

Construcción y explotación de un Almacén de Datos sobre información estadística de la NBA

Lourdes Cebrián Rodríguez
ETIS

Nom Consultor:
Jose Angel Martín Carballo

RESUMEN

Este proyecto se basa en el análisis, desarrollo y explotación de un Almacén de Datos para el diario *Sport New*; él cual ya dispone de una base de datos pero quiere mejorar el rendimiento y la explotación de la información.

El proyecto podríamos decir que tiene tres partes bien diferenciadas. Por un parte tenemos el análisis de la base de datos existente en MS ACCESS, la cual se ha de analizar y encontrar la solución más óptima para el nuevo modelo de datos.

Una segunda parte que sería todo el diseño técnico e implementación del modelo escogido, es decir, toda la construcción del almacén de Datos en Oracle.

Por último, nos encontramos con la explotación del almacén de Datos a partir del software de Oracle Discoverer, a partir del cual creemos desde el programa administrador creamos todo el modelo necesario para que desde el programa D. Desktop podemos realizar una serie de informes planificados para que el usuario lo pueda explotar.

INDICE

1. Introducción.....	4
1.1 Punto de partida del TFC.....	4
1.2 Objetivos del TFC.....	4
1.3 Fases del proyecto.....	5
1.4 Productos obtenidos.....	6
1.5 Planificación.....	7
1.6 Descripción de los capítulos.....	8
2. Análisis funcional.....	9
2.1 Situación de partida.....	9
2.1.1 Usuarios.....	9
2.1.2 Diagrama de modelo.....	9
2.1.3 Problemas y deficiencias base de datos Origen.....	10
2.1.4 Requisitos de usuario.....	11
2.2 Fundamentos del Almacén de Datos.....	12
2.3 Almacén de Datos.....	13
2.4 Modelo de Pceso.....	13
2.5 Modelo de Datos.....	14
2.6 Temporalidad.....	15
2.7 Informes.....	15
2.8 Usuarios.....	16
3. Arquitectura.....	17
3.1 Entorno de trabajo.....	17
3.2 Oracle.....	18
3.3 Software utilizado.....	18
3.3.1 MS ACCESS.....	18
3.3.2 Oracle Express 10g.....	19
3.3.3. Oracle SQL Developer.....	19
3.3.4 Oracle Discoverer.....	20
3.3.4.1 Oracle Discoverer Administrator.....	21
3.3.4.2 Oracle Discoverer Desktop.....	21
3.4 Requisitos del hard.....	21
4. Análisis Técnico.....	22
4.1 Origen de los datos.....	22
4.2 Modelo Estrella.....	23
4.2.1 Asociación entre Tablas finales y tabla de carga de origen.....	24
4.2.2 Campos nuevos.....	24
4.2.3 Tablas del modelo.....	25
4.3 Procesos de Carga en Oracle.....	28
4.3.1 Esquema resumen de usuarios creados en Oracle.....	28
4.3.2 Ejecución de Procesos.....	31
4.3.3 Detalle de los procedimientos.....	32
4.4 Creación entorno de trabajo de usuario.....	36
4.4.1 Discoverer Administrator- EUL.....	36
4.4.2 Discoverer Desktop-Informes.....	38
5. Escenario final de usuario.....	43
6. Posibles Mejoras o Ampliaciones.....	47
7. Conclusiones.....	48
8. Anexo	49
9. Bibliografía.....	50

1. Introducción

1.1 Punto de partida del TFC

Este proyecto se basa en el análisis y desarrollo de un Almacén de Datos para un diario; el cual ya dispone de una base de datos pero quiere mejorar el rendimiento y la explotación de la información.

Actualmente, existen muchas definiciones para el concepto de DW, yo he recogida una de ellas que dice “Un DW es una colección de datos orientados a temas, integrados, no-volátiles y variante en el tiempo, organizados para soportar necesidades empresariales”

Independientemente de la definición que se use; hay ciertas características comunes en todas ellas. En un DW siempre se busca la consolidación e integración de los datos; así como el poder disponer de una información fiable pero siempre de una manera fácil y directa para el usuario.

En nuestro caso el diario deportivo “Sport World”, ya estaba generando unas estadísticas semanales sobre la NBA.

Su base de datos origen esta en entorno MS Access y el cliente nos comenta que presenta alguna irregularidad.

Básicamente la Base de datos sirve para consultar información con la finalidad de poder realizar crónicas deportivas, consultas puntuales; así como elaborar estadísticas semanales.

A partir de la situación origen descrita, y de las necesidades de información que nos plantea la empresa, se diseñará un DW con un modelo de datos multidimensional, el cual podrá ser explotado por los usuarios con total confianza y solidez.

1.2 Objetivos del TFC

Por supuesto, el objetivo principal del proyecto es aprender ha analizar y desarrollar un DW adecuado a las necesidades de la empresa y en nuestro caso concreto, partimos de una base de datos previos que tendremos que transformar.

Este gran objetivo, descrito en el párrafo anterior, lo vamos a descomponer en muchos conceptos que paso a enumerar:

- Saber planificar todo un proyecto.
- Profundizar en los aspectos teóricos de análisis de modelos de datos.
- Transformar una base de datos relacional a multidimensional
- Desarrollar una solución óptima para los usuarios.
- Conocer de manera práctica el lenguaje SQL.
- Desarrollar de manera práctica un DW.

- Saber analizar la información obtenida y verificar que es correcta.
- Conocer diferentes herramientas de mercado para explotación de datos.
- Realizar una presentación clara y precisa donde se sintetice el proyecto.

1.3 Fases del Proyecto

Básicamente para descomponer el proyecto en diferentes fases, me he basado en las fechas de los entregables de la PACS, y por supuesto, en la fecha final de entrega.

Sobre estas fechas he ido construyendo las diferentes fases que creo son necesarias para llevar a cabo el proyecto; y por supuesto cada fase depende de su fase anterior para una correcta realización.

- **FASE 1**

Esta fase es de introducción al proyecto, y básicamente se refiere al leer el enunciado del TFC, en mi caso leer información sobre NBA y situarse en la asignatura y proyecto.

- **FASE 2**

Esta fase ha servido de base para el correcto desarrollo del proyecto.

Por una parte se ha elaborado un plan de trabajo, realizado a partir de haber analizado en global el proyecto.

Éste nos servirá de pauta a la hora de seguir avanzando de una manera correcta las fases y además nos marcará la planificación de proyecto; para poder entregarlo en plazo.

Por otra parte, también se ha realizado un análisis previo de las tablas actuales y el modelo de solución que damos para cubrir las necesidades de información y presentamos los diferentes informes a realizar.

- **FASE 3**

Aquí acabamos de completar toda la información que ya hemos analizado previamente en la fase anterior con respecto a la situación actual de tablas y solución aportada.

Una vez hecho el análisis definitivo, primero se elaboró diseño funcional, teniendo en cuenta las características que tiene que tener un DW y los requisitos de usuarios.

Una vez se haya cumplido con estos objetivos y tengamos el diseño funcional, hemos seguido por el diseño técnico, teniendo en cuenta todos los aspectos necesarios para poder llevar a cabo la implementación en la siguiente fase.

- **FASE 4**

Una vez se han analizado los aspectos funcionales y técnicos, solo nos queda implementar la solución.

Así que en esta fase se ha exportado Base de datos ACCESS al nuevo entorno y se construye el DW.

Una vez se ha realizado todas las pruebas y ya lo tenemos implementado, repasamos las tablas de errores para verificar la solidez la información.

Seguido a la construcción y comprobación del nuevo Dw, tendremos que implementar los diferentes informes.

Una vez todo probado, procedemos a la realización de los informes solicitados.

- **FASE 5**

En esta fase se ha de tratar de sintetizar y explicar todo el trabajo realizado en el proyecto, mediante la realización de esta memoria final y una presentación.

- **FASE 6**

En esta última fase, y una vez entregadas la memoria y la presentación, tenemos que estar disponibles para diferentes preguntas que le puedan surgir al tribunal de evaluación.

Por último se abrirá un debate sobre el TFC realizado.

1.4 Productos obtenidos

En el desarrollo del producto se han creado una serie de documentos que paso a describir a continuación

Fechas	Documento	Descripción
13/03/2007	Entregar PAC1	Plan de trabajo Análisis Previo
16/04/2007	Entregar PAC2	Análisis requerimientos Diseño funcional Diseño técnico
28/05/2007	Entregar PAC3	Implementación del DW Modelo EUL Explicación de procesos Pantallazos de informes realizados
11/06/2007	Entregar memoria y Presentación	Presentación pps Memoria Final

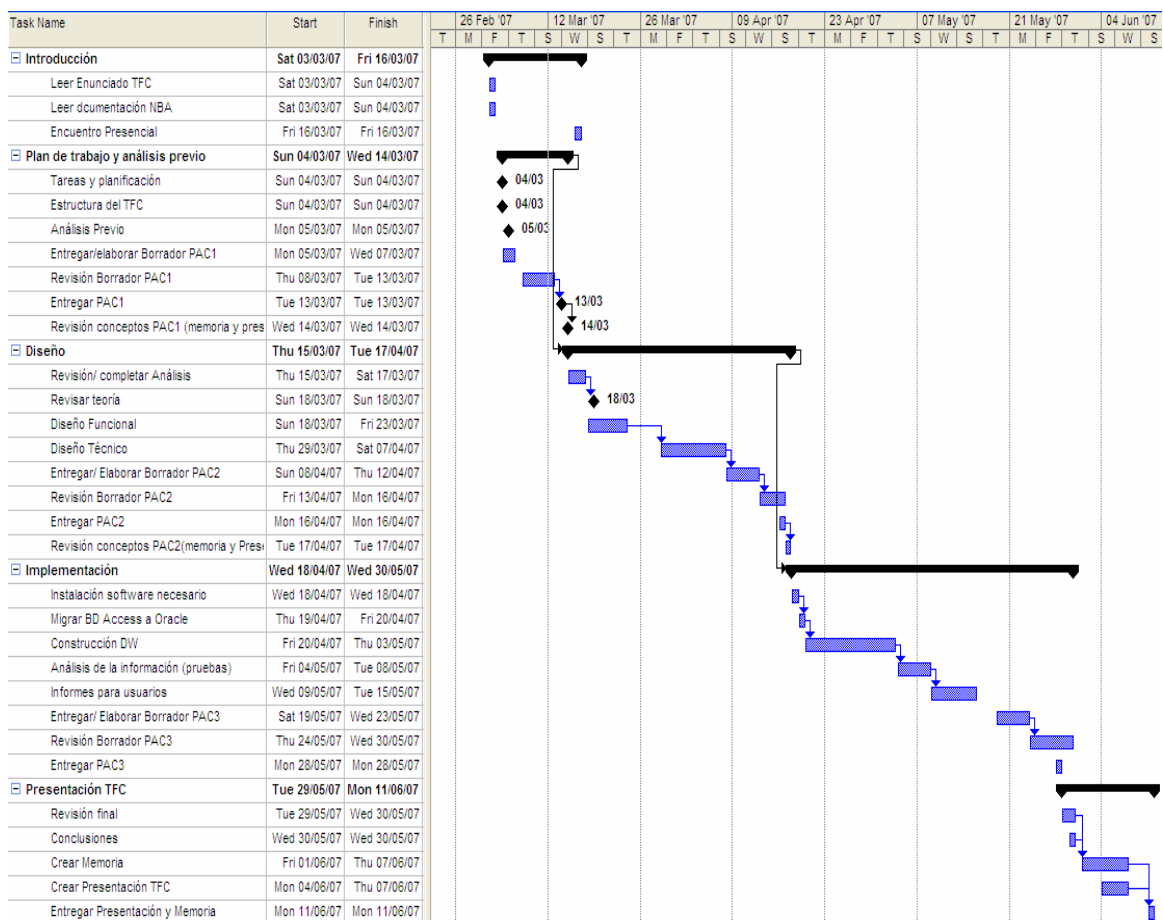
1.5 Planificación

A partir de las fases descritas en el apartado anterior y de las fechas de las Pac's a entregar se elaboró una planificación para entregar el proyecto en las fechas solicitadas.

Por una parte, he dividido las fases anteriores en tareas más específicas, las cuales son mucho más fáciles de planificar y asignarles el tiempo que creo voy a necesitar para realizarlas.

También hay que comentar, que debido a que no se ha podido dedicar el mismo tiempo al proyecto todos los días, se ha asumido alguna variación, pero se ha marcado como pautas las fechas de los entregables.

Diagrama de Gaant



Dibujo1-Diagrama de Gaant

1.6 Descripción de los capítulos

Este documento se ha dividido en diferentes capítulos, los cuales paso a describir a continuación.

Capítulo 2, el cual contiene todo lo relativo al análisis funcional del proyecto, tanto el análisis de la situación origen, problemas y deficiencias que presentaba la base de datos y la solución funcional.

Capítulo 3, en el cual describimos toda la arquitectura del proyecto, tanto los programas utilizados como los requerimientos técnicos necesarios.

Capítulo 4, en él esta todo el diseño técnico del análisis funcional, en él podremos ver la estructura de la implementación del proyecto.

