

MEMORIA

TFC F1



Construcció i explotació d'un magatzem de dades per a l'anàlisi estadístic dels resultats del Campionat de Fórmula 1

**Estudiant: Francesc Brullas Mús
Consultor: José Àngel Martín Carballo
Data Lliurament: 14 de juny de 2010**

**Enginyeria Tècnica en Informàtica de Gestió
Universitat Oberta de Catalunya Curs 2009-2010 – Segon semestre**



DEDICATORIA I AGRAIMENTS

A la meva dona i als meus fills, per la paciència que tenen amb mi.

Al meu tutor en la UOC, Juan Josep Cabré per animar-me a continuar en aquells moments en els que penses que estàs fent estudiant una carrera a la teva edat.

A tots els que creuen que el saber ens farà mes lliures.

El meu agraïment a totes aquelles persones que pengem la seva feina, experiència i coneixements a Internet, son el millor recurs en moments de desesperació.



RESUM

L'objectiu d'aquest projecte era adquirir coneixements en disseny, construcció i explotació d'un magatzem de dades, així com obtenir experiència en la gestió d'un projecte, desenvolupant les diferents fases anàlisis, planificació, implementació, etc. i la documentació associada a cada una de les diferents etapes.

En el devenir dels estudis de Enginyeria, fins ara s'han anat treballant una sèrie de assignatures en semestres diferents, amb objectius focalitzats en àrees determinades, però faltava potser fer una assignatura on poder aplicar el conjunt de totes les habilitats i competències adquirides. Aquesta assignatura es el treball de final de carrera.

La construcció de un magatzem de dades dels resultats de la Formula 1, partint d'unes dades proporcionades en MS Excel i d'uns requeriments d'informació mínima a obtenir, inclosos en el document de requeriments, ens ha permès aplicar una sèrie de estratègies i coneixements per portar el projecte a bon port, amb uns requisits tècnics suficients i aprofundint en una disciplina, els magatzems de dades, molt actual.

La temporalització i el esquema de treball proposat pel consultor ens ha ajudat a organitzar la feina, un dels grans problemes a l'hora de abordar qualsevol projecte informàtic. La necessitat de fer lliuraments parcials, obliga i ajuda a portar un ritme de feina constant i assolir una qualitat correcta en el treball proposat.

En resum, la valoració del projecte i els coneixements assolits, es al meu entendre suficient i m'ha permès enriquir-me professionalment i personalment.



INDEX

1.	INTRODUCCIÓ	6
1.1.	JUSTIFICACIÓ DEL TFC I CONTEXT EN EL QUAL ES DESENVOLUPA: PUNT DE PARTIDA I APORTACIÓ DEL TFC	6
1.2.	OBJECTIUS DEL TFC	6
1.3.	ENFOCAMENT I MÈTODE SEGUIT	7
1.4.	PLANIFICACIÓ DEL PROJECTE	7
1.5.	PRODUCTES OBTINGUTS	9
1.6.	DESCRIPCIÓ DELS ALTRES CAPÍTOLS DE LA MEMÒRIA	9
1.6.1	<i>Anàlisi</i>	9
1.6.2	<i>Disseny</i>	9
1.6.3	<i>Conclusions</i>	9
1.6.4	<i>Líneas d'evolució futura</i>	9
1.6.5	<i>Glossari</i>	9
1.6.6	<i>Bibliografia</i>	9
2.	ANÀLISI	10
2.1.	INTRODUCCIÓ	10
2.2.	ETL. EXTRACCIÓ, TRANSFORMACIÓ I CARREGA DE DADES	10
2.2.1	<i>Resultats</i>	10
2.2.2	<i>Poles</i>	11
2.2.3	<i>Pilots</i>	11
2.3.	MODEL CONCEPTUAL	11
2.3.1	<i>Gran Premi</i>	12
2.3.2	<i>Mundial</i>	15
3.	DISSENY	18
3.1.	INTRODUCCIÓ	18
3.2.	ARQUITECTURA DE SOFTWARE	18
3.3.	ARQUITECTURA DE HARDWARE	20
3.4.	DISSENY DE LA BASE DE DADES I MODEL FÍSIC	20
3.4.1	<i>Gran Premi</i>	21
3.4.2	<i>Mundial</i>	22
3.5.	DISSENY I DESCRIPCIÓ DELS INFORMES CREATS	23
3.6.	RELACIO ENTRE POLE I PNEUMATIC	28
4.	CONCLUSIONS	30
5.	LINEAS DE EVOLUCIÓ FUTURA	30
6.	GLOSSARI	31
7.	BIBLIOGRAFIA	31
7.1.	LLIBRES DE TEXT	31
7.2.	ENLLAÇOS D'INTERÈS A INTERNET	31
8.	ANNEXES	32
8.1.	REQUERIMENTS DEL ICEM	32



INDEX DE FIGURES

Figura 1 Planificació	8
Figura 2 Gran Premi	14
Figura 3 Mundial.....	17
Figura 4 Logo Windows.....	18
Figura 5 Oracle 10g.....	18
Figura 6 Discoverer Administrator	19
Figura 7 Discoverer Desktop	19
Figura 8 Arquitectura de Hardware.....	20
Figura 9 Abandonaments per tipus.....	23
Figura 10 Anys en actiu.....	24
Figura 11 Grans premis guanyats per pilot.....	24
Figura 12 Mundials guanyats per equip.....	25
Figura 13 Mundials guanyats per pilot.....	25
Figura 14 Mundials guanyats per marca de pneumàtic.....	26
Figura 15 Poles obtingudes per pilot	26
Figura 16 Posició mitjana per pilot.....	26
Figura 17 Punts obtinguts per Michael Schumacher per any	27
Figura 18 Velocitat màxima i mitjana per pilot	27
Figura 19 Nombre de poles per marca de pneumàtic	28
Figura 20 Poles per marca de pneumàtic per any	28
Figura 21 Anys de participació per marca.....	29
Figura 22 Poles per marca ponderades en funció dels anys de participació	30
Figura 23 Poles per marca ponderades en funció del nombre de grans premis.....	30
Figura 24 Taula de puntuacions	32



1. INTRODUCCIÓ

1.1. JUSTIFICACIÓ DEL TFC I CONTEXT EN EL QUAL ES DESENVOLUPA: PUNT DE PARTIDA I APORTACIÓ DEL TFC

La justificació del treball de final de carrera tal com s'explica en la informació de l'assignatura esta pensat "per a realitzar un treball de síntesi dels coneixements adquirits en altres assignatures de la carrera i que requereixi posar-los en pràctica conjuntament en un treball concret".

S'ha de tenir clar que en uns estudis d'enginyeria s'ha d'obtenir el que es demana, no es pot pensar en lliurar quelcom diferent per molt ben dissenyat que estigui. Això ens obliga a realitzar la recerca necessària per assolir els objectius en les àrees que en el moment de començar el treball no es coneixien amb la profunditat suficient.

El punt de partida a nivell de coneixements era tenir una base suficient en la gestió de projectes, bases de dades i manipulació de dades. Un possible handicap podia ser, com en el meu cas el no haver cursat les assignatures de magatzem de dades i per tant tenir poca experiència en aquest mon.

L'aportació del treball a estat, el poder treballar en un projecte on es pot veure tot el cicle de desenvolupament i posada en marxa d'una solució a unes necessitats. Conèixer una base de dades comercial (Oracle) i un producte de gestió de dades (Discoverer).

1.2. OBJECTIUS DEL TFC

Els objectius de aquest TFC, tal com indica l'enunciat es "adquirir experiència en el disseny, construcció i explotació d'un magatzem de dades a partir de la informació disponible en una base de dades transaccional". Això vol dir, fent servir els coneixement i competències suficients, lliurar una eina que permeti fer el que demanen els requeriments exposats en el enunciat, amb la temporalització demanada.

Vol dir també, el desenvolupar tenint com a punt de partida un Excel i unes necessitats un projecte que ens porti a lliurar l'eina que dona la solució als requeriments. Vol dir plantejar una temporalització del projecte, analitzar els requeriments i les dades lliurades, veure possibles problemes en la qualitat d'aquestes, construir un procés de carrega de dades i dissenyar l'eina per obtindre els informes demanats.

Si en centrem en el projecte en si mateix, podem veure que estem treballant un els punts mes importants de la informàtica de gestió, la presa de decisions partint de els dades. En un principi els desenvolupaments informàtics estaven pensats per registrar operacions d'una manera rapida i fiable. L'explotació d'aquesta informació era molt tancada i sovint no s'ha aprofitava tota.

Des de fa uns anys, s'ha imposat una nova realitat "la informació es poder", poder per prendre decisions, per millorar la gestió, per preveure el que passarà en els mercats, en l'economia, en definitiva en el mon. A més la globalització i la rapidesa dels canvis no ens permet analitzar aquesta informació d'una manera estàtica, els que avui es important dema pot no ser-ho.

