

## Memòria del Projecte

---

# **CONSTRUCCIÓ I EXPLOTACIÓ D'UN MAGATZEM DE DADES SOBRE INFORMACIÓ ESTADÍSTICA DE L'NBA**

**MANEL JOFRA LECHA**

Estudis : Enginyeria Tècnica en Informàtica de Gestió

Consultor : ÀNGEL MARTIN CARBALLO

12.06.2007

Quan un comença un viatge, no sap ni quan ni com acaba. En aquest llarg viatge ha estat important el recolzament de la gent que m'envolta, els meus amics, la meua família, la Mei, la Mariona i en Miquel.

Malauradament, en un llarg viatge, sempre hi ha algú que ens deixa a mig camí. Però estic segur, que des de algun lloc del univers infinit segueix les nostres passes i ens ajuda a arribar a bon port.

Manel

*Quan per Palafrugell començà a circular la notícia que Hermós havia abandonat la vida que sol fer-se habitualment en aquest món i que s'havia anat a viure sol, a les remotes llunyanies d'Agua-Xellida, una gran quantitat de persones ho trobaren absolutament plausible.*

*I és que, en aquella època, hi havia, en la meua vila natal, com en tot el litoral del meu país, molta gent que tenia la il·lusió de la vida lliure...*

*... la concreció de la vida lliure era la platja de Tamariu...*

*Josep Pla*

*Cinc Històries del mar*

*Un viatge frustrat*

*Qui viu temeros, no serà mai lliure  
(Horaci)*

## RESUM

El projecte **Construcció i explotació d'un magatzem de dades sobre informació estadística de l'NBA** consisteix en el disseny i implementació d'un Magatzem de Dades (Datawarehouse) a partir de les dades proporcionades pel nostre client, el diari Sport World i que ens ha de permetre facilitar informació estadística de l'NBA.

El projecte comença amb una anàlisi acurada de les dades que se'ns han proporcionat, a fi i a efecte de definir les Dimensions i Atributs per una costat i els Indicadors per una altre costat. Arribats a aquest punt, el següent pas es definir els Informes que facilitarem al nostre client i d'acord a les seves necessitats. Finalitzarem aquesta part amb una repassada de les fonts de dades.

El següent pas es el disseny es el Disseny del Magatzem de Dades. En aquesta part començarem amb el disseny de les Dimensions i Atributs dels elements que han de conformar el magatzem de dades i les Taules de Fets amb on recollirem la informació estadística segons els Indicadors que hem definit. Definirem i continuarem amb l'Extracció, Transformació i Càrrega de les dades (**ETL**). Un cop arribats a aquest punt, es conforma el Disseny Final del Magatzem de dades (GALAXY SCHEMA).

Finalment creem els informes que han de permetre al nostre client obtenir les dades estadístiques de l'NBA. Sorprenent el resultat final de DREAM TEAM.

Pel desenvolupament d'aquest projecte s'han utilitat les eines ofimàtiques OFFICE 2003 de la firma MICROSOFT i com a sistema de bases de dades s'ha treballat amb ORACLE 10g EXPRESS EDITION i com eina de creació d'àrea de treball i informes a facilitar al nostre client la ORACLE DEVELOPER SUITE.

# INDEX DE CONTINGUTS

1.	Introducció	6
1.1.	Motivació	6
1.2.	Situació	6
1.3.	El Projecte	8
1.4.	Planificació del projecte	9
1.4.1.	Pla de treball i anàlisi de requeriments	9
1.4.2.	Disseny	9
1.4.3.	Implementació	10
1.4.4.	Lliurament final	10
1.4.5.	Debat Virtual	10
1.5.	Productes obtinguts	12
1.6.	Breu descripció dels altres capítols de la memòria	13
2.	Elements d'anàlisi	14
2.1.	Dimensions i atributs	14
2.1.1.	Players	14
2.1.2.	Time	14
2.1.3.	Team	14
2.1.4.	Coaches	15
2.2.	Indicadors	15
2.2.1.	Regular Season	15
2.2.2.	Playoffs	16
2.2.3.	All-Star Games	16
2.2.4.	Coaches Season	17
2.2.5.	Team Season	17
2.3.	Informes	18
2.3.1.	Regular Season	18
2.3.2.	Playoffs	19
2.3.3.	All-Star Games	20
2.3.4.	NBA Drafts	21
2.3.5.	Dream Team	21
2.3.6.	Conclusió	21
2.4.	Font de dades	21
2.4.1.	Players	22
2.4.2.	Teams	22
2.4.3.	Coaches Season	23
2.4.4.	Player Regular Season	23
2.4.5.	Player Playoffs	24
2.4.6.	Player Allstar	24
2.4.7.	Draft	25
2.4.8.	Team Season	26
3.	Disseny	27
3.1.	Introducció	27

3.2.	Dimensions i Atributs	28
3.2.1.	Dimensió Time	28
3.2.2.	Dimensió Players	28
3.2.3.	Dimensió Team	28
3.2.4.	Dimensió Coaches	28
3.3.	Taules de fets	29
3.3.1.	Regular Season	29
3.3.2.	Playoffs	30
3.3.3.	All-Star Games	31
3.3.4.	NBA Drafts	32
3.3.5.	Dream Team	32
3.3.6.	Conclusió	33
3.4.	ETL: Extracció, Transformació i Carrega de les dades	33
3.4.1.	Introducció	33
3.4.2.	Extracció	33
3.4.3.	Transformació	35
3.4.4.	Carrega	39
3.5.	Disseny Final	39
4.	Creació d'informes	41
4.1.	Introducció	41
4.2.	Dades jugador i any	42
4.3.	Dades equip i any	43
4.4.	Dades partits guanyats i perduts per equips i entrenadors	43
4.5.	Drafts	44
4.6.	Dream Team	45
5.	Conclusions	49
6.	Línies d'evolució futur.	50
7.	Glossari	51
8.	Bibliografia	52
9.	Annexos	54
9.1.	Importació de dades	54

## INDEX DE FIGURES

Figura 1 :	Esquema Datawarehouse	7
Figura 2 :	Diagrama de Gantt	11
Figura 3 :	Taula Players (en format MS Access)	22
Figura 4 :	Taula Teams (en format MS Access)	22
Figura 5 :	Taula Coaches_season (en format MS Access)	23
Figura 6 :	Taula Player_Regular_Season (en format MS Access)	23
Figura 7 :	Taula Player_Playoff (en format MS Access)	24
Figura 8 :	Taula Player_Allstar (en format MS Access)	25
Figura 9 :	Taula Draft (en format MS Access)	25
Figura 10 :	Taula Team_Season (en format MS Access)	26
Figura 11 :	Regular Season Star Schema	29
Figura 12 :	Playoffs Season Star Schema	30
Figura 13 :	AllStar Star Schema	31
Figura 14 :	Draft Star Schema	32
Figura 15 :	NBA Galaxy Schema	40
Figura 16 :	Oracle Discoverer Administrator Carpetes	41
Figura 17 :	Oracle Discoverer Desktop Player Worksheet	42
Figura 18 :	Oracle Discoverer Desktop Team Worksheet	43
Figura 19 :	Oracle Discoverer Desktop Coach Worksheet	44
Figura 20 :	Oracle Discoverer Desktop Draft Worksheet	45
Figura 21 :	Oracle Discoverer Desktop Dream Team Worksheet 1	46
Figura 22 :	Oracle Discoverer Desktop Dream Team Worksheet 2	47
Figura 23 :	Oracle Discoverer Desktop Dream Team Worksheet 3	47
Figura 24 :	Oracle Discoverer Desktop Dream Team Worksheet 4	48

# 1. INTRODUCCIÓ

## 1.1. Motivació

Al llarg dels estudis de la UOC m'he trobat amb assignatures que feien referència al treball amb dades, Bases de Dades I, Bases de Dades II, Minería de Dades, etc.

Per raons de feina, estic indirectament envoltat de dades : facturació, clients, proveïdors, remeses, bancs, etc.

Estem en un món envoltat de dades i que genera contínuament dades i més dades. Mires un partit de bàsquet o qualsevol altre esport i ets bombardejat amb dades. Si els minuts reals jugats, si la possessió de la pilota, si les faltes personals al tercer quart, etc.

Un cop feta l'assignatura de Tècniques de Desenvolupament del Programari a on es viu un projecte en grup, arriba l'hora de posar en marxa un projecte en solitari, a on tens al repte de portar-lo a bon terme. Si aquest projecte pot forma part "del meu dia a dia" mes que millor.

## 1.2. Situació

Però, que es un Magatzem de Dades?

La resposta sembla trivial. Un magatzem de dades, tal com indica el seu nom, es un magatzem de dades. Però lamentablement cauríem en un parany.

En qualsevol organització dels nostres dies, de l'empresa mes petita a la mes gran, el volum de dades que si mou es inimaginable. Els nivells d'audiència de les diferents cadenes televisives es troben a partir d'analitzar milers de dades.

Quan un equip de bàsquet fitxar un jugador, el primer que es te en compte, apart del seu estat físic, son les seves estadístiques: percentatges d'encert, rebots, faltes personals, etc. A més a més, els aficionats a un esport volen saber dades. Volen estar informats.

Ja fa un quants anys, uns personatges tal com Ralph Kimball i W.H. Immon entre altres, es varen donar compte que aquest volum de dades

que es generava podia ser una important font d'informació. Només es tractava de agafar-la, elaborar-la i treure els resultats. Sembla fàcil. Aviat en sortien els primers entrebancs. Dades generades en sistemes operatius diferents, dades en formats diferents, informació redundant, etc.

Tort aquella informació es tenia que treballar i, a mes a mes, es tenia que desenvolupar un programari específic. Així, en sortirà una nova manera de llegir les dades, analitzar les dades i treure conclusions. Ens trobem davant els Magatzem de Dades, una de eines de la BI (Business Intelligence).

Els components d'un Magatzem de Dades estan detallats a la imatge següent:

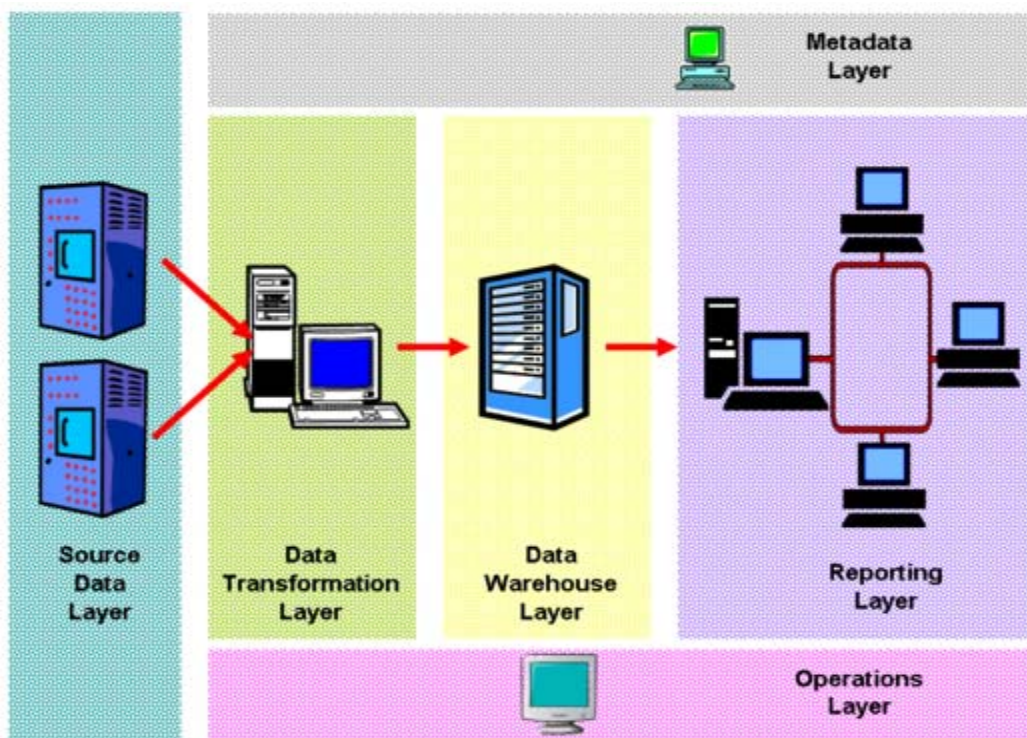


Figura 1 : Esquema Datawarehouse

I a on hi trobem les seves parts, des de les dades inicials, la transformació d'aquestes dades, la creació en si del magatzem de dades, i la posterior explotació i construcció d'informes a partir d'aquestes dades.