Capítulo 5, Aquí hago un resumen de las posibles ampliaciones o mejoras de cara al futuro.

Capítulo 6, Las conclusiones de la realización del TFC.

2. Análisis funcional

2.1.1 Situación de Partida

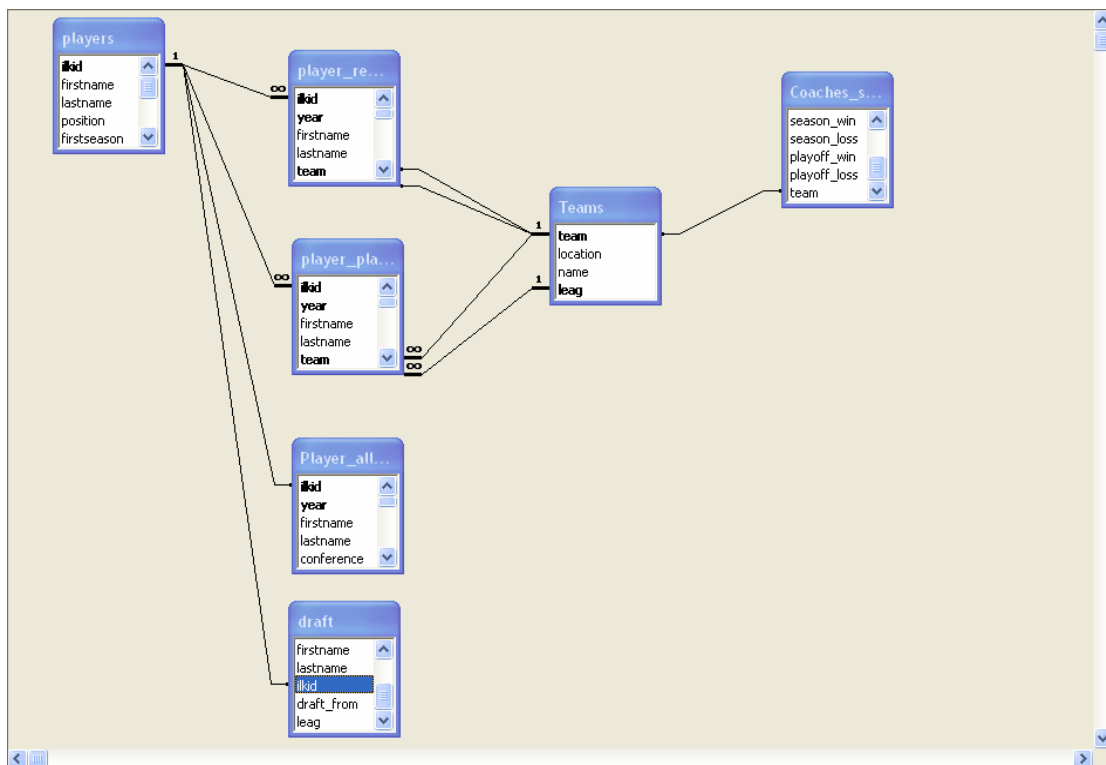
En este apartado vamos a mostrar el sistema de información origen del diario deportivo.

2.1.2 Usuarios

El diario no nos especificó el tema de control de acceso de usuarios, con lo el sistema origen base de datos en MS ACCESS era consultable por cualquier persona del diario que necesitará la información.

2.1.3 Diagrama del modelo de Datos

A partir del MS Access mostramos el diagrama de modelo de datos origen, que corresponde a un modelo Entidad-Relación.



Dibujo 2 –Modelo de Base de Datos Origen

Resumen de tablas del Diagrama:

- TEAM :Contiene la información de localidad, nombre y liga de cada equipo.
- TEAM_SEASON : Contiene la información que ha hecho cada equipo por temporada en la totalidad de los partidos jugados.
- PLAYERS : Contiene información personal sobre cada jugador.
- PLAYER_REGULAR_SEASON: Contiene la información sobre el rendimiento del jugador en cada equipo y temporada.
- PLAYER_PLAYOFFS: Contiene la información sobre el rendimiento del jugador en cada equipo y temporada si ha jugado en los playoffs.
- PLAYER_ALLSTAR: Contiene la información sobre el rendimiento del jugador por año si ha jugado en el Allstar
- DRAFT : Contiene la información sobre las rondas de selección.
- COACHES_SESSION: Contiene la información sobre las sesiones de entrenamiento de cada jugador

2.1.3.Problemas y deficiencias base de datos Origen

A partir del análisis de la base de datos origen, nos encontramos con los siguientes problemas:

- Tablas sin índices.
- No informado todos los campos
- Información redundante.
- Tablas no relacionadas
- Información desagregada. En el DW existen datos en distintos niveles de detalle y agregación
- Estructura relacional frente a la Multidimensional de un DW
- No orientada al análisis de la información (consulta)
- Orientada al análisis de la situación actual frente a la comparativa histórica.

2.1.4 Requisitos de usuario

A partir del documento entregado por la empresa, describimos en los siguientes puntos los requisitos necesarios que debe satisfacer nuestro proyecto.

- Proporcionar Estadísticas sobre la temporada regular, por diferentes indicadores como pueden ser jugador, equipo, entrenador.
- Estadísticas sobre Play-off.
- Estadísticas All-Star Games.
- Estadísticas Rondas de Selección.
- Posibilidad de crear Dream Team a partir de diferentes Indicadores.
- Poder tener una base de Datos fiable.
- Completar la base de Datos.
- Posibilidad de que los diferentes departamentos puedan explotar la información del DW .
- Guardar datos históricos.
- Información desagregada. En el DW existen datos en distintos niveles de detalle y agregación
- Estructura relacional frente a la Multidimensional de un DW
- No orientada al análisis de la información (consulta)
- Orientada al análisis de la situación actual frente a la comparativa histórica.
- Informes PlayOff , estos tendrán toda la información referente a los
- Informe AllStar, estos tendrán toda la información referente a los AllStar .
- Informe Rondas de selección, estos tendrán toda la información referente a las ronda de selección.
- Informe de temporada, estos tendrán toda la información referente a la temporada.
- Propuesta de Dream Team, se creará un DreamTeam cada año y el mejor Dream Team de la historia

2.2 Fundamentos del Almacén de Datos

Tras las dificultades de los sistemas tradicionales en satisfacer las necesidades informacionales, surge el concepto de Almacén de Datos, como solución a las necesidades informacionales globales de la empresa.

La ventaja principal de este tipo de sistemas se basa en su concepto fundamental: la estructura de la información. Este concepto significa el almacenamiento de información homogénea y fiable, en una estructura basada en la consulta, el tratamiento jerarquizado de la misma, y en un entorno diferenciado de los sistemas operacionales. Se caracteriza por ser:

Integrado

Los datos almacenados deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

Temático

Sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. De esta forma, las peticiones de información sobre un tema en concreto serán más fáciles de responder dado que toda la información reside en el mismo lugar.

Histórico

El tiempo es parte implícita de la información contenida en un Almacén de Datos. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el Data Warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el Data Warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

No volátil

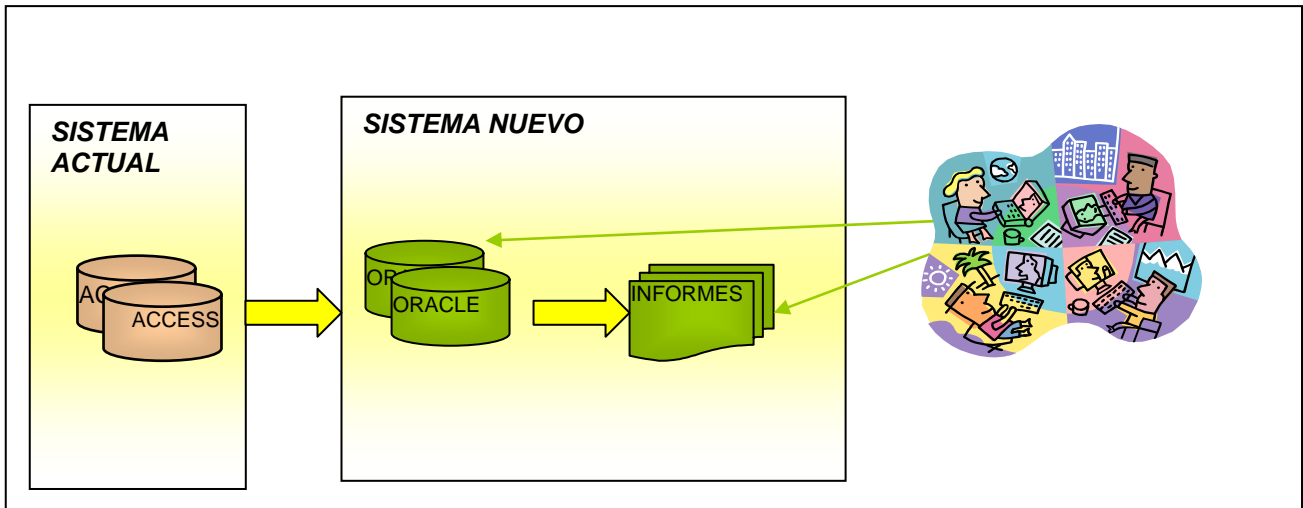
El almacén de información existe para ser leído, y no modificado. La información es por tanto permanente, significando la actualización del Data Warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

2.3 Almacén de Datos

A partir de las definiciones teóricas expuestas en el apartado anterior y de los informes finales se creará el Almacén de datos más adecuado a nuestras necesidades.

Nuestra nueva base de datos partirá de las tablas orígenes del sistema actual en MS Access y se cargarán en nuestro nuevo entorno Oracle.

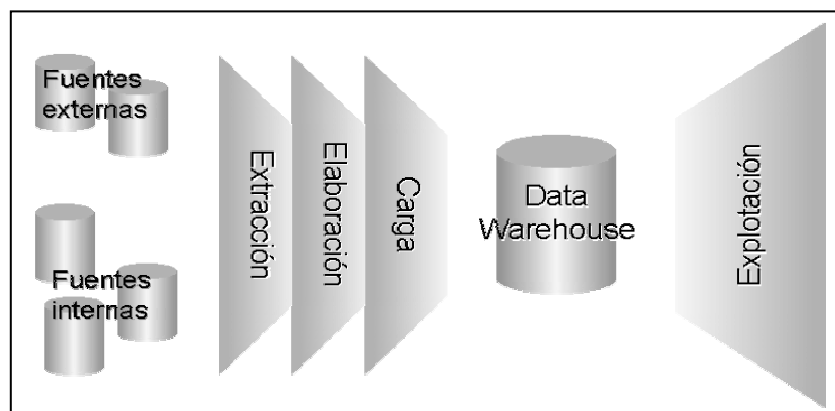
En él se llevarán a cabo una serie de procesos de transformaciones que verificarán que las tablas finales cumplan cada una de las características del apartado anterior.



Dibujo3 –Esquema general

2.4 Modelo de Procesos

Para comprender el concepto de Almacén de Datos, es importante considerar los procesos que lo conforman. A continuación, se describen dichos procesos clave en la gestión.



Dibujo4 –Modelo de Proceso

Extracción

Obtención de información de las distintas fuentes tanto internas como externas.

Elaboración

Filtrado, limpieza, depuración, homogeneización y agrupación de la información.

Carga

Organización y actualización de los datos y de la base de datos.

Explotación:

Extracción y análisis de la información en los distintos niveles de agrupación.

Desde el punto de vista del usuario, el único proceso visible es la explotación del almacén de datos, aunque el éxito del Almacén de Datos radica en los tres procesos iniciales que alimentan la información del mismo y suponen el mayor porcentaje de esfuerzo (en torno a un 80%) a la hora de desarrollar el almacén.

2.5 Modelo de Datos

El objetivo de esta fase es definir un modelo de datos consistente que satisfaga todas las necesidades de información del diario.

Para abordar esta cuestión de manera global, vamos a ser fragmentar el Almacén de Datos en partes conceptualmente. En nuestro caso esta división se ha efectuado para los jugadores y para los equipos.

A continuación, para cada parte deberemos identificar todas las medias y categorías que deseemos almacenar en nuestro Sistema. Estas se corresponden a las variables de clase y a las variables de análisis solicitadas por el usuario en cada área.

La idea principal consiste en acceder a través de las variables de clase mostrando como resultados las variables de análisis. A esta combinación de variables de clase que proporcionan una medida se le denomina *hecho*. El Almacén de datos a diseñar deberá soportar en la medida de lo posible estos *hechos* para cualquier combinación lógica de variables de clase y medidas. Estos *hechos* se almacenarán para el nivel inferior de detalle de las categorías en la denominada *Tabla de Hechos*.

Otro concepto importante en este punto son las *dimensiones*. Con este nombre denominamos a una combinación especial de parámetros o variables de clase que proporcionan una visión particular de los datos. Un ejemplo claro de variable de clase que constituye por si sola una *dimensión* es el tiempo.

El Modelo dimensional combina estos *hechos* con las *dimensiones* para obtener un esquema global.

Una vez definidos los *hechos* y *dimensiones* que queremos guardar en nuestro sistema, solo nos queda escoger una estructura de datos que nos permita almacenar y gestionar de forma eficiente los mismos. En la elección deberemos considerar el equilibrio entre el espacio de los datos y el tiempo de respuesta en su acceso.

