},Y2={nommat, tipoexo, créditos}, Y3={fecing, generacion, nomcar, tipoplan, plan, calinsc, fecha\_fin}

Resultado:

- Actividades''''(estci, carr, ciclo, mat, tact, fecha, per, tipoper, tipoactividad, tipoaprobacion, tiporesultado, nota)

#### 4.2 DD-ADDING N-1

Entrada:

- Actividades''''(estci, carr, ciclo, mat, tact, fecha, per, tipoper, tipoactividad, tipoaprobacion, tiporesultado, nota), Mat\_carr(carr, ciclo, mat, nommat, tipoexo, créditos)
- Calc\_creditos: calcula créditos obtenidos por el estudiante al realizar la actividad a partir de los créditos de la materia y el tipo de aprobación de la actividad
- {carr, ciclo, mat}
- false

Resultado:

- Actividades''''(estci, carr, ciclo, mat, tact, fecha, per, tipoper, tipoactividad, tipoaprobacion, tiporesultado, nota, créditos)

#### 2 TEMPORALIZATION

Entrada:

- Actividades''''(estci, carr, ciclo, mat, tact, fecha, per, tipoper, tipoactividad, tipoaprobacion, tiporesultado, nota, créditos)
- semestre : semestreyyyy

Resultado:

- Actividades''''(estci, carr, ciclo, mat, tact, fecha, semestre, per, tipoper, tipoactividad, tipoaprobacion, tiporesultado, nota, créditos)

### ***BD DESEMPEÑO***

#### 13 MINIDIMENSION BREAK-OFF

Entrada:

- Est\_carr'(estci, carr, ciclo, fecing, generacion, calinsc, creditos, fecha\_fin)
- {estci, carr, ciclo}
- {estci, carr, ciclo}

Resultado:

- Desempeño(estci, carr, ciclo)

#### 2 TEMPORALIZATION

Entrada:

- Desempeño(estci, carr, ciclo)
- semestre: semestreyyyy

Resultado:

- Desempeño'(estci, carr, ciclo, semestre)

#### 8 AGGREGATE GENERATION

Entrada:

- Actividades''''''(estci, carr, ciclo, mat, tact, fecha, semestre, per, tipoper, tipoactividad, tipoaprobacion, tiporesultado, nota, créditos)
- {estci, carr, ciclo, semestre}
- Cantidad de actividades aprobadas
- {Count(\*) where tiporesultado = 'Aprobado'}
- {mat, tact, fecha, per, tipoper, tipoactividad, tipoaprobacion, tiporesultado, nota, créditos}

Resultado:

- Act\_aprob(estci, carr, ciclo, semestre, cantaprob)

#### 4.3 DD-ADDING N-N

Entrada:

- Desempeño(estci, carr, ciclo, semestre), Act\_aprob(estci, carr, ciclo, semestre, cant\_aprob)
- Calc\_desemp(semestre, cant\_aprob) /  
Calc\_desemp: calcula el desempeño del estudiante a partir de la cantidad de aprobaciones
- {estci, carr, ciclo}
- {estci, carr, ciclo}

Resultado:

- Desempeño'(estci, carr, ciclo, semestre, desempeño)

### ***BD\_EGRESOS\_TITULOS***

#### 9 TABLE\_MERGE

Entrada:

- Def\_titulos( carr, ciclo, titulo, nomtit, tipotit )
- Egresos( carr, ciclo, titulo, estci, feceg, revalida )

Salida:

- Egresos\_titulo( carr, ciclo, titulo, nomtit, tipotit, estci, feceg, revalida )

