

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

Teresa Aixut Caveró

ETIS

José Ángel Martín Carballo

14 Juny 2010

Agraïments

No sé com, ni recordo quan va començar tot plegat, però fa temps que vaig superar el punt de no retorn. Ara estic aquí i vull agrair el suport i comprensió dels que han hagut de suportar els nervis, el cansament i les decepcions, i que també han compartit la il·lusió i les alegries en aquest periple. Sense la vostra ajuda tot seria més difícil, més gris, més amarg... Gràcies.

Computing is not about computers any more. It is about living.

~Nicholas Negroponte

Los ordenadores son inútiles. Sólo pueden darte respuestas.

~atribuida a Pablo Picasso

Resum

En els darrers anys les organitzacions han anat agafant consciència de la importància de la gestió de la seva informació. D'aquesta necessitat d'organitzar les pròpies dades d'una forma coherent i eficient per al seu anàlisi sorgeixen els magatzems de dades. El projecte "**Construcció i explotació d'un magatzem de dades per a l'anàlisi estadístic dels resultats del Campionat de Fórmula 1**" s'emmarca dins aquest context i té com a objectiu el disseny d'un magatzem de dades per a cobrir les necessitats del Institut Català d'Esports de Motor (ICEM) en la gestió de les seves dades referents a la Fórmula 1. En aquest projecte s'han executat les diferents fases necessàries per a la creació d'un nou magatzem. En primer lloc s'ha fet l'anàlisi de requeriments i de les dades disponibles. A continuació s'ha fet el disseny i la implementació física del magatzem de dades. Seguidament s'ha procedit amb l'extracció, transformació i càrrega de dades originals (ETL) i per finalitzar s'han creat els informes. Com a resultat s'han obtingut una sèrie d'informes per a ser consumits de manera immediata pels usuaris del ICEM. Aquests informes han de permetre l'anàlisi de les dades d'una forma senzilla i àgil. A més s'ha implementat un sistema d'actualització automàtic de les dades que per al manteniment de la informació del magatzem.

La implementació del magatzem de dades s'ha fet sobre una base de dades Oracle 10g Express Edition i els informes s'ha dissenyat amb l'eina Oracle Discoverer. Per a l'automatització de les dades s'ha fet servir a més visual basic script i l'eina de càrrega SQL*Loader.

Paraules clau: Magatzems de dades, ETL, Oracle, Discoverer, Fórmula 1.

En los últimos años las organizaciones han ido tomando conciencia de la importancia de la gestión de su información. De la necesidad de organizar los datos de una forma coherente y eficiente para facilitar su análisis surgen los almacenes de datos. El proyecto "**Construcción y explotación de un almacén de datos para el análisis estadístico de los resultados del Campeonato de Fórmula 1**" se enmarca en este contexto y tiene como objetivo el diseño de un almacén de datos para cubrir las necesidades del Institut Català d'Esports de Motor (ICEM) en la gestión de sus datos referentes a la Fórmula 1. En la realización del proyecto se han ejecutado todas las fases necesarias para la creación de nuevo almacén. En primer lugar se ha hecho el análisis de requisitos y de los datos. A continuación se ha diseñado y se ha implementado físicamente el almacén. Seguidamente se ha procedido con la extracción, transformación y carga de los datos originales (ETL). Para finalizar se han creado los informes. El resultado ha sido una serie de informes que pueden ser consumidos de forma inmediata por los usuarios del ICEM y que les permitirá el análisis de los datos de una forma sencilla y ágil. Además se ha implementado un sistema de actualización automático de los datos para el mantenimiento de la información del almacén.

La implementación del almacén de datos se ha hecho sobre una base de datos Oracle 10g Express Edition y los informes se han diseñado con la herramienta Oracle Discoverer. Para la automatización de datos se ha utilizado visual basic script y la herramienta de carga SQL*Loader.

Palabras clave: Almacén de datos, ETL, Oracle, Discoverer, Fórmula 1.

In recent years organizations have become aware of the importance of managing their information. Data Warehousing intends to help by organizing data in a coherent and efficient way to ease its analysis. The project "**Construction and operation of a data warehouse for statistical analysis of the results of the Championship of Formula 1**" is set within this context and aims to design a data warehouse to meet the needs of the Institut Català d'Esports Motor (ICEM) in the management of data relating to Formula 1. The different steps for the creation of a new data warehouse have been executed. First, the analysis of requirements and available data has been conducted. In the second step the data store has been designed and it's been implemented physically. Then it has been proceeded with the extraction, transformation and loading of the original data (ETL). Finally, series of reports have been created. As a result the users of ICEM have at their disposal a group of reports ready to be consulted. These reports will simplify the analysis of data making the whole process more agile. It has also been implemented an automatic update system data for maintaining the information up to date in the data store.

The implementation of data warehouse has been done on an Oracle Database 10g Express Edition and reports have been designed with Oracle Discoverer tool. For the data updating system it has been used visual basic script and the loading tool SQL * Loader.

Key words: Data Warehouse, ETL, Oracle, Discoverer, Formula 1.

Índex de continguts

Introducció	8
Historia i estat de l'art dels magatzems de dades.	8
Justificació i context	9
Objectius	9
Enfocament i metodologia	9
Planificació	10
Pla de treball i anàlisi preliminar	11
Anàlisi de requeriments i disseny conceptual i tècnic	11
Implementació	11
Memòria i presentació virtual	12
Productes	12
Estructuració de la memòria	12
Anàlisi	13
Anàlisi de requeriments	13
Orígens de dades	13
Consideracions sobre les dades disponibles	15
Informes a generar	15
Requeriments no funcionals	16
Casos d'ús	17
Model conceptual	18
Dimensions i atributs	18
Indicadors	20
Taules de Fets	21
Viabilitat del disseny	22
Disseny	23
Arquitectura	23
Arquitectura de programari	24
Arquitectura de maquinari	24
Disseny de la base de dades	26
Disseny lògic	26
Disseny físic	26
Disseny dels processos de càrrega	27
Càrrega de dades inicial	27
Traspàs de dades al magatzem	28
Informes	29
01_Grans Premis Guanyats	29
02_Posició Mitjana Pilot	29

03_Velocitat Màxima/Mitjana _____	30
04_Retiraments _____	30
05_Anys Actiu Pilot _____	30
06_Punts totals _____	30
07_Rànquing d'equips per mundial guanyat _____	31
08_Rànquing de pilots per mundial guanyat _____	31
09_Rànquing de pneumàtics per mundial guanyat _____	31
10_Rànquing de pneumàtics per pole aconseguida _____	31
Captures _____	32
01_Grans Premis Guanyats _____	32
02_Posició Mitjana Pilot _____	33
03_Velocitat Màxima/Mitjana _____	34
04_Retiraments _____	36
05_Anys Actiu Pilot _____	37
06_Punts totals _____	38
07_Rànquing d'equips per mundial guanyat _____	39
08_Rànquing de pilots per mundial guanyat _____	40
09_Rànquing de pneumàtics per mundial guanyat _____	41
10_Rànquing de pneumàtics per pole aconseguida _____	41
Conclusions i línies de futur _____	42
Glossari _____	43
Bibliografia _____	44
Libres de text _____	44
Enllaços d'interès _____	45
Annex _____	46
Script de transformació de full Excel a fitxer csv _____	46
Fitxers de control i crides des de la línia de comandes _____	46
Resultats.csv _____	46
Poles.csv _____	47
Pilots.csv _____	47
Programació del treball d'actualització en la base de dades _____	47
Registres pre-carregats en el magatzem de dades _____	47
Procediments emmagatzemats _____	48
F1_DW.ActualitzaMagatzem _____	48
F1_DW.actualitzaCircuit _____	48

F1_DW.actualitzaEquip	49
F1_DW.actualitzGP	49
F1_DW.actualitzaMotiu	49
F1_DW.actualitzaMundial	50
F1_DW.actualitzaPilot	50
F1_DW.actualitzaPneumatic	51
F1_DW.actualitzaResultatCursa	51
F1_DW.actualitzaResultatsMundial	53
F1_DW.getGP	53
LOAD_F1.netejaRegistres	54
F1_DW.getProfit	55
LOAD_F1.esborraTaules	56
LOAD_F1.isNumeric	56

Índex de figures

<i>Figura 1 - Metodologia de desenvolupament en cascada</i>	10
<i>Figura 2 - Taula de planificació del projecte</i>	10
<i>Figura 3 - Planificació anàlisi preliminar</i>	11
<i>Figura 4 - Planificació anàlisi de requeriments i disseny conceptual i tècnic</i>	11
<i>Figura 5 - Planificació implementació</i>	11
<i>Figura 6 - Planificació memòria i presentació virtual</i>	12
<i>Figura 7 - Taula amb el sistema de puntuació dels diferents mundials</i>	14
<i>Figura 8 - Diagrama de Cas d'ús Administrador</i>	17
<i>Figura 9 - Diagrama Cas d'ús Usuari</i>	17
<i>Figura 10 - Dimensió Gran Premi</i>	18
<i>Figura 11 - Dimensió Mundial</i>	18
<i>Figura 12 - Dimensió Pilot</i>	19
<i>Figura 13 - Dimensió Pneumàtic</i>	19
<i>Figura 14 - Dimensió Equip</i>	19
<i>Figura 15 - Dimensió Motiu</i>	20
<i>Figura 16 - Dimensió Circuit</i>	20
<i>Figura 17 - Taula de Fets Resultats Cursa</i>	21
<i>Figura 18 - Taula de Fets Resultats Mundial</i>	22
<i>Figura 19 - Esquema arquitectura</i>	23
<i>Figura 20 - Arquitectura de programari</i>	24
<i>Figura 21 - Arquitectura de maquinari</i>	25
<i>Figura 22 - Disseny lògic</i>	26
<i>Figura 23 - Diagrama de base de dades</i>	27
<i>Figura 24 - Selecció del llibre de treball</i>	32
<i>Figura 25 - Grans Premis guanyats agrupats per pneumàtic, equip i pilot</i>	32
<i>Figura 26 - Grans Premis guanyats per equip amb gràfic de pastís</i>	33
<i>Figura 27 - Posició mitjana d'un pilot agrupat per mundial i gran premi</i>	33
<i>Figura 28 - Posició mitjana pilot per gran premi / per mundial</i>	34
<i>Figura 29 - Comparativa de resultats entre velocitats mitjana i màxima</i>	34
<i>Figura 30 - Velocitat mitjana i màxima per equip</i>	35
<i>Figura 31 - Evolució de les velocitats mitjanes i màximes al llarg dels mundials</i>	35
<i>Figura 32 - Nombre d'abandonament seleccionant gran premi i mundial</i>	36
<i>Figura 33 - Nombre d'abandonaments per pilot</i>	36
<i>Figura 34 - Nombre d'abandonaments per tipologia</i>	37
<i>Figura 35 - Evolució del nombre d'abandonaments per mundial</i>	37
<i>Figura 36 - Anys en actiu d'un pilot</i>	37
<i>Figura 37 - Punts obtinguts per pneumàtic, equip i pilot seleccionant gran premi i mundial</i>	38
<i>Figura 38 - Punts totals per tipus de pneumàtic</i>	38
<i>Figura 39 - Punts totals per equip</i>	39
<i>Figura 40 - Rànquing d'equips per mundial guanyat</i>	39
<i>Figura 41 - Rànquing de pilots per mundial guanyat</i>	40
<i>Figura 42 - Rànquing de pneumàtics per mundial guanyat</i>	41
<i>Figura 43 - Rànquing pneumàtics per pole aconseguida</i>	41
<i>Figura 44 - Rànquing de poles per pneumàtic en Mònaco</i>	42
<i>Figura 45 - Rànquing de poles per pneumàtic en el mundial del 1984</i>	42

Introducció

Historia i estat de l'art dels magatzems de dades.

A finals dels anys 80 apareix el concepte magatzem de dades. En essència neix per descriure el model d'arquitectura necessari per suportar la presa de decisions ens entorns operacionals. La idea aborda els problemes associats als fluxos de dades del sistema de producció per al seu anàlisi. Aquests problemes encara són vigents en la actualitat i fins i tot han empitjorat amb la capacitat d'emmagatzemar i generar més informació per part de les empreses. La realitat pràctica de la majoria d'organitzacions és que la seva informació es troba repartida entre una sèrie de sistemes heterogenis que no estan integrats. Es necessita d'un repositori comú on emmagatzemar aquestes dades de manera que faciliti consultes complexes que permetin realitzar anàlisis d'alt nivell.

Els magatzems de dades (de l'anglès *data warehouse*) són repositoris de dades històriques provinents dels diferents sistemes integrats en una organització i dissenyades específicament per tractar amb grans volums de dades i poder consultar-les de forma eficient. Per treballar-hi existeixen diferents tipus d'eines, com ara les OLAP (*online analytical processing*) o les de mineria de dades, amb les que es poden realitzar anàlisis, estadístiques i informes. Aquests tipus d'eines ens poden mostrar els resultats de forma gràfica i multidimensional per facilitar la comprensió dels resultats obtinguts.

Per suportar els magatzems de dades podem triar entre sistemes ROLAP i sistemes MOLAP. En els sistemes ROLAP es fan servir bases de dades relacionals dissenyades explícitament per suportar els requeriments dels magatzems de dades. Per aconseguir-ho es fan servir unes estructures especials anomenades d'estrella i de floc de neu. El sistema MOLAP per contra treballen directament en estructures multidimensionals. A grans trets els avantatges i desavantatges d'un sistema respecte de l'altre són que els sistemes ROLAP poden treballar amb grans volums de dades, mentre que els sistemes MOLAP tenen un major rendiment. Per aquest motiu en els últims anys han sorgit sistemes híbrids anomenats HOLAP que permeten barrejar tots dos sistemes en funció del tipus de dades que s'han de fer servir.

La dificultat en el disseny d'aquest tipus d'estructures rau en conèixer a priori la informació que es vol explotar per poder organitzar les dades adientment. En molts casos aquesta informació es desconeix i pot arribar a canviar amb el temps. Per aquest motiu la informació no s'acostuma a estructurar segons la seva funcionalitat, sinó segons els temes d'interès. En el disseny s'hauran de considerar aquesta estructuració de la informació, el volum de dades a tractar, el nivell d'agregació desitjat i com i cada quan s'actualitza la informació.