2.6 Temporalidad

El Almacén de datos se actualizará cada semana, habiendo o no cambios que cargar.

Como no se solicita la información por semanas sino por temporada en la dimensión tiempo no hemos tenido en cuenta la semana, si en futuro se quisiera tener la información con la temporalidad histórica semanal, simplemente se añadiría la fecha de la semana en las tablas.

2.7 Informes

Una vez generado el Almacén de datos crearemos los informes solicitados en el documento de requisitos, estos informes se realizarán en formato Excel y se dejarán en una ubicación específica.

A continuación pasamos la lista de informes a realizar.

Nombre	Descripción	Tipo	Periodicidad
INP0001	Información para cada jugador	Playoff	Anual
INP0002	Información de cada equipo	Playoff	Anual
INP0003A	Ranking de los jugadores por puntos	Playoff	Anual
INP0003B	Ranking de los jugadores por asistencias	Playoff	Anual
INP0003C	Ranking de los jugadores por global	Playoff	Anual
INP0003D	Ranking de los jugadores por canastas	Playoff	Anual
INP0004A	Ranking de los equipos por puntos	Playoff	Anual
INP0004B	Ranking de los equipos por asistencias	Playoff	Anual
INP0004C	Ranking de los equipos por rebotes	Playoff	Anual
INP0004D	Ranking de los equipos por partidos ganados	Playoff	Anual
IND0001	Información rondas de selección	Draft	Anual
INS00001	Informe para cada jugador	Season	Semanal
INS00001	Informe para cada equipo	Season	Semanal
INS0003A	Ranking de los jugadores por puntos	Season	Semanal
INS0003B	Ranking de los jugadores por asistencias	Season	Semanal
INS0003C	Ranking de los jugadores por rebotes	Season	Semanal
INS0003D	Ranking de los jugadores por canastas	Season	Semanal
INS0004A	Ranking de los equipos por puntos	Season	Semanal
INS0004B	Ranking de los equipos por asistencias	Season	Semanal
INS0004C	Ranking de los equipos por rebotes	Season	Semanal
INS0004D	Ranking de los equipos por partidos ganados	Season	Semanal
INS0004E	Ranking de los equipos por partidos perdidos	Season	Semanal

IND0005	Informe por entrenador	Season	Semanal
INA0001	Información para cada jugador	AllStar	Anual
INA0003A	Ranking de los jugadores por puntos	AllStar	Anual
INA0003B	Ranking de los jugadores por asistencias	AllStar	Anual
INA0003C	Ranking de los jugadores por global	AllStar	Anual
INA0003D	Ranking de los jugadores por canastas	AllStar	Anual
INT00001	Creación de DreamTeam	Season	Anual

2.8 Usuarios

En un principio se nos han descrito dos perfiles de usuarios:

- **Administrador**

Que tendrá acceso de lectura y escritura al Almacén de Datos, que estará compuesto por el departamento de informática del diario *New Sport*. También tendrá acceso de lectura y escritura a los informes generados.

- **Usuario final**

Que tendrá acceso de lectura al almacén de datos y a los informes, en un principio cualquier departamento del diario tendrá este tipo de perfil.

La empresa no nos ha especificado más perfiles de usuario, pero podríamos encontrarnos con usuarios finales de diferentes perfiles.

3 Arquitectura del proyecto

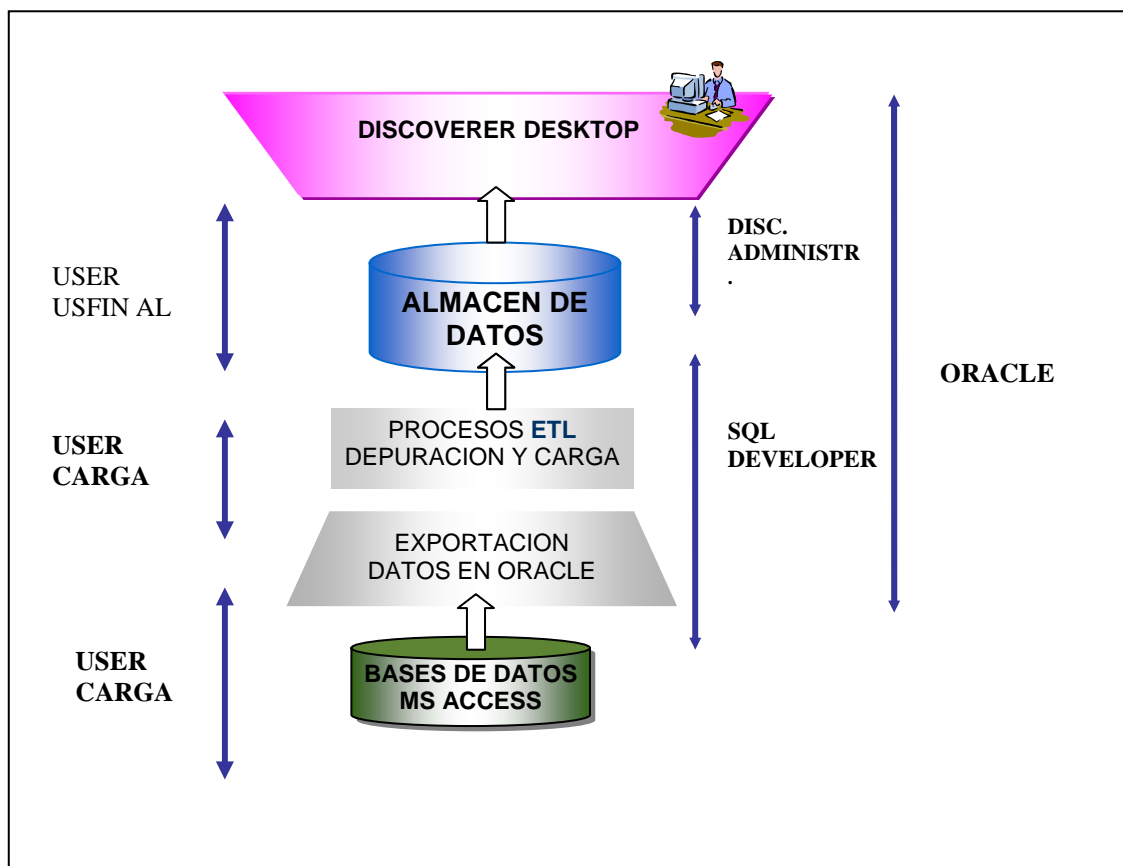
En este apartado se detallará la arquitectura necesaria para llevar a cabo el desarrollo del proyecto; así como los requerimientos técnicos para el correcto funcionamiento.

3.1 Entorno de trabajo

En este proyecto, por una parte tenemos los datos orígenes ubicados en MS ACCESS, los cuales solo tendremos acceso de lectura, el cliente es el que se encargará de actualizar estos datos, y por otra parte tenemos el entorno ORACLE que será nuestro entorno de trabajo.

La base de datos ORACLE la dividiremos en diferentes esquemas, por un lado, tendremos el entorno de real, es decir, el Almacén de datos propiamente dicho que verán los usuarios, y por otra parte tendremos dos entornos de trabajo, uno llamado *carga*, donde se ubicarán backups y las tablas exportadas desde ACCESS; y un usuario que se llamara *trans*; donde se llevan a cabo todos los procesos de transformación de los datos.

Una vez las cargas semanales actualizadas, el usuario podrá realizar consultas; así como podrá acceder a unos informes planificados, previamente diseñados por nosotros, que se actualizarán automáticamente



Dibujo5 –Arquitectura del proyecto

3.2 Oracle

Oracle, surgió a final de los años 70 y principio de los años 80, desde entonces ha ido evolucionando y se han ido desarrollando una serie de versiones, en nuestro caso utilizamos las 10g; en resumen, diremos que Oracle es básicamente una herramienta cliente/servidor para la gestión de Bases de Datos.

La Corporación Oracle ofrece este RDBMS como un producto incorporado a la línea de producción. Además incluye cuatro generaciones de desarrollo de aplicación, herramientas de reportes y utilitarios.

Oracle corre en pc's , microcomputadoras, mainframes y sistemas con procesamiento paralelo masivo. Soporta unos 17 idiomas, corre automáticamente en más de 80 arquitectura de hardware y software distinto sin tener la necesidad de cambiar una sola línea de código. Esto es porque más el 80% de los códigos internos de Oracle son iguales a los establecidos en todas las plataformas de sistemas.

Oracle soporta bases de datos de todos los tamaños, desde severas cantidades de bytes y gigabytes en tamaño.

Todo esto hace que sea el mayor y más usado Sistema de Base de Dato Relacional

Para desarrollar en Oracle utilizamos PL/SQL un lenguaje de 5ª generación, bastante potente para tratar y gestionar la base de datos.

3.3 Software utilizado

En este apartado explicamos todos los programas necesarios para el desarrollo técnico del proyecto.

3.3.1 MS ACCESS

Herramienta origen de los datos; hasta ahora el diario *Sport New* utilizaba esta herramienta como base para todas las operaciones sobre la información de la NBA, ya fuera actualizar como consultar.

A partir de ahora esta herramienta solo nos servirá para insertar los nuevos datos en las tablas; el diario no nos ha especificado como se actualizan estas tablas, pero si sabemos que se actualizan semanalmente; así en el proyecto se ha dejado el origen de los datos en MS ACCESS, por tanto, en nuestro proyecto se sigue utilizando esta herramienta para la persona que actualiza la base de datos, en ningún caso se utilizará como herramienta de usuario.

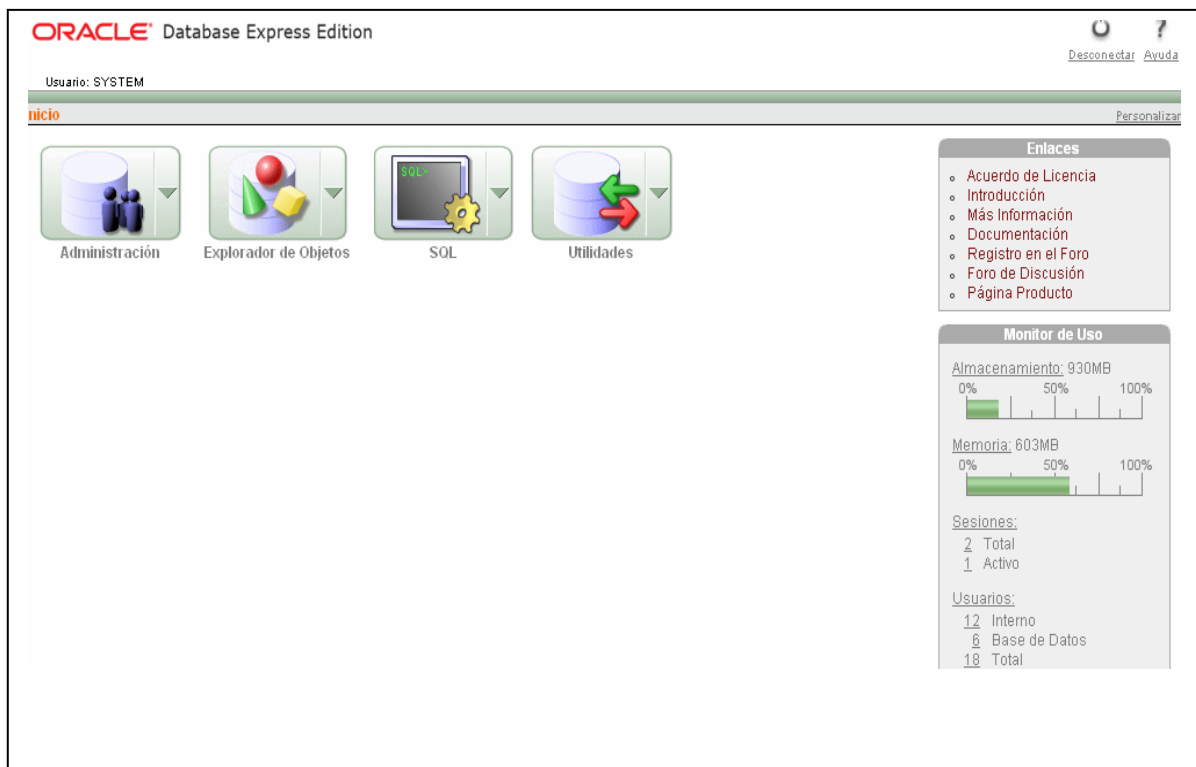
3.3.2 Oracle Express 10g

Este es el cliente Oracle que permite generar y administrar la base de datos Oracle.

Esta versión es gratuita y se puede descargar fácilmente, así se facilita la distribución a pequeñas empresas.

Las principales ventajas que tiene, a parte de la capacidad técnica, es el fácil uso para gestionar la base de datos.