### C. ESQUEMA RELACIONAL DATA MART BEDELÍA

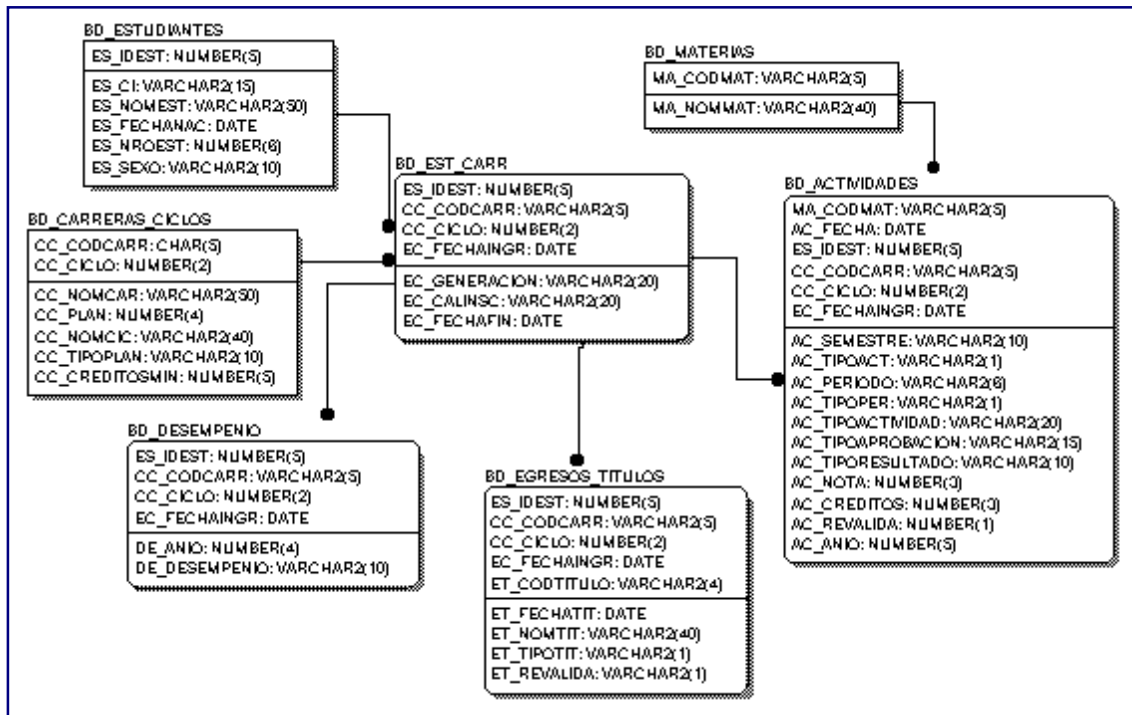


Figura 30: Esquema de Tablas Data mart Bedelía

## ➤ DISEÑO DE LA CARGA

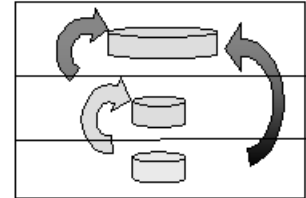
---

*Se realiza el diseño de las distintas rutinas de extracción y carga de datos al data mart.*

---

### A. PROCESO DE CARGA

El proceso de carga se divide en tres subprocesos a ejecutar en distintos momentos del año. Esta división se debe a pedidos específicos de tener la información reciente en el data mart para poder ser accedida.



Un proceso será el encargado de ingresar toda la información de ingresos a carreras (Ingresos), el segundo será el encargado de ingresar la información de actividades de los estudiantes (Actividades) y el tercero de calcular el desempeño de los estudiantes en las carreras que están cursando (Desempeño).

Se describen los procesos principales y se detallan a las cargas que los componen.

En el *Apartado 2, Metadata del Sistema* se encuentra el pseudocódigo de las distintas cargas.

#### **Subproceso Ingresos.**

CargaCarrerasCiclos;  
CargaEstCarreras;

#### **Subproceso Actividades.**

CargaMaterias;  
CargaCarrerasCiclos;  
CargaActividades;  
CargaEgresosTitulos;

#### **Subproceso Desempeño.**

CargaDesempeño;

---

### **CargaCarrerasCiclos**

**Funcionalidad:** Ingresa nuevos pares carreras-ciclo a la tabla *BD\_CARRERASCICLOS*.

**Especificación:** Se leen los datos de *carreras\_ciclos* de la bd fuente, donde las nuevas son ingresadas en *BD\_CARRERASCICLOS*.

**Limpieza:** Si el dato original no especifica un ciclo, se ingresa la carrera con ciclo 0.

### **CargaEstCarrera**

**Funcionalidad:** Ingresa nuevos ingresos de estudiantes a carreras en la tabla *BD\_ESTCARR*. Actualiza las inscripciones a carreras cuando las mismas son finalizadas.

**Especificación:** Se leen los datos de *estudiantes*, *carreras\_ciclos* e *ins\_carr*, considerando las inscripciones que se produjeron entre la carga anterior y la actual. Luego de validados y transformados los datos se ingresan en *BD\_ESTCARR*. Por último se actualizan todas las inscripciones en curso por si se invalidaron o finalizaron.