Les tendències que es poden observar actualment en l'evolució dels magatzems de dades no es resumeixen a la tria entre sistemes ROLAP i MOLAP sinó que existeixen línies de treball alternatives per mirar d'obtenir millors resultats. Un dels camps en els que s'està treballant és en el de les dades no estructurades. De fet, molta part de la informació de la que es disposa finalment no està estructurada de cap manera, per tant, és necessari començar a treballar amb aquest tipus de dades. La manera més intuïtiva de poder treballar amb aquest tipus de dades és mitjançant cerques, realitzant preguntes directes que ens retornin la informació desitjada. Els estudis en web semàntica van en aquesta direcció, de manera que les cerques siguin intel·ligents i retornin les dades que realment es volen recuperar. Aquest funcionament no només s'ha de poder utilitzar en els cercadors web, també s'han de poder aplicar a les dades d'una organització. Un altre dels punts en els que es treballa és en el de la freqüència d'actualització de les dades. Les dades han de ser el més recent possibles per a que els anàlisis pugin ser més acurats, així doncs es busca poder treballar amb les dades en temps real sense que hi hagi desfasament entre els entorns

operacionals i els analítics. Finalment, els magatzems de dades s'estan beneficiant de la proliferació d'arquitectures SOA (de l'anglès Service-Oriented Architecture) per a millorar els processos d'extracció, transformació i càrrega. Aquests tipus d'arquitectura permet integrar diferents tipus d'aplicacions i plataformes de forma transparent de manera que es permet un tractament més homogeneïtzada de les dades. Tot i que aquest tipus d'arquitectures ja es troben força integrades en els mateixos sistemes de magatzems de dades, s'hi continua treballant per millorar les seves prestacions.

Justificació i context

Aquest projecte s'emmarca dins el camp dels magatzems de dades i amb la seva realització es pretén abordar les diferents fases d'execució d'un projecte d'aquest àmbit. El propòsit del projecte és el de donar solucions a una entitat amb problemes de gestió de les seves dades basant la solució en la creació d'un magatzem de dades. Per a la implementació s'ha seleccionat un sistema ROLAP i dins els gestors de base de dades (SGBD) disponibles s'ha triat Oracle, un SGBD àmpliament estès en el mercat. Oracle proveeix tot un seguit d'eines que permeten a més realitzar les carregues i dissenyar informes, de manera que totes les eines utilitzades al llarg del projecte són d'Oracle.

Com a resultat s'ha obtingut un sistema completament funcional a punt per ser utilitzat pel client i que els pot servir com a base per a futurs projectes. És a dir, no només es presenta una solució a un problema concret sinó que se'ls proveeix amb una plataforma escalable que els hi doni solucions a llarg termini. També s'ha aconseguit una gran coneixement en l'execució d'aquest tipus de projecte des del seu inici passant per cadascuna de les fases d'anàlisi, disseny i desenvolupament.

Objectius

L'Institut Català d'Esports de Motor (ICEM) subministra dades referents als resultats obtinguts a les diferents competicions mundials d'esports de motor als diferents organismes públic i mitjans de comunicació catalans. L'elaboració dels seus informes es realitza de forma molt manual ja que actualment aquestes dades es troben emmagatzemades en fulls d'Excel. L'objectiu principal del projecte es proveir a l'ICEM d'un sistema àgil i eficient per poder realitzar els seus informes i que a més li permeti facilitar l'anàlisi de les seves dades, tot i que en aquest projecte, com a primera fase, només es recolliran els resultats dels mundials de Formula 1.

L'objectiu personal en la realització d'aquest TFC és l'aprofundiment en el coneixement de l'estructura i disseny de magatzems de dades, així com en l'ús de tècniques i eines disponibles.

Per acabar l'objectiu inherent a un treball de fi de carrera és el de realitzar un treball de síntesi dels coneixements adquirits al llarg dels estudis realitzats posant-los en pràctica en un projecte concret. Al llarg de la durada d'aquest treball es pretén millorar l'anàlisi d'un problema complex, aprendre a planificar correctament i estructurar-lo aplicant una metodologia adient. Com a resultat s'espera obtenir un solució realista al problema que se'ns ha presentat i l'elaboració d'una memòria i una presentació que ajudi a comprendre tot el desenvolupament que s'ha dut a terme.

Enfocament i metodologia

Aquest projecte dona solució a un problema concret d'una entitat, és tracta d'un projecte autònom i te una sèrie de requeriments clarament definits. Aquestes característiques fan que la metodologia de desenvolupament triada sigui la metodologia clàssica en cascada (si existeix molts dubtes sobre els requeriments i són difícils d'obtenir el més convenient es fer servir la metodologia en espiral).

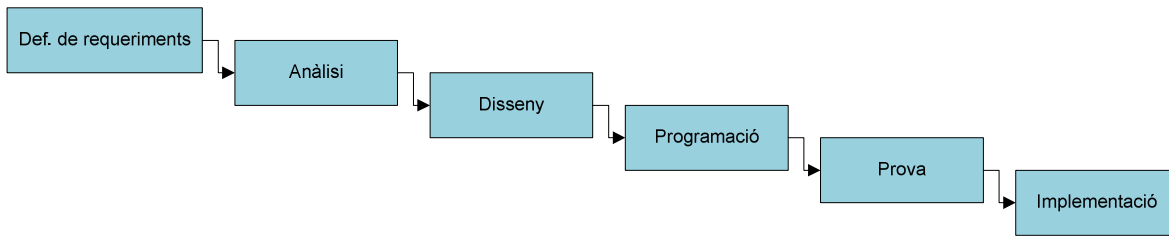


Figura 1 - Metodologia de desenvolupament en cascada

En aquest cas la definició de requeriments ja està feta, per tant aquesta fase no s'inclou en el desenvolupament del projecte. En la fase d'anàlisi, a partir dels requeriments obtinguts en la fase inicial, s'obtingran els components necessaris per al disseny posterior del magatzem de dades. En la fase de programació es prepararan les dades fent tasques de desnormalització, d'agregació d'informació, d'historificació, etc. També es crearan els informes segons les indicacions donades. La fase de prova no es farà directament amb el client, però si es rebrà un feedback de les entregues realitzades per poder fer canvis si s'escau abans de la implementació final. En la última fase s'entregarà la implementació de tot el projecte juntament amb la memòria i la presentació virtual.

Planificació

La planificació del projecte be fixada pel calendari de l'assignatura on es fixen una sèrie de fites amb un lliurament a cada fita. En la següent taula es mostren les diferents entregues que s'han d'entregar en cadascuna de les dates fixades:

Fita	Document	Descripció del document
10 Març 2010	PAC 1 Pla de treball i anàlisi preliminar de requeriments	Document amb la planificació del projecte i anàlisi inicial de requeriments del sistema.
13 Abril 2010	PAC2 Anàlisi de requeriments i disseny conceptual i tècnic	Document on després de realitzar un anàlisi més detallat dels requeriments s'obté el disseny conceptual i físic de la base de dades. També s'entregarà una descripció a alt nivell dels processos de càrrega.
17 Maig 2010	PAC3 Implementació	S'entrega la màquina virtual on s'ha desenvolupat el projecte. A més s'entrega un document per a indicar com accedir als diferents entorns, un altre on es descriu el funcionament dels processos de càrrega i finalment captures dels informes generats.
14 Juny 2010	Memòria i presentació virtual	Entrega de la màquina virtual amb les últimes modificacions. Entrega de la memòria i la presentació en format vídeo.

Figura 2 - Taula de planificació del projecte

Pla de treball i anàlisi preliminar

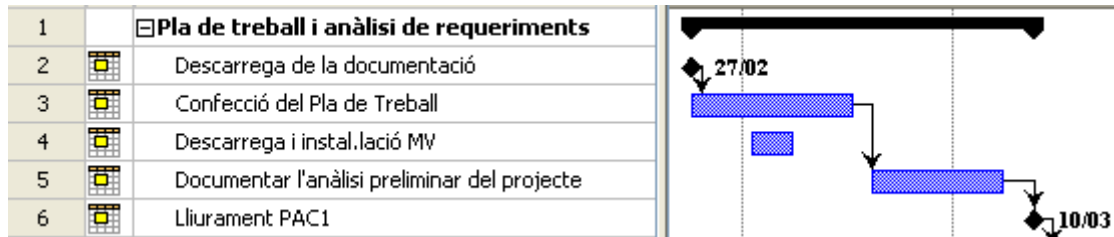


Figura 3 - Planificació anàlisi preliminar

Anàlisi de requeriments i disseny conceptual i tècnic

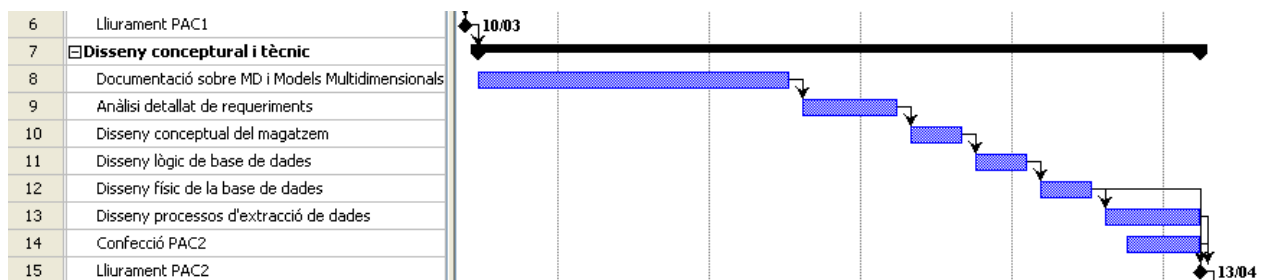


Figura 4 - Planificació anàlisi de requeriments i disseny conceptual i tècnic

Implementació

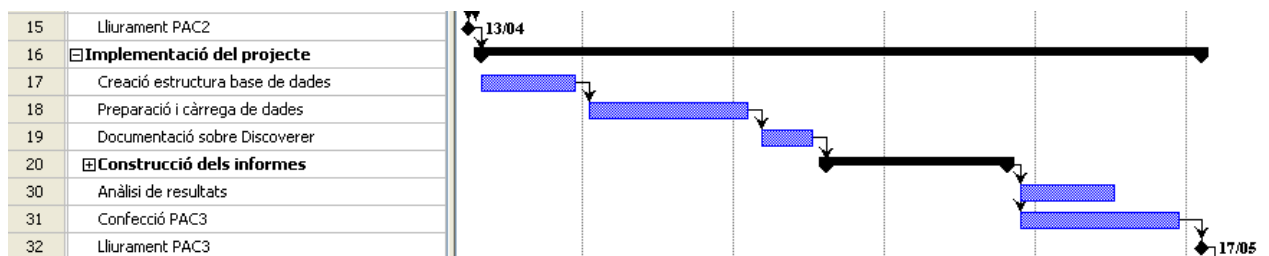


Figura 5 - Planificació implementació

Memòria i presentació virtual

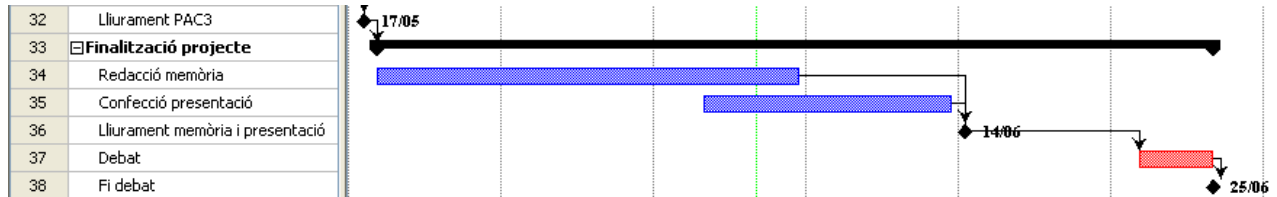


Figura 6 - Planificació memòria i presentació virtual

Productes

Al llarg d'un projecte es genera tota una sèrie de productes que es van entregant a les diferents fites del projecte. En aquest cas el llistat de productes generats ha estat:

- Pla de treball: document amb la planificació i un anàlisi preliminar del projecte.
- Anàlisi de requeriments i disseny conceptual i tècnic: document detallat dels components del projecte i les funcionalitats s'han de crear. Inclou també una descripció detallada de la solució proposada
- Accés eines: document explicatiu de l'accés a les diferents eines de la plataforma creada.
- Processos de càrrega: document on es detalla el funcionament dels processos de càrrega de dades implementats.
- Informes: document on es mostren captures de cadascun dels informes creats amb una breu descripció.
- Implementació: com a resultat final del projecte s'entrega un disc virtual on hi ha implementada una plataforma amb un magatzem de dades dedicat a la Fórmula 1. Aquesta plataforma inclou la base de dades, els llibres de treball i els processos de càrrega.
- Memòria: aquest document. Document detallat de tot el treball realitzat en el projecte.
- Presentació virtual: presentació multimèdia on es mostra breument tot el projecte.

Estructuració de la memòria

En aquest capítol s'ha fet un breu resum de la història dels magatzems de dades i les noves tendències en aquest camp per situar al lector en aquesta àrea. Després s'ha posat el projecte dins el context, s'han explicat els objectius, la metodologia i els resultats de manera que el lector pugui tenir una visió general de tot el projecte.

En el següent capítol s'entrarà en detall en l'anàlisi de requeriments i es mostrarà el disseny conceptual resultant d'aquest anàlisi.

A continuació es mostraran els diferents dissenys implementats. S'ensenyaran les propostes d'arquitectura tant de software com de hardware per a la implementació del model de magatzem de dades obtingut en la fase d'anàlisi així com el disseny lògic i físic de la base de dades on es crearà.

Per acabar aquest capítol també es mostrarà el recull d'informes finals generats per a ser utilitzats per l'usuari final. En el següent capítol es mostraran les captures realitzades d'aquests informes.

Finalment es mostraran les conclusions del projecte i les seves possibles línies de futur. En la memòria també s'hi inclou un glossari, la bibliografia i els annexos.