El que es demana en el projecte i les eines per implementar-ho, esta al ordre del dia, els directius, consultors, analistes, etc. volen poder extreure el que vulguin, quan vulguin i com vulguin dels gigabytes d'informació que pot tindre una companya, govern o entitat per poder prendré decisions al moment. Es per això que els magatzems de dades tenen actualment una gran importància en les organitzacions de arreu.



1.3. ENFOCAMENT I MÈTODE SEGUIT

L'enfocament ha estat eminentment pràctic, la temporalització i el temps del que es disposa no ha permès aprofundir en totes les parts del projecte, això enllaça en una de les virtuts que crec que te que tindrà un futur enginyer tècnic, la capacitat de donar solucions vàlides en un temps correcte i la capacitat de cercar-les si no disposa dels coneixements suficients en l'àrea corresponent.

S'ha plantejat el treball, seguint el que deia l'enunciat i fent la interiorització de que s'estava desenvolupant per una empresa que ens havia encarregat la feina. Això enriqueix el treball i a qui fa la feina, i permet un lliurament més professional.

El mètode de treball ha estat basat en un cicle de desenvolupament clàssic:

- Anàlisi previ
- Definició del pla de treball
- Anàlisi funcional
- Disseny funcional
- Disseny tècnic
- Implementació
- Proves
- Lliurament

Aquest mètode permet en la majoria dels casos obtenir els millors resultats, permet retrocedir i tornar a avançar si els requeriments canvien o es detecta algun punt no controlat en les fases anteriors. També es el més adient en un treball de final de carrera, al no poder interactuar amb l'usuari, seria complicat fer el exercici de gestionar el projecte d'una altra manera.

La manca de temps ha fet que en algunes fases no s'hagi pogut aprofundir tot el que s'hagués volgut. Això potser ha marcat el resultat obtingut i sobretot la documentació amb més profunditat de cadascuna de les aquestes fases.

1.4. PLANIFICACIÓ DEL PROJECTE

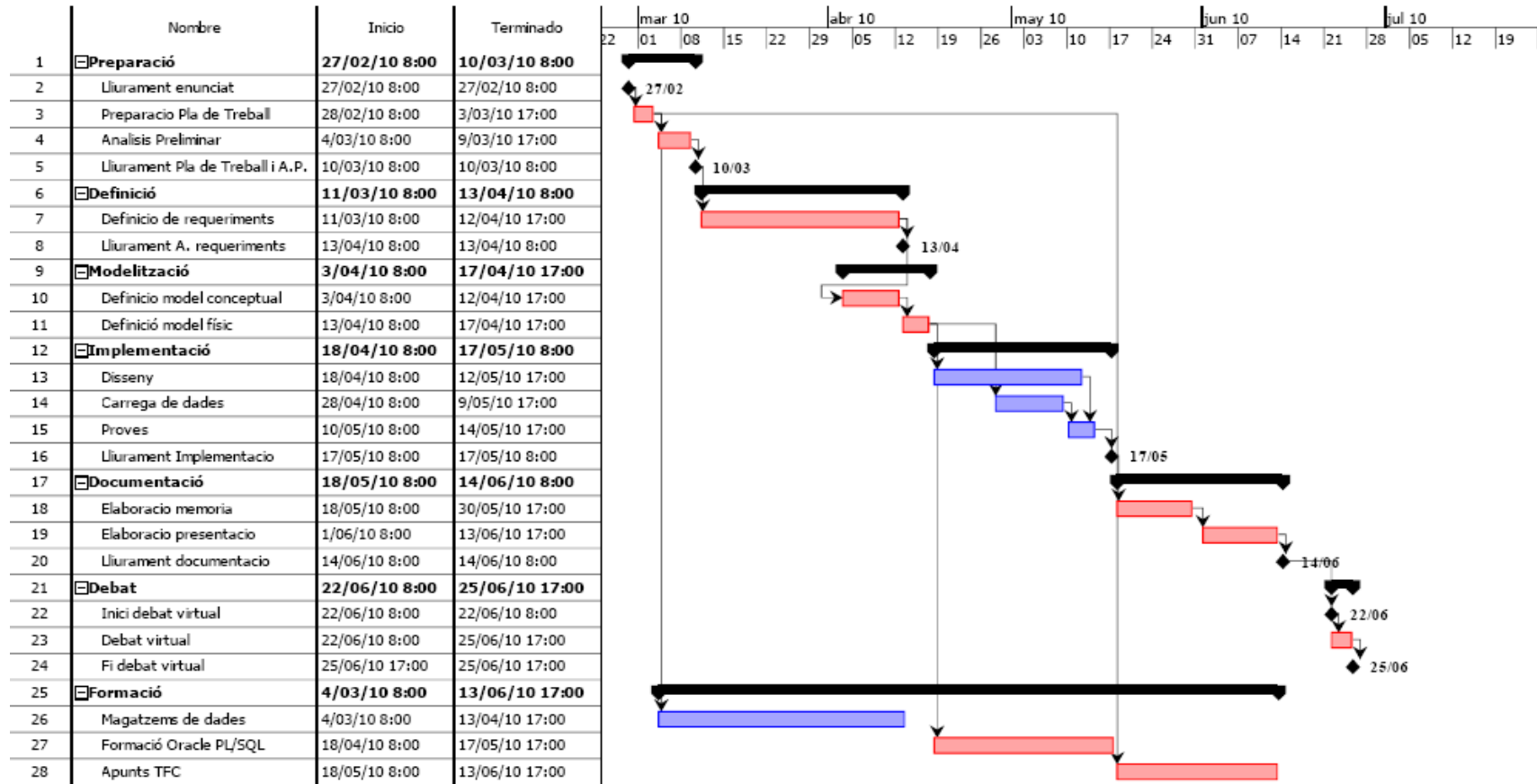
La planificació del projecte es va fer mitjançant el document "Pla de treball", lliurat en al PAC1, a grans trets s'han complert les dates i les fites, La planificació es va definir en dos grans eixos: el primer eix era el lliurament del treball i el segon era la planificació de totes les accions de formació, lectura del material docent, recerca i lectura de materials addicionals, etc.

Un dels grans problemes al que ens enfrontem al executar un projecte informàtica es la seva planificació i sobretot seu el compliment. Per desgracia segueix sent complicat obtenir totes les dades necessàries per poder planificar correctament, els usuaris no son capaços de explicar totes les seves necessitats i de vegades s'ha de esperar a la fase de proves per adonar-nos de que el que s'ha fet no es exactament el que es volia.

Per això es important la fase d'anàlisi i presa de requeriments per intentar captar totes els inquietuds i necessitats i així poder planificar correctament.



Figura 1 Planificació





1.5. PRODUCTES OBTINGUTS

Els productes obtinguts en aquest treball son els següents:

- Anàlisi preliminar. Document.
- Pla de treball. Document,
- Anàlisi de requeriments. Document
- Carrega de dades. Document, instruccions per fer la carrega.
- Discoverer. Document, instruccions per treballar amb Discoverer.
- Implementació. Document, instruccions i exemple d'informes.
- Creació de fitxers de carrega. Script SQL.
- Carrega d'estrelles. Script SQL.
- Magatzem de dades. XML estructura Discoverer.
- Gran Premi. CSV dades Excel lliurat, preparades per carregar
- Pilots. CSV dades Excel lliurat, preparades per carregar
- Pole. CSV dades Excel lliurat, preparades per carregar
- UOC_PRACT. Disc per a SUN VIRTUAL BOX, amb la practica operativa.

1.6. DESCRIPCIÓ DELS ALTRES CAPÍTOLS DE LA MEMÒRIA

1.6.1 ANÀLISI

En aquest capítol s'aborda totes les tasques de anàlisi i disseny funcional. S'ha analitzen les necessitats, es defineixen en profunditat els requeriments, es decideixen els fets, dimensions, etc. i s'estudien els dades proporcionades.

El resultat d'aquesta fase es el document d'anàlisis i de ETL (extracció, transformació i carrega de dades).

1.6.2 DISSENY

Desenvolupament de tot l'entorn tècnic, tan a nivell de programari com de maquinari, necessari per el bon funcionament de la solució. Carrega de les dades, elaboració dels informes, proves, validacions, etc.

El resultat d'aquesta fase es tot el programari necessari i l'estructura de dades carregada i operativa.

1.6.3 CONCLUSIONS

Conclusions finals del projecte. Valoració del projecte amb uan perspectiva global

1.6.4 LÍNEAS D'EVOLUCIÓ FUTURA

Exposició de millores, algunes sorgides en la fase de disseny i/o proves, així com reflexions de cap a on pot anar el producte desenvolupat.

1.6.5 GLOSSARI

Glossari de termes emprats en aquesta memòria

1.6.6 BIBLIOGRAFIA

Bibliografia



2. ANÀLISI

2.1. INTRODUCCIÓ

L' *Institut Català d'Esports de Motor (ICEM)* és l'entitat encarregada de proporcionar informació oficial als organismes públics i mitjans de comunicació catalans respecte els resultats obtinguts a les diferents competicions mundials d'esports de motor.