**Limpieza:** Se valida que el estudiante y la carrera ciclo existan. Se transforman las fechas a valores compatibles con el año 2000. A partir de la fecha de ingreso se calcula la generación del estudiante.

---

### **CargaMaterias**

**Funcionalidad:** Ingresa nuevas materias a la tabla *BD\_MATERIAS*.

**Especificación:** Se leen los datos de *materias* de la bd fuente, donde las nuevas son ingresadas en *BD\_MATERIAS*.

### **CargaActividades**

**Funcionalidad:** Ingresa las nuevas actividades realizadas por los estudiantes desde la fecha de carga anterior hasta la actual en la tabla *BD\_ACTIVIDADES*. Se ingresan las actividades para todas las carreras-ciclo donde se dicte la materia y esté en curso el estudiante. Se ingresan además todas las actividades aprobadas revalidadas al ingresar a una nueva carrera-ciclo.

**Especificación:** Se leen los datos de *actividades*, *matcarr* e *ingresos*, considerando las actividades que se produjeron entre la carga anterior y la actual. Luego de validados y transformados los datos se ingresan en *BD\_ACTIVIDADES*. Por último se recorren todas las actividades aprobadas de los estudiantes, y se ingresan asociadas a la carrera-ciclo donde el estudiante acaba de ingresar, si la materia se encuentra en la nueva carrera-ciclo. La fecha de actividad queda igual a la de ingreso a la carrera.

**Limpieza:** Se valida que el estudiante y la carrera ciclo existan. Se valida que el estudiante esté inscripto y en curso en la carrera-ciclo. Se transforman las fechas a valores compatibles con el año 2000.

Se calcula el tipo de actividad, el tipo de aprobación, el tipo de resultado. La nota se ingresa en 0 si la actividad no lleva nota o si fue reprobada. Los créditos se validan que estén en el rango adecuado y se ingresan en 0 si no se aprobó la materia o el valor original si se aprobó la misma.

Se calcula el semestre correspondiente a la fecha de la actividad.

### **CargaEgresosTítulos**

**Funcionalidad:** Ingresar las nuevas obtenciones de títulos por estudiantes a la tabla *BD\_EGRESOSTITULOS*

**Especificación:** Se leen los datos de *egresos* y *títulos*, considerando los egresos que se produjeron entre la carga anterior y la actual

**Limpieza:** Se valida que el egreso corresponda a una carrera-ciclo existente.

---

### **CargaDesempeño**

**Funcionalidad:** Ingresar el cálculo de desempeño en *BD\_DESEMPEÑO* de todos los estudiantes durante el último año escolar para todas las carreras en las que están en curso.

**Especificación:** Se cuentan las actividades por año, de los últimos 5 años de todos los estudiantes en cada carrera en la que se encuentre cursando. A partir de la cantidad de actividades por año se calcula el desempeño y se ingresa en la tabla *BD\_DESEMPEÑO* asociado al último año escolar pasado.

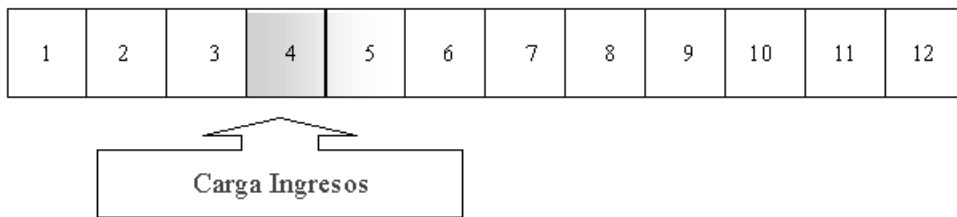
## **B. PERIODICIDAD DE LA CARGA**

La carga de ingresos debe realizarse anualmente al finalizar el período de inscripción. En general esto ocurre durante los meses de febrero y marzo, por lo que sería adecuado cargar los datos en abril.