Anàlisi

Anàlisi de requeriments

En la primera fase del projecte s'ha realitzat un anàlisi dels requeriments per detectar les necessitats del client i acotar els límits del projecte. S'ha fet un estudi de les dades de les que es disposa i els informes que es desitja poder consultar. També s'ha fet un recull dels requeriments no funcionals del projecte. Com a resultat d'aquest anàlisi s'han obtingut:

- les dimensions: una dimensió representa un punt de vista des del que una dada pot ser analitzada.
- els atributs: són el conjunt de valors que serveixen per descriure, seleccionar o agrupar una dimensió.
- els indicadors: mesures que permeten identificar un fet que es vol analitzar.

Orígens de dades

Es parteix d'un document Excel amb informació dels campionats de Formula 1 dels anys 1961 al 2007. En aquest tipus de format no se'ns garanteix la coherència de les dades: es poden trobar dades duplicades, valors nuls, canvis de formats, etc. que s'hauran de tractar adientment en realitzar les carregues. Les dades estan distribuïdes en tres fulls diferents:

1. Resultats: En aquest full hi ha la posició de cadascun dels pilots en les diferent carreres que s'han disputat en el període de temps que estem tractant. Cadascun dels camps indicats es troben en la seva corresponent columna:
 - Circuit/Data: En aquest camp es troba el nom del circuit i la data de la carrera. Es fa servir la barra al inici del camp i com a separador entre els dos camps. El format de la data és un format llarg en anglès.
 - Pilot: Nom i cognom del pilot.
 - Equip: Nom de l'equip.
 - Pneumàtic: Marca del pneumàtic.
 - Posició: Posició en la carrera en el cas que l'hagi acabat, sinó acaba la cursa no hi ha dada i tindrem, en canvi, el motiu de l'abandonament en la columna corresponent. Es tracta d'un valor numèric.
 - Voltes: Nombre total de voltes realitzades. Es tracta d'un valor numèric.
 - Temps: En aquest cas el valor varia en funció de la posició del corredor en la carrera. En el cas del primer es dona la durada de la seva carrera, a continuació la

- diferència de temps respecte al primer i quan la diferència supera la volta es passa a una mesura en voltes.
- Velocitat (Km/h): La velocitat mitjana del vehicle en la carrera, tot i que no sempre es disposa d'aquesta dada.
 - Motiu abandonament: En el cas d'abandonaments en aquesta casella es guarda el motiu.
2. Poles: En aquest full es troben les poles position aconseguides en cada carrera. Els camps estan ordenats en columnes seguint el següent ordre:
- Circuit: Nom identificatiu del circuit on s'ha fet la cursa.
 - País: País al que pertany el circuit.
 - Data: Data en que s'ha aconseguït la Pole position.
 - Pilot: El nom del pilot que ho ha aconseguït.
 - Temps: Temps en que se ha realitzat la volta.
3. Pilots: Finalment es mostra la informació de tots el pilots que han corregut en la Formula 1. Les dades es troben ordenades per columnes:
- País: País d'origen del pilot.
 - Pilot: Nom del pilot.
 - Anys en actiu: Anys en els que ha estat en actiu. Els diferents anys es troben separats per comes o per intervals separats amb guions si són consecutius.

També es troba disponible una taula amb els diferents sistemes de puntuacions que s'han fet servir des del 1950 fins al 2008. Aquesta taula ha de permetre calcular els rànquings dels diferents mundials dels que es tenen dades. S'haurà de tenir en compte que la taula només arriba al 2008 i a mesura que en el sistema es vulguin afegir nous mundials s'hauran d'actualitzar els punts en funció de les noves normatives que s'apliquin.

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

Figura 7 – Taula amb el sistema de puntuació dels diferents mundials

Consideracions sobre les dades disponibles

Després de realitzar un petit anàlisi sobre les dades disponibles s'han vist una sèrie de factors que s'hauran de tenir en compte posteriorment en la càrrega de dades i que cal remarcar.

En primers lloc, s'han detectat una sèrie d'errors en les dades: valors que no es corresponen amb el que s'espera en la columna, dades incompletes que no aporten informació i registres repetits.

Pel que fa als formats de les dades es pot veure que els formats data canvien entre les diferents pàgines. Es veu, a més, com les mateixes dades en diferents fulles s'escriuen i es mostren també de forma diferent (ex: noms del pilots). També es pot comprovar com algunes de les dades arriben en la mateixa columna separades amb caràcters especials. Caldrà controlar també els espais en blanc al inici de les cadenes de caràcters.

Referent a les dades de Resultats s'ha observat que en un mateix mundial hi poden haver diferents curses en un mateix país, per tant, s'haurà de tenir en compte a l'hora de crear les dimensions i fets corresponents.

Per finalitzar, de la fulla de pilots s'han de descartar aquells que han estat actius únicament abans de 1961, ja que el tractament de les dades es fa a partir de 1961. Per una altra banda, cal contrastar les dades d'aquesta fulla amb les de resultats per comprovar que siguin correctes.

Informes a generar

En el document de requeriments es demana poder generar com a mínim el següent conjunt d'informes:

- N° de grans premis guanyats per equip, tipus de pneumàtic i/o pilot.
A partir de les dades de gran premi s'ha de poder trobar el total de títols guanyats podent seleccionar l'equip, el tipus de pneumàtic i/o el pilot. En total es tracta de fins a 7 tipus de consultes diferents. Es farà servir com indicador la posició, en aquest cas posició = 1 equival a haver guanyat la cursa.
- Posició mitjana d'un pilot per mundial i/o gran premi.
S'agruparà per gran premi o per mundial o tots dos alhora. Com a indicador es farà servir la posició.
- Velocitats màxima i mitjana per mundial, gran premi, equip i/o pilot.
S'utilitzarà la velocitat com a indicador per realitzar aquests informes. Cal aclarir que el valor de velocitat que hi ha a l'origen de dades fa referència a la velocitat mitja de cada cursa, per tant, la velocitat màxima serà la màxima de les velocitats mitjanes. Es podrà seleccionar entre mundial, gran premi, equip i/o pilot. En total es podrà disposar fins a 19 informes diferents.
- N° de retiraments de cursa per mundial, gran premi, equip, pilot i/o motiu.
Com a indicador es farà servir el booleà abandonament, que ens indica si hi hagut retirament quan pren el valor '1'. En aquest cas podem agrupar per mundial, per gran premi, equip, pilot i/o motiu.
- N° d'anys en actiu d'un pilot.
En aquest cas no cal fer servir cap mesura. Tan sols es vol comprovar si en un mundial un pilot ha estat o no en actiu i per comprovar-ho no cal cap mesura associada, n'hi ha prou amb la existència de la instància de la taula de fets que es faci servir per a obtenir aquest resultat.

- Punts per mundial, gran premi, equip, tipus de pneumàtic i/o pilot.
Es faran servir els punts com a indicador i podem agrupar per mundial, gran premi, equip, tipus de pneumàtic i/ o pilot.
- Rànquing d'equips per nombre de mundials guanyats.
En aquest cas es farà servir el booleà mundial guanyat per trobar si un equip ha guanyat el mundial o no. L'equip guanyador serà el del pilot guanyador del mundial.
- Rànquing de pilots per nombre de mundials guanyats.
Com a indicador es tornarà a utilitzar el booleà mundial guanyat. Sumant el total de punts per pilot aconseguit al llarg d'un mundial es pot trobar el pilot guanyador com aquell que n'ha aconseguit més.
- Rànquing de tipus de pneumàtic per nombre mundials guanyats.
A l'igual que en els dos informes anteriors es farà servir el booleà mundial guanyat. En aquest cas es suposarà que el pneumàtic guanyador de cada mundial és el que ha fet servir el pilot guanyador al llarg del mundial, si n'ha fet servir més d'un constarà aquell amb el que hagi obtingut més punts (si el nombre de punts fos igual constarien tots dos).

Per últim es demana com a part dels requeriments trobar una relació entre aconseguir la Pole Position i el tipus de pneumàtic. Aquesta informació es pot obtenir a partir del rànquing de pneumàtics per Pole Position. La relació entre els pneumàtics i poles es pot extreure de la relació entre pilots, cursa i pole. Per a realitzar aquest últim informe es farà servir el booleà pole position que indicarà quan un pilot ha aconseguit la pole en una cursa.

A partir de l'anàlisi dels informes a realitzar ja es pot extreure quines han de ser les dimensions i els indicadors que poblaran les taules de fets.

Requeriments no funcionals

A nivell de programari un dels requeriments que es demanen en aquest projecte és treballar sobre una màquina virtual de VirtualBox. El sistema operatiu instal·lat en aquesta màquina virtual és Microsoft Windows XP i la base de dades amb la que treballar Oracle 10g Express Edition. Tota la programació es realitzarà amb PL/SQL i per realitzar els informes es farà servir Oracle Discoverer. En aquest cas, el fet d'utilitzar la versió Express Edition limita en el disseny físic de la base de dades en la utilització d'índexs de tipus bitmap, ja que no es troben disponibles en aquesta versió.

Per finalitzar, no s'ha fet cap indicació sobre els rols dels usuaris que han de fer servir el sistema, per tant, només es farà una autenticació bàsica d'usuari en tots els informes generats.

Casos d'ús

Un cop es disposa de tota la informació de dades disponibles i necessitats de client es mostraran els diagrames de casos d'ús amb una breu descripció de cadascun. En la definició de requeriments no s'ha donat indicacions dels tipus de rols que accediran al sistema, per tant, s'assumirà que només hi haurà dos tipus: un rol Administrador que s'encarregarà del manteniment del sistema (càrrega de dades, crear nous llibres de treball, noves funcionalitats...) i un rol usuari que només pot accedir als llibres de treball per treballar-hi.

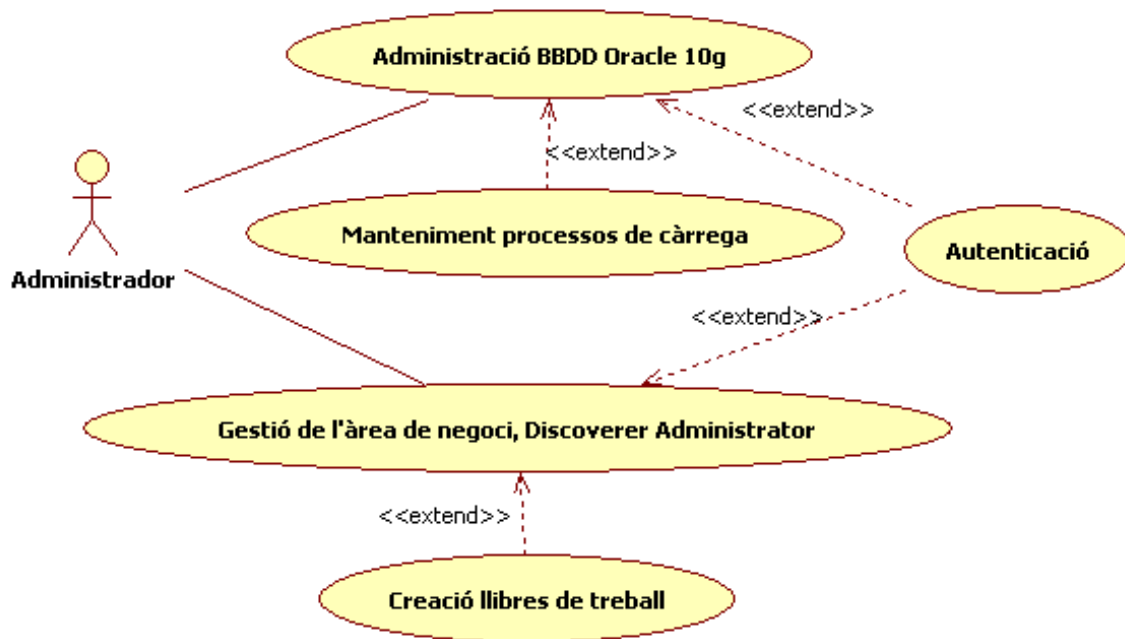


Figura 8 - Diagrama de Cas d'ús Administrador

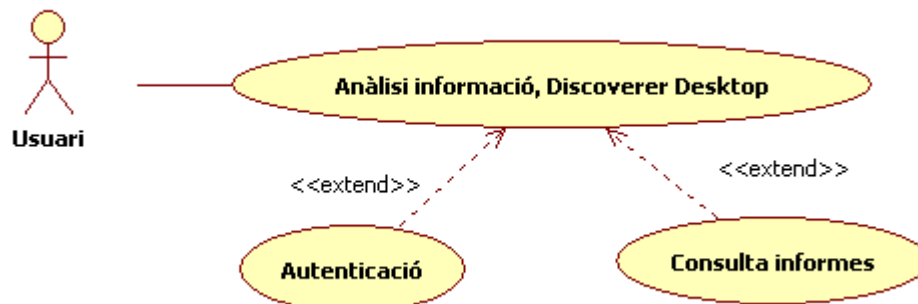


Figura 9 - Diagrama Cas d'ús Usuari

Model conceptual

En base a l'anàlisi d'informes a generar en aquest apartat es citaran, en primer lloc, totes les dimensions detectades amb els seus atributs associats i a continuació els indicadors que es faran servir. A partir de les dimensions i els indicadors, es dissenyaran els fets i es definirà la granularitat.

Dimensions i atributs

Gran Premi

Per poder seleccionar o agrupar segons Gran Premi s'ha de crear com a dimensió. Com a atributs només es farà servir el nom. Tot i que es pogués estar temptat d'afegir com a atributs el circuit i el país es pot veure en l'anàlisi de dades que els Grans Premis no sempre es celebren en el mateix circuit i fins i tot, n'hi ha alguns que canvien de país (ex: Gran Premi d'Europa).