La base de datos XE de Oracle se puede instalar en cualquier pc; pero hay que tener en cuenta que XE almacenará hasta 4GB de datos de usuario y utilizará hasta 1GB de

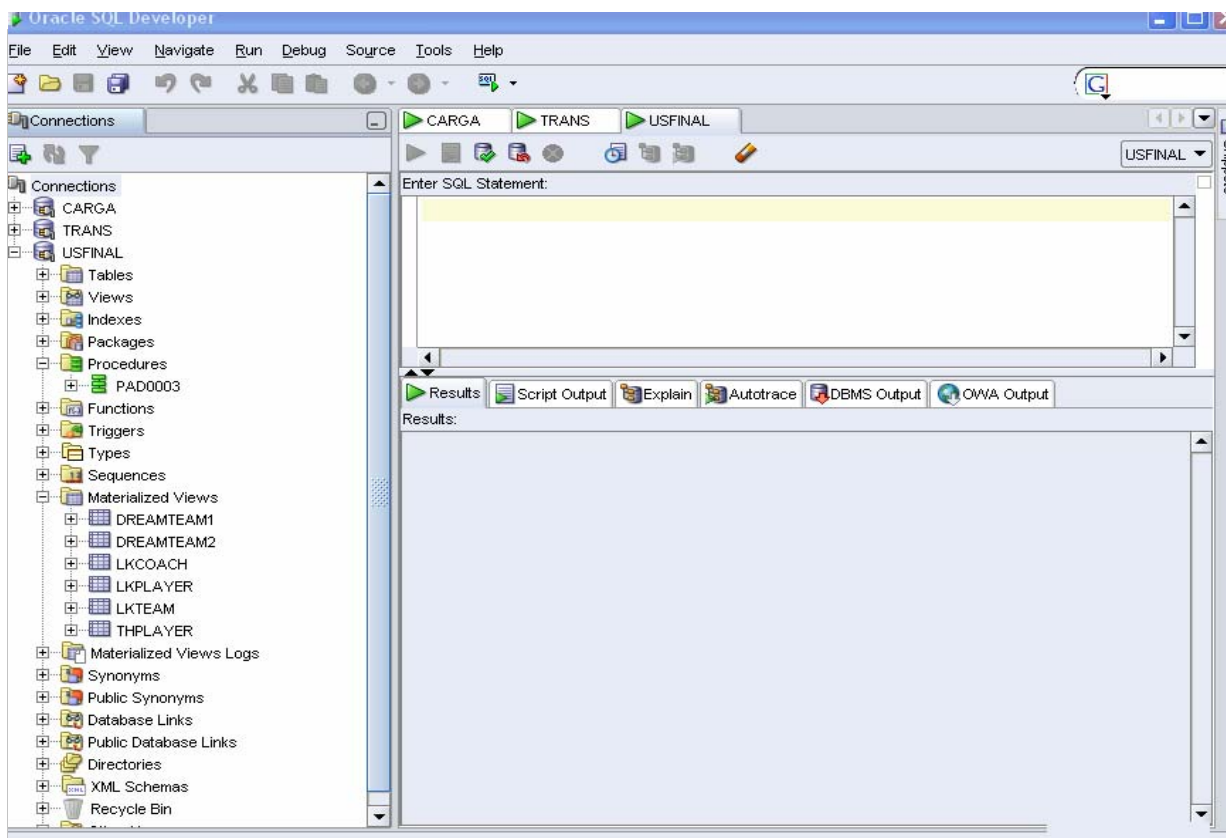


Dibujo 6 –Pantalla OracleExpress 10g

3.3.3. Oracle SQL Developer

Oracle SQL Developer es una herramienta gráfica para construir bases de datos en Oracle. Ésta es la herramienta que hemos utilizado para crear la base de datos, y probar todos los procedimientos; en definitiva has sido nuestra herramienta de desarrollo.

Con esta herramienta, podemos ver todos los objetos de la base de datos, así como ejecutar sentencias o sripts SQL , también nos permite debugar y ejecutar sentencias PL/SQL.



Dibujo 7- Pantalla SQL Developer

3.3.4 Oracle Discoverer

Oracle Business Intelligence Discoverer es un paquete que permite al usuario realizar consultas ad-hoc, reporting, e incluso exportar a excel o publicar en la Web información de la base de datos.

Oracle Business Intelligence Discoverer permite crear diferentes niveles de usuarios; así podemos tener usuarios finales que solo necesiten los iformes predefinidos o bin podemos tener usuarios más avanzados que puedan realizar sus propias consultas.

Esta herramienta tiene muchas aplicaciones que permiten crear el llamado EUL desde el cual podemos crear el escenario final que queremos sea visible para los diferentes usuarios

3.3.4.1 Oracle Discoverer Administrator

Esta es la herramienta que permite al Administrador definir el escenario para cada usuario, así como definir las áreas de negocio.

Con esta herramienta creamos el EUL (End User Layer), que sería el metadato de nuestro modelo de datos; es decir, creamos un EUL que contiene información de nuestras tablas; aquí podemos definir etiquetas para las tablas, crear jerarquías e incluso crear variables nuevas.

El EUL será el resultado final que verá cada usuario, se pueden definir tanto EUL como perfiles de usuario haya, así cada usuario tendrá visibilidad a el área que le corresponda; de esta manera podemos administrar debidamente los accesos de los usuarios a la información.

3.3.4.2 Oracle Discoverer Desktop

Discoverer Desktop es la herramienta para explotar el EUL que hemos creado con el Oracle Discoverer Administrator .

Desde esta herramienta vamos a crear todos los informes predefinidos que tendrá planificado el usuario

También permite al usuario generar consultas de una manera rápida, así como crear variables nuevas, exportar la información a Excel, o crear una serie de cálculos estadísticos sobre las tablas, sin necesidad de saber como funciona el lenguaje SQL, esta es una herramienta muy visual de cara a usuarios finales.

3.4 Requisitos de Hardware

- **Oracle Express 10g**

Es recomendable mínimo 1 Gb de Memoria y mínimo 5 Gb de espacio en el servidor para que pueda funcionar correctamente.

- **Para el resto de herramientas Oracle**

Se recomiendan que cada pc tenga unos 512 de memoria RAM y cualquier sistema operativo.

4 Análisis técnico

En este apartado se detallará el diseño técnico de los procesos que se han implementado para cargar los datos desde el sistema origen hasta transfórmalos en nuestro almacén de datos definido en el análisis funcional.

El objetivo principal a la hora de diseñar es crear procesos genéricos i reutilizables para poder simplificar el mantenimiento y futuras ampliaciones

El Almacén de Datos se alimentará de la información de los sistemas de origen , en nuestro caso las tablas Ms Access.

Crearemos procesos específicos para cada una de las tareas que se van a realizar como son la carga, transformación y certificación de la información.

Por último se definirán unos informes y se dejará un modelo de datos para consultable para el usuario final.

4.1 Origen de los datos

Para poder migrar los datos del Access a nuestro sistema Oracle, crearemos Una conexión creando un ODBC desde Windows, con esta de interfase de aplicaciones api se puede acceder a datos en sistemas manejadores de bases de datos tanto relacionales como no relacional, utilizando para ello SQL lenguaje de consulta estructurado.

Una vez creado el ODBC, crearemos una macro visual desde el MS ACCESS la cual nos exportará los datos al usuario carga de Oracle.

Así cada semana haremos esta exportación de tablas que será el inicio de los procesos ETL.

```
Option Compare Database
```

```
Sub TFC_NBA()
```

```
DoCmd.TransferDatabase acExport, "Bases de datos ODBC",
"ODBC;DSN=TFC;UID=trans;PWD=trans;SERVER=XE", acTable, "players", "players", False
DoCmd.TransferDatabase acExport, "Bases de datos ODBC",
"ODBC;DSN=TFC;UID=trans;PWD=trans;SERVER=XE", acTable, "draft", "draft", False
DoCmd.TransferDatabase acExport, "Bases de datos ODBC",
"ODBC;DSN=TFC;UID=trans;PWD=trans;SERVER=XE", acTable, "team_season", "team_season", False
DoCmd.TransferDatabase acExport, "Bases de datos ODBC",
"ODBC;DSN=TFC;UID=trans;PWD=trans;SERVER=XE", acTable, "Teams", "Teams", False
DoCmd.TransferDatabase acExport, "Bases de datos ODBC",
"ODBC;DSN=TFC;UID=trans;PWD=trans;SERVER=XE", acTable, "Coaches_season", "Coaches_season", False
DoCmd.TransferDatabase acExport, "Bases de datos ODBC",
"ODBC;DSN=TFC;UID=trans;PWD=trans;SERVER=XE", acTable, "Player_allstar", "Player_allstar", False
DoCmd.TransferDatabase acExport, "Bases de datos ODBC",
"ODBC;DSN=TFC;UID=trans;PWD=trans;SERVER=XE", acTable, "players", "players", False
DoCmd.TransferDatabase acExport, "Bases de datos ODBC",
"ODBC;DSN=TFC;UID=trans;PWD=trans;SERVER=XE", acTable, "player_playoffs", "player_playoffs", False
DoCmd.TransferDatabase acExport, "Bases de datos ODBC",
"ODBC;DSN=TFC;UID=trans;PWD=trans;SERVER=XE", acTable, "player_regular_season", "players", False
```

4.2 Modelo Estrella

En nuestro caso hemos escogido el modelo datos estrella, ya que este combina una o varias Tabla de Hechos con las distintas dimensiones. En este caso cada dimensión se almacena en una única tabla.

Las características principales de este modelo son que es altamente desnormalizado con mucha redundancia de datos lo que supone que el número de Joins para acceder a un dato será mínimo.

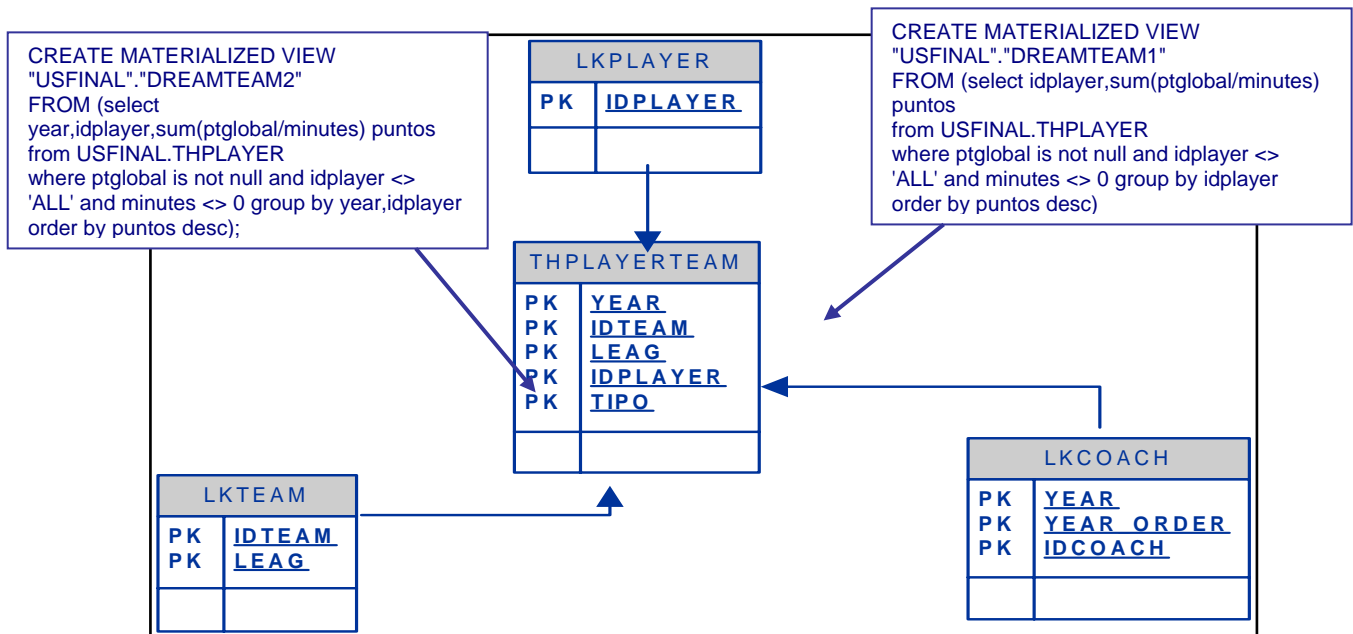
Este modelo se encuentra entre los modelos Snowflake y los modelos sin normalizar y es de uso frecuente en el diseño de sistemas de almacén de Datos.

En nuestro caso hemos optado por una única tabla de hechos, THPLAYER; esta tabla agregada contiene información a nivel de equipo y a nivel de jugador. Se he creado un campo tipo que nos indicará el nivel de detalle de la tabla.

Por otro lado se han creado 3 tablas de dimensiones, una de características del jugador, otra similar pero para los equipos y otra para el entrenador.

Por último para poder crear propuestas de Dream Team se han generado dos vistas Sobre la tabla de hechos.

En el siguiente diagrama podemos ver el diseño de nuestro almacén de Datos .



Dibujo8 – Modelo Estrella

4.2.1 Asociación entre Tablas finales y tabla de carga de origen

En este punto vamos analizar las transformaciones de los datos origen con las tablas finales, hemos creado la siguiente tabla a modo esquemático, donde se muestran las transformaciones que se llevarán a cabo.