Actualment, l'*ICEM* proporciona aquesta informació de manera molt manual ja que tots els resultats històrics es troben recollits en documents Excel. Això implica la realització d'una feina molt feixuga i la necessitat de disposar d'un elevat nombre de recursos humans.

Per resoldre aquest problema, l'*ICEM* ha decidit encarregar-nos, la creació d'un magatzem de dades on en una primera fase només es recolliran resultats del Campionat del Món de Pilots de Fórmula 1 des de l'any 1961 fins el 2007. A continuació s'exposarà una presentació tant del contingut d'aquest document com del projecte que descriu.

Analitzades les dades i escoltades les necessites s'arriba a la conclusió de que es tenen que definir dos estrelles per cobrir les necessitats actuals i futures a mig termini de l'*ICEM*:

- GRAN PREMI
- MUNDIAL

2.2. ETL. EXTRACCIÓ, TRANSFORMACIÓ I CARREGA DE DADES

Després de estudiar el Excel proporcionat comprovem que esta dividit amb tres fulles, que la primera columna de cadascuna d'elles conte el nom del camp i detectem que hi ha algun errors greus:

- Registres erronis
- Manca de dades en alguns registres
- Diferents tipus de dades en el mateix camp
- Etc.

En el cas dels registres erronis s'opta per no fer la carrega i en els altres casos s'incorporen dades generals per poder mantenir la congruència de les dades. Es planteja demanar per properes carregues la millora d'aquestes.

Resumim ara el contingut de cadascuna de les fulles proporcionades.

2.2.1 RESULTATS

Aquesta fulla conte els resultats de totes les curses.

Primera columna, el circuit i la data del gran premi, en el procés de carrega es tindran que separar les dues dades.

Segona columna, pilot, es tindrà que validar amb el mestre de pilots que es definirà i es carregarà de la fulla de pilots.

Tercera columna, l'escuderia o equip.

Quarta columna, esta informat el pneumàtic utilitzat, semblen dades correctes, però en algunes files davant del nom hi ha un espai en blanc, es tindrà que tindre present per que no provoqui problemes al carregar.



Cinquena columna, esta la posició, is el pilot ha abandonat no hi ha classificació. Es farà una prova de integritat per evitar que en una mateixa cursa es trobin posicions iguals (excepte en el cas de abandonament que no ni ha.

Sisena columna, nombre de voltes.

Setena columna, temps emprat. Aquesta columna es tindrà que estudiar amb profunditat, en algunes files esta per temps, en altres per temps amb un signe “+” i en altres per voltes.

Vuitena columna, velocitat màxima.

Novena columna, els comentaris en cas de abandonament.

2.2.2 POLES

Aquesta fulla conte les dades de les pole.

Primera columna, el gran premi

Segona columna, el país

Tercera columna, la data

Quarta columna, el pilot que ha obtingut la pole

Cinquena columna, el temps de la volta rapida

2.2.3 PILOTS

Aquesta fulla conte les dades dels pilots.

Primera columna, el país del pilot

Segona columna, el nom del pilot

Tercera columna, els anys en actiu. Les dades estan en diferents formats: “any inicia l- any final”, “any, any”, “any”, etc. S’aprofundirà en el anàlisis i es definiran les estratègies de procés de aquesta informació.


2.3. MODEL CONCEPTUAL

Identificats els fet en la fase anàlisis preliminar, passarem ara a fer una definició amb mes profunditat de cadascun d’ells.

El magatzem de dades de l’ICEM contindrà una sèrie cubs de dades, estrelles, per poder explotar d’informació d’una manera àgil i flexible.

Per la definició de aquestes estrelles emprarem una metodologia que ajuda a simplificar la seva construcció, aquesta metodologia segueix un camí de nou passos:

- Triar el fet.
- Trobar el grànul escaient.
- Escollir les dimensions que s’utilitzaran en el anàlisis.
- Trobar els atributs de cada dimensió.

- 
- Distingir entre descriptors i jerarquies d'agregació.
 - Decidir quines son les mesures que interessin.
 - Definir les Cel·les.
 - Explicar les restriccions d'integritat.
 - Estudiar la viabilitat de l'estrella

Com en molts casos les dimensiona a utilitzar seran les mateixes es decideix que en la definició de cadascuna de les estrelles es definirà però el seu desenvolupament complet només es farà en la primera estrella que la faci servir.

2.3.1 GRAN PREMI

Es el fet de que s'ha fet una cursa de F1 del calendari oficial. Es celebra en un circuit determinat, en una data i hi ha una sèrie de participants que assoleixen una classificació al final de la cursa.

2.3.1.1 TRIAR EL FET

Se ha celebrat una cursa de formula 1, en un lloc determinat en una data i amb uns participants.

2.3.1.2 TROBAR EL GRÀNUL ESCAIENT

El nivell màxim de detall es el de participació d'un pilot en una cursa determinada.

2.3.1.3 ESCOLLIR LES DIMENSIONS

Les dimensions candidates son:

- Circuit
- Temps
- Pilot
- Escuderia
- Pneumàtics
- Abandonament
- Pole
- Motiu de abandonament. Dimensió deixalla que serà el motiu del abandonament i en el cas de que el motiu sigui un accident s'associarà amb el altre pilot implicat.

2.3.1.4 TROBAR ELS ATRIBUTS DE CADA DIMENSIÓ

- Circuit: La relació de circuits en els que es celebren grans premis. País.
- Temps. La data de realització de la cursa. Dia, Mes, Any.
- Pilot. La relació de pilots que participen en la cursa, País.
- Escuderia. La relació de escuderies sota la que corren els pilots. No s'associa la pilot degut a que un pilot pot anar canviant d'escuderia cada any.
- Pneumàtics. La marca de pneumàtics que portes els cotxes en la cursa. No s'associa ni a pilot ni a escuderia perquè pot anar canviant cada any.
- Abandonament. Tindrà 2 valors un indicarà que no ha abandonat i l'altre que ha abandonat. Es una dimensió degenerada, per tant, encara que a nivell conceptual tingui una taula associada, a nivell físic no la tindrà al ser una dimensió amb dos valors.
- Pole. Tindrà 2 valors un indicarà que el pilot en aquella cursa ha obtingut la pole i el altre que no la ha obtingut. Es una dimensió degenerada, per tant, encara que a nivell conceptual tingui una taula associada, a nivell físic no la tindrà al ser una dimensió amb dos valors.



2.3.1.5 DISTINGIR ENTRE DESCRIPTORS I JERARQUIES D'AGREGACIÓ

Tots els atributs de cada dimensió serviran com a descriptors i per les agregacions podrem fer servir-les totes, en funció del tipus d'informació que vulguem explotar.

2.3.1.6 DEFINIR LES MESURES

- Posició. Posició obtinguda en una cursa, numèric, cas de que abandonament serà 0. No additiva
- Voltes. Voltes realitzades.
- Velocitat. Velocitat mitjana del pilot en la cursa. No transitiva
- TempsCursa. Temps emprat (en el cas del guanyador), diferencia de temps en el cas dels que entren en la mateixa volta, de la resta de participants no tindrem dades.
- Punts obtinguts. Nombre de punts obtinguts, seguim el criteri que marca la taula del plec de condicions
- TempsPole. Temps emprat (en el cas del pilot que ha obtingut la pole), 0 en els altres pilots

2.3.1.7 DEFINIR LES CEL·LES

Totes les mesures pertanyen a la mateixa granularitat i per tant estaran a la mateixa cel·la

2.3.1.8 RESTRICCIONS D'INTEGRITAT

S'ha de anar en compte amb totes les mesures, la majoria no poden agregar-se directament.

2.3.1.9 ESTUDIAR LA VIABILITAT DE L'ESTRELLA

La estrella a nivell de volum de dades y rendiment es totalment viable.

Volums:

Estem parlant de unes 15 proves al any amb una participació de uns 25 pilots per prova això dona uns 375 registres any. Si pensem amb una longitud de registre de 256 bytes aproximadament estarem parlant de uns 100 MB es a dir en 10 anys 1 GB. Quantitat que no ens tindria que preocupar amb la tecnologia de la que disposem.