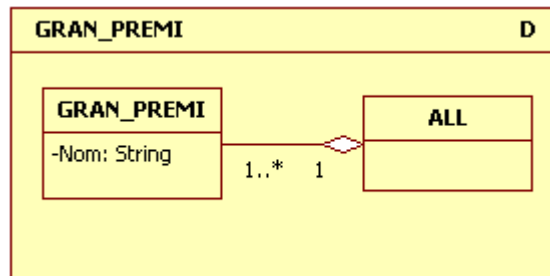


Figura 10 - Dimensió Gran Premi

Mundial

De la mateixa manera que amb els Grans Premis s'ha creat una dimensió per poder realitzar agrupacions per mundials. En aquest cas l'únic atribut que hi haurà és l'any, ja que cada mundial es celebra al llarg d'un únic any.

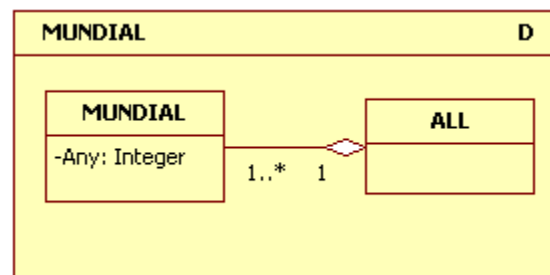


Figura 11 - Dimensió Mundial

Pilot

S'ha creat la dimensió pilot que fa referència, als pilots que han participat en les carreres que s'estan tractant. En aquesta dimensió s'emmagatzemarà el nom del pilot i el país del que prové (tot i que és una dada que no s'ha demanat explícitament, es podria aprofitar-la en un futur). Existeix la temptació d'afegir l'equip com a atribut, però un pilot pot formar part de diferents equips al llarg de la seva carrera professional, per tant, és lògic mantenir dos dimensions separades.

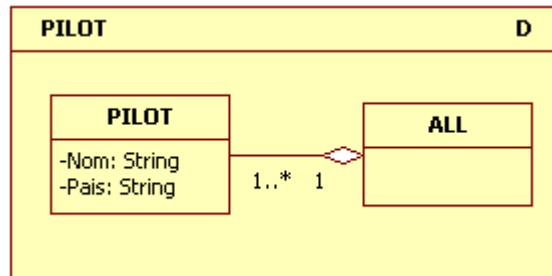


Figura 12 - Dimensió Pilot

Pneumàtic

La dimensió pneumàtic fa referència als diferents tipus de pneumàtics que s'han fet servir. D'aquesta dimensió només s'ha emmagatzemat el nom.

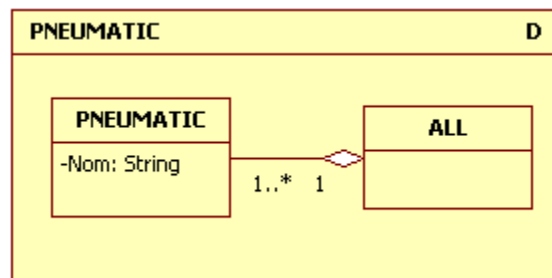


Figura 13 - Dimensió Pneumàtic

Equip

La dimensió equip mostra tots els equips que participen en la Fórmula 1. Només s'ha emmagatzemat el nom com atribut d'aquesta dimensió.

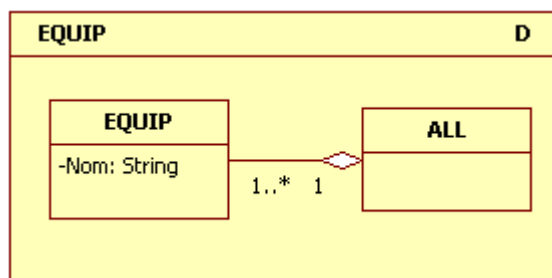


Figura 14 - Dimensió Equip

Motiu

S'ha creat una dimensió on s'emmagatzemaran els tipus de motiu que han provocat els diferents abandonaments en les carreres. Com a atribut es disposarà d'un text descriptiu del motiu.

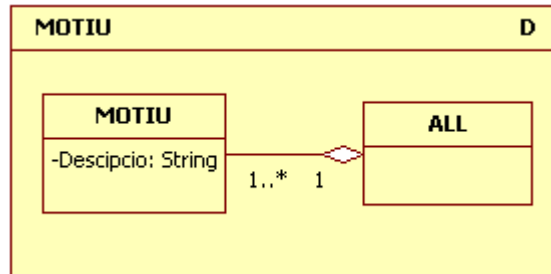


Figura 15 - Dimensió Motiu

Circuit

Tot i que ens els informes que s'han demanat no s'utilitza la dimensió circuit per realitzar cap consulta, s'ha creat aquesta dimensió per poder assignar correctament les curses als grans premis corresponents i a més, pot servir en el futur per generar nous tipus d'informe (d'aquesta manera s'evita haver de refer l'estructura).

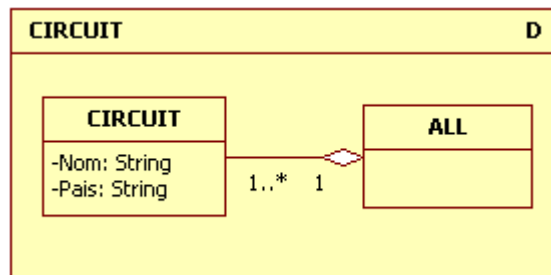


Figura 16 - Dimensió Circuit

Indicadors

A continuació es detallen els indicadors que s'han detectat al llarg de l'estudi de requeriments:

- Posició: Aquest valor permet, a més de conèixer la posició d'un pilot en una cursa saber si el pilot ha guanyat, en aquest cas pren el valor 1.
- Mundial guanyat: Es tracta d'un booleà indicatiu de si s'ha guanyat o no el mundial (calculat a partir de la suma de puntuacions dels grans premis que formen el mundial).
- Velocitat: Es tracta de la velocitat mitjana de cada pilot en una cursa.
- Punts: Puntuació obtinguda en cada premi. Aquest valor s'obté a partir de la posició en la carrera d'un pilot i el sistema de puntuació aplicat en l'any en que s'ha corregut.
- Abandonament: Indicador si hi ha hagut un abandonament en la cursa.

- Pole Position: Booleà que indica si el pilot ha obtingut la pole position en una cursa.

Per acabar, per obtenir els anys en actiu d'un pilot no hi ha cap tipus de mesura concreta que es pugui aplicar. En aquest cas cal definir el fet i amb la instància de cel·la que defineixi aquest esdeveniment n'hi ha prou.

Taules de Fets

Les dimensions i els indicadors s'agrupen en funció dels esdeveniments als que fan referència i, d'aquesta manera es facilita l'explotació de dades posterior. Com a resultat d'aquesta agrupació i dels informes que s'han de generar s'han obtinguts les següents taules de fets.

Resultats Cursa

El primer dels fets que es vol representar és el resultat d'una cursa. Les mesures associades a aquest fet són la posició final en la cursa, que a més serveix per obtenir les curses guanyades, la velocitat mitjana en la cursa, els punts obtinguts, si hi ha hagut abandonament, i un booleà que indica si a més s'ha aconseguit la pole. Totes les mesures es trobaran en la mateixa cel·la ja que tenen la mateixa granularitat. La representació de la taula de fets envoltada de les seves dimensions com es veu en la figura també es coneix com a estrella.

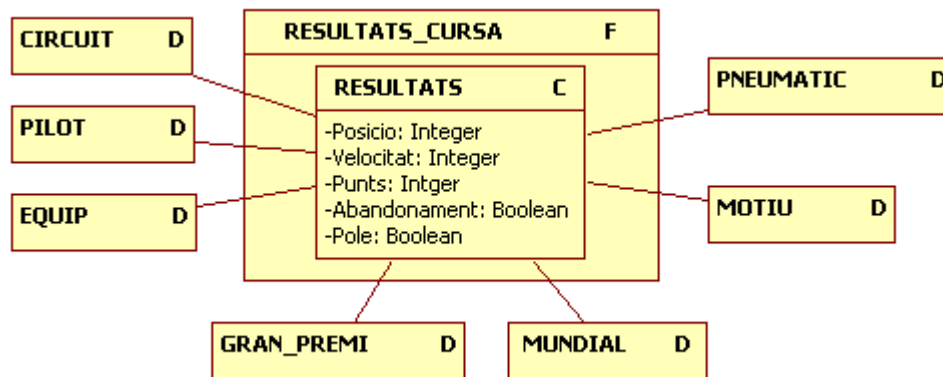


Figura 17 - Taula de Fets Resultats Cursa

Aquesta taula de fets és la que té la granularitat més petita de les que s'han dissenyat, ja que les mesures estan associades als resultats d'un pilot en una cursa que és la unitat mínima proporcionada. Per trobar la cursa concreta només cal trobar la intersecció entre les dimensions mundial i gran premi.

Resultats Mundial

En aquesta figura es representa el resultat d'un pilot en un mundial. Com a mesura associada es disposa d'un únic valor de tipus booleà que indica si s'ha guanyat o no el mundial. El fet en si també pot donar informació, en aquest cas dels anys en actiu d'un pilot ja que l'aparició d'un pilot en un any ja indica que aquell any va estar actiu.

Aquesta taula té un gràdul més gran ja que s'agrupa a nivell de mundial i per tant es perd tota referència amb les curses que hi formen part, però per realitzar els rànquings de mundials guanyats aquesta estructura facilita les consultes i millora la eficiència.

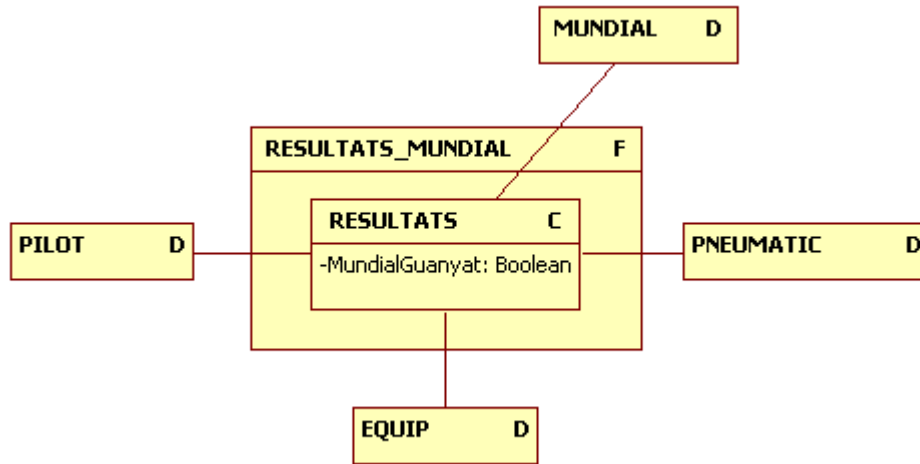


Figura 18 - Taula de Fets Resultats Mundial

Viabilitat del disseny

Un cop acabat el disseny conceptual cal realitzar una estimació de l'espai que ocuparan totes les dades per comprovar que la implementació sigui viable. Per dur a terme aquest estudi es considerarà només el que pot arribar a ocupar les instàncies dels fets.

En la primera de les estrelles, *Resultats Cursa*, el nombre d'instàncies del fet ha de coincidir amb el nombre de registres vàlids de la fulla de resultats del llibre d'Excel que ens han donat amb les especificacions del projecte. El volum de cel·les amb el que es tractarà està al voltant de les 15.000. En aquesta taula de fets tenim 3 mesures que són enters i dos booleans. Suposem que cada enter ocupa 32 bits i el booleà només 8 (en Oracle no existeix el tipus booleà, per tant, es farà servir un char(1)), en total fan 112 bits. També s'ha de considerar els identificadors de cada dimensió que hi haurà en cada cel·la. En aquest cas hi ha 7 dimensions, suposant que com a identificador es farà servir un RowId (o clau substituïda) que ocupa 4 bytes la mida és de 224 bits. En total tenim 336 bits per 15000 cel·les que implica una grandària aproximada de 615 Kbytes. La mida final sembla més que raonable per al seu tractament amb el sistema que ha d'explotar les dades.

En la segona taula el nombre de cel·les disminueix, en aproximadament un factor 10 respecte a la primera, ja que agrupem a nivell de mundial. Tenint en compte que en aquesta taula només es guarda com a mesura un booleà i a més disminueix el nombre de dimensions no cal realitzar els càlculs per veure que no hi ha cap problema en el seu emmagatzemament.

Disseny

Arquitectura

Per a donar suport la presa de decisions a una entitat, en aquest cas l'ICEM, no n'hi ha prou amb el disseny d'un magatzem de dades. Aquest magatzem de dades s'ha d'integrar amb un conjunt de components que giren al seu entorn. Tot el conjunt es coneix com a factoria d'informació corporativa (FIC).

L'arquitectura total d'una FIC recull molts components que en aquest cas no s'han contemplat degut a les condicions específiques de requeriments del projecte:

- Les dades tractades estan orientades a un tema específic i només dona solucions a un grup d'usuaris concret.
- El volum de dades amb el que es tracta i la previsió de creixement és baix.
- El nivell d'actualització de les dades es força baix (com a màxim es poden tenir actualitzacions mensuals).
- No es necessita una historicitat detallada de la informació.