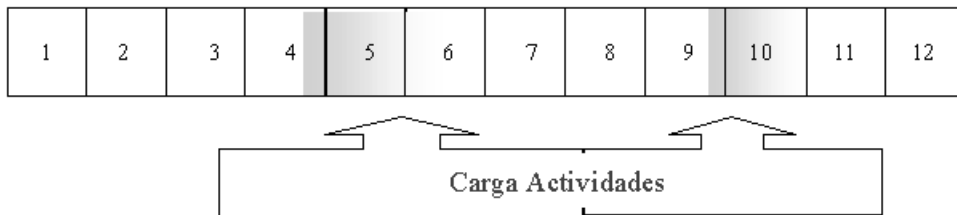
La carga de actividades tiene sentido realizarla después de finalizados los períodos de exámenes para así cargar los resultados de exámenes y cursos realizados en el semestre. Los períodos en general de exámenes son durante los meses de julio y agosto (primer período), noviembre y diciembre (segundo período) y febrero y marzo (tercer período).

Como los resultados del segundo período pueden no estar hasta fechas durante el tercer período (enero no laborable), la carga de los últimos dos períodos conviene realizarla en conjunto. Por lo tanto se cargan actividades luego de finalizado el primer período y luego de finalizado el tercer período, siempre y cuando se de un tiempo de unos 20 a 30 días para asegurar que estén todos los resultados.

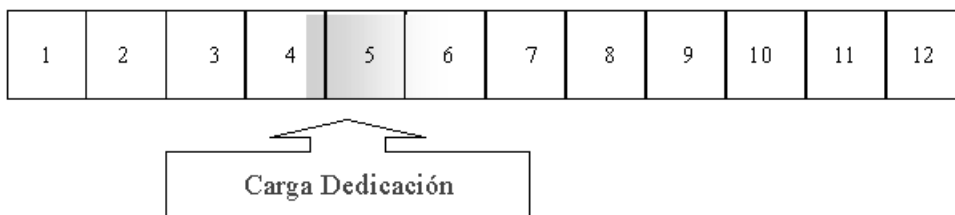
La carga de dedicación tiene sentido realizarla al finalizar el año escolar, luego de haber ingresado las actividades del último período del año.



*Figura 31: Carga de Ingresos durante el año*



*Figura 32: Carga de Actividades durante el año*



*Figura 33: Carga de Dedicación durante el año*



## 2.5 IMPLEMENTACIÓN

---

Como resultado del desarrollo del data mart de Bedelía se obtuvieron los siguientes componentes principales: bd del data mart, procesos de carga y metadata. Se detalla a continuación información de la bd del data mart y de los procesos de carga. La metadata se documenta en el *Apartado 2, Metadata del Sistema*.

### Base de datos del data mart.

Nombre: 'bede98'  
DBMS: Oracle 8  
Ubicación: servidor 'Felipe' del INCO  
Contenido: información pertinente al seguimiento de estudiantes  
Usuario: dw\_bede/dw\_bede

### Procesos de carga.

Procesos de interface:

Nombre: 'Cargas\_Datamarts.exe'  
Ubicación: 'C:\DWTaller5\Vb' del servidor 'Tornasol' del INCO  
Leng.Prog.: Visual Basic  
Funcionalidad: Interface central para ejecución de procesos de carga.  
Permite disparar los tres procesos de carga de bedelía: carga de inscripciones, carga de actividades y carga de desempeño.

Proceso Carga de Inscripciones:

Nombre: 'carga\_inscripciones'  
Ubicación: base de datos 'bede98'  
Leng.Prog.: PLSQL  
Funcionalidad: carga de inscripciones de estudiantes a carreras

Proceso Carga de Actividades:

Nombre: 'carga\_actividades'  
Ubicación: base de datos 'bede98'  
Leng.Prog.: PLSQL  
Funcionalidad: carga de resultados de actividades

Proceso Carga de Desempeño:

Nombre: 'carga\_desempeño'  
Ubicación: base de datos 'bede98'  
Leng.Prog.: PLSQL  
Funcionalidad: carga del desempeño de estudiantes

## 3 APÉNDICE

### TÉCNICA PARA IDENTIFICACIÓN DE DATOS REQUERIDOS

Esta técnica permite visualizar qué elementos del modelo conceptual multidimensional (información requerida) se encuentran en los sistemas fuente. Se mapean las Dimensiones y Medidas del modelo multidimensional con los objetos del modelo conceptual de cada sistema fuente.

Se realiza un cuadro donde se mapean las Dimensiones requeridas contra los objetos del modelo conceptual. Se marca una 'x' cuando el objeto del modelo conceptual, representa información de utilidad para cargar la Dimensión. Se realiza éste mapeo para cada sistema fuente (Figura 34).