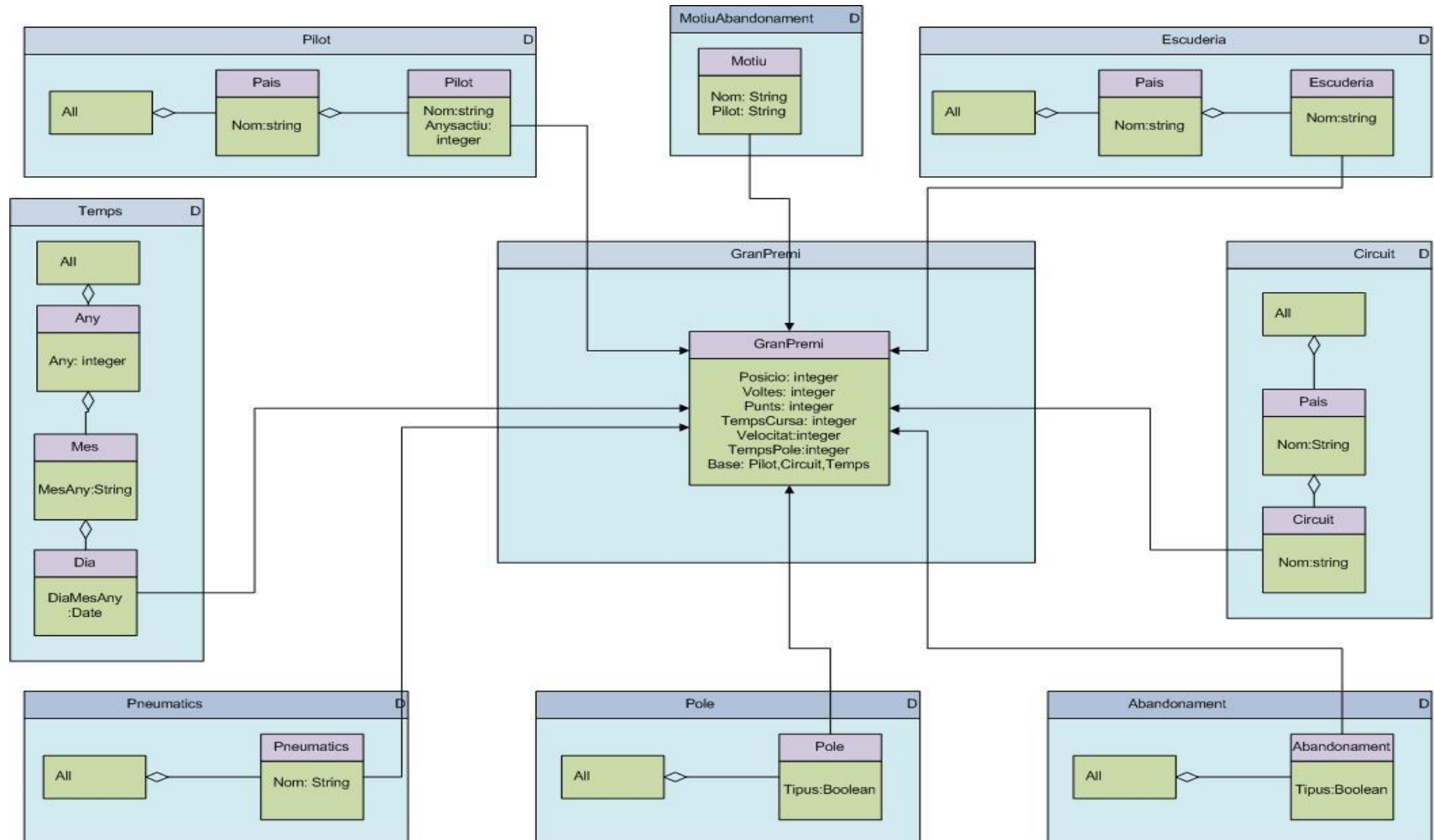
Rendiment:

Estem parlant de un volum de dades molt petita amb el que el rendiment no te que ser un problema per l'explotació de les sol·licituds que podrem fer a l'estrella.



2.3.1.10 DIAGRAMA MULTIDIMENSIONAL

Figura 2 Gran Premi





2.3.2 MUNDIAL

Es el fet de que s'ha celebrat un mundial en un any determinat i ha esta guanyat per un pilot que pertany a una escuderia amb un tipus de pneumàtics. El mundial es guanya per haver fet el màxim nombre de punts.

2.3.2.1 TRIAR EL FET

Se ha celebrat un mundial.

2.3.2.2 TROBAR EL GRÀNUL ESCAIENT

El nivell màxim de detall es el de participació d'un pilot en un mundial determinat.

2.3.2.3 ESCOLLIR LES DIMENSIONS

Les dimensions candidates son:

- Pilot
- Escuderia
- Pneumàtics
- Any

2.3.2.4 TROBAR ELS ATRIBUTS DE CADA DIMENSIÓ

- Pilot. La relació de pilots que participen en la cursa, País. Anys en actiu. Nombre de grans premsi guanyats.
- Escuderia. La relació de escuderies sota la que corren els pilots. No s'associa la pilot degut a que un pilot pot anar canviant d'escuderia cada any. Nombre de grans premsi guanyats.
- Pneumàtics. La marca de pneumàtics que portes els cotxes en la cursa. No s'associa ni a pilot ni a escuderia perquè pot anar canviant cada any. Nombre de grans premsi guanyats.
- Any. Any de execució del mundial

2.3.2.5 DISTINGIR ENTRE DESCRIPTORS I JERARQUIES D'AGREGACIÓ

Tots els atributs de cada dimensió serviran com a descriptors i per les agregacions podrem fer servir-les totes, en funció del tipus d'informació que vulguem explotar.

2.3.2.6 DEFINIR LES MESURES

- Aquesta estrella no te mesures. Es podran obtindre fent drill across amb l'estrella GranPremi.

2.3.2.7 DEFINIR LES CEL·LES

No es defineix al ser una estrella de relació.

2.3.2.8 RESTRICCIONS D'INTEGRITAT

Cap.

2.3.2.9 ESTUDIAR LA VIABILITAT DE L'ESTRELLA

La estrella a nivell de volum de dades y rendiment es totalment viable.

Volums:



Estem parlant de un registre per any. Si pensem amb una longitud de registre de 100 bytes aproximadament estarem parlant en 10 anys 1 KB. Quantitat que no ens tindria que preocupar amb la tecnologia de la que disposem.

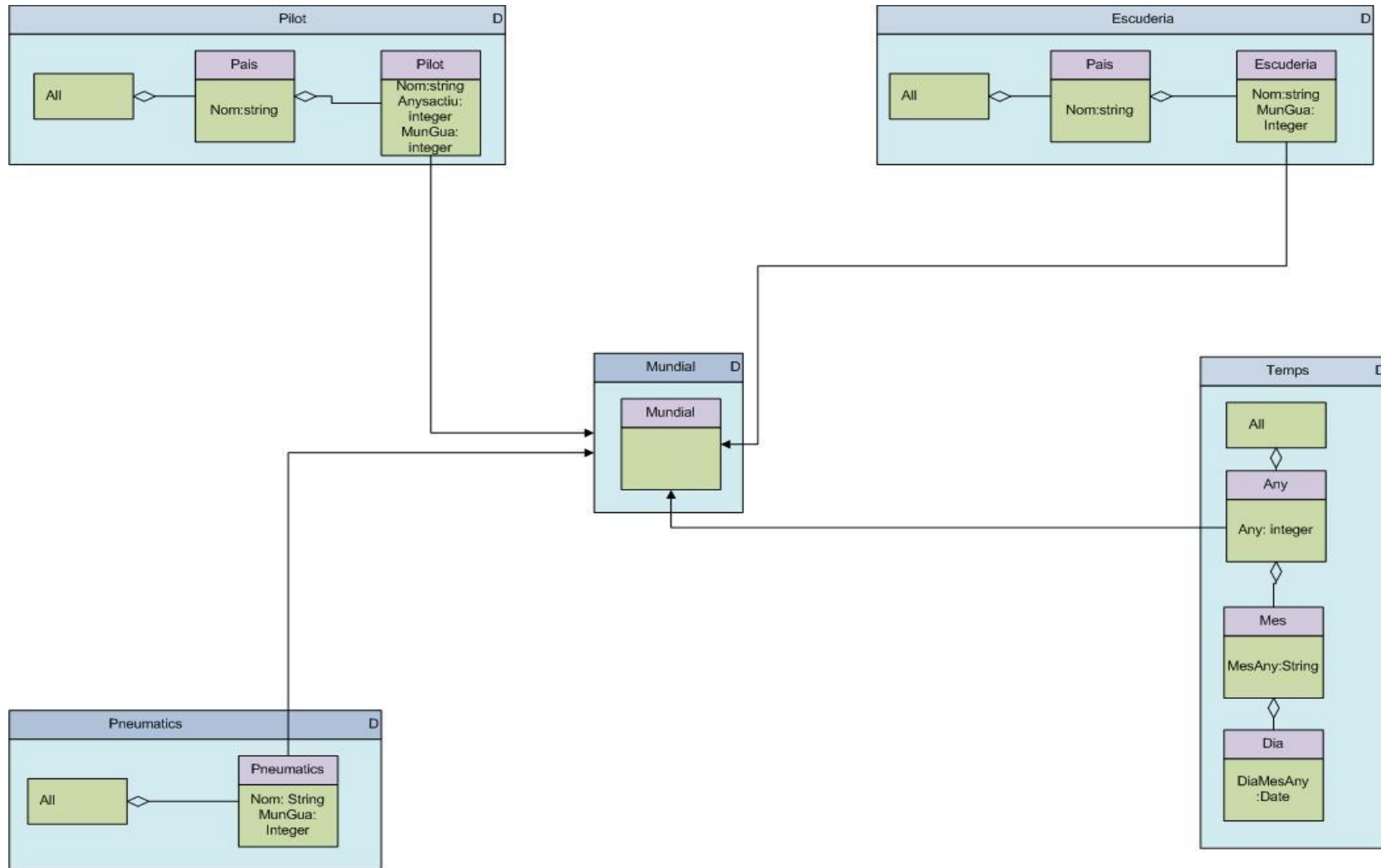
Rendiment:

Estem parlant de un volum de dades molt petita amb el que el rendiment no te que ser un problema per l'exploració de les sol·licituds que podem fer a l'estrella.