Tabla Final	Tabla origen	Cambios en campos	Campo nuevo	Campo Eliminado
LKPLAYER	PLAYERS	llkid => idplayer H_feet=> mfeet h_inches=>minches weight=>mweight		
LKTEAM	TEAM	team=> idteam		
LKCOACH	COACH_SEASON	Coachid => Idcoach		
THPLAYER	PLAYER_REGULAR_SEASON PLAYER_PLAYOFFS PLAYER_ALLSTAR DRAFT TEAM_SEASON	llkid => idkid team=> Idteam Todos los campos de puntos => totxxx team=> Idteam o_xx => totxxx_	Tipo pct_fg pct_ft pct_tp ptglobal	draft (utilizamos year)

4.2.2 Campos nuevos:

Tabla Final	Campo Nuevo	Cálculo
THPLAYER	Tipo	Si " es detalle de equipo Si existe PLAYER_REGULAR_SESION='S' Si existe PLAYER_PLAYOFFS='P' Si existe PLAYER_ALLSTAR='A' Si existe DRAFT='D'
	pct_fg	totfgm/totfga
	pct_ft	totftm/totfta
	pct_tp	tottpm/tottpa
	ptglobal	totasts+totpts+totreb +totstl+totblk

4.2.3 Tablas del modelo

Tabla de dimensiones:

- **Tabla LKPLAYER**

Hemos creado esta tabla con las características de cada jugador, el índice de esta tabla será el identificador de cada jugador.

Esta tabla tendrá un único registro por jugador, es decir, siempre tendrá la información actualizada del jugador.

LKPLAYER		
Campos	Descripción	Obs
<u>Idplayer</u>	Identificador Jugador	Not null
firstname	1er Apellido	Not null
lastname	2on Apellido	Not null
position	Posición	
firstseason	Primera temporada	
lastseason	Última temporada	
m_feet	medida pie	
m_inches	Altura	
m_weight	Peso	
college	Colegio	
birthdate	Fecha Nacimiento	
Idteam	Equipo Actual	Not null

- **Tabla LKTEAM**

Descripción: Características del equipo, esta tabla tendrá un único registro por equipo, es decir, la información siempre será la foto actual.

LKTEAM		
Campos	Descripción	Obs
<u>Idteam</u>	Equip	Not null
location	Localidad	Not null
name	Nombre completo	Not null
leag	Liga	Not null

- **Tabla LKCOACH**

Descripción: Características del entrenador

En esta tabla se había estudiado la posibilidad de eliminar la información sobre partidos ganados y calcularlo a partir de la tabla de hechos de equipo pero se ha descartado por su complejidad.

Claves foràneas: Idteam -Lkteam

LKCOACH		
Atributos	Descripción	Obs
<u>Idcoach</u>	Identificador entrenador	Not null
<u>year</u>	Año	Not null
<u>yr_order</u>	Orden de año	Not null
firstname	1er Apellido	Not null
lastname	2on Apellido	Not null
Season_win	Partidos ganados	Not null
Season_loss	Partidos perdidos	
Playoff_win	Partidos ganados Playoff	
Playoff_loss	Partidos perdidos Playoff	
Idteam	Equipo Actual	Not null

Tablas de Hechos :

- **Tabla THPLAYERTeam**

Descripción: Contiene la información agregada a diferentes niveles, tanto a nivel de jugador como equipo.

Claves foràneas: Idplayer-Lkplayer
Idteam, leg -Llteam

THPLAYERTeam			
Atributos	Descripción	obs	tipo
<u>Year</u>	año temporada	Not null	Categoría
<u>IdTeam</u>	Nombre equipo	Not null	Categoría
<u>leag</u>	liga	Not null	Categoría
<u>idplayer</u>	Identificador jugador	not null	Categoría
<u>tipo</u>	tipo registro	not null	Categoría
conference	Conferencia		Categoría
draft_round	Ronda Seleccion		Categoría

draft_from	Origen de Ronda		Categoría
o_fgm	Canastas ofensivas		análisis
o_fga	Canastas falladas ofensivas		análisis
o_ftm	tiros libres ofensivas		análisis
o_fta	Tiros libres fallados ofensivas		análisis
o_oreb	Rebotes ofensivos		análisis
o_dreb	Rebotes defensivos		análisis
o_ast	Asistencias ofensivas		análisis
o_stl	Robos ofensivos		análisis
o_to	Giros ofensivos		análisis
o_blk	Bloqueos ofensivos		análisis
o_3pm	Triple ofensivos		análisis
o_3pa	Triples fallados		análisis
o_pts	Puntos ofensivos		análisis
pace	Partidos empatados		análisis
won	Partidos ganados		análisis
lost	Partidos perdidos		análisis
Pct_fg	Ratio canastas ofensivas	calculado	análisis
Pct_ft	Ratio tiros libres	calculado	análisis
Pct_tp	Ratio de triples	calculado	análisis
ptglobal	Puntuación global	calculado	análisis
gp	Partidos jugados		Análisis
minutes	minutos jugados		Análisis
totpts	Puntos		Análisis
totdreb	Rebotes defensivos		Análisis
totoreb	Rebotes ofensivos		Análisis
totreb	Rebotes		Análisis
totasts	Asistencias		Análisis
totstl	Robos		Análisis
totblk	Bloqueos		Análisis
totto	Giros		Análisis
totfga	canastas fallados		Análisis
totfgm	canastas		Análisis
totfta	tiros libres fallados		Análisis
totftm	tiros libres		Análisis
tottpa	Triples fallados		Análisis
tottpm	Triples		Análisis

4.3 Procesos de Carga en Oracle

En este apartado hemos identificado todos los procesos que deberemos realizar para poder cargar las tablas de análisis final.

Se han elaborado cinco procesos los cuales se planificarán una vez por semana para actualizar el Almacén de datos.

Estos procesos se planificarán de manera que cada uno dependerá del correcto fin del proceso anterior para empezar a ejecutarse.

- Proceso de carga de los datos: Se planificará el domingo por la noche, una vez actualizado en el ACCESS los datos de la semana, exportación de la macro Visual.
- Proceso de Transformación de los datos: se realizarán las transformaciones complejas que no se pueden realizar directamente en la migración.
- Proceso de comprobación de la integración de los datos e identificación de errores.
- Migración de los datos y backup : Se cargarán los datos directamente para que estén accesibles para los usuarios.
Antes de actualizar las tablas de usuarios se realizará un backup de las tablas actuales
- Diseño de Informes.

4.3.1 Esquema resumen de usuarios creados en Oracle :

En nuestro proyecto hemos creado tres esquemas/usuarios; se ha decidido crearlos para que la actualización semanal no tenga impacto en el usuario; así puede seguir trabajando sin problemas mientras pasan los procesos.

1) CARGA Usuario donde se guarda todas las tablas que no manipulamos.

- **Tablas:**

1. Se ubicarán el import de las tablas Acces.
2. Se creará un vistas materializadas de las tablas de real, un backup al principio del proceso.

- **Procedures:**

1. Actualización de las vistas de backup
2. Borrado tablas Access.

USUARIO CARGA	
TABLAS	
draft	export desde MS ACC
teams	export desde MS ACC
team_season	export desde MS ACC
players	export desde MS ACC
coaches_season	export desde MS ACC
player_regular_season	export desde MS ACC
Player_allstar	export desde MS ACC
player_playoff	export desde MS ACC
LKCOACH_BKP	Vistas materializadas de USFINAL.LKCOACH
LKTEAM_BKP	Vistas materializadas de USFINAL.LKTEAM
LKPLAYER_BKP	Vistas materializadas de USFINAL.LKPLAYER
THPLAYER_BKP	Vistas materializadas de USFINAL.THPLAYER
PROCESOS	
PAD0000	Actualización tablas *_BKP
PAD0004	Borrado tablas MS Acces

2) CARGA Usuario principal donde están todas las tablas que se transforman; así como los procedures principales de carga y transformación.

- **Tablas:**
 1. Se crean todas las tablas de transformacions merge entre tablas carga y actuales del usuario trans (insert/updates)
 2. Se crean todas las tablas de Errores (en los insert/updates).
- **Procedures:**
 1. Están todos los procedimientos principales

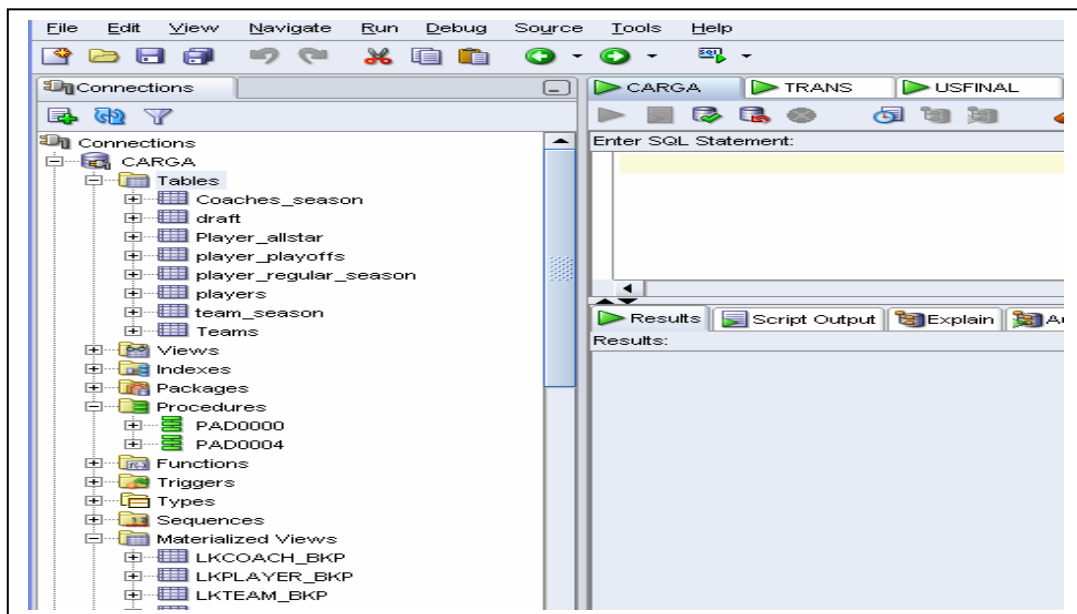
USUARIO TRANS	
TABLAS	
EXCEPTIONS	Errores
LKCOACH	Tablas actualizadas
LKTEAM	Tablas actualizadas
LKPLAYER_	Tablas actualizadas
THPLAYER	Tablas actualizadas
LKCOACH_ERR	Tablas actualizadas errores
LKTEAM_ERR	Tablas actualizadas errores
LKPLAYER_ERR	Tablas actualizadas errores
THPLAYER_ERR	Tablas actualizadas errores
PROCESOS	
PAD0001	Procedimiento que encapsula al resto
PR1DESCONTRAINTS	Deshabilitar Constraints

PR2CARGALKPLAYER	Carga LKPLAYER
PR3CARGALKTEAM	Carga LKTEAM
PR4CARGALKCOACH	Carga LKCOACH
PR5CARGATHPLAYER	Carga THPLAYER
PR7VALIDAR	Validación de claves
PR8CONSTRAINTS	Habilitar Constraints

3) USFINAL Usuario que contiene las tablas finales de usuario.

- **Tablas:**
 1. Se crean vistas materializadas de las tablas del TRANS, que serán las tablas finales de usuario.
- **Procedures:**
 1. Actualización vistas.

USUARIO USFINAL	
TABLAS FINALES DE USUARIO	
LKCOACH	Vistas materializadas de TRANS.LKCOACH
LKTEAM	Vistas materializadas de TRANS..LKTEAM
LKPLAYER	Vistas materializadas de TRANS.LKPLAYER
THPLAYER	Vistas materializadas de TRANS.THPLAYER
DREAMTEAM1	Vistas materializadas de USFINAL.THPLAYER
DREAMTEAM2	Vistas materializadas de USFINAL.THPLAYER
PROCESOS	
PAD0003	Actualización Vistas



Dibujo9-SQL DEVELOPER user CARGA

4.3.2 Ejecución de Procesos

Nuestros procesos de ETL se ejecutarán mediante el planificador de Windows, todo y que el cliente no nos lo ha especificado se podría utilizar cualquier tipo de planificador. .

En nuestro proceso hemos creado un fichero *.bat que ejecuta todos los procesos de carga.

Fichero ejecutar_tot.bat

```
sqlplus TRANS/TRANS@XE @C:\TFC_NBA\ejecutar_tot.txt
```

Fichero ejecutar_tot.txt

```
BEGIN
-- Backup Tablas Finales de la carga anterior
  CARGA.PAD0002;
  COMMIT;

-- Carga de las tablas de trabajo
  TRANS.PAD0001;
  COMMIT;

-- Analize tablas de carga
FOR N IN (SELECT DISTINCT OWNER FROM ALL_TABLES WHERE OWNER =
'TRANS') LOOP
DBMS_STATS.GATHER_SCHEMA_STATS
(ownname
=>N.OWNER,
estimate_percent
=> 20,
options
=> 'GATHER STALE' );
COMMIT;
END LOOP;

-- Paso de la información a tablas de usuario final
USFINAL.PAD0003;
  COMMIT;

-- Analize tablas finales
FOR N IN (SELECT DISTINCT OWNER FROM ALL_TABLES WHERE OWNER =
'USFINAL' AND (TABLE_NAME like 'LK%' OR TABLE_NAME like 'TH%') ) LOOP
DBMS_STATS.GATHER_SCHEMA_STATS
(ownname
=>N.OWNER,
estimate_percent
=> 20,
ptions
=> 'GATHER STALE' );
COMMIT;
END LOOP;
END;
```

La función `dbms_stats.gather_schema_stats` realiza el computo de las estadísticas de todas las tablas para el esquema indicado. El refresco de las estadísticas de las tablas mejora el rendimiento de índices y consultas sobre las mismas, permitiendo al motor gestor de la base de datos seleccionar el plan de acceso óptimo a los datos.