Además, se realiza un cuadro donde se mapean las Medidas requeridas contra los objetos del modelo conceptual. Se marca una 'x' cuando el objeto del modelo conceptual, representa información de utilidad para calcular las medidas. Esto se realizará por cada sistema fuente (Figura 34).

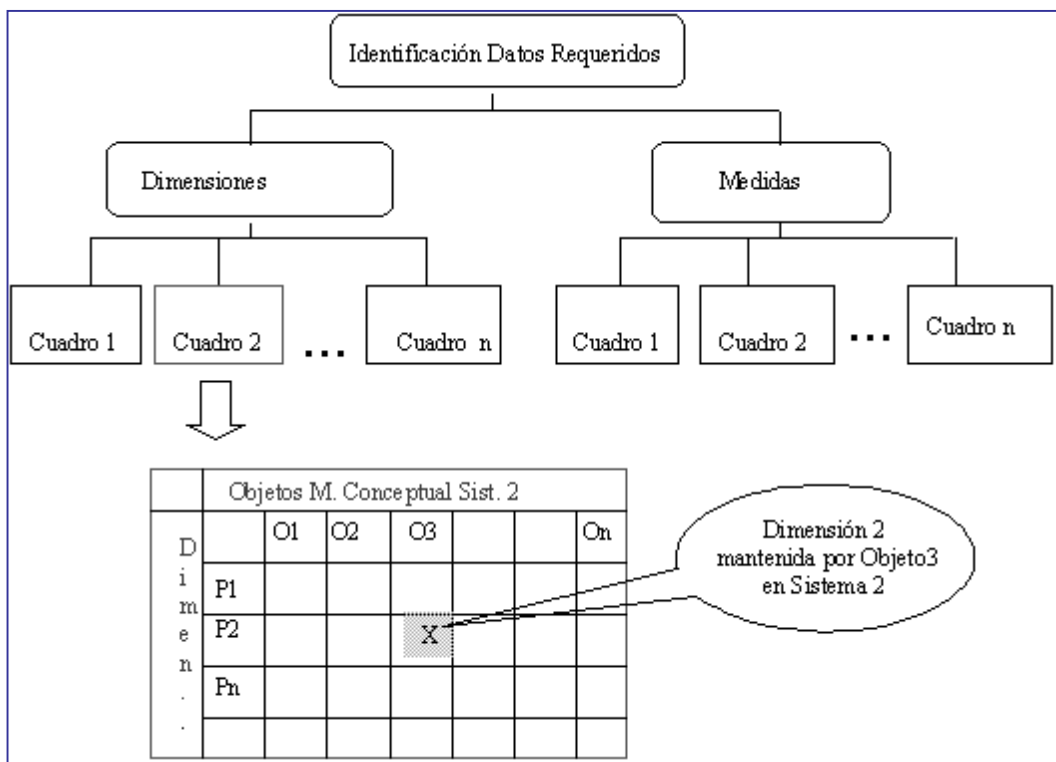


Figura 34 : Identificación de dimensiones y medidas

En el cuadro de dimensiones, cuando la dimensión no tiene categorías, sólo se marcan los objetos que directamente aportan información a ella. En cambio, cuando la dimensión tiene distintas categorías, se deben marcar los objetos que aportan información a cada categoría y además, los objetos que aportan información de la relación de cada categoría con la categoría superior.

Para identificar los objetos conceptuales que mantienen información acerca de los cruzamientos entre dimensiones, se realiza otro cuadro. El cuadro mapea dimensiones con dimensiones, identificando los cruzamientos válidos según el modelo conceptual multidimensional. Para aquellos cruzamientos válidos, se identifica el conjunto de objetos conceptuales que aportan información al cruzamiento. Todo cruzamiento, identificado como válido que no tenga asociado un conjunto de objetos conceptuales, representa información ausente en el sistema fuente analizado. Esto se debe realizar para cada sistema fuente (*Figura 35*).

		DIMENSIONES						
		Dim1	Dim2	Dim3	Dim4			
D I M E N S I O N E S	Dim1			{01,02}	{02,04}			
	Dim2			{05,02}				
	Dim3				{01,02}			
	Dim4							

*Figura 35: Identificación de cruzamientos*