Per aquests motius s'ha decidit que tot i que en un futur pot resultar força interessant per l'ICEM la creació d'un magatzem de dades corporatiu, en aquest projecte només es construirà un magatzem de dades departamental (en aquest tipus de magatzem a diferència del corporatiu només es guarden les dades que interessen al departament o grup d'usuaris, tant a nivell de tipus de dades com de la seva història). En la següent figura es pot observar els components que componen l'arquitectura d'aquest projecte:

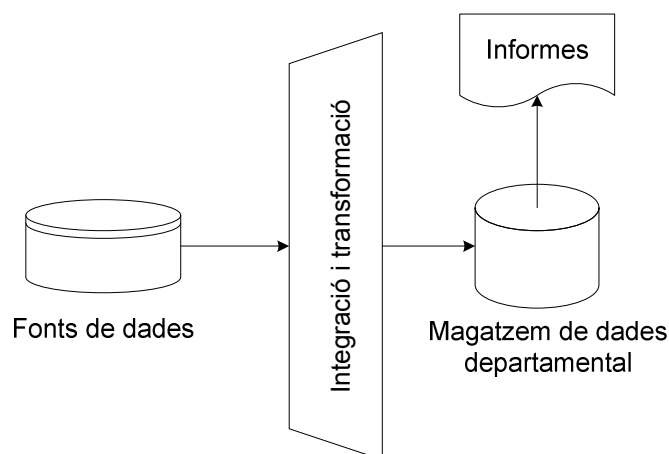


Figura 19 - Esquema arquitectura

Arquitectura de programari

A continuació es mostren els components de programari que formen part de l'arquitectura del projecte:

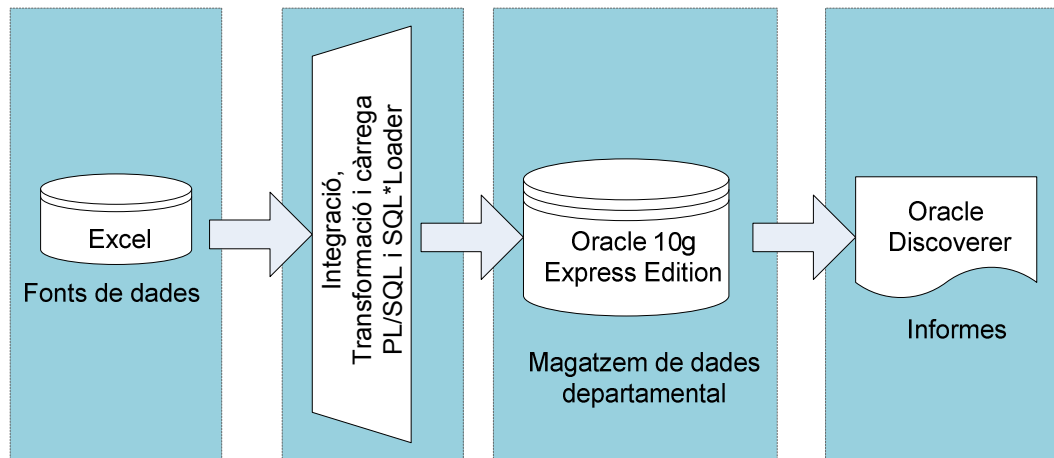


Figura 20 - Arquitectura de programari

Les dades arriben en format Excel i a través de l'eina SQL*Loader es carreguen en la base de dades. Un cop dins la base de dades i fent ús de procediments programats en PL/SQL s'emmagatzemen en l'estructura del magatzem de dades departamental. Aquestes dades posteriorment seran consumides a través d'informes amb Oracle Discoverer.

A nivell de programari serà necessari:

- Oracle SQL-Developer: O qualsevol altra eina per programar procediments sobre la base de dades Oracle, com pot ser APEX, TOAD, ...
- Oracle 10g Express Edition: En aquest SGBD s'emmagatzemarà el magatzem de dades. Es tracta de la versió gratuïta d'Oracle i té certes limitacions, si es volgués crear un entorn escalable, probablement caldria utilitzar una de les seves versions comercials.
- Oracle Discoverer: Conjunt d'eines que permeten explotar el magatzem de dades per al seu anàlisi. La versió Administrator s'utilitza per a l'administració de l'estructura de treball i la versió Desktop és la que faran servir els usuaris i que caldrà instal·lar en tots aquells ordinadors des dels que es vulgui consultar les dades.

Arquitectura de maquinari

En aquest cas els requeriments de maquinari no són molt exigents degut a la senzillesa del projecte. Si es volgués aprofitar la base de dades Oracle per crear nous magatzems departamentals, o fins i tot, el magatzem corporatiu caldria replantejar la estructura i muntar la base de dades sobre una cabina de discs i d'aquesta manera aconseguir més robustesa i escalabilitat.

A continuació es mostra l'arquitectura de maquinari proposada:

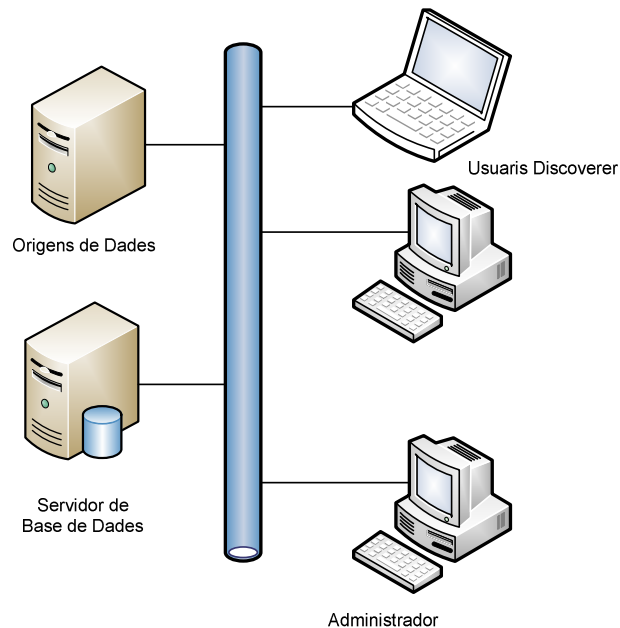


Figura 21 - Arquitectura de maquinari

El servidor de base de dades es situaria en un sector de xarxa diferent a la dels terminals dels usuaris. Els orígens de dades han d'acomplir com a única condició necessària que es puguin deixar els fitxers en format csv en el servidor de base de dades.

Els requeriments mínims de maquinari són:

- Servidor de base de dades:
 - Sistema Operatiu: Linux Red Hat Enterprise Linux RHEL3 and RHEL4 o Windows 2000 Server SP4
 - RAM: 1G (la versió Express està limitada a 1G de RAM)
 - Espai disc dur: 20G (la versió XE te limitat l'espai de base de dades en 4G)
 - Oracle 10g Express Edition
- Estació de treball administrador:
 - Sistema Operatiu: Windows XP Professional SP1
 - RAM: 2G
 - Espai disc dur: 100G
 - Client administrador base de dades Oracle
 - Oracle Discoverer Administrator
 - Eina de desenvolupament amb Oracle (SQL Developer, TOAD,...)
- Estació de treball usuaris:
 - Sistema Operatiu: Windows XP Professional SP1
 - RAM: 1G
 - Espai disc dur: 50G
 - Oracle Discoverer Desktop

Disseny de la base de dades

Disseny lògic

Un cop obtingut el disseny conceptual el següent pas en el disseny dels magatzems de dades és passa al model lògic. Tot i que existeixen diferents tipus d'implementacions dels models multidimensionals, un dels requeriments del projecte és la realització del magatzem en un sistema ROLAP.

Per al disseny lògic de la estrella s'ha de tenir en compte que la taula de fets estarà lligada amb les taules de dimensió que formen part de l'estrella mitjançant claus foranes. La clau primària de la taula estarà formada per un subconjunt de les claus foranes de les seves dimensions. Per reduir la grandària de la taula de fets la clau primària de les dimensions es forma mitjançant substituïts, tret en el cas de mundial on l'any ens serveix com a clau.

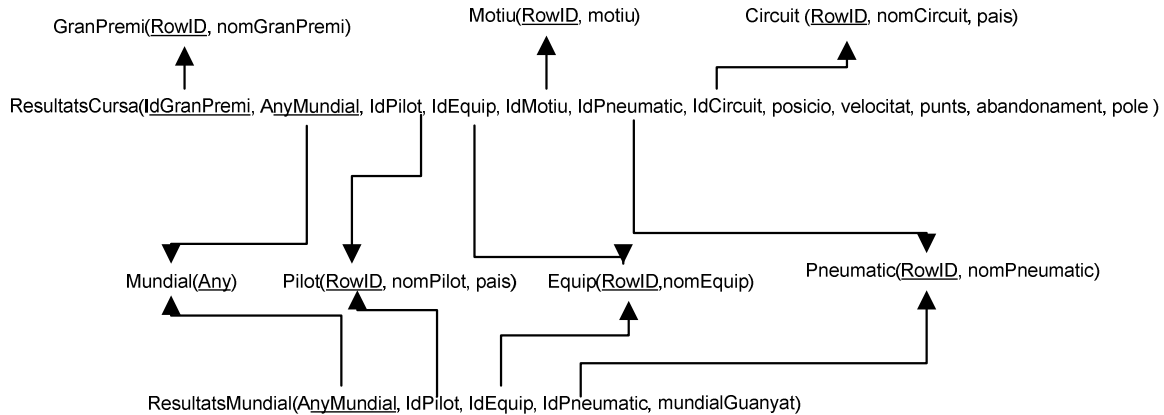


Figura 22 - Disseny lògic

Com es pot apreciar en aquest esquema hi ha dimensions que es reutilitzen entre les dos estrelles. Aquesta compartició de dimensions permet realitzar *Drill-across* (moviment entre taules de fets que comparteixen dimensions, es canvia el fet, però es mantenen les dimensions) entre les dos estrelles.

Cal indicar també que l'ús de claus substituïdes a més d'estalviar espai permeten fer canvis en les dimensions o corregir errors que s'han realitzat en la càrrega.

Disseny físic

Un cop que les relacions ja estan definides cal realitzar el disseny físic sempre tenint present que l'objectiu d'aquest disseny és la eficiència en el temps de resposta de les consultes que es volen realitzar.

En la figura següent es mostra l'esquema de base de dades, però d'una manera més propera a com es visualitzarà en la base de dades amb les taules, les seves claus primàries i foranes. El paràmetre PK representa la clau primària de la taula, FK indica les claus foranes i els valors en negreta indiquen que el valor és obligatori i per tant, no pot ser nul.

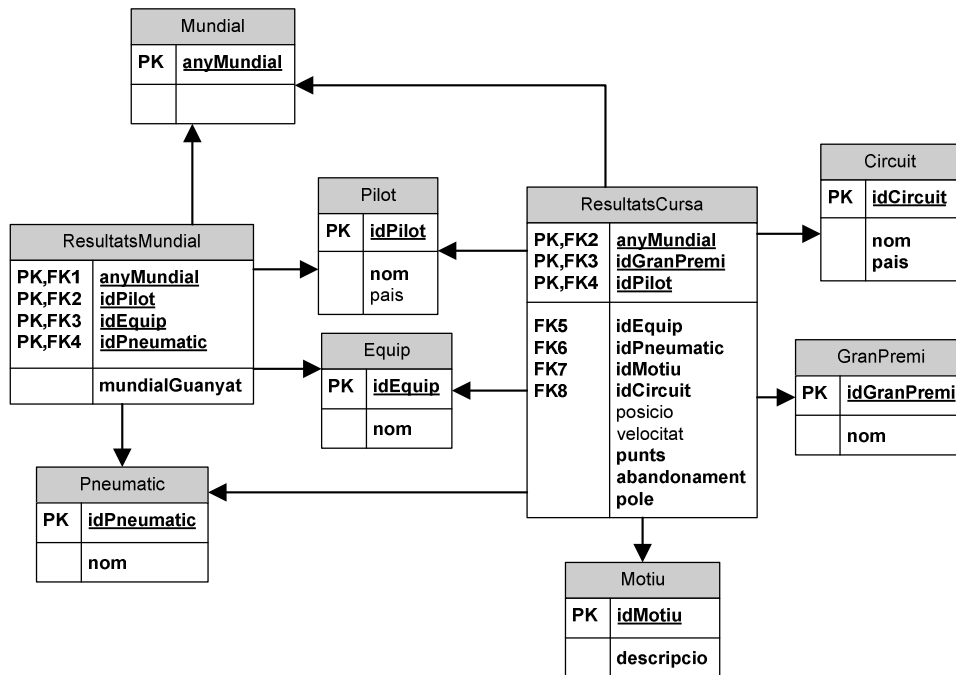


Figura 23 - Diagrama de base de dades

Quan es cerca la manera de millorar l'accés de les consultes a les dades relacionades entre diferents taules, com passa en els magatzems de dades, és mitjançant índexs de combinació. En aquest cas la clau primària de les taules de fets ja és un índex de combinació, però l'ordre en que definim els atributs d'aquest tipus d'índex és important ja que dependrà del tipus de consulta que es faci sobre la taula. Per aquest motiu l'índex està definit en primer lloc sobre mundial, ja que creiem que la majoria de consultes discriminaran per mundial. S'ha decidit que l'índex sobre les dos taules resultats sigui un índex agrupat. El seu cost és baix (suposem que les dades només s'actualitzen d'any en any, per tant, sempre s'inseriran al final) i en canvi, per determinades consultes pot millorar la cerca.

Sobre els atributs idEquip, idPneumatic i idMotiu, de les taules de fets s'ha decidit a més crear un índex de mapa de bits. En els tres casos el nombre de valors possibles és baix i tenen una gran selectivitat. El seu cost en espai és molt baix, però permet millorar molt l'accés a les dades.

Disseny dels processos de càrrega

En el primer apartat s'explicarà com es realitza la càrrega de dades des del fitxer Excel cap a la base de dades. A continuació es mostraran els scripts que s'han generat per fer les càrregues inicials i els que es faran servir per integrar les noves dades així com la seva automatització.

Càrrega de dades inicial

El primer punt que s'ha de tractar és com carregar les dades del fitxer Excel en la base de dades per al seu tractament posterior. Oracle consta d'una eina que permet la càrrega d'una manera senzilla i bastant potent des de fitxers csv cap a la base de dades, per tant, el primer pas consistirà

en automatitzar la generació de csv a partir de fitxers Excel. Per realitzar aquesta transformació s'ha de servir visual basic script, que d'una manera senzilla converteix cada pàgina d'un excel en un fitxer csv amb el nom de la fulla (en els annexes es pot veure el codi).

Després d'executar l'script es generaran 3 fitxers en el mateix directori: Resultats.csv, Poles.csv i Piltos.csv. Per passar-los cap a la base de dades s'ha fet servir SQL*Loader. Cada csv generat s'ha carregat a la seva taula corresponent. Per fer-ho cal definir el fitxer de control adient indicant quins són els separadors, la taula de destí i la relació entre camps. Tant les crides com els scripts es poden trobar en els annexes del projecte.

Tot plegat ens permetria automatitzar la càrrega, fent les crides des d'una aplicació o fins i tot en un bat on s'inclourien totes les instruccions juntament amb l'esborrat de fitxers. Aquest mateix fitxer es podria programar amb les tasques programades de windows per a que s'executés segons la freqüència amb la que finalment es passin les dades cap al magatzem de dades.