Un magatzem de dades o Datawarehouse es una col·lecció de dades, orientada a un domini, integrada, no volàtil, que varia en el temps i que



és de gran utilitat a l'hora de la presa de decisions de les empreses i de facilitar informació acurada a qui la demani.

En el cas que ens ocupa, por ser de molta utilitat per el Director General d'un club a l'hora de planificar els fitxatges per a la propera temporada, com pels seguidors, a l'hora de conèixer quines son les millors estadístiques dels seus admirats jugadors.

Els magatzems de dades solen tenir habitualment molta informació que es parteix en unitats lògiques mes petites.

Tal com comentat, no es tan fàcil. No es tracta només les dades que tenim dels diferents sistemes informàtics de l'empresa, sinó que ens veiem obligats a manipular aquest dades per que siguin coherents amb totes les dades del magatzem. Hem de normalitzar les dades. Per fer aquest procés, tenim unes eines per a la Extracció, Transformació i Carrega (Eines ETL) de les dades a la nostre magatzem.

Un magatzem de dades ha de remetre la informació correcte a la persona o persones indicades en el moment precís i en el format escaient. Els usuaris finals podran fer consultes al seu Magatzem de Dades sense el perill de provocar el funcionament incorrecte del sistema general.

### **1.3. El Projecte**

Tal com esmentat abans el nombre de dades que se genera i que disposem es molt alt. I d'aquestes dades en podem treure una informació molt valuosa.

En el cas a estudiar, disposem de les dades que ens ha facilitat el diari esportiu SPORT WORLD sobre les seves estadístiques de bàsquet de la NBA i que han estat creades i ampliades durant anys.

Des de SPORT WORLD se'ns demana el disseny, construcció i explotació d'un magatzem de dades a on figurin aquestes dades i a on els responsables de la secció NBA podran realitzar consultes fàcilment, proporcionar estadístiques, realitzar projeccions i prediccions de resultats, anàlisis des de els diferents punts de vista, etc.

Partirem de les dades generades, els analitzarem i procedirem a la implementació del magatzem de dades corporatiu, sempre tenint en compte les necessitats bàsiques dels responsables de la secció NBA del diari.

## **1.4. Planificació del Projecte**

### **1.4.1. Pla de treball i anàlisi de requeriments**

Període del 1 de març al 14 de març de 2007

- Distribució del temps entre les fites.
- Creació del document "Pla de treball".
- Creació del documents "Anàlisi de requeriments".

En aquesta fase es pren un primer contacte amb el projecte i es crea un pla de treball. L'objectiu principal del pla de treball és organitzar el temps per aconseguir complir amb les dates de presentació de les PACS, i que l'aplicació estigui a punt en la data del lliurament.

Per a portar a bon terme aquesta fita, es fa una recerca de la informació necessària per a comprendre que és un Magatzem de Dades.

A més a més es prepara el document "Anàlisi de Requeriments", a on es descriuen quins són els requeriments a analitzar, els seus elements, els informes que podran obtenir els usuaris del Magatzem de Dades i es fa una anàlisi de les dades proporcionades pel diari SPORT WORLD, ja que és a partir d'aquestes dades que es construirà el Magatzem de Dades.

### **1.4.2. Disseny**

Període del 15 de març al 16 d'abril de 2007

En aquesta fase es procedirà a preparar el model dimensional que donarà suport a les necessitats dels usuaris, segons l'anàlisi previ realitzat i el disseny dels procediments d'extracció de dades a alt nivell (processos, pseudocodi, etc.).

Serà en aquesta fase quan decidirem quin model dimensional es el mes apropiat pel nostre Magatzem de Dades i cercarem quins són els processos mes escaients per a la "Extracció, Transformació i Carrega" de les dades proporcionades pel diari SPORT WORLD al Magatzem de dades.

### **1.4.3. Implementació**

Període del 17 d'abril al 28 de maig de 2007

En aquesta fase es prenen totes les decisions que afecten a la implementació i que constarà de les següents tasques:

- Construcció del magatzem de dades corporatiu: base de dades, càrregues, etc.
- Instal·lació de l'eina de explotació de dades.
- Construcció dels informes.

Un cop treballats els conceptes de ELT procedim a la implementació del projecte amb Oracle Express com a base de dades per al magatzem de dades i Discoverer (també l'Oracle) com a eina d'explotació d'aquest magatzem de dades.

### **1.4.4. Lliurament Final**

Període del 29 de maig al 11 de juny de 2007

- Acabar la Memòria
- Fer la presentació del treball

En aquesta fase es farà la memòria amb els documents de l'especificació, el disseny i un petit manual d'usuari de l'aplicació.. Per acabar es farà una presentació del Treball en PowerPoint.

### **1.4.5. Debat Virtual Final**

Període del 25 de juny al 29 de juny de 2007

En aquesta fase el tribunal plantejarà preguntes i situacions a l'alumne per mitjà del correu electrònic. Aquests missatges s'hauran de contestar en un termini de 24 hores.

## Diagrama de Gantt

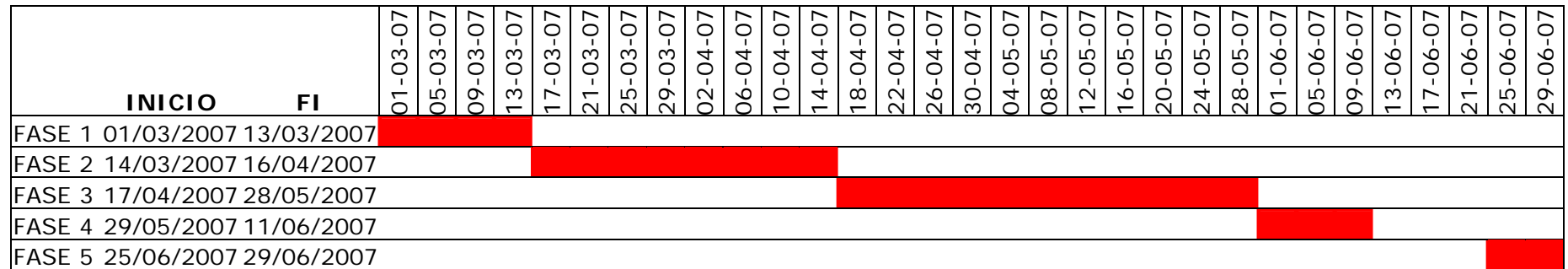


Figura 2 : Diagrama de Gantt

- Fase 1 : Pla de treball i anàlisi de requeriments
- Fase 2 : Disseny
- Fase 3 : Implementació
- Fase 4 : Lliurament final
- Fase 5 : Debat virtual final

## 1.5 Productes obtinguts

Seguidament es detallant els productes que s'han obtingut durant l'elaboració del projecte.

PLA DE TREBALL	Descripció inicial del projecte a on s'indiquen els objectius del mateix
ANALISI DE REQUERIMENTS	Descripció detallada dels components del projecte i del seu funcionament
DISSENY	Descripció del projecte amb els seus components.
PROTOTIP	Petit prototip efectuat amb MS Excel del funcionament del disseny
INFORMES	Informes obtinguts a partir del disseny final
PROCESSOS DE CARREGA	Document a on s'indica com instal·lar la base de dades i la EUL final.
MEMORIA	Aquest document
PRESENTACIO DEL PROJECTE	Presentació multimèdia a on es destacant els punts principals dels projecte
PRODUCTE FINAL	Export d'Oracle amb la Base de Dades i la EUL final del projecte

## 1.6 Breu descripció dels altres capítuls de la memòria.

A continuació es fa un petit resum del contingut de la resta de capítols de la memòria.

**Capítol 2: Elements d'anàlisi.** En aquest capítol es fa una repassada important a les dades que ens han proporcionat i de que ens permetran establir les dimensions, atributs i indicadors del nostre Magatzem de Dades i els informes que en traurem

**Capítol 3: Disseny.** En aquest capítol es defineixen les taules de fet i es procedeix a l'Extracció, Transformació i Càrrega de les Dades (ETL). Un cop efectuat aquest procés es defineix el Disseny Final (GALAXY SCHEMA)

**Capítol 4: Creació d'informes.** En aquest capítol es creen els informes que estaran a disposició del nostre client.

Els últims capítols presenten una conclusió final, les possibles línies de desenvolupament futur, per millorar i ampliació dels informes obtinguts, un glossari de termes utilitzats i la bibliografia i referències d'Internet que s'han fet servir per al desenvolupament d'aquest TFC.

## 2. ELEMENTS D'ANÀLISI

Els elements d'anàlisi són aquells elements del model conceptual que han de donar resposta a les necessitats dels responsables de la secció sobre l'NBA del diari esportiu SPORT WORLD.

Tal com ens demana el director d'informàtica del diari, haurem de treballar les dades que se'ns faciliten i proporcionar estadístiques sobre la Regular Season, els Playoffs i els All-Star Games. A més a més proporciona informació sobre els NBA Drafts i proposar un Dream Team segons diferents variables.

Un cop analitzades les dades proporcionades, passem a descriure les dimensions amb els corresponent atributs.

### 2.1. Dimensions i atributs

#### 2.1.1. Players

DIMENSIO : PLAYERS	
ATRIBUTS	Descripció
Illkid	Codi jugador
Firstname	Nom
Lastname	Cognom
Position	Posició
Firstseason	Primer any NBA
Lastseason	Ultim any NBA
High	Alçada
Weight	Pes
College	Collegi
Birthdate	Data naixement

#### 2.1.2. Time

DIMENSIO : TIME	
ATRIBUTS	Descripció
Year	Any

#### 2.1.3. Team

DIMENSIO : TEAM
-----------------

<b>ATRIBUTS</b>	<b>Descripció</b>
Team	Equip
Location	Població
Name	Nom
Leag	Lliga

### 2.1.4. Coaches

<b>DIMENSIO : COACHES</b>	
<b>ATRIBUTS</b>	<b>Descripció</b>
Coachid	Codi entrenador
Firstname	Nom
Lastname	Cognom

Un cop determinades les dimensions i els atributs passem a descriure els indicadors que ens permetran analitzar treure les estadístiques i informació que ens demanen.

## 2.2. Indicadors

### 2.2.1. Regular Season

<b>PLAYER_REGULAR_SEASON</b>	
<b>INDICADORS</b>	<b>Descripció</b>
Gp	Partits jugats
Minutes	Minuts
Pts	Punts
Oreb	Rebots ofensius
Dreb	Rebots defensius
Reb	Rebots totals
Asts	Assistències
Stl	Pilotes recuperades
Blk	Blocatges
Turnover	Pilotes perdudes
Pf	Faltes personals
Fga	Llançaments de 2 punts intentats
Fgm	Llançaments de 2 punts fets
Fta	Llançaments de 1 punt intentats
Ftm	Llançaments de 1 punt fets
Tpa	Llançaments de 3 punts intentats
Tpm	Llançaments de 3 punts fet



**2.2.2. Playoffs**

<b>PLAYER_PLAYOFFS</b>	
<b>INDICADORS</b>	<b>Descripció</b>
Gp	Partits jugats
Minutes	Minuts
Pts	Punts
Oreb	Rebots ofensius
Dreb	Rebots defensius
Reb	Rebots totals
Asts	Assistències
Stl	Pilotes recuperades
Blk	Blocatges
Turnover	Pilotes perdudes
Pf	Faltes personals
Fga	Llançaments de 2 punts intentats
Fgm	Llançaments de 2 punts fets
Fta	Llançaments de 1 punt intentats
Ftm	Llançaments de 1 punt fets
Tpa	Llançaments de 3 punts intentats
Tpm	Llançaments de 3 punts fet

**2.2.3. All-Star Games**

<b>DIMENSIO : PLAYER_ALLSTAR</b>	
<b>INDICADORS</b>	<b>Descripció</b>
Gp	Partits jugats
Minutes	Minuts
Pts	Punts
Oreb	Rebots ofensius
Dreb	Rebots defensius
Reb	Rebots totals
Asts	Assistències
Stl	Pilotes recuperades
Blk	Blocatges
Turnover	Pilotes perdudes
Pf	Faltes personals
Fga	Llançaments de 2 punts intentats
Fgm	Llançaments de 2 punts fets
Fta	Llançaments de 1 punt intentats
Ftm	Llançaments de 1 punt fets
Tpa	Llançaments de 3 punts intentats
Tpm	Llançaments de 3 punts fet