2.3.2.10 DIAGRAMA MULTIDIMENSIONAL

Figura 3 Mundial



3. DISSENY

3.1. INTRODUCCIÓ

En aquest capítol s'aborden totes les tasques de disseny necessàries per poder aplicar la solució proposada.

3.2. ARQUITECTURA DE SOFTWARE

Per poder implementar la solució als requeriments proposats, s'ha pensat en fer servir productes comercials amb una gran implantació i de empreses de gran solvència. D'aquesta manera es garanteix la evolució de l'eina i facilita la contractació de evolutius, preventius i correctius sense dependre de una determinada empresa de desenvolupament.

En referència al sistema operatiu s'ha pensat en Windows de Microsoft, aquest sistema, tot i els seus detractors, dona un rendiment òptim i una gran escalabilitat. Es un sistema robust, testejat i amb elevat nivell de qualitat.

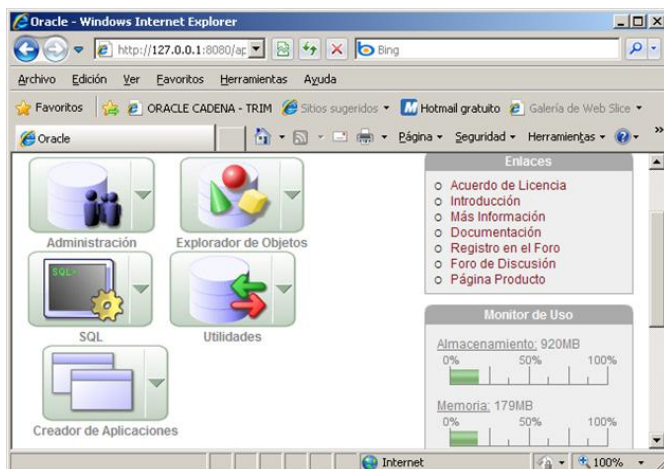
Figura 4 Logo Windows



Al pensar en el producte de gestió de la base de dades que te que suportar el magatzem de dades Oracle es una decisió encertada. Aquest producte es un estàndard de mercat, per no dir que es el referent del mercat. Amb una possibilitat de creixement pràcticament il·limitada un suport tècnic a nivell mundial i una gran quantitat d'eines al seu voltant, permet assegurar una estabilitat al magatzem de dades.

Oracle permet una gestió eficaç de les dades amb un rendiment i seguretat fora de dubtes, quan el rendiment es important i el temps de resposta un factor crític Oracle permet garantir ambdues coses. Tanmateix es important recalcar el baix cost de propietat que te el producte.

Figura 5 Oracle 10g





Per gestionar el magatzem, s'ha triat Discoverer, producte que permet la creació d'un magatzem de dades d'una manera fàcil i intuïtiva i permet deslligar la seva definició de la explotació. Això evita possibles problemes de seguretat i confereix a la instal·lació un plus de estabilitat.

Es contracten inicialment els següents mòduls:

- Discoverer Administrator. Eina per definir el model de dades, els fets, les dimensions, relacions, etc. Té un interfase gràfic molt fàcil de fer servir i permet anar depurant el disseny del magatzem i fer-lo creixi en funció de noves necessitats. Permet també definir els usuaris que tindran accés a les dades i amb quin rol les podran explotar.
- Discoverer Desktop: És el producte que permetrà als analistes i usuaris en general explotar les dades: Un entorn fàcil i agradable, molt integrat amb altres eines informàtiques, de ràpid aprenentatge, ajuda des del primer moment a obtenir dades fiables. Permet tots els tipus de navegació que s'espera en una eina d'aquest tipus (Slice, Dice, Roll up, Drill down, Drill across).

Figura 6 Discoverer Administrator

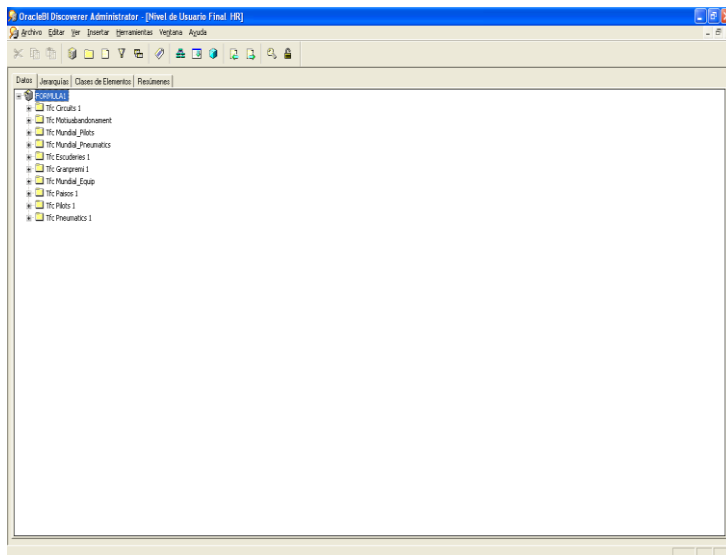
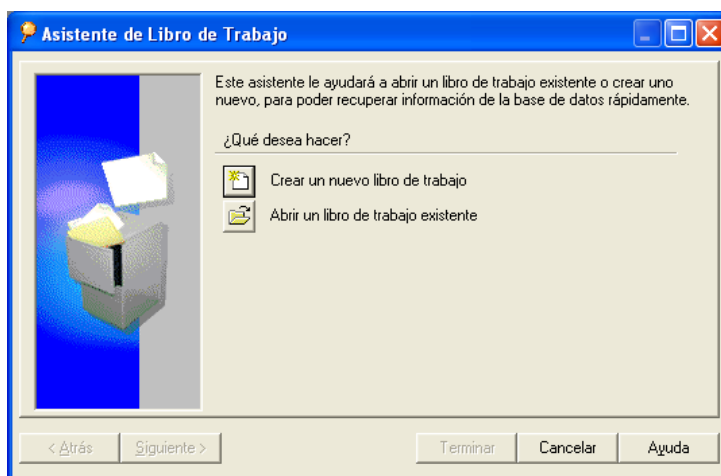


Figura 7 Discoverer Desktop



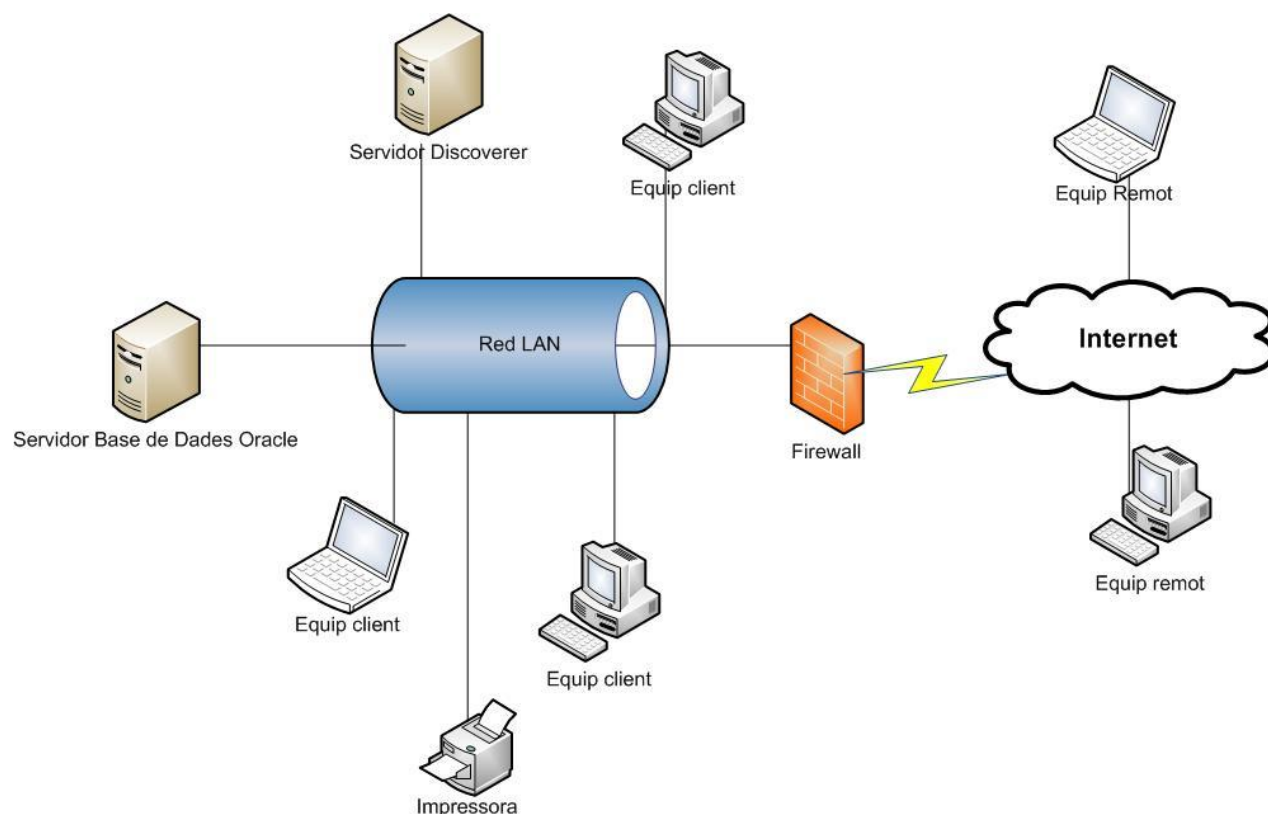
3.3. ARQUITECTURA DE HARDWARE

Pensant en un model de hardware que permeti un rendiment òptim, una alta disponibilitat i una relació preu rendiment acurada. Es presenta un model amb tres capes: base de dades, aplicació i clients. Encara que en aquesta fase s'instal·larà aplicació i base de dades en un únic servidor.