Traspàs de dades al magatzem

Un cop les dades ja es troben carregades en taules a la base de dades cal fer el traspàs a les taules definitives. Per realitzar aquest traspàs s'ha creat un treball en la base de dades, configurat amb una periodicitat setmanal (el criteri de programació vindrà donat pels requeriments de client i s'ha de sincronitzar amb la càrrega dels fitxers) que executa un procediment encarregat de fer les crides necessàries per actualitzar totes les taules del magatzem.

El primer pas d'aquest procediment, serà eliminar registres de les taules que no compleixin una sèrie de criteris definits. Per fer-ho es realitzarà una crida a LOAD_F1.netejaRegistres que eliminarà els registres que es considerin erronis i els traspassarà a taules auxiliars per si és volgués fer una segona comprovació manual.

Per dur a terme la neteja corresponent de la taula Resultats es buscarà que:

- En la primera columna hi ha d'haver el nom del circuit que ha de coincidir amb el nom que s'ha obtingut de la taula poles. També hi ha d'haver la data en un format correcte (l'any obtingut s'ha de trobar dins l'interval marcat en la taula Mundial).
- En la segona columna s'ha de poder extreure el nom del pilot que ha de coincidir amb el que ja s'ha extret de la taula pilots.
- La resta de columnes no són necessàries per fer la càrrega, ja que podem suplir-les amb valors comodí, però igualment cal fer una revisió dels valors que hi ha entrats i prendre les mesures necessàries:
 - El nom de l'equip ha de trobar-se dins la taula Equip que ja s'ha carregat (sinó està disponible se li pot assignar el registre corresponent a 'Unknown')
 - El nom del tipus de pneumàtics també s'ha de trobar dins la taula Pneumatic (sinó està disponible se li pot assignar el registre corresponent a 'Unknown')
 - La posició pot no estar disponible, però si hi és ha de ser un valor numèric positiu i no superior a 30 (per posar-li un límit superior)
 - La velocitat pot no estar disponible, però en cas de ser-hi ha de ser un valor numèric positiu.
 - El motiu de l'abandonament pot estar buit, en aquest cas en funció de si el valor de posició està buit o no es poden produir dos situacions: si el valor de posició està buit suposarem que efectivament hi ha hagut un abandonament, però es desconeix el motiu, per tant, se li assignarà el registre 'Unknown', en cas contrari s'entendrà que ha acabat la cursa i per tant se li assignarà el

registre 'No Retirement'. Si el motiu no està en blanc ha de coincidir amb algun dels registres de la taula motiu.

- La resta de columnes que no han estat referides no es fan servir en el magatzem de dades, per tant, s'ignoraran.

En la càrrega inicial de dades i abans de traspasar les dades a les taules de dimensions es crearan una sèrie de registres que es consideren adients per reflectir la falta d'algunes dades en la càrrega, com ara el tipus de pneumàtic o el motiu de l'abandonament. També es crearan els registres corresponents als grans premis celebrats que no es corresponen directament amb el país on es celebren.

Seguidament s'aniran cridant els procediments que omplen les diferents taules del magatzem, un per cada taula. En totes les taules amb una clau substituïda s'ha creat un trigger que omplirà aquest valor automàticament i facilitarà el procés de càrrega. Sempre s'actualitzaran primer les taules de dimensions i després les de fets. Per fer la càrrega de la taula ResultatCursa s'han creat una taula (GPExcepcions) i una funció auxiliar (getGP) per obtenir el nom correcte del gran premi en cada cursa.

Per finalitzar, quan s'hagin executat tots s'esborraran els registres de les taules de càrrega preparant-les per a la propera actualització.

En l'annex s'ha afegit tot el codi generat per cadascun dels procediments creats i que permeten realitzar tant la primera càrrega de dades inicial (sempre i quan s'hagi previst introduir els registres anteriorment indicats) com la resta d'actualitzacions.

Informes

S'han generat fins a 10 llibres de treball diferents, segons els requeriments inicials. En alguns d'ells s'ha creat més d'una fulla amb vistes diferents per facilitar la feina a l'usuari que ho vulgui consultar:

01_Grans Premis Guanyats

Es demanava poder visualitzar el nombre de grans premis guanyats per equip, tipus de pneumàtic i/o pilot. Per realitzar aquest informe s'ha optat per col·locar el element Gran Premi en la capçalera de l'informe. D'aquesta manera podem cercar les dades de cada gran premi individualment. A continuació s'han agrupat la resta d'elements agrupats segons l'ordre: pneumàtic, equip i pilot. De tota manera l'usuari té llibertat per plegar, desplegar i canviar l'ordre de tots els elements.

En molts casos es vol visualitzar el nombre de premis guanyats agrupat segons un dels elements. Per aquest motiu, s'han generat diferents fulls on s'han anat eliminant elements per deixar només aquells que puguin interessar.

02_Posició Mitjana Pilot

En el segon informe es demana mostrar la posició mitjana d'un pilot per mundial i/o gran premi. S'han creat tres fulles diferents en aquest llibre per mostrar les diferents agrupacions possibles. En la primera prèvia selecció del pilot es mostra la posició mitjana d'un pilot en cada mundial en el que ha participat i el detall dels grans campionats per on ha passat. En la segona es mostra la posició mitjana d'un pilot en tots els grans premis en els que ha participat i finalment la seva posició mitjana en cada mundial. S'ha volgut indicar de les posicions que representen podi marcant-les de color verd.

03_Velocitat Màxima/Mitjana

En aquest informe es demana mostrar les velocitats màxima i mitjana per mundial, gran premi, equip i/o pilot. Com en els llibres anteriors s'han creat diferents fulles mostrant, eliminant i canviant elements de posició, obtenint múltiples vistes i resultats en funció de les agrupacions creades. Se n'han creat fins a 7 fulles diferents, canviant i posant elements en la capçalera.

En una de les vistes s'ha creat un codi de colors per mostrar les velocitats marcant en colors vermellós/rosats velocitats altes i en blaus velocitats baixes d'aquesta manera mentre es navega es pot tenir un referent de com d'alta o baixa és una velocitat respecte als totals recollits en el magatzem. S'han fet vuit divisions amb salts de 10 en 10 entre els límits superiors i inferiors. Els límits que s'han fet servir per a la velocitat mitjana i per la màxima ha estat diferents, ja que la velocitat màxima sempre serà igual o superior.

En aquest informe s'ha adjuntat també un gràfic on es mostra l'evolució de les velocitats màximes i mínimes al llarg dels diferents mundials.

04_Retiraments

S'ha demanat mostrar el número de retiraments de cursa per mundial, gran premi, equip, pilot i/o motiu. Com s'ha fet en els informes anteriors s'han generat varies fulles per representar les diferents combinacions dels elements que en formen part, concretament sis. Les variables que es volen fixar per analitzar s'han col·locat en la capçalera de l'informe.

En la vista on es mostren el total d'abandonaments d'un pilot s'ha volgut remarcar aquells que han tingut més de 5 abandonaments amb una gradació del taronja al vermell a mesura que augmenta el nombre d'abandonaments. D'aquesta manera es veuen ràpidament els pilots que han patit més contratemps.

A banda dels informes s'ha creat un gràfic del nombre d'abandonaments que s'han produït al llarg de la història de la formula 1, agrupant les dades a nivell de mundial.

05_Anys Actiu Pilot

En aquest informe es demanava el nombre d'anys en actiu d'un pilot. Aquesta informació s'ha tret de la taula ResultatsMundial contant el nombre d'anys distintos en que apareix un registre del pilot. Cal remarcar que no es pot contar directament el nombre de registres d'un pilot, ja que es pot donar el cas que un pilot en un mundial hagi corregut per dos escuderies diferents.

06_Punts totals

Un altre dels informes requerits és el nombre de punts per mundial, gran premi, equip, tipus de pneumàtic i/o pilot. En aquest cas s'han generat 4 fulls diferents (s'ha considerat que consultar el nombre de punts total d'un mundial o d'un gran premi sense intervenir cap altre element no té massa sentit, i si fos necessari l'usuari se'l podria crear ell mateix). S'ha de tenir en compte, que al llarg d'un mundial un equip pot córrer amb pneumàtics diferents, així que depenent de la consulta poden aparèixer equips i pilots repetits i el total de punts no seran els que es mostren en un registre sinó la suma dels registres on apareixen. Per aquest motiu l'usuari pot canviar l'ordre dels elements de la pàgina.

En aquest informe també s'ha adjuntat un gràfic on es mostra el nombre total de punts aconseguits segons la marca de pneumàtics. Si tots els elements de la capçalera de la pàgina es seleccionen amb el total, el resultat és el nombre de punts aconseguits per les diferents marques al llarg de tots els anys que s'han introduït.

07_Rànquing d'equips per mundial guanyat

Els últims informes que es demanaven fan referència a rànquings de mundials guanyats. En aquest cas en concret es mostra el resultat de rànquing per equips on es mostra el total de mundial guanyats per cada equip ordenat de forma descendent (ens dona un total de 47 mundials que són precisament el nombre de mundials carregats).

08_Rànquing de pilots per mundial guanyat

Seguint amb els rànquings en aquest informe es demanava el de pilots per nombre de mundials guanyats. L'informe segueix el mateix patró que l'anterior però en aquest cas s'ha seleccionat l'element pilot per realitzar la llista.

09_Rànquing de pneumàtics per mundial guanyat

Com a últim informe de rànquing que s'ha demanat es mostra el de tipus de pneumàtic per nombre mundials guanyats, on es mostren les diferents marques en ordre descendent per mundial guanyat.

10_Rànquing de pneumàtics per pole aconseguida

Finalment es demanava com a part dels requeriments buscar una relació entre aconseguir la Pole Position i el tipus de pneumàtic. Per trobar aquesta relació s'ha generat un informe on es mostra el percentatge de poles aconseguides per cada tipus de pneumàtic i també s'ha afegit el percentatge d'ús per part dels pilots dels diferents tipus de pneumàtic (evidentment si només es fa servir un tipus de pneumàtic és lògic que la pole s'aconsegueixi amb aquest tipus de pneumàtic).

Aquest informe però no és suficientment acurat ja que la dada que realment ens dona aquesta relació es saber per cada carrera quants tipus de pneumàtics diferents al que ha guanyat la Pole hi havia. Per aquest motiu s'ha creat una funció que calcula per cada pole aconseguida el valor d'haver-ho aconseguit respecte a la competència i en el cas d'haver-la perdut valorar aquesta pèrdua. Per realitzar aquesta funció es comprova quants pilots han participat en la cursa i quants d'aquests duen el mateix tipus de pneumàtic. Si s'aconsegueix la pole es divideix el total d'altres tipus de pneumàtics entre el total de participants (és a dir, si tots fan servir el mateix tipus de pneumàtic no te mèrit en la obtenció de la victòria, de forma contrària te més valor si és l'únic participant a fer servir aquell tipus de pneumàtic). Si per contra no s'obté la pole el valor que es dona és un valor negatiu entre el nombre de participants amb el pneumàtic analitzat i el total de pilots (si no s'ha aconseguit la pole i el nombre total de pneumàtics és elevat penalitza més que si només n'hi ha un). Amb aquest valor podem fer la mitja en un gran premi o mundial per veure com afecta, en qualsevol cas, els valors zero no es comptabilitzen ja que no te sentit valorar una victòria o una derrota quan no hi ha rival.

Si es fa la mitja de tots els mundials i grans premis introduïts el resultat indica que els pneumàtics que donen millor resultat són els de Goodyear, seguit de Michelin i els que donen el pitjor són Pirelli i Avon amb molt poca diferència. Si es busquen dades més detallades i es mira els resultats per al gran premi de Mònaco s'obté que el millor pneumàtic és el de Michelin amb força diferència respecte als seus seguidors. Així doncs, es bastant possible que en funció de les característiques del circuit pugui canviar la relació del millor pneumàtic per obtenir la pole. Si es miren les dades a nivell de mundial també hi ha variacions respecte a la resposta obtinguda, així en el 1984 Michelin apareix com a millor opció mentre que Goodyear surt com la pitjor de les opcions i és que en més de 40 anys de fórmula poden haver-hi molts canvis.

Caldria fer aquests tipus d'anàlisis acotant per espais temporals o en funció de les tipologies de circuit, d'aquesta manera s'obtinarien resultats més fiables a l'hora de buscar aquest tipus de relacions. A nivell general, però, i basat en totes les dades disponibles, les dos marques Goodyear i Michelin surten com les millor opcions per calçar els fórmules si es vol obtenir la pole.