**2.2.4. Coaches Season**

<b>COACHES_SEASON</b>	
<b>INDICADORS</b>	<b>Descripció</b>
Season_win	Partits lliga guanyats
Season_loss	Partits lliga perduts
Playoff_win	Partits playoff guanyats
Playoff_loss	Partits playoff perduts

**2.2.5. Team Season**

<b>TEAM_SEASON</b>	
<b>INDICADORS</b>	<b>Descripció</b>
O_pts	Punts a favor
O_oreb	Rebots ofensius a favor
O_oreb	Rebots defensius a favor
O_reb	Rebots totals a favor
O_ast	Assistències a favor
O_stl	Pilotes recuperades a favor
O_blk	Blocatges a favor
O_to	Pilotes perdudes a favor
O_pf	Faltes personals a favor
O_fga	Llançaments de 2 punts a favor intentats
O_fgm	Llançaments de 2 punts fets a favor
O_fta	Llançaments de 1 punt intentats a favor
O_ftm	Llançaments de 1 punt fets a favor
O_tpa	Llançaments de 3 punts intentats a favor
O_tpm	Llançaments de 3 punts fet a favor
D_pts	Punts en contra
D_oreb	Rebots ofensius en contra
D_oreb	Rebots defensius en contra
D_reb	Rebots totals en contra
D_ast	Assistències en contra
D_stl	Pilotes recuperades en contra
D_blk	Blocatges en contra
D_to	Pilotes perdudes en contra
D_pf	Faltes personals en contra
D_fga	Llançaments de 2 punts en contra intentats
D_fgm	Llançaments de 2 punts fets en contra
D_fta	Llançaments de 1 punt intentats en contra
D_ftm	Llançaments de 1 punt fets en contra
D_tpa	Llançaments de 3 punts intentats en contra
D_tpm	Llançaments de 3 punts fet en contra

Won	Partits guanyat
Lost	Partits perduts

## 2.3. Informes

Els informes a realitzar segons el que ens demanen els departaments del diari SPORT WORLD són:

### 2.3.1. Regular Season

- Partits jugats per jugador i any
- Minuts jugats per jugador i any
- Punts per jugador i any
- Rebots per jugador i any
- Rebots ofensius per jugador i any
- Rebots defensius per jugador i any
- Assistències per jugador i any
- Recuperacions per jugador i any
- Blocatges per jugador i any
- Pilotes perdudes per jugador i any
- Faltes personals per jugador i any
- Llançaments de 3 punts per jugador i any
- Llançaments de 2 punts per jugador i any
- Llançaments de 1 punts per jugador i any
- Cistelles de 3 punts per jugador i any
- Cistelles de 2 punts per jugador i any
- Cistelles de 1 punts per jugador i any
- Percentatge d'encert de cistelles de 3 punts per jugador i any
- Percentatge d'encert de cistelles de 2 punts per jugador i any
- Percentatge d'encert de cistelles de 1 punts per jugador i any
  
- Punts per equip i any
- Rebots per equip i any
- Rebots ofensius per equip i any
- Rebots defensius per equip i any
- Assistències per equip i any
- Recuperacions per equip i any
- Blocatges per equip i any
- Pilotes perdudes per equip i any
- Faltes personals per equip i any

- Llançaments de 3 punts per equip i any
  - Llançaments de 2 punts per equip i any
  - Llançaments de 1 punts per equip i any
  - Cistelles de 3 punts per equip i any
  - Cistelles de 2 punts per equip i any
  - Cistelles de 1 punts per equip i any
  - Percentatge d'encert de cistelles de 3 punts per equip i any
  - Percentatge d'encert de cistelles de 2 punts per equip i any
  - Percentatge d'encert de cistelles de 1 punts per equip i any
- 
- Partits guanyats per entrenador i any
  - Partits perduts per entrenador i any

### 2.3.2. Playoffs

- Partits jugats per jugador i any
  - Minuts jugats per jugador i any
  - Punts per jugador i any
  - Rebots per jugador i any
  - Rebots ofensius per jugador i any
  - Rebots defensius per jugador i any
  - Assistències per jugador i any
  - Recuperacions per jugador i any
  - Blocatges per jugador i any
  - Pilotes perdudes per jugador i any
  - Faltes personals per jugador i any
  - Llançaments de 3 punts per jugador i any
  - Llançaments de 2 punts per jugador i any
  - Llançaments de 1 punts per jugador i any
  - Cistelles de 3 punts per jugador i any
  - Cistelles de 2 punts per jugador i any
  - Cistelles de 1 punts per jugador i any
  - Percentatge d'encert de cistelles de 3 punts per jugador i any
  - Percentatge d'encert de cistelles de 2 punts per jugador i any
  - Percentatge d'encert de cistelles de 1 punts per jugador i any
- 
- Punts per equip i any
  - Rebots per equip i any
  - Rebots ofensius per equip i any
  - Rebots defensius per equip i any
  - Assistències per equip i any
  - Recuperacions per equip i any

- Blocatges per equip i any
- Pilotes perdudes per equip i any
- Faltes personals per equip i any
- Llançaments de 3 punts per equip i any
- Llançaments de 2 punts per equip i any
- Llançaments de 1 punts per equip i any
- Cistelles de 3 punts per equip i any
- Cistelles de 2 punts per equip i any
- Cistelles de 1 punts per equip i any
- Percentatge d'encert de cistelles de 3 punts per equip i any
- Percentatge d'encert de cistelles de 2 punts per equip i any
- Percentatge d'encert de cistelles de 1 punts per equip i any
  
- Partits guanyats per entrenador i any
- Partits perduts per entrenador i any

### **2.3.3. All-Star Games**

- Partits jugats per jugador i any
- Minuts jugats per jugador i any
- Punts per jugador i any
- Rebots per jugador i any
- Rebots ofensius per jugador i any
- Rebots defensius per jugador i any
- Assistències per jugador i any
- Recuperacions per jugador i any
- Blocatges per jugador i any
- Pilotes perdudes per jugador i any
- Faltes personals per jugador i any
- Llançaments de 3 punts per jugador i any
- Llançaments de 2 punts per jugador i any
- Llançaments de 1 punts per jugador i any
- Cistelles de 3 punts per jugador i any
- Cistelles de 2 punts per jugador i any
- Cistelles de 1 punts per jugador i any
- Percentatge d'encert de cistelles de 3 punts per jugador i any
- Percentatge d'encert de cistelles de 2 punts per jugador i any
- Percentatge d'encert de cistelles de 1 punts per jugador i any
  
- Punts per conferència i any
- Rebots per conferència i any
- Rebots ofensius per conferència i any

- Rebots defensius per conferència i any
- Assistències per conferència i any
- Recuperacions per conferència i any
- Blocatges per conferència i any
- Pilotes perdudes per conferència i any
- Faltes personals per conferència i any
- Llançaments de 3 punts per conferència i any
- Llançaments de 2 punts per conferència i any
- Llançaments de 1 punts per conferència i any
- Cistelles de 3 punts per conferència i any
- Cistelles de 2 punts per conferència i any
- Cistelles de 1 punts per conferència i any
- Percentatge d'encert de cistelles de 3 punts per conferència i any
- Percentatge d'encert de cistelles de 2 punts per conferència i any
- Percentatge d'encert de cistelles de 1 punts per conferència i any

#### **2.3.4. NBA Drafts**

- Jugadors per selecció i ronda i any
- Origen per selecció i ronda i any

#### **2.3.5. Dream Team**

- Selecció per millors percentatges
- Selecció per millors puntuacions.
- Selecció combinada

#### **2.3.6. Conclusió**

El nombre final de informes estarà subjecte a la qualitat de les dades proporcionades. Per exemple, el nombre de rebots defensius i ofensius no van ésser comptabilitzats duran molts d'anys, eren rebots i prou.

### **2.4. Fonts de dades**

Les dades aportades per SPORT WORLD tenen la estructura i propietats detallades en els apartats següents.

### 2.4.1. Players




The screenshot shows the 'players' table structure in MS Access. The table has the following fields and data types:

Nombre del campo	Tipo de datos
ilkid	Texto
firstname	Texto
lastname	Texto
position	Texto
firstseason	Numérico
lastseason	Numérico
h_feet	Numérico
h_inches	Texto
weight	Numérico
college	Texto
birthdate	Fecha/Hora

Figura 3 : Taula Players (en format MS Access)

Ens aporta dades bàsiques sobre els jugadors. Les dades que conformen la clau primària son correctes. Aquesta clau primària (ilkid), també es troba a altres taules. Les dades aportades son força correctes.

### 2.4.2. Teams




The screenshot shows the 'Teams' table structure in MS Access. The table has the following fields and data types:

Nombre del campo	Tipo de datos
team	Texto
location	Texto
name	Texto
leag	Texto

Figura 4 : Taula Teams (en format MS Access)

Ens aporta dades bàsiques sobre els equips. Les dades que conformen la clau primària son correctes. En aquest cas, aquest, el camp "team" també es troba en altres taules. Manquen dades en alguns camps i hi ha equips repetits. A l'hora de efectuar la ETL, s'hauran d'efectuar les correccions oportunes.

### 2.4.3. Coaches Season




Coaches_season : Taula		
	Nombre del campo	Tipo de datos
🔑	coachid	Texto
🔑	year	Numérico
🔑	yr_order	Numérico
	firstname	Texto
	lastname	Texto
	season_win	Numérico
	season_loss	Numérico
	playoff_win	Numérico
	playoff_loss	Numérico
	team	Texto
		Season Wins
		Season Losses
		Playoff Wins
		Playoff Losses

Figura 5 : Taula Coaches\_season (en format MS Access)

Hi trobem les dades dels entrenadors i les derrotes i victòries dels equips que han estat entrenant.

D'aquestes dades en traurem la DIMENSIÓ COACH

### 2.4.4. Player Regular Season



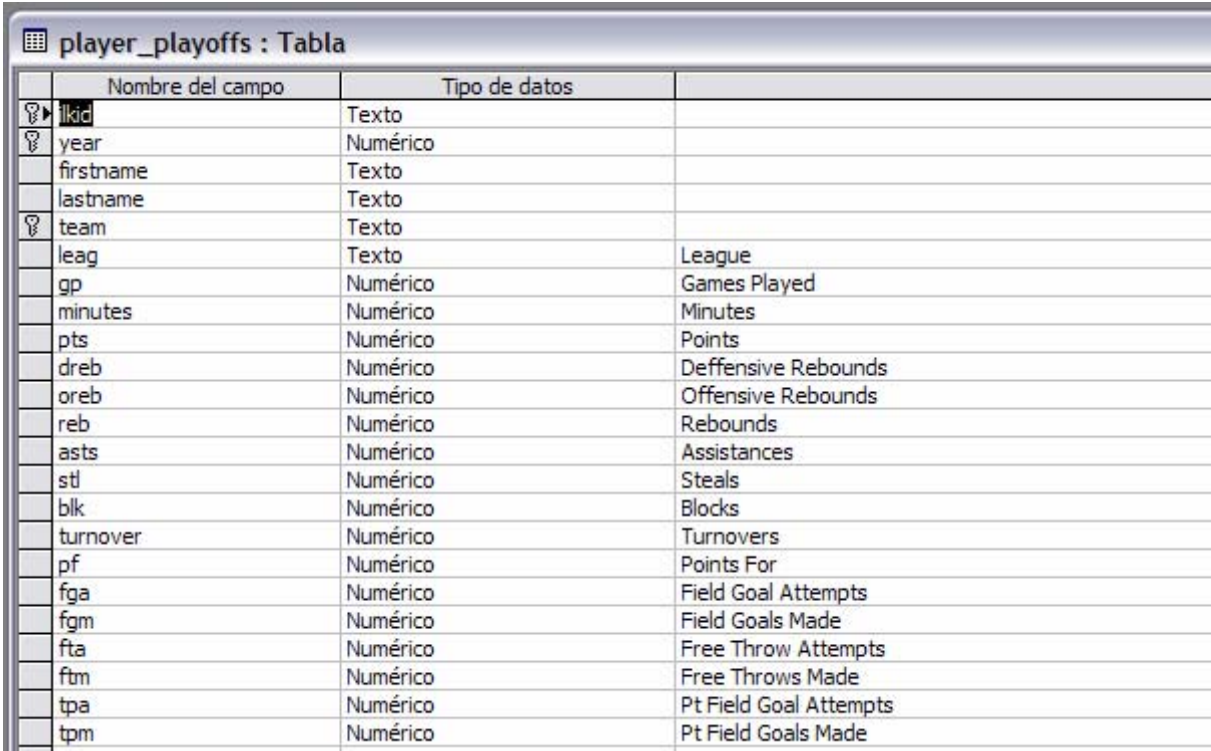
player_regular_season : Taula		
	Nombre del campo	Tipo de datos
🔑	ilkid	Texto
🔑	year	Numérico
	firstname	Texto
	lastname	Texto
🔑	team	Texto
	leag	Texto
	gp	Numérico
	minutes	Numérico
	pts	Numérico
	oreb	Numérico
	dreb	Numérico
	reb	Numérico
	asts	Numérico
	stl	Numérico
	blk	Numérico
	turnover	Numérico
	pf	Numérico
	fga	Numérico
	fgm	Numérico
	fta	Numérico
	ftm	Numérico
	tpa	Numérico
	tpm	Numérico
		League
		Games Played
		Minutes
		Points
		Offensive Rebounds
		Deffensive Rebounds
		Rebounds
		Assistances
		Steals
		Blocks
		Turnovers
		Points For
		Field Goal Attempts
		Field Goals Made
		Free Throw Attempts
		Free Throws Made
		Pt Field Goal Attempts
		Pt Field Goals Made