Es planteja també definit tres entorns:

- Test. Per testejar el model per part del equip de desenvolupament
- Pre producció. Entorn per que l'usuari faci les proves abans de posar el desenvolupament en productiu
- Producció. Entorn de treball per els usuaris finals.

Figura 8 Arquitectura de Hardware



3.4. DISSENY DE LA BASE DE DADES I MODEL FÍSIC

El disseny de la base de dades s'ha fet pensat en aprofitar totes les possibilitats que ens dona l'arquitectura Oracle, treballar seguint les instruccions i recomanacions del fabricant permet obtenir el millor rendiment i evita problemes de funcionament, garantint també el suport en cas de qualsevol problema.

Per deslligar la clau primària de totes les taules de les dades, que potser poden variar, aquesta clau primària s'ha definit com auto numèrica i automàtica, això permet poder canviar les dades que defineixen les dimensions i facilita possibles canvis en la definició de aquestes dimensions.

En els apartats següents es desenvolupa el model físic.

3.4.1 GRAN PREMI

3.4.1.1 TAULES

Taula	GranPremi		
Atribut	Tipus	Clau primària	Clau forànea
Pilot	String	SI	Pilots
Escuderia	String	SI	Escuderies
Pneumàtic	String	SI	Pneumàtics
Circuit	String	SI	Circuits
Data	Date	SI	Temps
Abandonament	Boolea	SI	Abandonaments
Pole	Boolea	SI	Poles
Posició	Integer		
Voltes	Integer		
Punts	Integer		
Velocitat	Integer		
TempsCursa	Integer		
TempsPole	Integer		

Taula	Pilots		
Atribut	Tipus	Clau primària	Clau forànea
Pilot	String	SI	
Nom	String		
País	String		Països
AnysActiu	Integer		

Taula	Escuderies		
Atribut	Tipus	Clau primària	Clau forànea
Escuderia	String	SI	
Nom	String		
País	String		Països

Taula	Pneumàtics		
Atribut	Tipus	Clau primària	Clau forànea
Pneumàtic	String	SI	
Nom	String		



Taula	Circuits		
Atribut	Tipus	Clau primària	Clau forànea
Circuit	String	SI	
Nom	String		
País	String		Països

Taula	MotiuAbandonament		
Atribut	Tipus	Clau primària	Clau forànea
Pilot	String	SI	Pilots
Escuderia	String	SI	Escuderies
Pneumàtic	String	SI	Pneumàtics
Circuit	String	SI	Circuits
Data	Date	SI	Temps
Nom	String	SI	
Pilot	String		Pilots

3.4.1.2 ÍNDEXS

Índex	GranPremi1
Atribut	Tipus
Pilot	String
Circuit	String
Data	Date
Abandonament	Boolea

Índex	GranPremi2
Atribut	Tipus
Escuderia	String
Circuit	String
Data	Date
Abandonament	Boolea

Índex	GranPremi3
Atribut	Tipus
Pneumàtic	String
Circuit	String
Data	Date
Abandonament	Boolea

3.4.2 MUNDIAL

3.4.2.1 TAULES

Taula	Mundial		
Atribut	Tipus	Clau primària	Clau forànea
Pilot	String	SI	Pilots
Escuderia	String	SI	Escuderies
Pneumàtic	String	SI	Pneumàtics
Any	Integer	SI	Temps

3.4.2.2 ÍNDEXS

Índex	Mundial1
Atribut	Tipus
Pilot	String



Índex	Mundial2
Atribut	Tipus
Pneumàtic	String

Índex	Mundial3
Atribut	Tipus
Escuderia	String

3.5. DISSENY I DESCRIPCIÓ DELS INFORMES CREATS

Els informes creats, s'adjunten en el annexa. Per la seva construcció s'ha fet servir DISCOVERER DESKTOP. S'ha seguit en tots els informes la mateixa estructura de disseny per que el projecte tingui una coherència estètica.

S'han estructurat els informes tipus taula dinàmica amb la possibilitat de que totes les dimensions serveixin a l'hora per filtra i ordenar, arrastrant senzillament la dimensió de l'àrea de selecció a l'àrea de dades podem donar la volta, fer drill down, etc.

També s'han dissenyat pensant amb la possibilitat de poder agrupar, des agrupar i que sigui fàcil i ràpid poder baixar a veure el detall de les operacions que componen una dada determinada.


Figura 9 Abandonaments per tipus



ABANDONAMENTS

Data: Year: 2000	Nomcircuit: <Todo>	Nompilot: <Todo>	Nomescuderia: <Todo>	Nompneumatic: <Todo>
	Abandonaments			
Engine	32,00			
Collision	29,00			
Spun off	28,00			
Gearbox	17,00			
Electrical	12,00			
Hydraulics	8,00			
Suspension	6,00			
Transmission	5,00			
Brakes	4,00			
Disqualified	3,00			
Fuel pressure	2,00			
Alternator	1,00			
Clutch	1,00			
Fuel system	1,00			
Halfshaft	1,00			
Oil leak	1,00			
Oil pressure	1,00			
Throttle	1,00			
Wheel	1,00			


Figura 10 Anys en actiu



ANYS ACTIU

	Nompilot	Anysactiu
1	Oscar Galvez	1
2	Pablo Birger	2
3	Alessandro de Tomaso	2
4	Jorge Daponte	1
5	Roberto Bonomi	1
6	Onofre Marimon	3
7	Esteban Tuero	1
8	Ricardo Zunino	3
9	Miguel Angel Guerra	1
10	Clemer Bucci	2
11	Nasif Estefano	1
12	Jose Froilan Gonzalez	1
13	Juan Manuel Fangio	7
14	Alberto Rodriguez Larreta	1
15	Carlos Reutemann	11
16	Adolfo Schwelm-Cruz	1
17	Roberto Mieres	3
18	Carlos Menditeguy	3
19	Gaston Mazzacane	2
20	Oscar Larrauri	1
21	Jesus Iglesias	1
22	Norberto Fontana	1
23	Argentina: 22	
24	Vern Schuppan	2
25	Frank Gardner	2
26	David Brabham	2

Figura 11 Grans premis guanyats per pilot



GRANS PREMIS GUANYATS

Nompneumatic: <Todo>	Nomescluderia: <Todo>	Nomcircuit: <Todo>	Data: Year: <Todo>
GUANYATS			
Michael Schumacher	91,00		
Alain Prost	51,00		
Ayrton Senna	40,00		
Nigel Mansell	31,00		
Jackie Stewart	27,00		
Jim Clark	25,00		
Niki Lauda	25,00		
Nelson Piquet	23,00		
Damon Hill	22,00		
Mika Hakkinen	20,00		
Fernando Alonso	19,00		
Kimi Raikkonen	15,00		
Emerson Fittipaldi	14,00		
Graham Hill	14,00		
David Coulthard	13,00		
Alan Jones	12,00		

Figura 12 Mundials guanyats per equip




NOMBRE MUNDIALS GUANYATS PER EQUIP

Codians: <Todo>

	Guanyats
Ferrari	15,00
McLaren	9,00
Williams	9,00
BRM	3,00
Lotus	3,00
Brabham	2,00
Renault	2,00
Tyrrell	2,00
Benetton	1,00
Matra	1,00

Figura 13 Mundials guanyats per pilot



NOMBRE MUNDIALS GUANYATS PER PILOT

Codians: <Todo>

	Guanyats
Michael Schumacher	7,00
Alain Prost	6,00
Graham Hill	3,00
Jackie Stewart	3,00
Nelson Piquet	3,00
Ayrton Senna	2,00
Emerson Fittipaldi	2,00
Fernando Alonso	2,00
Jim Clark	2,00
Mika Hakkinen	2,00
Niki Lauda	2,00
Alan Jones	1,00
Damon Hill	1,00
Denny Hulme	1,00
Jack Brabham	1,00
Jacques Villeneuve	1,00
James Hunt	1,00
Jochen Rindt	1,00
Jody Scheckter	1,00
Keke Rosberg	1,00
Kimi Raikkonen	1,00
Mario Andretti	1,00
Nigel Mansell	1,00
Phil Hill	1,00