Captures

En primer lloc es mostra la pantalla de selecció en el Discoverer dels llibres de treball disponibles:

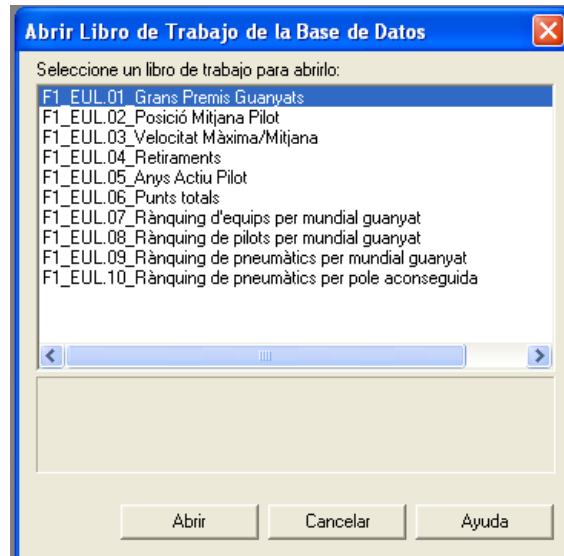
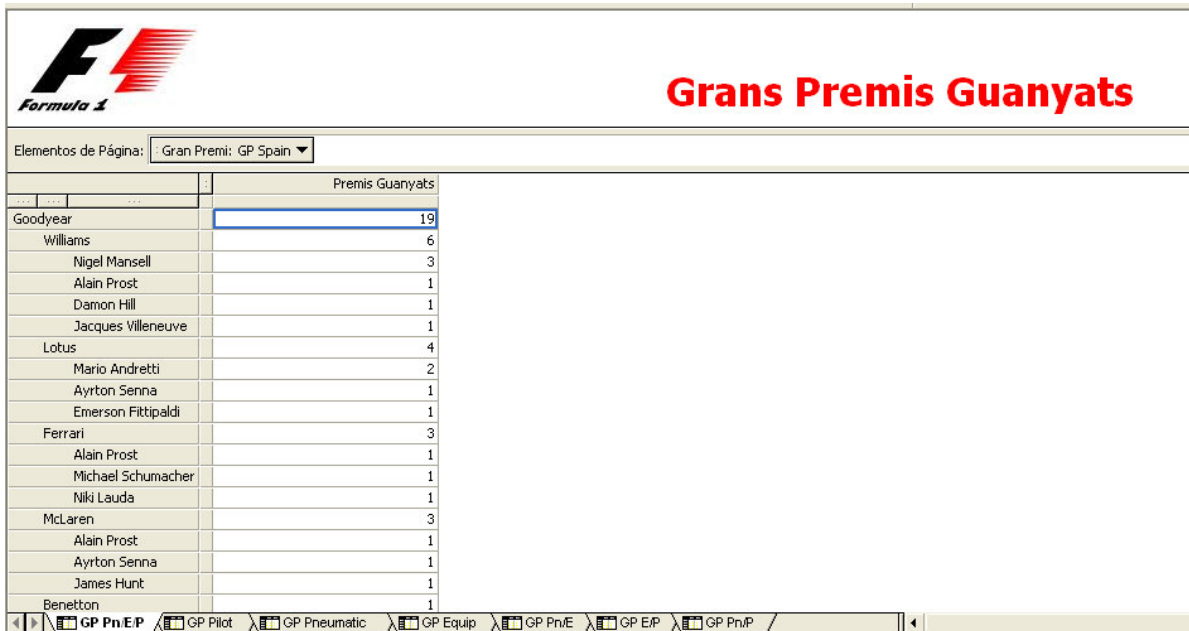


Figura 24 - Selecció del llibre de treball

01_Grands Premis Guanyats

En aquesta captura es mostra el resultat de victòries del gran premi d'Espanya agrupat per pneumàtic, equip i finalment pilot.



Grands Premis Guanyats	
Elementos de Página: Gran Premi: GP Spain	
	Premis Guanyats
Goodyear	19
Williams	6
Nigel Mansell	3
Alain Prost	1
Damon Hill	1
Jacques Villeneuve	1
Lotus	4
Mario Andretti	2
Ayrton Senna	1
Emerson Fittipaldi	1
Ferrari	3
Alain Prost	1
Michael Schumacher	1
Niki Lauda	1
McLaren	3
Alain Prost	1
Ayrton Senna	1
James Hunt	1
Benetton	1

Figura 25 - Grands Premis guanyats agrupats per pneumàtic, equip i pilot

En aquest cas es mostra un rànquing de tots els grans premis guanyats per equip amb un gràfic on es representa visualment:

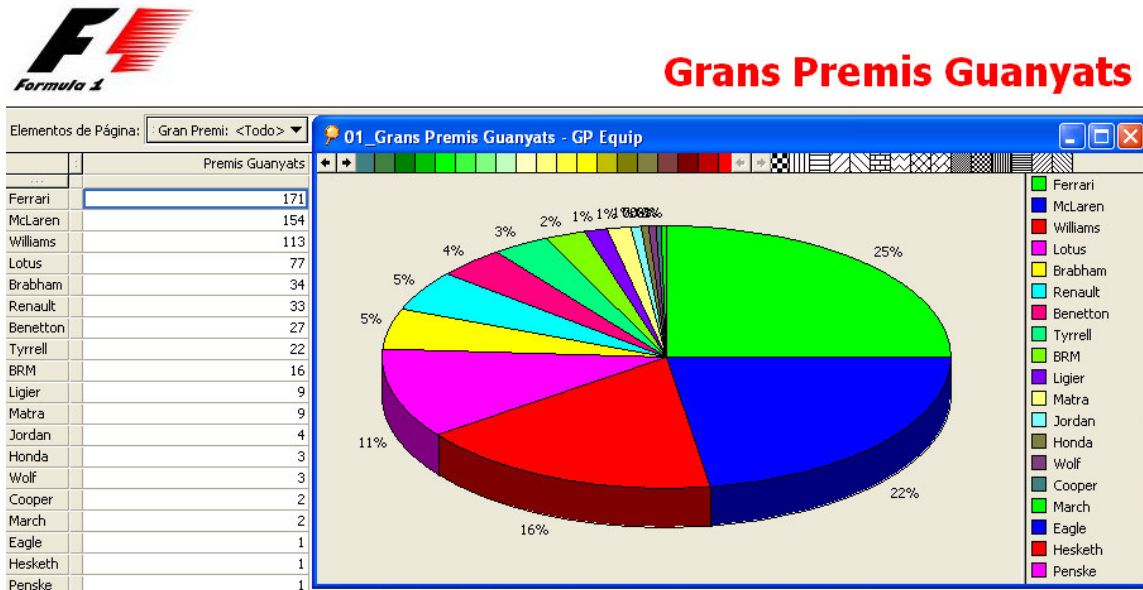


Figura 26 - Grans Premis guanyats per equip amb gràfic de pastís

02_Posició Mitjana Pilot

En aquesta captura es mostra les posicions mitjanes del pilot per gran premi i mundial. S'ha volgut destacar amb color verd aquelles posicions que indicarien podi.

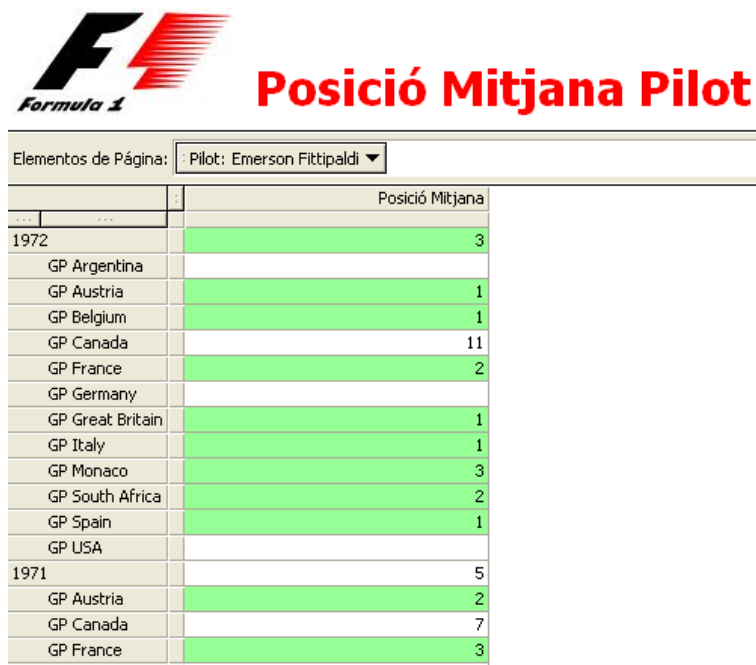


Figura 27 - Posició mitjana d'un pilot agrupat per mundial i gran premi

En la següent es mostra la posició mitjana d'un pilot en tots els gran premis en els que ha participat i de la mateixa manera i al costat es pot veure la posició mitjana en cada mundial en el que ha corregut.

Elementos de Página: Pilot: Michael Schumacher ▼		Posició Mitjana
GP Argentina	1	
GP Australia	3	
GP Austria	3	
GP Bahrain	2	
GP Belgium	1	
GP Brazil	3	
GP Canada	2	
GP China	7	
GP Europe	2	
GP France	2	
GP Germany	3	
GP Great Britain	3	
GP Hungary	4	

Elementos de Página: Pilot: Mika Hakkinen ▼		Posició Mitjana
2001	4	
2000	2	
1999	2	
1998	2	
1997	5	
1996	5	
1995	5	
1994	5	
1993	3	
1992	6	
1991	11	

Figura 28 - Posició mitjana pilot per gran premi / per mundial

03_Velocitat Màxima/Mitjana

Captura de pantalla del mateix informe però amb la selecció d'any mundial diferents per veure com la diferència de colors ens permet comparar ràpidament unes velocitats amb altres.

Elementos de Página: Gran Premi: <Todo> ▼ Any Mundial: 1963 ▼		
	Velocitat Mitja	Velocitat Màxima
Brabham	168	201
Dan Gurney	165	198
David Prophet		
Jack Brabham	171	201
BRM	164	203
Graham Hill	160	198
Lorenzo Bandini	169	170
Maurice Trintignant	197	197
Moises Solana		
Richie Ginther	161	203
BRP	169	184
Innes Ireland	169	184
Cooper	151	202
Bruce McLaren	161	202
Jo Bonnier	142	199
John Love	143	143

Elementos de Página: Gran Premi: <Todo> ▼ Any Mundial: 2005 ▼		
	Velocitat Mitja	Velocitat Màxima
BAR	201	244
Anthony Davidson	186	186
Jenson Button	206	244
Takuma Sato	197	241
Ferrari	199	242
Michael Schumacher	197	242
Rubens Barrichello	202	242
Jordan	195	237
Narain Karthikeyan	194	229
Tiago Monteiro	196	237
McLaren	204	247
Alexander Wurz	208	208
Juan Pablo Montoya	204	247
Kimi Raikkonen	203	246
Pedro de la Rosa	205	205

Figura 29 - Comparativa de resultats entre velocitats mitjana i màxima de dos mundials diferents

En la següent captura es veu la mateixa informació però, s'ha reduït un nivell i es mostren les dades agrupades a nivell d'equip:

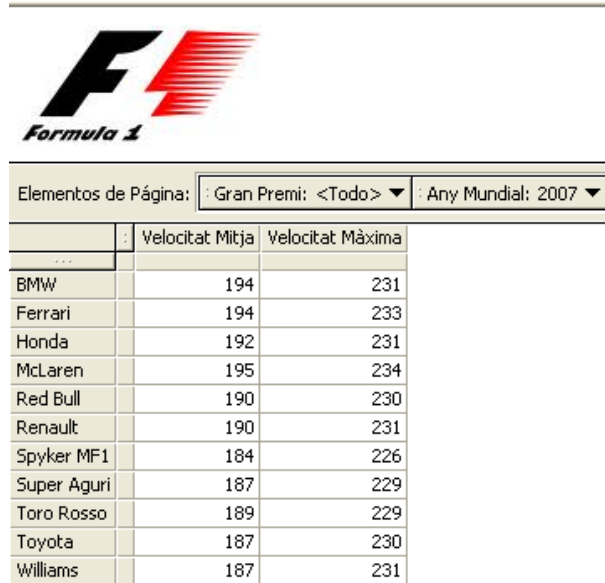


Figura 30 - Velocitat mitjana i màxima per equip

En aquesta imatge es veu gràficament l'evolució de les velocitats màximes i mínimes al llarg dels diferents mundials. En aquest cas l'informe mostra les velocitats només a nivell de mundial:

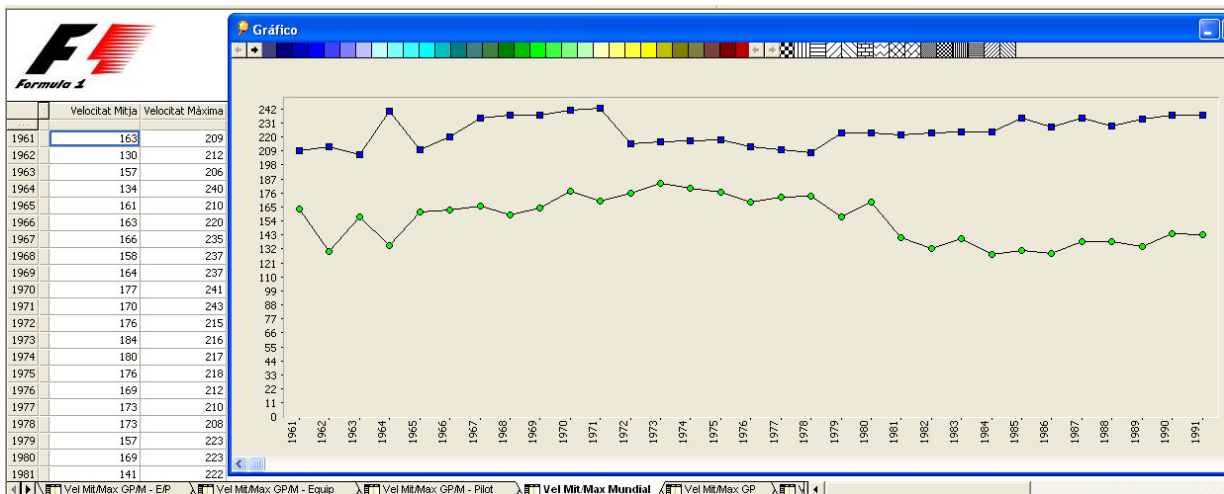


Figura 31 - Evolució de les velocitats mitjanes i màximes al llarg dels mundials

04_Retiraments

En la següent captura es mostra el número de retiraments produïts en el 61 agrupats per tipologia d'abandonament, equip i pilot.