Figura 6: Taula Player\_Regular\_Season (en format MS Access)



Un dels fitxers que ens aporten més dades. S'hauran de comprovar que siguin correctes. Per exemple, "pf" no correspon a "points for", sinó a "personal faults". La suma total de punts tampoc es correcta. A l'hora de efectuar la ETL, s'hauran d'efectuar les correccions oportunes

### 2.4.5. Player Playoffs



	Nombre del campo	Tipo de datos	
🔍	lkid	Texto	
🔍	year	Numérico	
	firstname	Texto	
	lastname	Texto	
🔍	team	Texto	
	leag	Texto	League
	gp	Numérico	Games Played
	minutes	Numérico	Minutes
	pts	Numérico	Points
	dreb	Numérico	Deffensive Rebounds
	oreb	Numérico	Offensive Rebounds
	reb	Numérico	Rebounds
	asts	Numérico	Assistances
	stl	Numérico	Steals
	blk	Numérico	Blocks
	turnover	Numérico	Turnovers
	pf	Numérico	Points For
	fga	Numérico	Field Goal Attempts
	fgm	Numérico	Field Goals Made
	fta	Numérico	Free Throw Attempts
	ftm	Numérico	Free Throws Made
	tpa	Numérico	Pt Field Goal Attempts
	tpm	Numérico	Pt Field Goals Made

Figura 7 : Taula Player\_Playoff (en format MS Access)

Un altre dels fitxers que ens aporten més dades. S'hauran de comprovar que siguin correctes. Per exemple, "pf" no correspon a "points for", sinó a "personal faults". A l'hora de efectuar la ETL, s'hauran d'efectuar les correccions oportunes.

### 2.4.6. Player All-Star

Player_allstar : Taula		
	Nombre del campo	Tipo de datos
	ilkid	Texto
	year	Numérico
	firstname	Texto
	lastname	Texto
	conference	Texto
	leag	Texto
	gp	Numérico
	minutes	Numérico
	pts	Texto
	dreb	Texto
	oreb	Texto
	reb	Texto
	asts	Texto
	stl	Texto
	blk	Texto
	turnover	Texto
	pf	Texto
	fga	Texto
	fgm	Texto
	fta	Texto
	ftm	Texto
	tpa	Texto
	tpm	Texto

Figura 8 : Taula Player\_Allstar (en format MS Access)

Un altre dels fitxers que ens aporten més dades. Manquen dades en alguns camps. A l'hora de efectuar la ETL s'hauran de efectuar les correccions oportunes.

### 2.4.7. Draft

draft : Taula		
	Nombre del campo	Tipo de datos
	draft	Numérico
	draft_round	Numérico
	selection	Numérico
	team	Texto
	firstname	Texto
	lastname	Texto
	ilkid	Texto
	draft_from	Texto
	leag	Texto

Figura 9 : Taula Draft (en format MS Access)

S'haurà de comprovar la validesa de les dades. Es pot relacionar amb PLAYERS.

### 2.4.8. Team Season



	Nombre del campo	Tipo de datos
▶	team	Texto
?	year	Numérico
	leag	Texto
	o_fgm	Numérico
	o_fga	Numérico
	o_ftm	Numérico
	o_fta	Numérico
	o_oreb	Numérico
	o_dreb	Numérico
	o_reb	Numérico
	o_ast	Numérico
	o_pf	Numérico
	o_stl	Numérico
	o_to	Numérico
	o_blk	Numérico
	o_3pm	Numérico
	o_3pa	Numérico
	o_ots	Numérico

Figura 10 : Taula Team\_Season (en format MS Access)

Dades corresponents a equip i any. S'hauran de comprovar que siguin correctes. A l'hora de efectuar la ETL, s'hauran d'efectuar les correccions oportunes. La taula de dades més extensa. Ens servirà per contrastar dades amb altres fitxers.

## 3. DISSENY

### 3.1. Introducció

Un cop efectuada l'anàlisi de requeriments a partir de les dades que se'ns han proporcionat i un cop establertes les dimensions, els atributs i els indicadors, el següent pas és el de determinar quins són els models que s'ajusten més a la solució més adequada per als nostres clients.

Tenint en compte les dades proporcionades, definim les següents dimensions:

- . PLAYERS
- . TEAMS
- . COACHES
- . TIME

El pas següent es determina quins són els atributs que formaran part de cadascuna de les dimensions i sobre els que ens interessi obtenir informació.

El model és susceptible de ser ampliat. A part dels informes definits a l'anàlisi de requeriments, en un futur, ens poden interessar saber qui són els jugadors més alts, a quin equip juguen i quins anys. El ventall de possibles informes és molt ampli.

També tenim en compte que el que ens demana el nostre client SPORT WORLD són les dades estadístiques. En aquest disseny hi farem costar les dades que en aquest moment són susceptibles de ser utilitzades.

A més a més també tenim en compte el procés utilitzat a l'hora d'efectuar l'extracció, transformació i càrrega de les dades (ETL). Aquest procés serà fonamental a l'hora d'establir el disseny definitiu del magatzem de dades.

Inicialment disposem de les dimensions i atributs detallats següentment.

## 3.2. Dimensions i atributs

### 3.2.1. Dimensió Time

<b>DIMENSIO : TIME</b>	
<b>ATRIBUTS</b>	<b>Descripció</b>
Year	Any

### 3.2.2. Dimensió Players

<b>DIMENSIO : PLAYERS</b>	
<b>ATRIBUTS</b>	<b>Descripció</b>
Illkid	Codi jugador
Firstname	Nom
Lastname	Cognom
Position	Posició
Firstseason	Primer any NBA
Lastseason	Ultim any NBA
High	Alçada
Weight	Pes
College	Collegi
Birthdate	Data naixement

### 3.2.3. Dimensió Team

<b>DIMENSIO : TEAM</b>	
<b>ATRIBUTS</b>	<b>Descripció</b>
Team	Equip
Location	Població
Name	Nom
Leag	Lliga

### 3.2.4. Dimensió Coaches

<b>DIMENSIO : COACHES</b>	
<b>ATRIBUTS</b>	<b>Descripció</b>
Coachid	Codi entrenador
Firstname	Nom
Lastname	Cognom

## 3.3. Taules de Fets

### 3.3.1. Regular Season

Un cop analitzades les dades aportades per el nostre client SPORT WORLD, la taula de fets que ens permet proporcionar estadístiques sobre la REGULAR SEASON, ens queda:

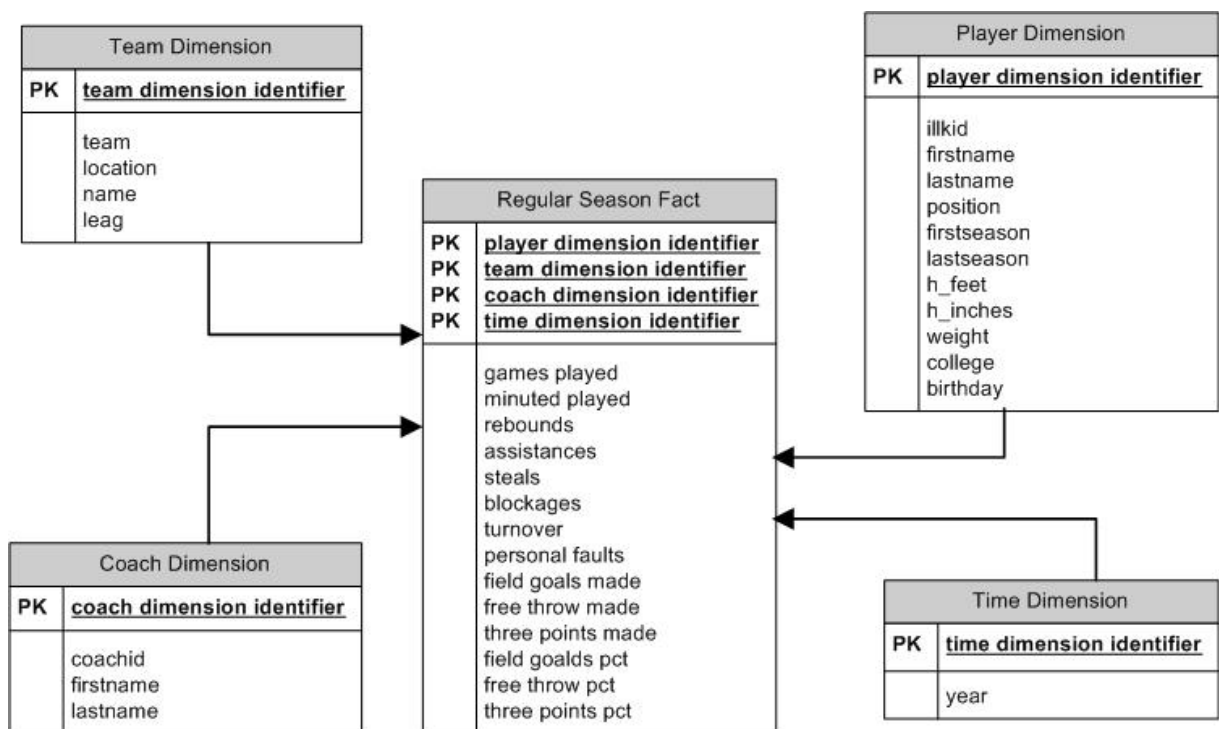


Figura 11 : Regular Season Star Schema

S'escull un esquema en forma de estrella (STAR SCHEME) per a la seva senzillesa a l'hora de formular les consultes.