4.3.3 Detalle de los procedimientos

Nombre	PAD0000
Descripción	<ul style="list-style-type: none"> Esta será nuestro primer proceso planificado en Oracle 1. Se realizará un refresh de las vistas materializadas en el usuario CARGA
Panificación	Depende de la exportación de Access-Oracle
Tablas	LKCOACH_BKP, LKTEAM_BKP, LKPALYER_BKP, THPLAYERTEAM_BKP
Entorno de trabajo	Esquema CARGA

Nombre	PAD0001
Descripción	<ul style="list-style-type: none"> Esta será el procedimiento principal que contendrá el resto de procedimientos del esquema trans
Panificación	Depende del finok PAD0000
Tablas	Todas las tablas TRANS
Entorno de trabajo	Esquema TRANS

```

CREATE OR REPLACE PROCEDURE "TRANS"."PAD0001" AS
  v_sql VARCHAR2 (450);
BEGIN

  -- Desactivar Constraints
  PR1DESCONSTRAINTS('LKTEAM');
  PR1DESCONSTRAINTS('LKPLAYER');
  PR1DESCONSTRAINTS('LKCOACH');
  PR1DESCONSTRAINTS('THPLAYER');

  -- Borrado Inicial de las tablas de Error
  v_sql := 'TRUNCATE TABLE THPLAYER_ERR DROP STORAGE';
  EXECUTE IMMEDIATE v_sql;
  v_sql := 'TRUNCATE TABLE LKCOACH_ERR DROP STORAGE';
  EXECUTE IMMEDIATE v_sql;
  v_sql := 'TRUNCATE TABLE LKPLAYER_ERR DROP STORAGE';
  EXECUTE IMMEDIATE v_sql;
  v_sql := 'TRUNCATE TABLE LKTEAM_ERR DROP STORAGE';
  EXECUTE IMMEDIATE v_sql;

  -- Carga Tablas TRANS
  PR2CARGALKPLAYER;
  PR3CARGALKTEAM;
  PR4CARGALKCOACH;
  PR5CARGATHPLAYER;
  PR7VALIDAR('LKTEAM');
  PR7VALIDAR('LKPLAYER');
  PR7VALIDAR('LKCOACH');
  PR7VALIDAR('THPLAYER');
END;

```


Nombre	PR1DESCONSTRAINTS
Descripción	Se trata de deshabilitar todas las constraints de las tablas que tenemos guardadas de la semana anterior y que están actualizadas. Se llama al procedimiento para cada tabla
Panificación	Llamada desde PAD0001
Tablas	LKCOACH, LKTEAM, LKPALYER, THPLAYERTEAM
Entorno de trabajo	Esquema TRANS

Nombre	PR2CARGALKTEAM
Descripción	Hacemos un merge con las tablas de CARGA.team y TRANS.LKTEAM; si existe el registro hará un UPDATE de los campos, si no insertaremos un nuevo registro en la tabla TRANS.LKTEAM
Panificación	Llamada desde PAD0001
Tablas	TRANS. LKTEAM, CARGA.TEAM
Entorno de trabajo	Esquema TRANS

Nombre	PR3CARGALKCOACH
Descripción	Hacemos un merge con las tablas de CARGA.coach_season y TRANS.LKCOACH; si existe el registro hará un UPDATE de los campos, si no insertaremos un nuevo registro en la tabla TRANS.LKCOACH
Panificación	Llamada desde PAD0001
Tablas	TRANS. LKCOACH, CARGA.coaches_season
Entorno de trabajo	Esquema TRANS

```

CREATE OR REPLACE PROCEDURE "TRANS"."PR3CARGAKCOACH" AS
BEGIN
MERGE INTO LKCOACH T
USING "CARGA"."Coaches_season" V
ON(T.IDCOACH=TRIM(UPPER(V."coachid")) and T.YEAR=TRIM(V."year") and
T.YR_ORDER=TRIM("yr_order"))
WHEN MATCHED THEN
UPDATE SET
T.IDTEAM=UPPER(V."team"),
T.FIRSTNAME=V."firstname"||','||V."lastname",
T.SEASON_WIN=V."season_win",
T.SEASON_LOSS=V."season_loss",
T.PLAYOFF_WIN=V."playoff_win",
T.PLAYOFF_LOSS=V."playoff_loss"
WHEN NOT MATCHED THEN INSERT
(IDCOACH,YEAR,YR_ORDER,FIRSTNAME,SEASON_WIN,SEASON_LOSS,PLAYOFF_
WIN,PLAYOFF_LOSS,IDTEAM)
VALUES
(TRIM(UPPER(V."coachid")),TRIM(V."year"),TRIM(V."yr_order"),V."firstname"||','||V."lastna
me",V."season_win",V."season_loss"
,V."playoff_win",V."playoff_loss",TRIM(UPPER(V."team")));

EXCEPTION
WHEN OTHERS
THEN
BEGIN
DBMS_OUTPUT.PUT_LINE('ERROR - CARGA_LKCOACH');
END;
END;

```

Nombre	PR4CARGALKPLAYER
Descripción	Hacemos un merge con las tablas de CARGA.players y TRANS.LKPLAYER; si existe el registro hará un UPDATE de los campos, si no insertaremos un nuevo registro en la tabla TRANS.PLAYER
Panificación	Llamada desde PAD0001
Tablas	TRANS. LKPLAYER, CARGA.players
Entorno de trabajo	Esquema TRANS

Nombre	PR5CARGATHPLAYER
Descripción	Hacemos un merge con las tablas de CARGA restantes y TRANS.THPLAYER; si existe el registro hará un UPDATE de los campos, si no insertaremos un nuevo registro en la tabla TRANS.THPLAYER. En el caso de la tabla DRAFT, solo realizamos update con la información.
Panificación	Llamada desde PAD0001
Tablas	TRANS. THPLAYERTEAM, CARGA.team_season,CARGA.player_playoff, CARGA.draft, CARGA.player_allstar, CARGA.player_regular_season
Entorno de trabajo	Esquema TRANS

Nombre	PR7VALIDAR
Descripción	Validamos cada una de las tablas de TRANS. Cualquier tipo de error tipo constraints se informan en unas tablas de errores que el usuario podrá ver. Llamamos para cada una de las tablas.
Panificación	Llamada desde PAD0001
Tablas	LKCOACH, LKTEAM, LKPALYER, THPLAYER, LKCOACH_ERR, LKTEAM_ERR, LKPLAYER_ERR, THPLAYER_ERR,
Entorno de trabajo	Esquema TRANS

Nombre	PR8CONSTRAINTS
Descripción	Habilita de nuevo todas las constraints
Panificación	Llamada desde PR7VALIDAR
Tablas	LKCOACH, LKTEAM, LKPALYER, THPLAYER
Entorno de trabajo	Esquema TRANS

Nombre	PAD0003
Descripción	Realiza un refresh de las vistas materializadas
Panificación	Depende del finok PAD0001
Tablas	LKCOACH, LKTEAM, LKPALYER, THPLAYER,DRAMTEMA1,DREAMTEAM2
Entorno de trabajo	Esquema USFINAL

Nombre	PAD0004
Descripción	Realiza un borrado de las tablas cargadas desde el ACCESS
Panificación	Depende del finok PAD0003
Tablas	Draft, players,team,coach_season,player_playoff, player_regular_season, Team_season,players_allstar
Entorno de trabajo	Esquema CARGA

4.4 Creación entorno de trabajo de usuario

En este paso vamos a crear el entorno de trabajo de los usuarios.

4.4.1 Discoverer Administrator- EUL

Creamos el Eul de usuario, en nuestro caso creamos un único usuario con permisos de lectura y de consulta, pero no podrá crear informes planificados.

Creamos dos áreas de negocio :

1. Datos pdtes de Validar

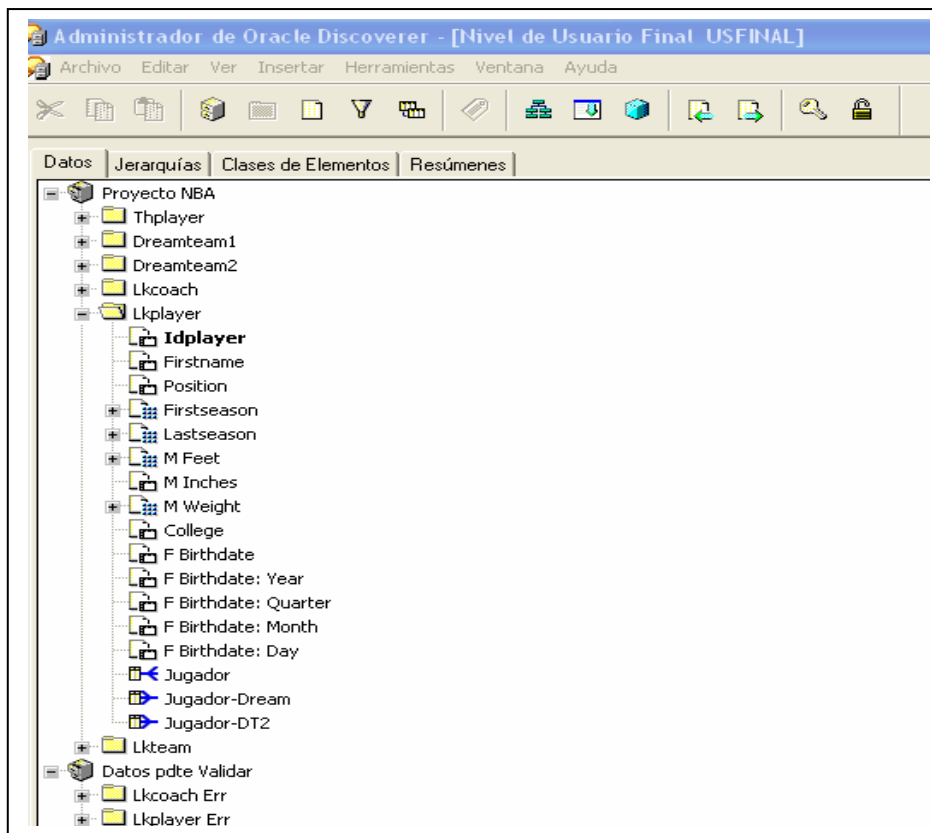
Que contiene las tablas de Errores, con las cuales haremos informes para que el usuario pueda comprobar los errores de carga, es decir, todos aquellos registros que tienen alguna información incoherente, si el usuario lo encuentra oportuno, se puede rectificar el error actualizando las tablas MS ACCES y volver a hacer carga en Oracle.

2. Información NBA

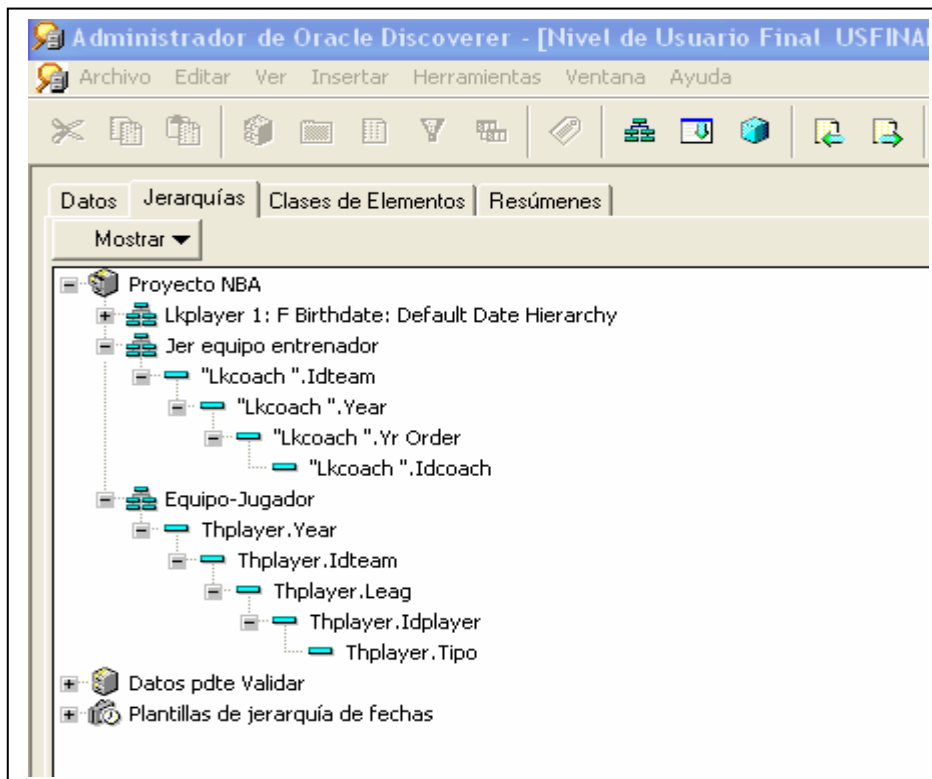
Aquí se encuentra nuestro modelo de Datos con las cuatro tablas finales, LKTEAM, LKPLAYER, LKCOACH, THPLAYER y las dos vistas, DREAMTEAM1, DREAMTEAM2

Se han creado dos jerarquías, una por Equipo-jugador y otra por Equipo-entrenador.