Figura 14 Mundials guanyats per marca de pneumàtic

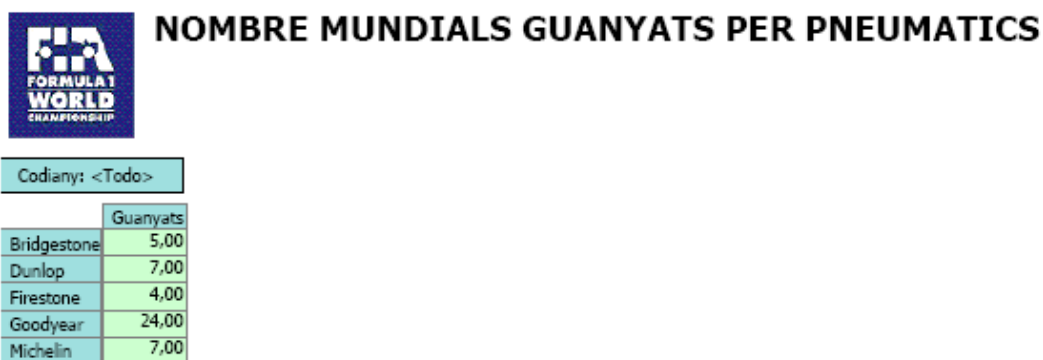


Figura 15 Poles obtingudes per pilot

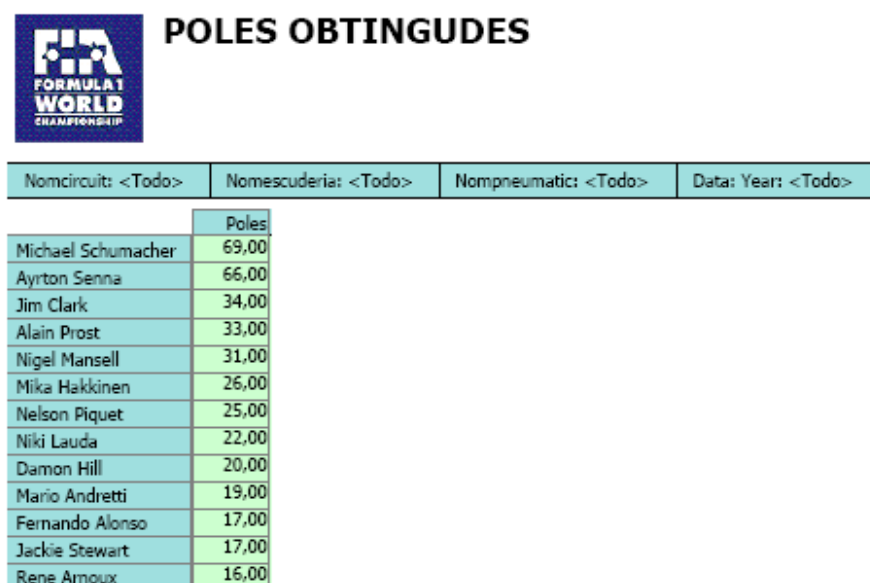


Figura 16 Posició mitjana per pilot

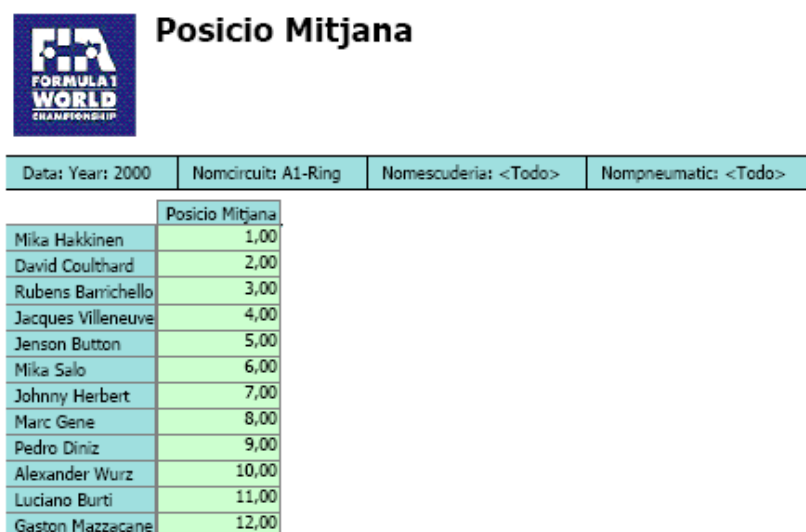


Figura 17 Punts obtinguts per Michael Schumacher per any



PUNTS OBTINGUTS

Nompilot: Michael Schumacher	Nomescureria: <Todo>	Nompneumatic: <Todo>	Nomcircuit: <Todo>
	Punts	Cursas	Mitjana
1991	4,00	3,00	1,33
1992	53,00	11,00	4,82
1993	52,00	9,00	5,78
1994	92,00	10,00	9,20
1995	102,00	12,00	8,50
1996	59,00	9,00	6,56
1997	78,00	13,00	6,00
1998	86,00	12,00	7,17
1999	44,00	7,00	6,29
2000	108,00	13,00	8,31
2001	123,00	15,00	8,20
2002	144,00	17,00	8,47
2003	93,00	15,00	6,20
2004	148,00	16,00	9,25
2005	62,00	12,00	5,17
20-MAR-2005	2,00	1,00	2,00
24-ABR-2005	8,00	1,00	8,00
22-MAY-2005	2,00	1,00	2,00
29-MAY-2005	4,00	1,00	4,00
12-JUN-2005	8,00	1,00	8,00
19-JUN-2005	10,00	1,00	10,00
03-JUL-2005	6,00	1,00	6,00
10-JUL-2005	3,00	1,00	3,00
24-JUL-2005	4,00	1,00	4,00
31-JUL-2005	8,00	1,00	8,00
25-SEP-2005	5,00	1,00	5,00
09-OCT-2005	2,00	1,00	2,00
2006	121,00	16,00	7,56

Figura 18 Velocitat màxima i mitjana per pilot



VELOCITATS MAXIMA I MITJANA

Nomcircuit: <Todo>	Nomescureria: <Todo>	Data: Year: <Todo>
	Velocitat MAX	Mitjana
Adrian Campos	209	105,36
Adrian Sutil	226	186,50
Aguri Suzuki	230	131,49
Al Pease	69	69,00
Alain Prost	240	166,80
Alan Jones	220	162,47
Alessandro Nannini	229	124,62
Alessandro Pesenti-Rossi	185	178,50
Alessandro Zanardi	237	124,79
Alex Caffi	227	131,30
Alex Ribeiro	202	182,00
Alex Yoong	223	127,83
Alexander Wurz	225	170,20
Allan McNish	221	152,09
Allen Berg	173	98,67
Andre Pilette	37	37,00
Andrea Chiesa	32	32,00

3.6. RELACIO ENTRE POLE I PNEUMATIC

Al començar a analitzar les dades sembla clar que la marca de pneumàtics que ha guanyat mes poles es Goodyear.

Figura 19 Nombre de poles per marca de pneumàtic

Marca	Pole
Goodyear	349
Bridgestone	116
Michelin	106
Dunlop	67
Firestone	45
Pirelli	5
Suma	688

Però, al fer un drilldown baixant a nivell d'any observem que Goodyear va tindre durant molts Mundials una posició de quasi exclusivitat.

Figura 20 Poles per marca de pneumàtic per any

	Bridgestone Pole	Dunlop Pole	Firestone Pole	Goodyear Pole	Michelin Pole	Pirelli Pole
1961		8				
1962		9				
1963		10				
1964		11				
1965		10				
1966		4	5			
1967			8	3		
1968			10	2		
1969		11				
1970		4	8	1		



1995				17		
1996				16		
1997				17		
1998	13			3		
1999	16					
2000	17					
2001	13				4	
2002	10				7	
2003	8				8	
2004	12				6	
2005	1				18	
2006	8				10	
2007	17					
Suma	116	67	45	349	106	5



Això ens fa pensar que potser tindríem que creuar el nombre de poles amb el nombre de participacions en Mundials per marca, potser el resultat variava.