### 3.3.2. Playoffs

Un cop analitzades les dades aportades per el nostre client SPORT WORLD, la taula de fets que ens permet proporcionar estadístiques sobre els PLAYOFFS, ens queda:

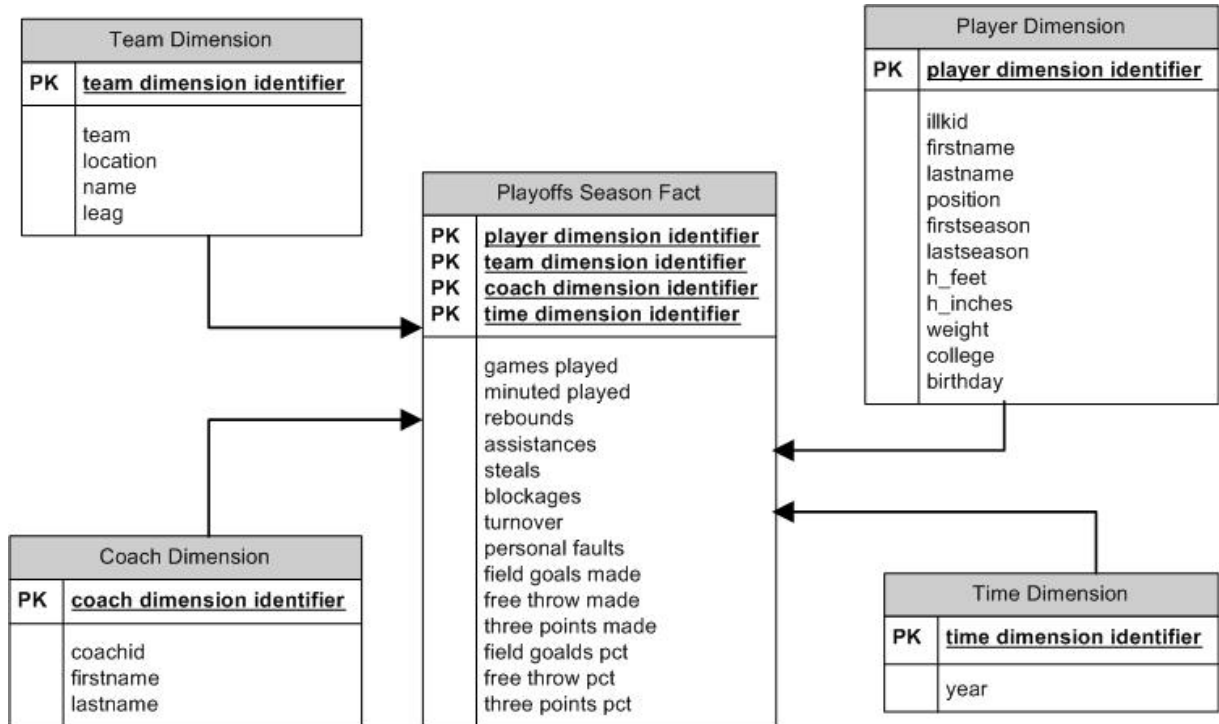


Figura 12 : Playoffs Season Star Schema

S'escull un esquema en forma de estrella (STAR SCHEME) per a la seva senzillesa a l'hora de formular les consultes.

### 3.3.3. All-Star Games

Un cop analitzades les dades aportades per el nostre client SPORT WORLD, la taula de fets que ens permet proporcionar estadístiques sobre els ALLSTAR GAMES, ens queda:

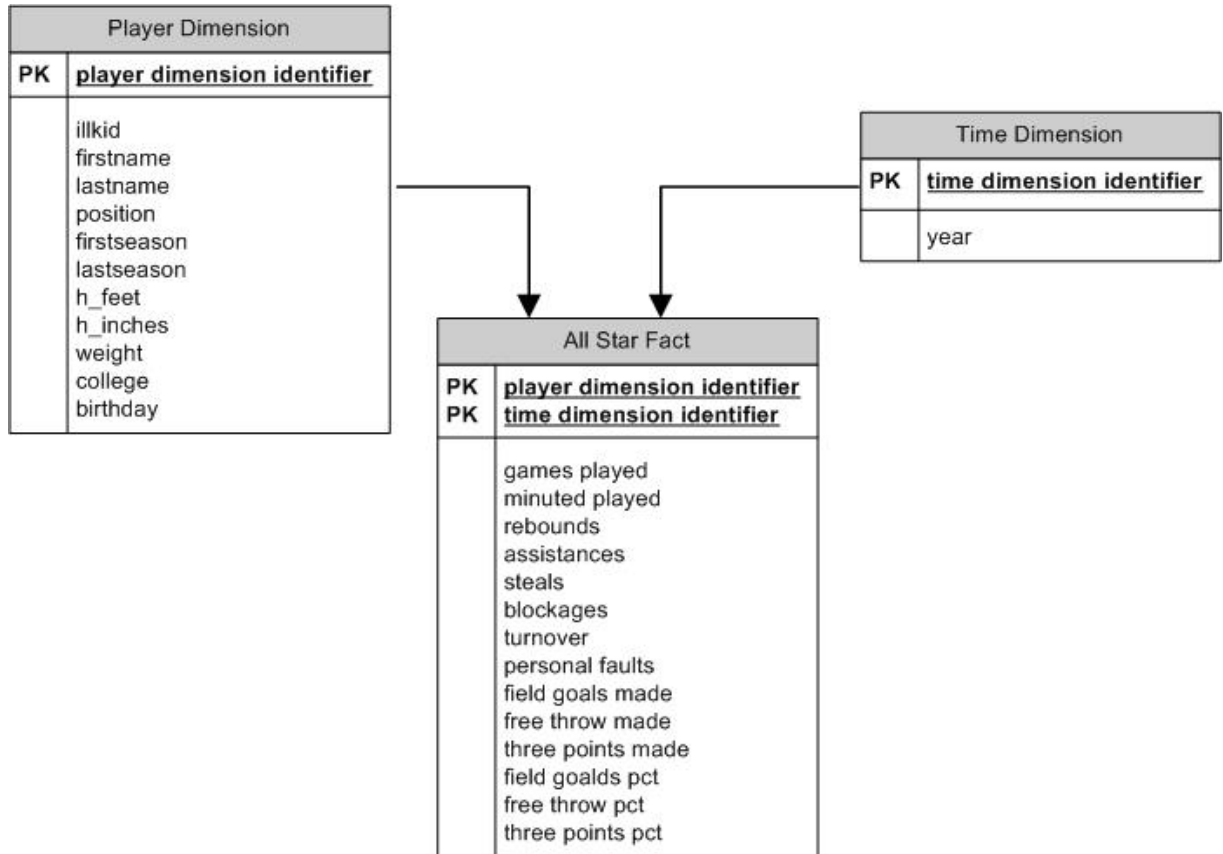


Figura 13 : AllStar Star Schema

S'escull un esquema en forma de estrella (STAR SCHEME) per a la seva senzillesa a l'hora de formular les consultes.



### 3.3.4. NBA Drafts

Un cop analitzades les dades aportades per el nostre client SPORT WORLD, la taula de fets que ens permet proporcionar estadístiques sobre els DRAFTS, ens queda:

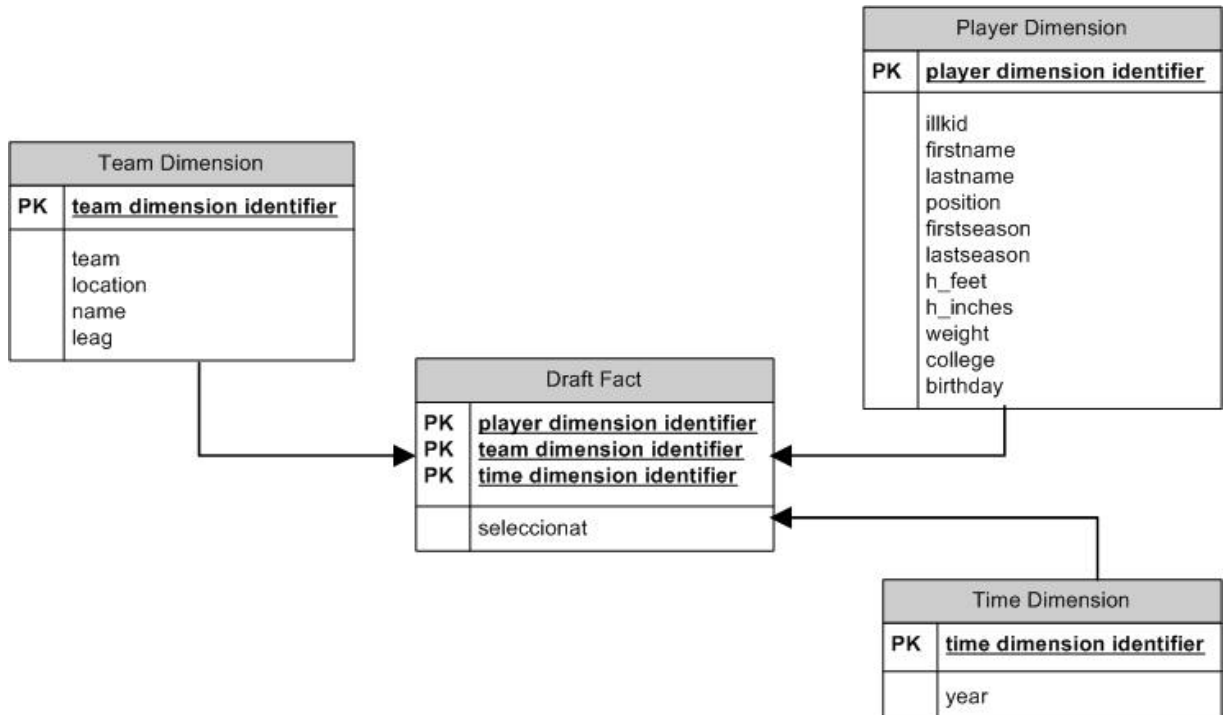


Figura 14 : Draft Star Schema

S'escull un esquema en forma de estrella (STAR SCHEME) per a la seva senzillesa a l'hora de formular les consultes.

### 3.3.5. Dream Team

En aquest cas, a l'hora d'escollir els nostres DREAM-TEAMS i d'acord a l'anàlisi de requeriments, treballaríem sobre les dades que obtindrem a la REGULAR SEASON FACT y els PLAY OFF FACT.

També buscarem les dades a partir de la posició que ocupen els jugadors en el camp :

- . G : GUARD > BASE
- . F : FORWARD > ESCOLTA
- . C : CENTER > PIVOT

i de cada un de les posicions hi aportarem :

G : 1 JUGADOR

F : 2 JUGADORS

C : 2 JUGADORS

### **3.3.6. Conclusió**

Com a tot projecte de DATAWAREHOUSE, el disseny del mateix esta subjecte a la retroalimentació del mateix, ja que van sortint noves possibilitats de obtenció d'estadístiques i informació, i a les que hem d'estar preparats per donar-hi resposta.

Tal com hem indicat el procés de ETL tindrà una importància cabdal a l'hora del disseny definitiu.

## **3.4. ETL : Extracció, Transformació i Carrega de les dades**

### **3.4.1. Introducció**

Una de les parts mes costoses en el disseny d'un magatzem de dades, tant en temps com en recursos, es l'extracció, transformació i càrrega de les dades.

En el cas d'estudi, les dades ens han estat donades en una base de dades en format ACCESS. Les passes que hem seguit ens han portat a convertir aquestes dades en llistes de dades en EXCEL, a on els hem manipulades i transformades. Un cop transformades, han estat carregades a l'Oracle XE mitjançant un import dels fitxers EXCEL en format .csv.

### **3.4.2. Extracció**

El diari Sport World ens ha facilitat una base de dades ACCESS amb 8 taules.

- PLAYERS
- TEAMS
- COACHES\_SEASON
- DRAFT
- TEAM\_SEASON

- PLAYER\_REGULAR\_SEASON
- PLAYER\_PLAYOFFS
- PLAYER\_ALLSTAR

D'aquestes 8 taules, tenim dues que conformen dues de les dimensions del magatzem de dades, son PLAYERS i TEAMS. A més a més, podem treure de la taula COACHES\_SEASON, una de les altres dimensions: COACHES.

El primer pas es obtenir la taula COACHES.

Es evident que les taules PLAYER\_REGULAR\_SEASON, PLAYER\_PLAYOFFS i PLAYER\_ALLSTAR son dades històriques dels jugadors i la taula TEAM\_SEASON son dades històriques dels equips.

El següent pas es comprovar si els jugadors i els equips que figuren a PLAYER\_REGULAR\_SEASON, PLAYER\_PLAYOFFS i PLAYER\_ALLSTAR, figuren a les taules PLAYERS i TEAMS.

S'entén també que els total de les dades de la taula PLAYER\_REGULAR\_SEASON en comparació amb les dades de la taula TEAM\_SEASON haurien de ser força similars ( s'utilitza el terme força en lloc de totalment iguals, ja que les diferències trobades son notables).

Es detecta una repetició de dades important. A la taula PLAYER\_REGULAR\_SEASON, ens trobem que les jugadors que han jugat a dos equips a la mateixa temporada, tenen les dades doblades. S'eliminen aquests registres.

Es detecta que no es disposen de les dades dels entrenadors dels partits dels PLAYOFF.

A més a més, es troben petites errades puntuals, com un equip inexistent, un mateix jugador amb dos noms, etc. .

Un cop eliminades i/o corregits certs registres, es preparen les dades que hauran de ser carregades al magatzem de dades.