Se han creado las uniones correspondientes entre tablas.



Dibujo 10-Pantalla Administrator -EUL



Dibujo 11-Pantalla Jerarquías -EUL

4.4.2 Discoverer Desktop

Desde esta herramienta que también servirá al usuario final para explotar la información de la NBA, construimos los informes a través del Eul definido con el Administrator.

1 .Informe por Entrenador

A partir de la tabla LKCOACH creamos este informe por año y también se puede escoger el equipo; se le añaden variables calculadas de ratios.

Datos Entrenador por año y Equipo						
Elementos de Página: Year: 2003 ▾ Equipo: <Todo> ▾						
	Partidos perdidos	Partidos Ganados	% Partidos	Playoff Ganados	Playoff Perdidos	
Don,Nelson	30	52	63,4%	1	4	
Flip,Saunders	24	58	70,7%	10	8	
Gregg,Popovich	50	57	69,5%	12	8	
Hubie,Brown	32	50	61,0%	0	4	
Jeff,Bzdelik	78	43	104,9%	2	8	
Jeff,Van Gundy	37	45	54,9%	1	4	
John,Carroll	22	14	38,9%	0	4	

Dibujo 12-Informe Entrenador

2 .Informe por jugador:

Mediante la jerarquía tenemos la información para cada jugador por Playoff, temporada Regular y AllStar, eliminado los registros a nivel de equipo.

También podemos escoger el año y el equipo.

se ha de filtrar la tabla por tipo '<>' para escoger los registros a nivel de jugador.

Se crean diferentes informes por jugador:

- 1 . Datos jugador por años : Las variables que se han utilizado son las del jugador y se le han añadido ratios, como canastas acertados sobre fallos.
2. Datos jugador por posición de juego: Diferentes datos por la posición de juego de los jugadores.
2. Rondas de selección : Información de las rondas de selección
3. Ranking de jugadores : Se crean una serie de informes por ranking de jugador según alguna característica y dividiéndola por los minutos jugados en la temporada
 - Ranking por rebotes
 - Ranking por puntos
 - Ranking por ratio de aciertos
 - Rankig por asistencias

Datos Jugador por Año:														
Información => Tipos :A-allstar; P-Playoff; R- Regular														
Elementos de Página: Year: 2000 Tipo: P Name: <Todo>														
	Partidos	Puntos	Minutos	Rebotes	Asistencias	Robos	Bloqueos	Canastas	Canasta Fallo	% Canastas	Tiros Libres	Tiembres Fallo	%Tiros libres	Triples
A.c., Green	3	3	21	4	2	1	0	1	3	50,0%	1	1	50,0%	
Al,Harrington	3	5	40	4	3	0	0	2	13	50,0%	1	2	33,3%	
Allan,Houston	5	104	189	9	7	5	1	38	64	63,3%	22	22	50,0%	
Alonzo,Mourning	3	35	91	16	3	0	5	12	25	38,7%	11	19	36,7%	
Alvin,Williams	12	165	486	35	50	15	8	69	160	73,4%	17	25	40,5%	
Andrew,Declercq	4	20	54	16	1	2	2	8	14	50,0%	4	8	33,3%	
Antonio,Harvey	2	0	14	6	0	0	0	0	1	#DIV/0!	0	0	#DIV/0!	
Austin,Croschere	4	43	129	20	6	4	2	14	35	48,3%	13	15	46,4%	
Avery,Johnson	13	76	281	16	41	10	1	34	88	69,4%	8	15	34,8%	
Bobby,Jackson	8	56	182	26	18	8	0	21	48	60,0%	10	14	41,7%	

Dibujo 13-Informe Jugador

Ranking jugadores por Canastas			
Información => tipo A:AllStar R:Regular P:Playoff			
Elementos de Página: Year: 2001 Tipo: R			
	Minutos	Totfgm SUM	%canastas
Shawn,Marion	3110	654	21,0%
Carl,Malone	3040	635	20,9%
Keith,Vanhorn	2462	471	19,1%
Steve,Nash	2837	525	18,5%

Dibujo 14-Rankig por Aciertos

3 .Informes por equipo:

Sobre la THPLAYER, se ha escogido los registros a nivel de equipo y se muestran casi todas las variables analíticas; también se han creado ratios, se ha de filtrar la tabla por tipo=' ' para escoger los registros totales por equipo..

Se crean diferentes informes por equipo:

- 1 . Datos equipo por años : Las variables que se han utilizado son las del equipo y se le han añadido ratios, como canastas acertados sobre fallos.
2. Comparativa de equipos: Comparativa de la evolución de cada equipo por años.
4. Ranking de equipos : Se crean una serie de informes por ranking de equipos según alguna característica y dividiendola por partidos jugados en la temporada
 - Ranking por partidos ganados
 - Ranking por aciertos
 - Ranking por partidos perdidos

Comparativa de Equipos									
	Perdidos	Ganados	%Partidos	Asistencias	Rebotes	CAstas	Tiros Libres	TRiples	
Rockets	244	216	47,0%	22429	41057	37637	18193	2019	
1998	19	31	62,0%	2174	4313	3683	1731	36	
1997	41	41	50,0%	3794	6916	6264	3149	59	
1996	25	57	69,5%	4218	7517	6632	3154	81	
1988	37	45	54,9%	4128	7495	6985	3810	17	
1982	68	14	17,1%	4183	7374	6979	3183	6	
1977	54	28	34,1%	3932	7442	7094	3166		
TOTALES	244	216	47,0%	22429	41057	37637	18193	2019	
Rebets	30	30	50,0%	494	0	1674	903		

Dibujo 15-Comparativa de Equipos

Ranking por partidos ganados			
Elementos de Página: Year: 2002			
	Gandos	Perdidos	%partidos
	N	N	N
Mavericks	60	22	73,2%
Spurs	60	22	73,2%
Kings	59	23	72,0%
Timberwolves	51	31	62,2%
Lakers	50	32	61,0%
Pistons	50	32	61,0%
Trailblazers	50	32	61,0%
Nets	49	33	59,8%
Pacers	48	34	58,5%
76ers	48	34	58,5%
Warriors	47	35	57,0%

Dibujo 16-Ranking partidos ganados

2) Propuestas Dream Team:

A partir de las dos vistas creadas de la tabla THPLAYER, creamos dos propuestas de Dream Team , una para el mejor Dream Team de la historia y la otra para el mejor Dream Team del año.

Para escoger los mejores jugadores se ha utilizado una variable nueva calculada PTGLOBAL, que es la suma de varias características del jugador; esta variable se ha dividido por los minutos jugados para obtener un ratio con el cual se obtienen los mejores jugadores.

DREAM TEAM de LA HISTORIA								
	F.Nacimiento	Puntos	1 Temporada	Ultima temporada	Position	Colegio	Pie	Pulgadas
Hakeem,Olujuwon	21/01/63	36	1984	2001	C	Houston	6	10.0
Julius,Erving	22/02/50	37	1971	1986	F	Massac	6	6.0
Kareem,Abdul-jabbar	16/04/47	42	1969	1988	C	UCLA	7	2.0
Karl,Malone	24/07/63	41	1985	2003	F	Louisian	6	9.0
Michael,Jordan	17/02/63	37	1984	2002	G	North Ca	6	6.0

Dibujo 17-Dream Team de la historia

Dream Team por Años							
	Year	Nombre	1 Temporada	2 Temporada	Colegio	F Birthdate	Posición Gk
▶ 1	2004	Ibrahim,Kutluay	2004	2004		01-JUL-1974	8,3%
▶ 2	2004	Maurice,Baker	2004	2004	Oklahoma St.	28-JUL-1979	22,2%
▶ 3	2004	Pavel,Podkolzine	2004	2004		15-ENE-1985	30,0%
▶ 4	2004	Linton,Johnson	2003	2004	Tulane	13-JUN-1980	26,7%
▶ 5	2004	Britton,Johnsen	2003	2004	Utah	08-JUL-1979	31,0%
▶ 6	2003	Popeye,Jones	1993	2003	Murray State	17-JUN-1970	10,0%
▶ 7	2003	Pat,Garrity	1998	2004	Notre Dame	23-AGO-1976	13,6%
▶ 8	2003	Ruben,Boumtje-Boum	2001	2003	Georgetown	20-MAY-1978	26,9%
▶ 9	2003	Mark,Pope	1997	2003	Kentucky	11-SEP-1972	30,0%
▶ 10	2003	Michael,Curry	1993	2004	Georgia Southern	12-AGO-1968	30,3%
▶ 11	2002	Chris,Dudley	1987	2002	Yale	22-FEB-1965	18,2%

Dibujo 18-Dream Team por año

3) Informes de Errores

Se define un informe de error para cada tabla obteniendo la clave del registro para poder identificar el error del dato.

- Informe para LKCOACH
- Informe para LKTEAM
- Informe para LKPLAYER
- Informe para THPLAYER

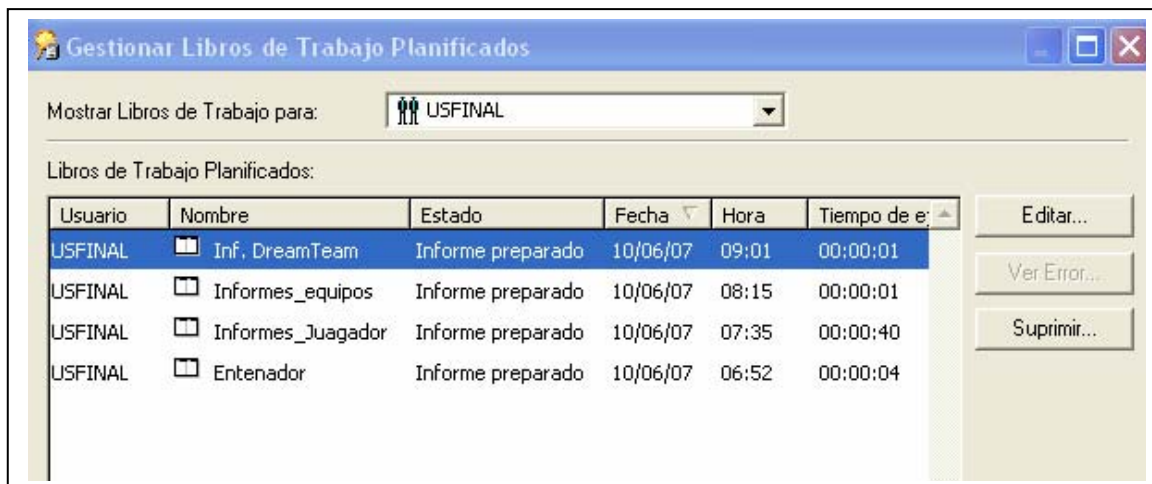
THPLAYER-Infomres errores					
	Year	Idteam	Leag	Idplayer	Tipo
1	1990	TOT	N	BROWNT001	R
2	1991	TOT	N	BROWNT001	R
3	2000	TOT	N	BRUNSRI01	R
4	1976	TOT	N	BROKAGA01	R
5	1994	TOT	N	BROOKSC01	R

Dibujo 19-Infomre Errores THPLAYER

Planificación de Informes

Todos los informes desarrollados por nosotros se planifican para que se vayan refrescando semanalmente.

El propio discoverer nos permite planificar los informes y desde el Discoverer Administrator los podremos gestionar.



Dibujo20-Administrator Palnificador

5. Escenario final para Usuario

Una vez acabado todos los procesos de carga , toda la elaboración del EUL y todos los informes predefinidos y creados el usuario ya puede explota la base de datos.

En nuestro caso los usuarios finales tienen acceso de lectura en la base de datos, pero no podrán planificar informes, de momento, este tema se deja en manos del administrador, en el caso que se optará por abrir el perfil del usuario se deberían cambiar los permisos desde el Discoverer Administrator.