Figura 21 Anys de participació per marca

	Avon	Bridgestone	Dunlop	Firestone	Goodyear	Michelin	NOT AVAILABL E	Pirelli
	AnysParticipacio	AnysParticipacio	AnysParticipacio	AnysParticipacio	AnysParticipacio	AnysParticipacio	AnysParticipacio	AnysParticipacio
1961			1					
1962			1					
1963			1				1	
1964			1		1			
1965			1		1		1	1
1966			1	1	1		1	1
1967				1	1			1
1968			1	1	1		1	1
1969			1				1	1
1970			1	1	1		1	
1971				1	1		1	1
1972				1	1		1	
1973				1	1			
1974				1	1		1	
1975				1	1			
1976		1	1		1			
1977		1	1		1	1		
1978					1	1		
1979					1	1		
1980					1	1		
1981	1				1	1	1	1
1982	1				1	1		1
1983					1	1		1
1984					1	1		1
1985					1			1
1986					1			1
1987					1			
1988					1		1	
1989					1			1
1990					1		1	1
1991					1			1
1992					1		1	
1993					1			
1994					1			
1995					1			
1996					1			
1997		1			1			
1998		1			1			
1999		1						
2000		1						
2001		1				1		
2002		1				1		
2003		1				1		
2004		1				1		
2005		1				1		
2006		1				1		



Ara calcularem la mitjan de poles per any de participació (s'han eliminat les maques que no han guanyat cap Gran Premi)

Figura 22 Poles per marca ponderades en funció dels anys de participació

	Bridgestone	Dunlop	Firestone	Goodyear	Michelin	Pirelli
Participacions	13	11	9	34	14	15
Pole	116	67	45	349	106	5
Mitja	8,92	6,09	5,00	10,26	7,57	0,33

Goodyear segueix estan en primer lloc però ara les distancies s'han reduït

Seguirem ara veient de contrastar el nombre de participacions en grans premsi i la seva relació amb les poles aconseguides:

Figura 23 Poles per marca ponderades en funció del nombre de grans premis

	Bridgestone	Dunlop	Firestone	Goodyear	Michelin	Pirelli
Nombre	205	110	108	482	215	148
Pole	116	67	45	349	106	5
Mitja	56,59%	60,91%	41,67%	72,41%	49,30%	3,38%

Sembla clar que Goodyear es la marca líder en el referent a les poles.

4. CONCLUSIONS

El sistema dissenyat es fàcil de fer servir, permet obtenir informes a la carta amb un bon rendiment i fent servir eines estàndard del mercat. Fàcil de mantindre i amb una bona estabilitat. Compleix amb escriu les expectatives dels usuaris.

A nivell didàctic a estat un projecte engrescador, amb una dificultat tècnica mitjana i que a nivell personal ens ha permès conèixer una sèrie d'eines que fins aquest moment no havia treballat. Personalment estic content del producte obtingut i del desenvolupament tant de la part de documentació com del disseny obtingut.

5. LINEAS DE EVOLUCIÓ FUTURA

A nivell de les dades que es proporcionen, es pot millorar molt la seva qualitat, es important fer-ho per assegurar que la informació que se esta transmeten es correcta i fiable.

Punt que poden evolucionar i permetre obtindre molta mes informació son:

El temps i les voltes estan barrejats en els resultats de les curses això fa que no es puguin manegar d'una manera coherent. Es cert, que la informació que proporciona la FIA es així, però es podria estudiar algun sistema per poder analitzar con ha anat una cursa en general, nombre de voltes fetes, diferencia de voltes entre el primer i el últim, etc.

Adaptar el procés de carrega desenvolupat per poder afegir dades dels anys següents, el sistema actual nomes permet fer una carrega complerta i des de zero.

Desenvolupar una lloc Web on poder publicar els dades i penjar també alguns del informes preparats perquè estiguin disponibles per a tothom.



6. GLOSSARI

Magatzem de dades

Es una base de dades amb la informació històrica d'una organització dissenyada i estructurada per a realitzar-hi consultes eficientment.

Slice

Permet fer fa un tall al cub de manera que es redueix el nombre de dimensions. Es tracta simplement de fixar un valor a una de les dimensions del cub, de manera que passem a tenir un cub amb $n-1$ dimensions en què totes les cel·les fan referència al valor que hem triat.

Dice

Permet seleccionar un subespai del cub original, sense reduir el nombre de dimensions. Això s'aconsegueix seleccionant un subconjunt de valors en cadascuna de les dimensions.

Roll up

Permet reduir el detall amb què veiem les dades. Agrupa les cel·les del cub seguint una certa jerarquia d'agregació. Per a obtenir les dades de cada cel·la aplica una certa operació matemàtica (com ara la suma o la mitjana) a les dades de les cel·les que formen part de cadascun dels grups.

Drill down

Permet augmentar el detall amb què veiem les dades. És l'operació inversa al *Roll-up*. En comptes de pujar en una jerarquia d'agregació, baixem, desfem els grups.

Drill across

Permet canviar el tema d'anàlisi. Després d'aplicar aquesta operació, continuem tenint el mateix espai n -dimensional que teníem, però ara les cel·les contindran dades que corresponen a un tipus de fet diferent.

ETL

Extract, Transform and Load. Extreure, transformar i carregar. Desenvolupament de la estratègia per carregar una aplicació no va amb les dades del sistema anterior, per evitar introducció manual de dades.

7. BIBLIOGRAFIA

7.1. LLIBRES DE TEXT

The Data Warehouse Toolkit. (2nd Edition). Ralph Kimball and Margy Ross.

Building the Data Warehouse (3rd Edition). W. H. Inmon.

The Data Warehouse ETL Toolkit. Joe Caserta and Ralph Kimball.

Oracle SQL Developer. Oracle (Llibre virtual).

Doing SQL from PL/SQL: Best and Worst Practices. Oracle (Withe paper)

Implementación de datawarehouse con Oracle Discoverer. Gustavo Teran. (Llibre virtual)

7.2. ENLLAÇOS D'INTERÈS A INTERNET

<http://www.inmoncif.com>

<http://www.rkimball.com>

<http://www.dwreview.com>

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



<http://www.intranetjournal.com/features/datawarehousing.html>
http://ca.wikipedia.org/wiki/Fórmula_1

8. ANNEXES

El sistema dissenyat es fàcil de fer servir, permet obtenir informes a la carta amb un bon rendiment i fent servir eines estàndard del mercat. Fàcil de mantindre i amb una bona estabilitat. Compleix amb escriu les expectatives dels usuaris.

8.1. REQUERIMENTS DEL ICEM

L' *Institut Català d'Esports de Motor (ICEM)* és l'entitat encarregada de proporcionar informació oficial als organismes públics i mitjans de comunicació catalans respecte els resultats obtinguts a les diferents competicions mundials d'esports de motor.

Actualment, l' *ICEM* proporciona aquesta informació de manera molt manual ja que tots els resultats històrics es troben recollits en documents Excel. Això implica la realització d'una feina molt feixuga i la necessitat de disposar d'un elevat nombre de recursos humans.

Per resoldre aquest problema, l' *ICEM* ha decidit encarregar-nos a nosaltres, com a consultoria externa a l' Institut, la creació d'un magatzem de dades on en una primera fase només es recolliran resultats del Campionat del Món de Pilots de Fórmula 1 des de l'any 1961 fins el 2007.

Es desitja poder obtenir, al menys, la següent informació:

- N° de grans premis guanyats per equip, tipus de pneumàtic i/o pilot
- Posició mitjana d'un pilot per mundial i/o gran premi
- Velocitats màxima i mitjana per mundial, gran premi, equip i/o pilot
- N° de retiraments de cursa per mundial, gran premi, equip, pilot i/o motiu
- N° d'anys en actiu d'un pilot

I, tenint en compte el següent sistema de puntuació

Figura 24 Taula de puntuacions

Posició	1950-1959	1960	1961-1990	1991-2002	2003-2008
1er	8 punts	8 punts	9 punts	10 punts	10 punts
2on	6 punts	5 punts	6 punts	6 punts	8 punts
3er	4 punts	4 punts	4 punts	4 punts	6 punts
4rt	3 punts	3 punts	3 punts	3 punts	5 punts
5è	2 punts	2 punts	2 punts	2 punts	4 punts
6è	1 punt	1 punt	1 punt	1 punt	3 punts
7è					2 punts
8è					1 punt

obtenir:

- Punts per mundial, gran premi, equip, tipus de pneumàtic i/o pilot
- Rànking d'equips per nombre de mundials guanyats
- Rànking de pilots per nombre de mundials guanyats
- Rànking de tipus de pneumàtic per nombre mundials guanyats

Adicionalment, hem de proporcionar un conjunt predefinit d'informes on es mostri la informació demanada i qualsevol altra que creiem que pot resultar útil per a l'anàlisi estadístic dels resultats esportius.



Finalment, també ens demanen que expliquem si trobeu que hi ha una relació entre el fet d'obtenir la *Pole Position* i el tipus de pneumàtic utilitzat, detallant quins criteris s'han seguit per arribar a la conclusió donada.

Ens comenten que les dades per carregar el magatzem de dades estan recollides en un llibre Excel que ens proporcionarà l'ICEM. Aquest Excel consta de tres fulls:

- Resultats: proporciona els resultats de les curses.
- Poles: llistat de pilots que van aconseguir la *Pole Position*.
- Pilots: llistat de pilots amb la seva nacionalitat i els anys que han estat en actiu (com a enumeració o com a franja any inicial-any final, ambdós inclosos).