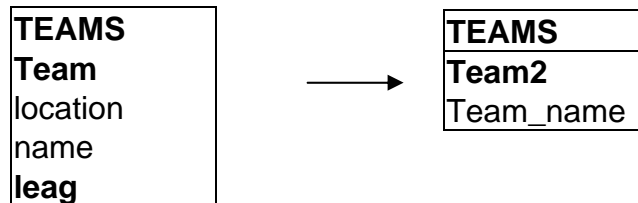
La manca de fiabilitat de les dades proporcionades, junt amb el fet de la mateixa història de l'NBA, lligues paral·leles, franquícies, etc, condicions força el disseny definitiu

### 3.4.3. Transformació

El següent pas consistirà a preparar les taules que conformen el nostre magatzem de dades.

#### TEAMS

La taula inicial queda transformada en :

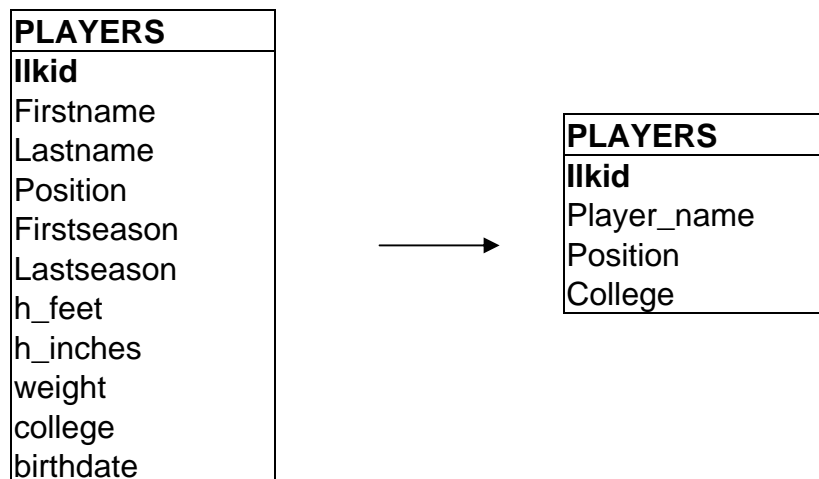


A on Team2 ens ve donat amb la concatenació dels camps 'leag&team' i team\_name amb la concatenació de 'location&" "&name'

Amb això traguem possibles repeticions i una sola clau primària

#### PLAYERS

La taula inicial venia donada per :



A on aconseguim el camp player\_name amb la concatenació de 'lastname &" "&firstname'.

Amb això traguem possibles repeticions i tenim una sola clau primària.

Guardem el camp Position, que ens permetrà confeccionar el nostre equip ideal, segons la posició del jugador.

## COACH

L'obtenim de la taula COACHES\_SEASON i ens queda

<b>COACHES</b>
<b>Coach_id</b>
Coach_name

Tal com esmentat anteriorment, no disposem dels entrenadors de la lliga ABA, lliga que va coexistir alguns anys amb la NBA. Inserim el registre "ZZZZZZZZ, Not available". També manquen les dades del any 2004.

## TIME

Creem la taula TIME, amb el disseny :

<b>TIME</b>
<b>Year</b>

I que conforma una de les dimensions principals.

## TAULES DE FETS

El següent pas es preparar les taules que conformaran les taules de fets.

Per un costat tenim que les taules PLAYER\_REGULAR\_SEASON, PLAYER\_PLAYOFF i PLAYER\_ALLSTAR tenen la mateixa estructura. Els unim i hi afegim un camp (ROL), a on indiquem si aquest registre correspon a la REGULAR SEASON (R), PLAYOFF (P) o ALLSTAR (A).

Inserim a la taula TEAMS els 2 registres :

Team2	Team_name
East	East Conference
West	West Conference

Un cop efectuades les transformacions esmentades ens queda la taula PLAYER\_TOTAL :

<b>PLAYER_TOTAL</b>
Ilkid
Year
Rol
Team2
Gp
Minutes
Pts
Dreb
Oreb
Reb
Asts
Stl
Blk
Turnover
Pf
Fga
Fgm
Fta
Ftm
Tpa
Tpm

I per un altre costat preparam la taula COACHES\_TOTAL amb l'estructura:

<b>COACHES_TOTAL</b>
Coached
Year
yr_order
Win
Loss
Rol
Team2

Si tenim en compte que un equip te un sol entrenador per equip, ens permet tenir les dades de partits guanyats i perduts tant per entrenador com equip.

## **DRAFT**

Hi guardem les dades a on el id del jugador també hi consta a la taula PLAYERS.

## **CONCLUSIÓ**

Un cop efectuades l'extracció, manipulació i carrega de les dades (ETL), ens queden les taules :

- PLAYERS
- TEAMS
- COACHES
- TIME
- PLAYER\_TOTAL
- COACH\_TOTAL
- DRAFT

I amb les que poblarem el magatzem de dades

### **3.4.4. Càrrega**

Les taules abans definides son guardades en format .csv, quedant els fitxers :

PLAYERS.csv  
TEAMS.csv  
COACHES.csv  
TIME.csv  
PLAYER\_TOTAL.csv  
COACH\_TOTAL.csv  
DRAFT.csv

Mitjançant t el recurs Import d'Oracle XE, es procedeix a la càrrega de les dades al magatzem de dades, sota l'esquema i usuari NBA de l'ORACLE XE.

## **3.5. Disseny Final**

Un cop arribats a aquest punt, ens manca redefinir les taules de fets i que son :

PLAYERS\_FACT  
COACH\_TEAM\_FACT  
FRACT\_FACT

I amb les dimensions ja establertes abans de PLAYERS, TEAMS, COACHES i TIME.

El esquema escollit passa a ser de "GALAXY SCHEMA"



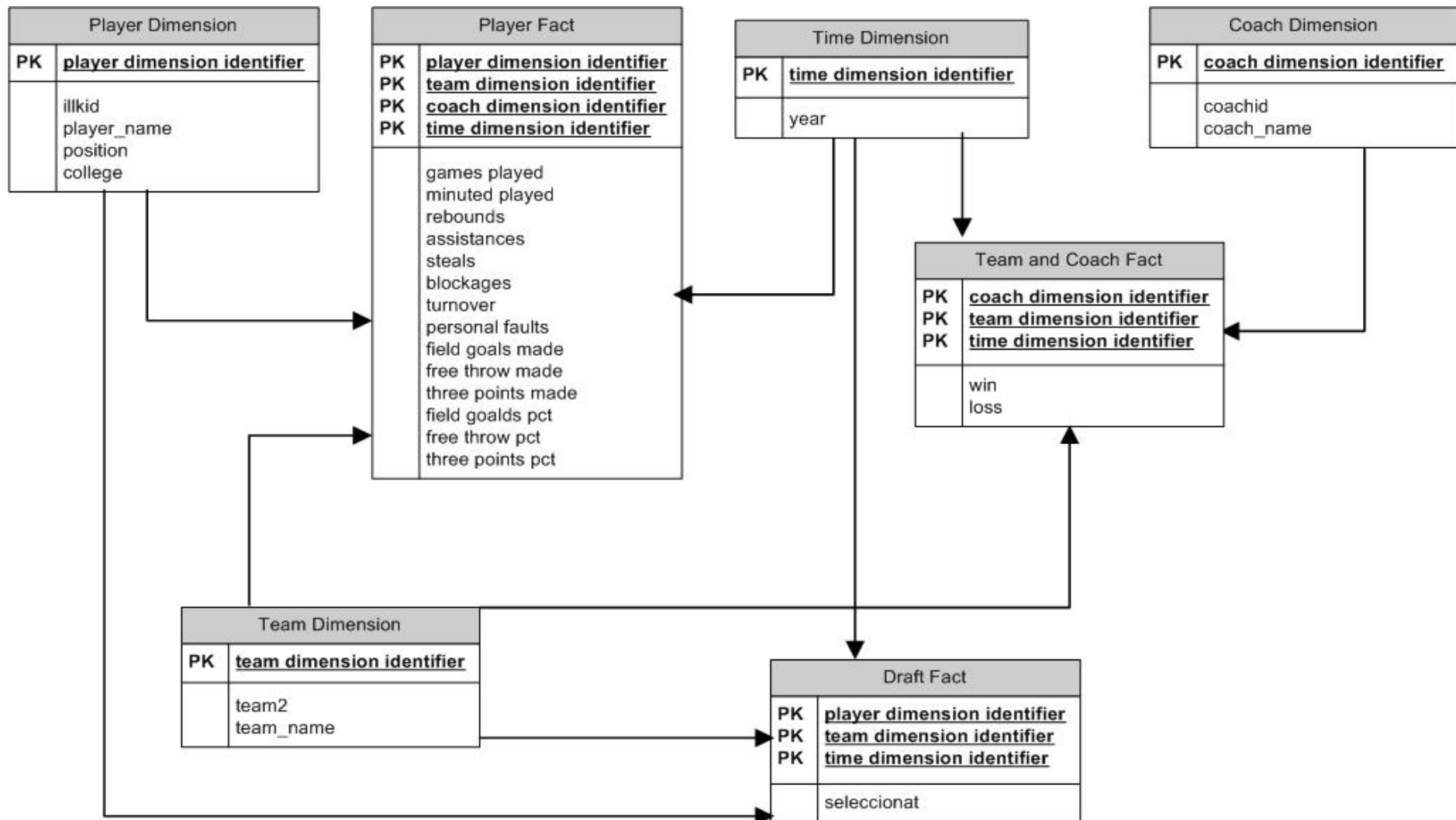


Figura 15 : NBA Galaxy Schema

## 4. CREACIO DE INFORMES

### 4.1. Introducció

Un cop posat en marxa el ORACLE XE i el ORACLE DISCOVERER, el que s'ha fet ha estat crear un compte d'usuari a on importar les dades ja transformades.

Amb referència al ORACLE DISCOVERER, utilitzant el ORACLE ADMINISTRATOR es crea una EUL (End User Layer). Un cop creada la EUL, ens permet treballar amb al ORACLE DESKTOP i preparar els informes demanats.

A dintra la EUL es munta una Business Area, que porta el nom de NBA.

Dintre d'aquesta Business Area de nom NBA, es monten 3 carpetes complexes i es a on, des de el DISCOVERER DESKTOP es tindrà accés a les dades estadístiques de l'NBA.

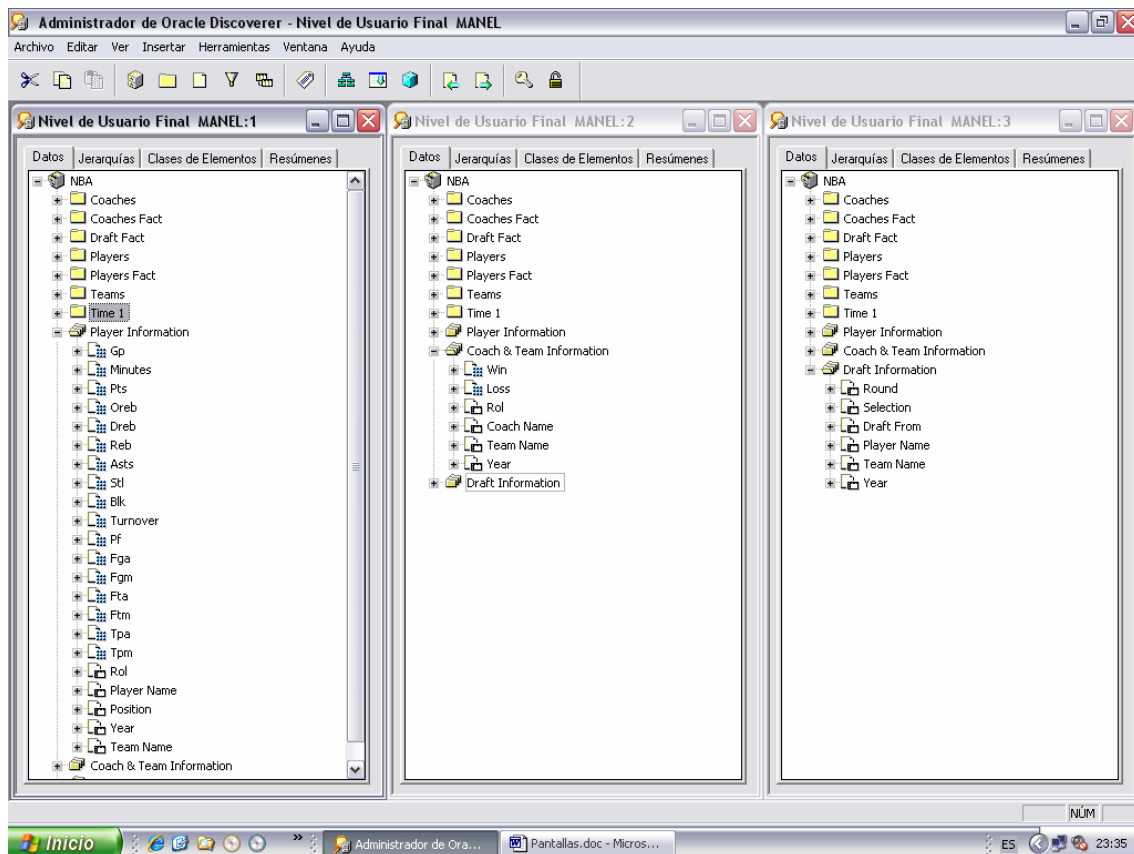


Figura 16 : Oracle Discoverer Administrator Carpetes

Un cop analitzades les dades, extretes, transformades i carregades al nostre magatzem de dades, ha arribat l'hora de generar els informes que ens ha demanat el diari Sport World.