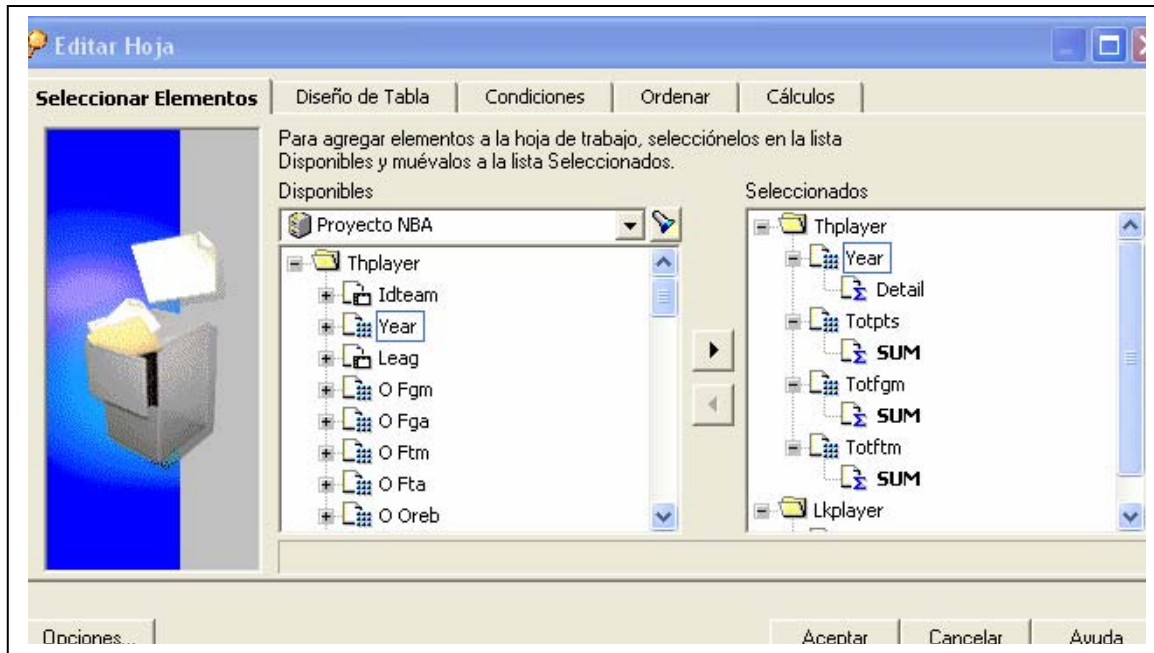
En el siguiente apartado mostramos unos pantallazos de la herramienta de cómo realizar una simple consulta sobre una tabla.

1. Escogemos la opción deseada, en este caso una simple tabla.



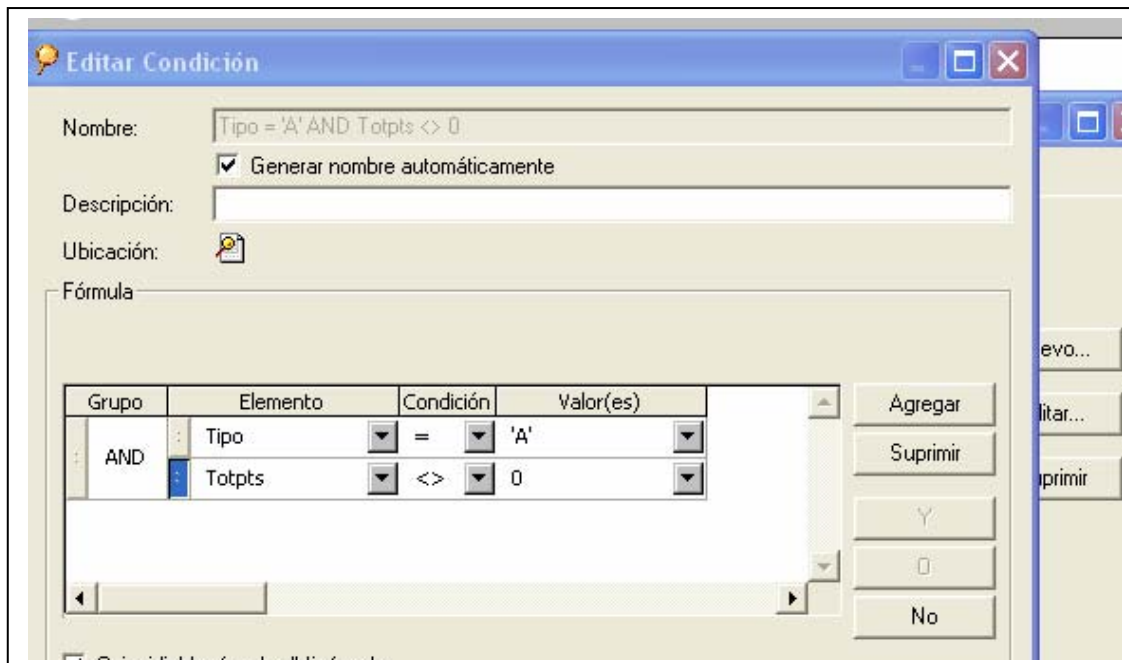
Dibujo21-Discoverer Desktop-Crear tabla

2. Escogemos la consulta que queremos realizar, en el caso del ejemplo sobre la THPLAYER escogeremos características del jugador.



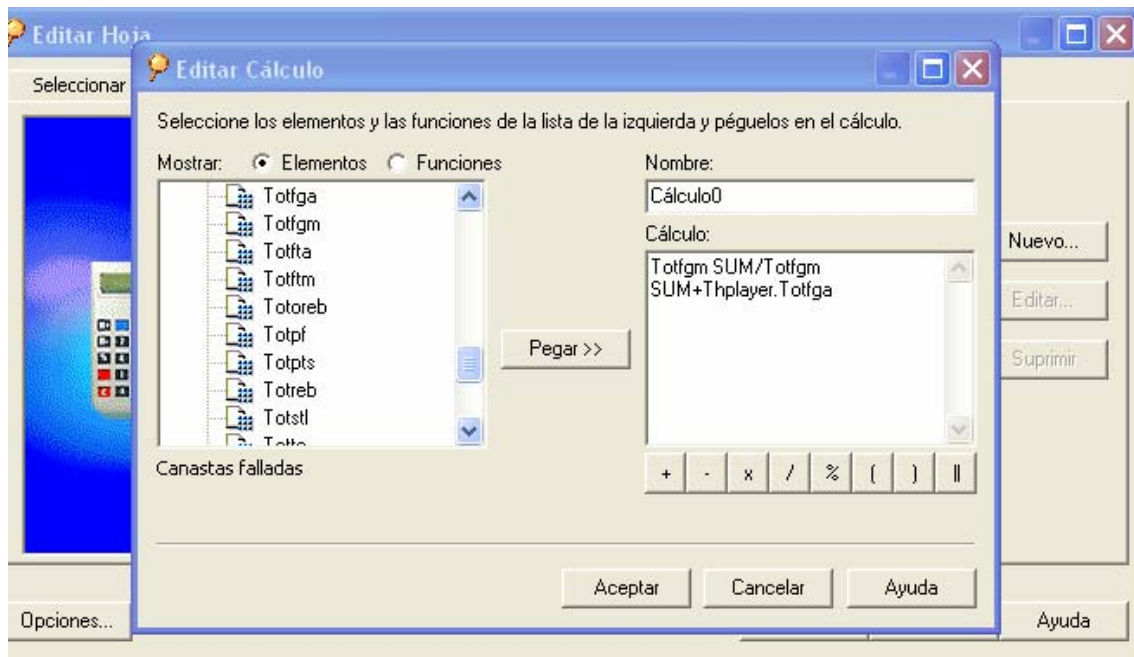
Dibujo 22- Discoverer Desktop -Escoger consulta

3. Creamos un filtro sobre la consulta.



Dibujo 23 Discoverer Desktop -Poner filtro

5 Añadimos una columna calculada, en este caso un ratio de aciertos.



Dibujo 24 Discoverer Desktop -Crear cálculo

6. Ejecutamos la consulta.

	Year	Firstname	Totpts SUM	Totfgm SUM	Totftm SUM	Cálculo1
▶ 1	1952	Bob,David	9	3	3	0,30
▶ 2	1968	Warren,D	6	3	0	0,27
▶ 3	1969	Dave,Deb	10	5	0	0,33
▶ 4	1971	Dave,Deb	8	4	0	0,33
▶ 5	1989	Clyde,Dre	7	2	2	0,25
▶ 6	1995	Clyde,Dre	11	5	0	0,38
▶ 7	2002	Tim,Dunc	19	8	3	0,31
▶ 8	2003	Tim,Dunc	14	6	2	0,35
▶ 9	1973	Jim,Eakin	2	1	0	0,20
▶ 10	1951	Thomas,E	2	1	0	0,25

Dibujo 25 Discoverer Desktop -Consulta creada

6. Posibles Mejoras o Ampliaciones

La parte que se podría mejorar del proyecto es todo lo relacionado a la exportación del ACCES y el hecho de tener la base de datos origen en este sistema.

El ACCESS es una herramienta muy sencilla, cualquier usuario puede usuario, pero nos presenta la dificultad que no hay ningún control de uso de la base de datos, y cuando se actualizan datos no hay ningún control de errores, esto produce que sea una plataforma inestable que no puede garantizar los datos.

Así que una solución importante pasaría por incorporar la base de datos origen a otro sistema con más control o intentar controlar el uso de la información de la NBA.

De todas maneras, como se ha podido comprobar en el proyecto, todos estos problemas de integridad de los datos quedan resueltos en la carga en Oracle y con el tratamiento que se hace de las constraints.

Otro punto a mejorar sería la gestión que se hace de los errores detectados en la carga; en la situación actual se crean unas tablas de errores con todos los registros que han dado problemas del tipo que las claves de las tablas no coinciden, estas tablas se cargan en el EUL y desde aquí las puede ver el usuario vía informes creados.

Para que esta gestión fuera más rápida, se podría incorporar en los procesos el enviar vía email a los usuarios que actualizan las tablas en MS ACCESS los errores detectados, para poder rectificarlos de forma rápida y así en el EUL no haría falta carga esta área de negocio.

Una posible ampliación pasaría por incorporar más información sobre la NBA al modelo de datos, o bien diseñar nuevos informes que necesiten de la creación de nuevas vistas. Como nuestro proceso se ha parametrizado, una nueva incorporación de tablas tendría una repercusión pequeña en el proceso ETL.

7. Conclusiones

Después de desarrollar este proyecto desde principio a fin, creo que he llegado a los objetivos marcados al principio e incluso he adquirido más conocimientos.

Por una parte el hecho de crear un almacén de datos no desde cero, si no, a partir de uno ya creado, ha permitido que más capacidad de análisis ya que nos hemos tenido que adaptar a lo que ya existía e intentar buscar la mejor solución para los usuarios.

Se ha pasado de tener 8 tablas a tener únicamente 4 , con lo cual a temas de mantenimiento y de optimización a la hora de hacer consulta se ha ganado con el almacén de Datos.

El hecho de hacer el proyecto en Oracle ofrece una gran libertad a la hora de crear un almacén de Datos, pero al mismo tiempo conserva la integridad y solidez de los datos.

Las herramientas utilizadas, son fáciles de usar tanto para el desarrollo de informes que hemos realizado, como para el usuario final; por tanto, la solución final mejora la capacidad de consulta y además de una manera muy sencilla.

8. Anexo

INDICE DE IMAGENES

Dibujo1-Diagrama de Gaant.....	7
Dibujo 2 –Modelo de Base de Datos Origen.....	9
Dibujo3 –Esquema general.....	13
<i>Dibujo4 –Modelo de Proceso.....</i>	<i>13</i>
Dibujo5 –Arquitectura del proyecto.....	17
Dibujo 6 –Pantalla OracleExpress 10g.....	19
Dibujo 7- Pantalla SQL Developer.....	20
<i>Dibujo8 – Modelo Estrella.....</i>	<i>23</i>
<i>Dibujo9-SQL DEVELOPER user CARGA.....</i>	<i>30</i>
Dibujo 10-Pantalla Administrator –EUL.....	37
Dibujo 11-Pantalla Jerarquías –EUL.....	37
Dibujo 12-Informe Entrenador.....	38
<i>Dibujo 13-Informe Jugador.....</i>	<i>39</i>
Dibujo 14-Rankig por Aciertos.....	39
Dibujo 15-Comparativa de Equipos.....	40
Dibujo 16-Ranking partidos ganados.....	40
Dibujo 17-Dream Team de la historia.....	41
<i>Dibujo 18-Dream Team por año.....</i>	<i>41</i>
Dibujo 19-Infomre Errores THPLAYER.....	41
Dibujo20-Administrator Palnificador.....	42
Dibujo21-Discoverer Desktop-Crear tabla.....	43
<i>Dibujo 22- Discoverer Desktop -Escoger consulta.....</i>	<i>44</i>
Dibujo 23 Discoverer Desktop -Poner filtro.....	44
Dibujo 24 Discoverer Desktop -Crear cálculo.....	45
Dibujo 25 Discoverer Desktop -Consulta creada	45

9. Bibliografía

- Apuntes de la UOC, sobre Almacén de Datos.
- **“Más allá del Business Intelligence”** (2005)
Luis mendez del Río , Ed. Gestión 2000
- <http://www.nba.com/espanol/>
- <http://www.sas.com>
- <http://www.oracle.com/index.html>
- <http://www.monografias.com/trabajos16/warehousing-discoverer/warehousing-discoverer.shtml>
- http://www.devjoker.com/asp/indice_contenido.aspx?co_grupo=PLSQ&as_categoria=2
- <http://www.programacion.com/tutorial.php?id=oracle>