## 4.2. Dades Jugador i Any

En aquest cas treballem amb les dades de la carpeta complexa PLAYER INFORMATION.

Edición de Escritorio de Oracle Discoverer - [Libro de Trabajo2]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 8 B U

Elementos de Página: Player Name: Bryant Kobe Rol: R

	G Los Angeles Lakers								
	1996	1997	1998	1999	2000	2001	2002	2003	2004
Gp SUM	71	79	50	66	68	80	82	65	66
Minutes SUM	1103	2056	1896	2524	2783	3064	3401	2447	2690
Pts SUM	539	1220	996	1485	1938	2019	2461	1557	1819
Oreb SUM	47	79	53	108	104	112	106	103	95
Dreb SUM	85	163	211	308	295	329	458	256	297
Reb SUM	132	242	264	416	399	441	564	359	392
Asts SUM	91	199	190	323	338	438	481	330	398
SH SUM	49	74	72	106	114	118	181	112	86
Blk SUM	23	39	50	62	43	35	67	28	53
Turnover SUM	112	157	157	182	220	223	288	171	270
Pf SUM	102	180	153	220	222	228	218	176	175
Fga SUM	422	913	779	1183	1510	1597	1924	1178	1324
Fgm SUM	176	391	362	554	701	749	868	516	573
Fta SUM	166	457	292	403	557	589	713	533	664
Ftm SUM	136	363	245	331	475	488	601	454	542
Tpa SUM	136	220	101	144	200	132	324	217	387
Tpm SUM	51	75	27	46	61	33	124	71	131
PCT 1 POINT	0,82	0,79	0,84	0,82	0,85	0,83	0,84	0,85	0,82
PCT 2 POINTS	0,42	0,43	0,46	0,47	0,46	0,47	0,45	0,44	0,43
PCT 3 PONTS	0,38	0,34	0,27	0,32	0,31	0,25	0,38	0,33	0,34

Para obtener ayuda, pulse F1

Inicio Pantallas.doc - Micro... Documento2 - Micro... fin2.doc - Microsoft... Edición de Escritorio... ES 00:00

Figura 17 : Oracle Discoverer Desktop Player Worksheet

Aquí hi trobem les dades corresponents al nombre de punts, partits jugats, rebots, etc; i percentatges d'encert en els llançaments.

### 4.3. Dades Equip i Any

En aquest cas treballarem amb les dades de la carpeta complexa PLAYER INFORMATION.

Edición de Escritorio de Oracle Discoverer - [Libro de Trabajo2]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Tahoma 8 B z U

Elementos de Página: Team Name: Memphis Grizzlies Rol: R

	2001	2002	2003	2004
Gp SUM	752	811	844	830
Minutes SUM	19781	19930	19880	19705
Pts SUM	7372	7995	7930	7661
Oreb SUM	930	943	1047	887
Dreb SUM	2468	2466	2381	2307
Reb SUM	3398	3409	3428	3194
Asts SUM	1785	1892	1915	1716
Stl SUM	646	651	795	699
Blk SUM	488	500	565	476
Turnover SUM	1290	1214	1180	1145
Pf SUM	1571	1678	1903	1910
Fga SUM	6535	6743	6657	6271
Fgm SUM	2851	3049	2963	2802
Fta SUM	1883	1936	2143	2024
Ftm SUM	1334	1430	1557	1526
Tpa SUM	1096	1279	1314	1486
Tpm SUM	336	467	447	531
PCT 1 POINT	0,71	0,74	0,73	0,75
PCT 2 POINTS	0,44	0,45	0,45	0,45
PCT 3 POINTS	0,31	0,37	0,34	0,36

Hoja 1

Inicio Pantallas.doc - Micro... Documento2 - Micro... fin2.doc - Microsoft... Edición de Escritorio... ES 00:03

Figura 18 : Oracle Discoverer Desktop Team Worksheet

Aquí hi trobarem les dades corresponents al nombre de punts, partits jugats, rebots, etc; i percentatges d'encert en els llançaments.

### 4.4. Dades partits guanyats i perdut per equips i entrenadors.

En aquest cas treballarem amb les dades de la carpeta complexa COACH & TEAM INFORMATION.

Edición de Escritorio de Oracle Discoverer - [Libro de Trabajo4]

Archivo Editar Ver Hoja Formatear Herramientas Gráfico Ventana Ayuda

Min Max % + - x ÷ = ≠ < ≤ > ≥

Tahoma 8 B U

Elementos de Página: Coach Name: Levane Andrew Rol: R

	Milwaukee Hawks		New York Knicks		St. Louis Hawks	
	Win SUM	Loss SUM	Win SUM	Loss SUM	Win SUM	Loss SUM
1952	27	44				
1953	11	35				
1958			40	32		
1959			8	19		
1961					20	40

Hoja 1

Inicio Pantallas.doc - Micro... Documento2 - Micro... fin2.doc - Microsoft ... Edición de Escritorio... ES 00:22

Figura 19 : Oracle Discoverer Desktop Coach Worksheet

Aquí hi trobarem les dades corresponents al nombre de partits guanyats per equips.

## 4.5. Drafts

En aquest cas treballarem amb les dades de la carpeta complexa DRAFT INFORMATION.

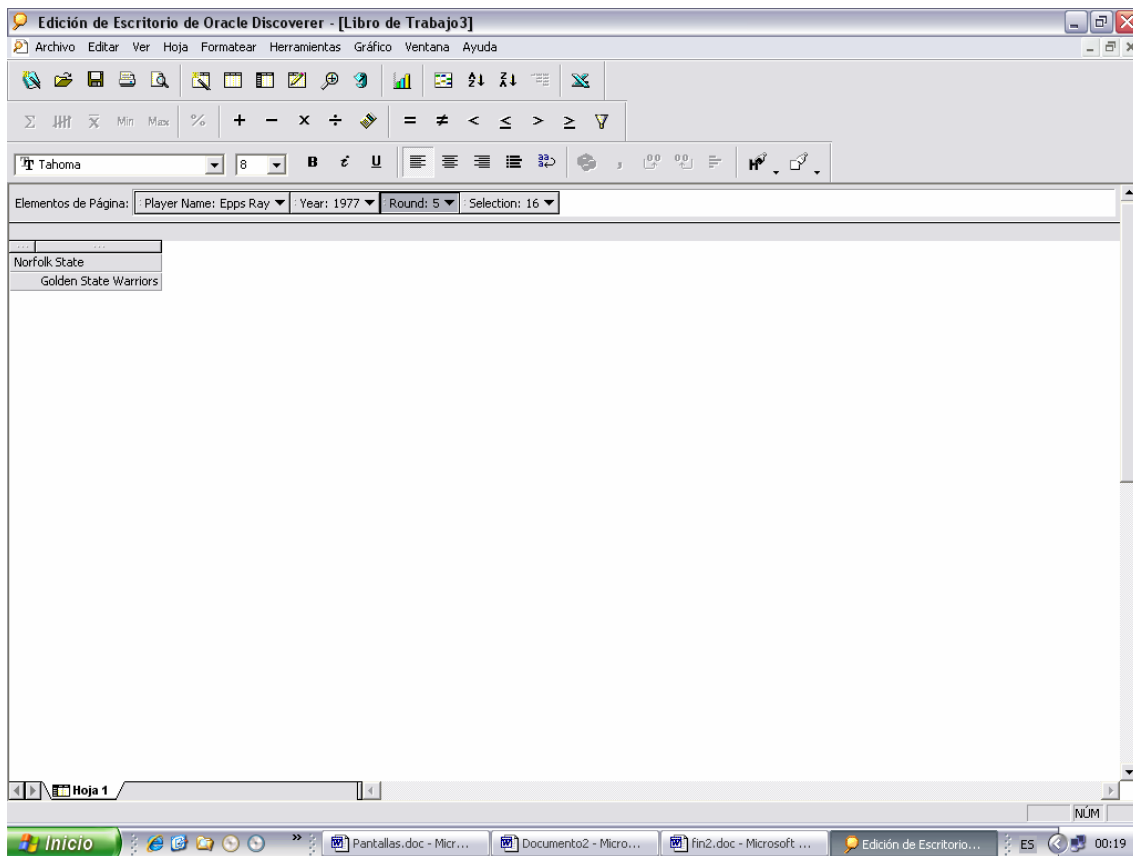


Figura 20 : Oracle Discoverer Desktop Draft Worksheet

En Ray Epps quan va ser seleccionat i de quina universitat.

## 4.6. Dream Team

En aquest cas treballarem amb les dades de la carpeta complexa PLAYER INFORMATION.

Primer mostrem 3 pantalles a on trobem informació sobre els jugadors que han fet més punts, segons la seva posició.

D'aquí en traurem un primer equip.

El següent pas mostrarem una pantalla a un traurem el criteri de la posició. Escollirem els 5 jugadors que han fet mes punt a la història de l'NBA.

	Pts SUM	PCT 2 POINTS
Malone Karl	41834	0,51
Erving Julius	34919	0,50
Havlicek John	30350	0,44
Hayes Elvin	29626	0,45
Barry Rick	28281	0,45
Wilkins Dominique	28197	0,46
English Alex	27343	0,51
Baylor Elgin	26990	0,43
Barkley Charles	26703	0,54
Bird Larry	25805	0,49
Dantley Adrian	24794	0,54
Pettit Bob	23344	0,43
Pippen Scottie	22667	0,47
Chambers Tom	21788	0,47
Cummings Terry	21150	0,49
Walker Chet	20804	0,47
Mchale Kevin	20578	0,55
Schayes Dolph	20463	0,38
King Bernard	20381	0,52
Johnson Eddie	20370	0,47
Robinson Clifford	20262	0,44
Aguirre Mark	20241	0,48
Worthy James	19416	0,52
Hudson Lou	19306	0,49
Rice Glen	19269	0,45
Mcginnis George	19218	0,45
Howell Bailey	19203	0,48
Mullin Chris	18942	0,51

Figura 21 : Oracle Discoverer Desktop Dream Team Worksheet 1

	Pts SUM	PCT 2 POINTS
Abdul-jabbar Kareem	44413	0,56
Chamberlain Wilt	35217	0,54
Malone Moses	31923	0,49
Olajuwon Hakeem	30818	0,51
Issel Dan	30518	0,50
O'neal Shaquille	28251	0,58
Ewing Patrick	27734	0,50
Gilmore Artis	26831	0,58
Parish Robert	26235	0,53
Robinson David	23153	0,51
Bellamy Walt	21839	0,51
Mcadoo Bob	20593	0,50
Lanier Bob	20562	0,52
Sikma Jack	18799	0,46
Russell Bill	17315	0,44
Beaty Zelmo	17302	0,49
Edwards James	15892	0,49
Thurmond Nate	15434	0,42
Cowens Dave	15276	0,46
Laimbeer Bill	15172	0,50
Adams Alvan	14990	0,50
Divac Vlade	14874	0,49
Paultz Billy	14581	0,49
Smits Rik	14418	0,51
Mourning Alonzo	14218	0,52
Thompson Mychal	13942	0,50
Cartwright Bill	13821	0,52
Daniels Mel	13811	0,47

Figura 22 : Oracle Discoverer Desktop Dream Team Worksheet 2

	Pts SUM	PCT 2 POINTS
Jordan Michael	38541	0,50
West Jerry	29809	0,47
Gervin George	28989	0,50
Robertson Oscar	28856	0,48
Miller Reggie	28291	0,47
Drexler Clyde	25254	0,47
Greer Hal	23592	0,45
Payton Gary	22936	0,47
Stockton John	22228	0,51
Johnson Magic	21582	0,52
Thomas Isaiah	21256	0,45
Richmond Mitch	21003	0,46
Davis Walter	20938	0,51
Goodrich Gail	20664	0,45
Ellis Dale	20040	0,48
Theus Reggie	19275	0,47
Cousy Bob	19125	0,37
Monroe Earl	18965	0,46
Murphy Calvin	18900	0,48
Wilkins Lenny	18888	0,43
Bing Dave	18845	0,44
Boone Ron	18820	0,46
Blackman Rolando	18804	0,49
Iverson Allen	18731	0,42
Johnson Dennis	18728	0,44
Free World	18446	0,45
Jones Sam	18361	0,45
Malone Jeff	18199	0,48

Figura 23 : Oracle Discoverer Desktop Dream Team Worksheet 3



La millor defensa es el millor atac. Qui fa més punts probablement serà qui guany el partit.

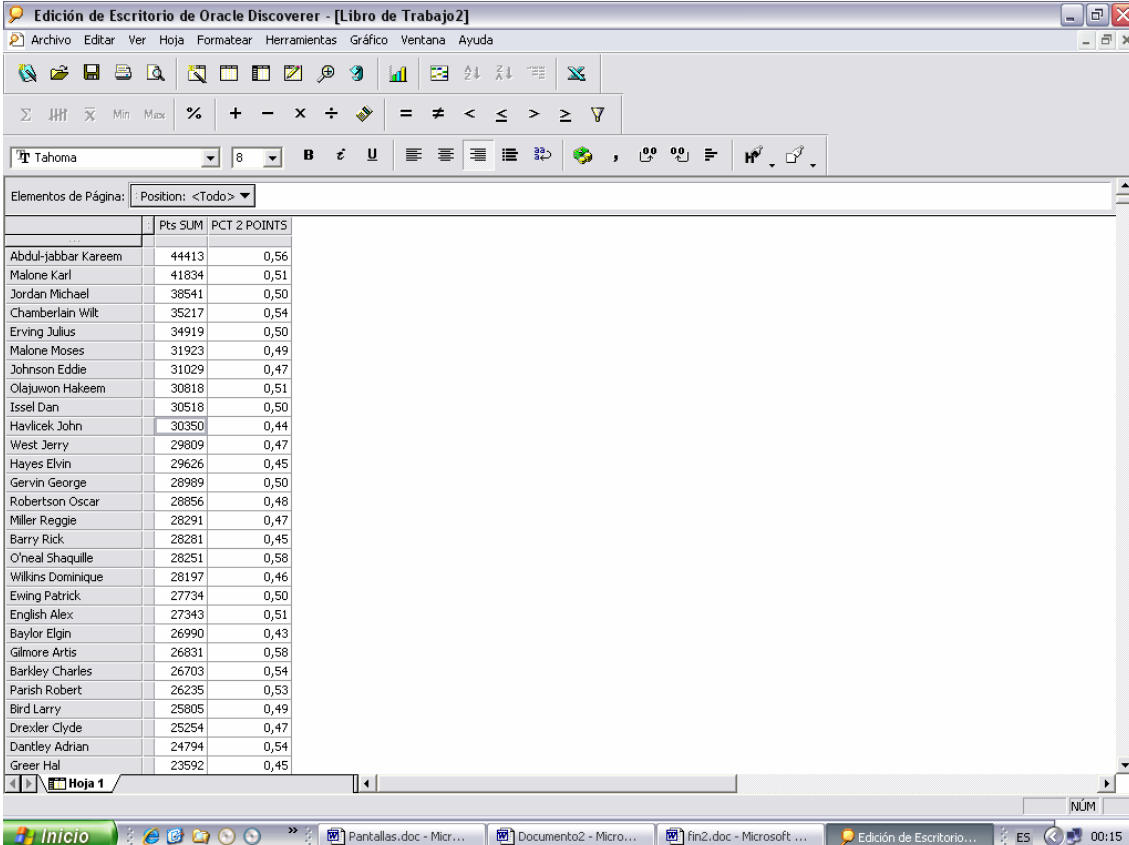
Seleccionem els que han fet met punts, però segons el lloc : 1 base, 2 pivots i 2 escoltes

BASE : MICHAEL JORDAN

ESCOLTES : KARL MALONE – JULIUS ERVING

PIVOTS : KARIM ABDUL-JABAAR – WILT CHAMBERLAIN

Tal com comentat abans, ara escollirem un segon equip, eliminant el criteri de la posició



The screenshot shows the Oracle Discoverer Desktop interface. The main window displays a worksheet titled 'Edición de Escritorio de Oracle Discoverer - [Libro de Trabajo2]'. The worksheet contains a table with the following data:

	Pts SUM	PCT 2 POINTS
Abdul-jabbar Kareem	44413	0,56
Malone Karl	41834	0,51
Jordan Michael	38541	0,50
Chamberlain Wilt	35217	0,54
Erving Julius	34919	0,50
Malone Moses	31923	0,49
Johnson Eddie	31029	0,47
Olajuwon Hakeem	30818	0,51
Issel Dan	30518	0,50
Havlicek John	30350	0,44
West Jerry	29809	0,47
Hayes Elvin	29626	0,45
Gervin George	28989	0,50
Robertson Oscar	28856	0,48
Miller Reggie	28291	0,47
Barry Rick	28281	0,45
O'neal Shaquille	28251	0,58
Wilkins Dominique	28197	0,46
Ewing Patrick	27734	0,50
English Alex	27343	0,51
Baylor Elgin	26990	0,43
Gilmore Artis	26831	0,58
Barkley Charles	26703	0,54
Parish Robert	26235	0,53
Bird Larry	25805	0,49
Drexler Clyde	25254	0,47
Dantley Adrian	24794	0,54
Greer Hal	23592	0,45

Figura 24 : Oracle Discoverer Desktop Dream Team Worksheet 4

Els jugador escullits son : KARIM ABDUL-JABAAR, KARL MALONE, MICHAEL JORDAN, WILT CHAMBERLAIN i JULIUS ERVING.

ELS MATEIXOS JUGADORS !!!!!!!!

## 5. CONCLUSIONS

La realització del projecte **Construcció i explotació d'un magatzem de dades sobre informació estadística de l'NBA** ha significat la possibilitat de portar a terme un projecte a on bolcar els coneixements adquirits durant un quants anys i desenvolupar un projecte de caire professional.

Es un motiu de satisfacció arribar fins aquest punt, ja que s'han aconseguit els objectius fixats al inici del projecte.

Actualment estem vivint una embranzida important del mon dels magatzem de dades, degut al potencial d'informació que hi trobem. Tant des de el mon de la empresa (conèixer quines son les preferències dels clients segons el seu historial de compres i segons quines zones), fins al món del esport (a l'hora de fer un fitxatge, saber quins sons els millors percentatges d'encert cara a cistella). El fet de tenir l'oportunitat de desenvolupar-ne un, permet conèixer millor quines son les necessitats dels nostres possibles clients.

Com a molts projectes, sempre es queda una mica curt de temps, o vies noves sense explorar. En el capítol següent es donen unes quantes línies d'actuació a desenvolupar en el futur.

## 6. LÍNIES D'EVOLUCIÓ FUTUR

Quan una empresa o una societat, la que sigui, munta un magatzem de dades, aprofita per fer ordre i neteja al sistema de gestió informàtica de la mateixa.

Una dels problemes mes importants que he trobat a l'hora de muntar el magatzem de dades ha estat la fiabilitat de les dades que ens han facilitat.

Una de les possibilitats sospesades a l'hora de extraure les dades era la de fer un DATALINK amb la base de dades proporcionada (MS ACCESS) amb la base de dades de ORACLE. Però degut al nombre de dades a manipular, s'ha preferit fer el treball des de el MS Excel. Un cop aquestes dades han estat extretes i transformades, s'han carregat mitjançant una connexió OCDB a la base de dades creada a ORACLE.

Una tasca futura a efectuar seria la de adequar la introducció de les dades dels periodistes del Sport World a una sèrie de formularis, de tal forma que, una vegada aquesta dades fossin bolcades a la seva base de dades, fos fàcil establir un DATALINK amb el Magatzem de dades i la seva recarrega automàtica.

Una altre tasca futura seria la ampliació dels informes a realitzar, des de conèixer el nombre de jugadors nascuts al 1960.

## 7. GLOSSARI

**DATAWAREHOUSE** : Es un magatzem de dades orientades a un domini, , integrat, no volàtil, que varia en el temps i que ens ajuda a la pressa de decisions.

**DIMENSIONS** : Son els elements rellevats a estudiar. En el nostre cas son els jugadors (players), equips (teams), entrenadors (coaches).

**ATRIBUTS** : Son dades particulars e cadascuna de les dimensions (a la dimensió Players, un atribut es el nom del jugador)

**INDICADORS** : Son les dades que son analitzades (per exemple, els punts assolits per un equip durant tota una temporada).

**STAR SCHEMA** : Es el sistema mes senzilla de representació d'un magatzem de dades, amb la taula de fets al mig i les dimensions als costats.

**GALAXY SCHEMA** : Varis Star Schemas, amb dimensions comunes.

**ORACLE** : Multinacional informàtica, de la que utilitzem la seva ORACLE DEVELOPER SUITE i la seva BASE DE DATOS ORACLE 10g EXPRESS EDITION.

**MICROSOFT** : Multinacional informàtica, de la que utilitzem la seva OFFICE SUITE.

## 8. BIBLIOGRAFIA

William H. Immon. **Building the Datawarehouse**. Fourth Edition. *Wiley Publishing Inc.*

Ralph Kimball. **The Datawarehouse ETL Toolkit**. *Wiley Publishing Inc.*

Michael Armstrong-Smith. **ORACLE DISCOVERER 10g HANDBOOK**. *McGraw-Hill*

<http://st-curriculum.oracle.com/tutorial/DBXETutorial/index.htm>

**Oracle Database 10g Express Edition Tutorial**. Plana Web amb tutorial d'Oracle

<http://www.basketball-reference.com/>

**Basketball statistics, analysis, and history**. Plana web amb informació de l'NBA.

<http://www.inmoncif.com>

**Corporate Information Factory**. Plana web d'un dels mentors dels Magatzems de Dades

<http://www.ralphkimball.com>

**Kimball Group**. Plana web d'un dels altres mentors del Magatzems de dades.

<http://www.dwreview.com>

**Data Warehouse Reviewing**. Plana web amb informació genèrica dels Magatzems de Dades

<http://www.1keydata.com/datawarehousing/datawarehouse.html>

**Data Warehousing and Business Intelligence**. Plana web amb informació genèrica dels Magatzems de Dades

<http://www.intranetjournal.com/features/datawarehousing.html>

**Intranet Journal.** Plana web amb informació genèrica dels Magatzems de Dades

[http://en.wikipedia.org/wiki/Data\\_warehouse](http://en.wikipedia.org/wiki/Data_warehouse)

**Data Warehouse – Wikipedia, the free encyclopedia.** Plana web amb informació genèrica dels Magatzems de Dades

## 9. ANNEXOS

### 9.1 IMPORTACIÓ DE DADES

Un cop posat en marxa el ORACLE XE i el ORACLE DISCOVERER, el que s'ha fet ha estat crear un compte d'usuari a on importar les dades ja transformades.

Amb referència al ORACLE DISCOVERER, utilitzant el ORACLE ADMINISTRATOR es crea una EUL (End User Layer). Un cop creada la EUL, ens permet treballar amb al ORACLE DESKTOP i preparar els informes demanats.

Mitjançant l'aplicació EXP.EXE, es crea un doble fitxer .DMP a on son bolcades les dades del magatzem de dades i de la EUL.

Per tornar passar aquestes dades a ORACLE XE i muntar de nou la EUL, caldrà utilitzar l'aplicació IMP.EXE, a on els paràmetres a posar son:

Per l'ORACLE XE	NBA/NBA@XE
Per la EUL	MANEL/JOFRA@XE

A dintre la EUL es munta una Business Area, que porta el nom de NBA.

Dintre d'aquesta Business Area de nom NBA, es munten 3 carpetes complexes i es a on, des de el DISCOVERER DESKTOP es tindrà accés a les dades estadístiques de l'NBA.