...	Numero Abandonaments
Accident	12
BRM	1
Graham Hill	1
Cooper	4
Jack Brabham	1
John Surtees	2
Walt Hansgen	1
Ferrari	4
Giancarlo Baghetti	1
Willy Mairesse	1
Wolfgang von Trips	2
Lotus	3
Gerry Ashmore	1
Henry Taylor	1
Jim Clark	1
Brakes	1
Lotus	1

Figura 32 - Nombre d'abandonament seleccionant gran premi i mundial

A continuació es mostra el nombre d'abandonaments per pilot podent fixar en la capçalera el gran premi, el mundial, el motiu i l'equip. Quan hi ha més de 5 abandonaments es marca la fila d'un altre color per remarcar-ho:

...	Numero Abandonaments
Alain Prost	4
Alan Jones	3
Andrea de Cesaris	8
Ayrton Senna	9
Christian Danner	2
Derek Warwick	9
Eddie Cheever	12
Elio de Angelis	5
Francois Hesnault	4
Gerhard Berger	7
Huib Rothengatter	5
Ivan Capelli	1
Jacques Laffite	8
Jonathan Palmer	7

Figura 33 - Nombre d'abandonaments per pilot

En un informe similar es mostra el nombre d'abandonaments per diferents motius que s'han produït segons la selecció de mundial, gran premi, equip i pilot:

Elementos de Página: Gran Premi: GP Spain Mundial: 2000 Equip: <Todo> Pilot: <Todo>

	Numero Abandonaments
Collision	2
Engine	1
Gearbox	1
Hydraulics	1
Spun Off	1

Figura 34 – Nombre d'abandonaments per tipologia

Finalment es mostra l'evolució del nombre d'abandonaments de forma gràfica que s'han produït al llarg de la història de la formula 1, agrupant les dades a nivell de mundial:

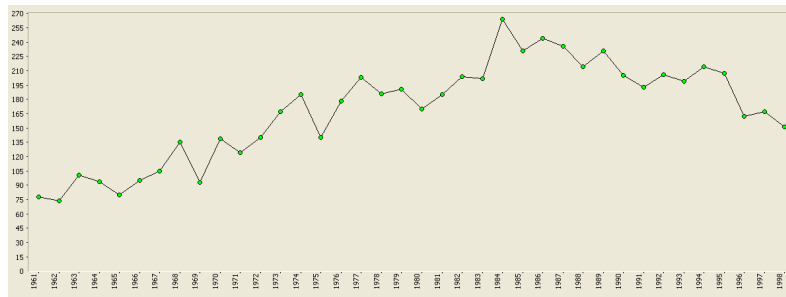


Figura 35 - Evolució del nombre d'abandonaments per mundial

05_Anys Actiu Pilot

En aquesta captura es mostra un rànquing de pilots per any en actiu:

	Anys en Actiu
Riccardo Patrese	17
Michael Schumacher	16
Andrea de Cesaris	15
Graham Hill	15
Nigel Mansell	15
Rubens Barrichello	15
Chris Amon	14
David Coulthard	14
Gerhard Berger	14
Mario Andretti	14
Michele Alboreto	14
Nelson Piquet	14
Alain Prost	13
Jacky Ickx	13
Jacques Laffite	13
Jean Alesi	13
Niki Lauda	13
Giancarlo Fisichella	12

Figura 36 - Anys en actiu d'un pilot

06_Punts totals

En aquesta captura es pot veure el nombre total de punts aconseguits al gran premi de Luxemburg al llarg dels anys agrupat per pneumàtic, equip i pilot (un pilot que hagi corregut més d'un cop en aquest gran premi en diferents equips sortirà més d'un cop amb els punts aconseguits per a cada equip).

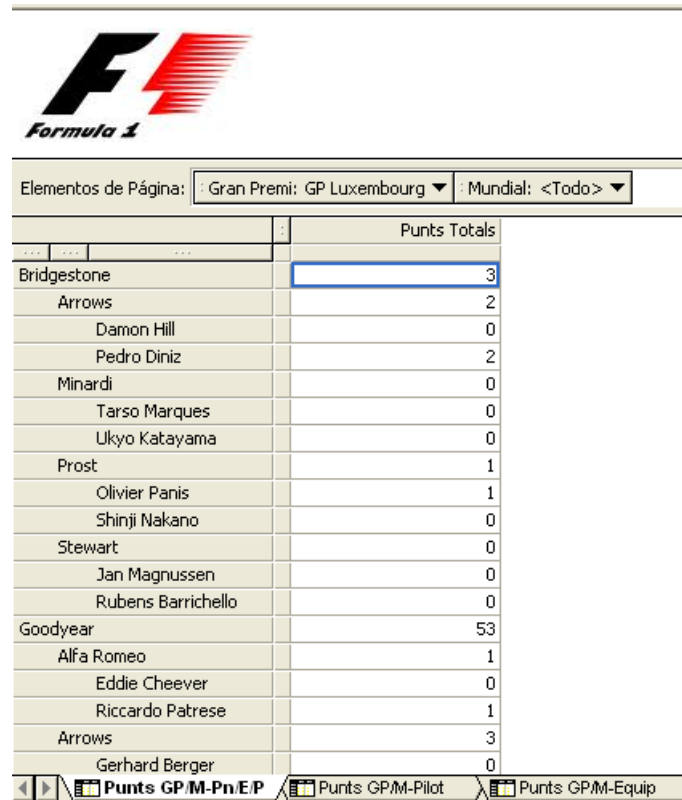


Figura 37 - Punts obtinguts per pneumàtic, equip i pilot seleccionant gran premi i mundial

En aquesta captura es pot veure el total de punts aconseguits segons la marca de pneumàtics al llarg de tots els anys que s'han introduït. Amb el gràfic de la dreta es mostra en forma de percentatge i es pot veure com Goodyear s'emporta pràcticament la meitat del pastís.

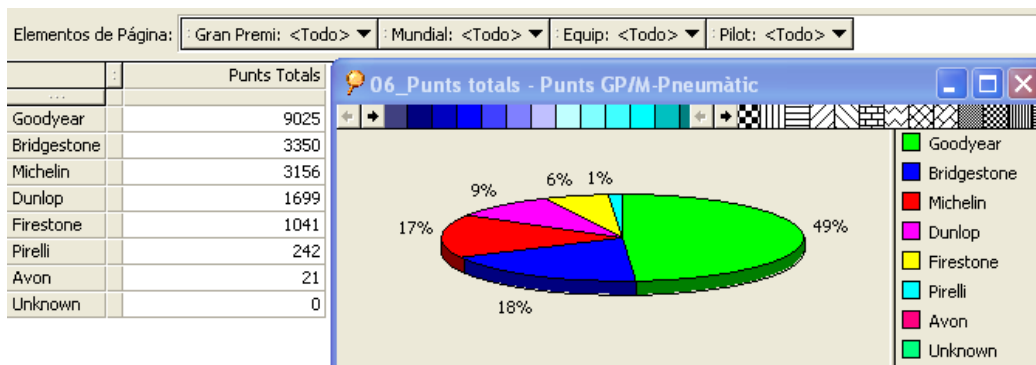
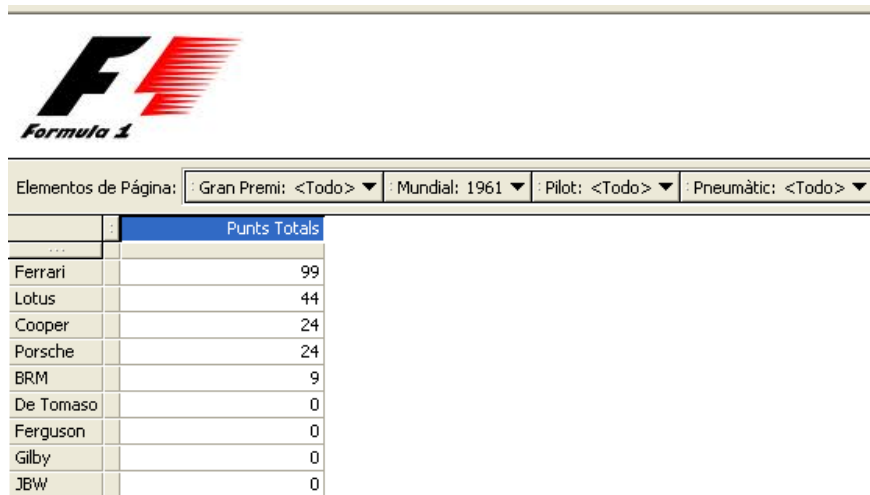


Figura 38 - Punts totals per tipus de pneumàtic

S'ha repetit l'informe però mostrant els punts obtinguts per equip:



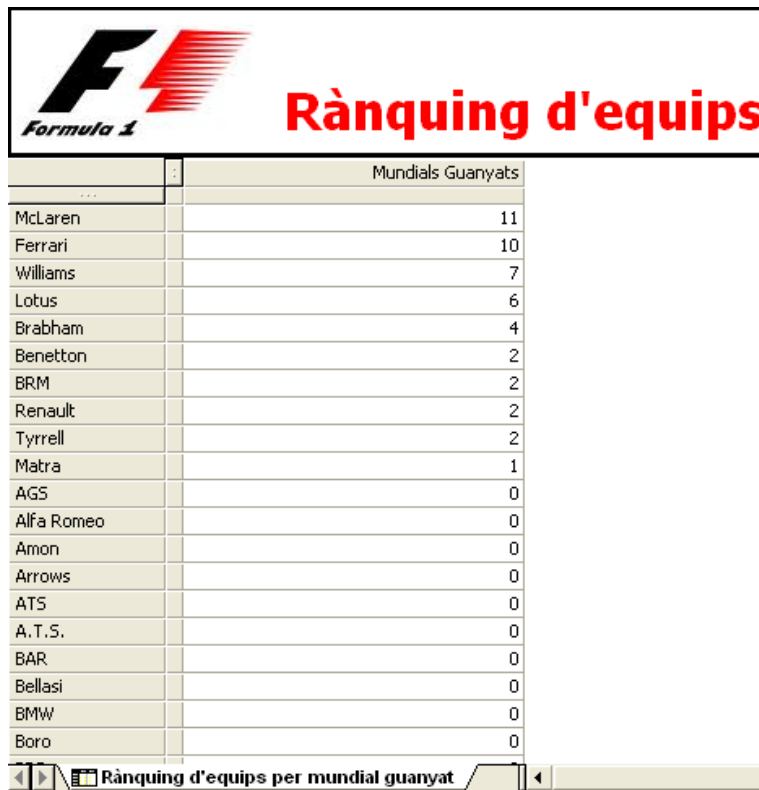
The screenshot shows the Formula 1 logo at the top left. Below it is a navigation bar with filters: 'Elementos de Página: Gran Premi: <Todo> Mundial: 1961 Pilot: <Todo> Pneumàtic: <Todo>'. The main table is titled 'Punts Totals' and lists the following data:

Equip	Punts Totals
Ferrari	99
Lotus	44
Cooper	24
Porsche	24
BRM	9
De Tomaso	0
Ferguson	0
Gilby	0
JBW	0

Figura 39 - Punts totals per equip

07_Rànquing d'equips per mundial guanyat

En aquesta captura es mostra el llistat total d'equips endreçats pel nombre de mundials guanyats:




The screenshot shows the Formula 1 logo at the top left. To its right is the title 'Rànquing d'equips' in red. Below is a table titled 'Mundials Guanyats' listing the following data:

Equip	Mundials Guanyats
McLaren	11
Ferrari	10
Williams	7
Lotus	6
Brabham	4
Benetton	2
BRM	2
Renault	2
Tyrrell	2
Matra	1
AGS	0
Alfa Romeo	0
Amon	0
Arrows	0
ATS	0
A. T. S.	0
BAR	0
Bellasi	0
BMW	0
Boro	0

Figura 40 - Rànquing d'equips per mundial guanyat

08_Rànquing de pilots per mundial guanyat

En aquesta captura es mostra el llistat de pilots endreçat per nombre de mundials guanyats:



Rànquing de pilots

	Mundials Guanyats
Michael Schumacher	7
Alain Prost	6
Graham Hill	3
Jackie Stewart	3
Nelson Piquet	3
Ayrton Senna	2
Emerson Fittipaldi	2
Fernando Alonso	2
Jim Clark	2
Mika Hakkinen	2
Niki Lauda	2
Alan Jones	1
Damon Hill	1
Denny Hulme	1
Jack Brabham	1
Jacques Villeneuve	1
James Hunt	1
Jochen Rindt	1
Jody Scheckter	1
Keke Rosberg	1
Kimi Raikkonen	1
Mario Andretti	1
Nigel Mansell	1
Phil Hill	1
Adrian Campos	0
Adrian Sutil	0

Figura 41 - Rànquing de pilots per mundial guanyat

09_Rànquing de pneumàtics per mundial guanyat

Com a últim informe de rànking es mostra el de tipus de pneumàtic per nombre mundials guanyats:

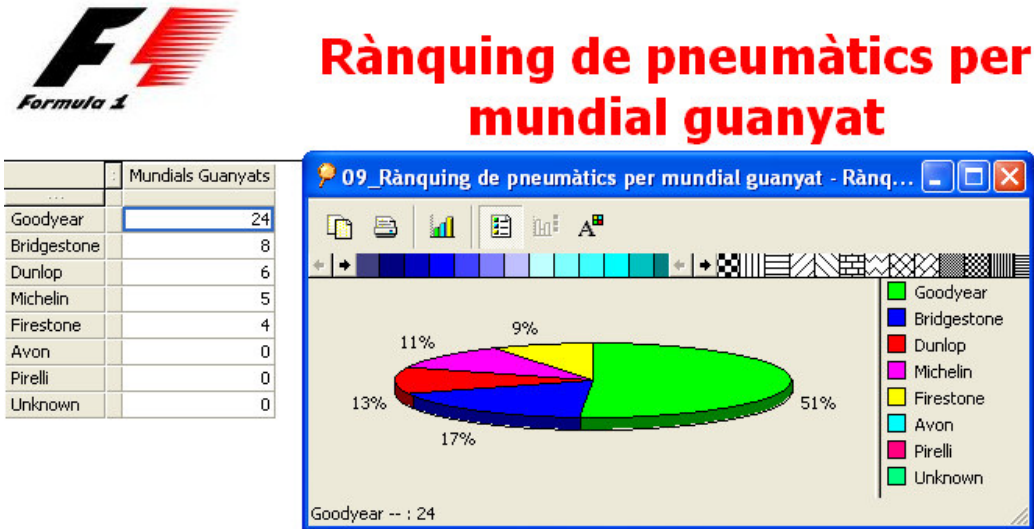


Figura 42 - Rànquing de pneumàtics per mudial guanyat

10_Rànquing de pneumàtics per pole aconseguida

En aquesta captura es mostra un llistat de pneumàtics on es mostra el nombre total de poles aconseguides, així com el resultat de la funció creada per obtenir una valoració del resultat de la cursa en funció del tipus de pneumàtic. En aquesta imatge es mostra el resultat general obtingut per a tots el anys i curses i podem veure com Goodyear encapçala el rànking seguit de Michelin:

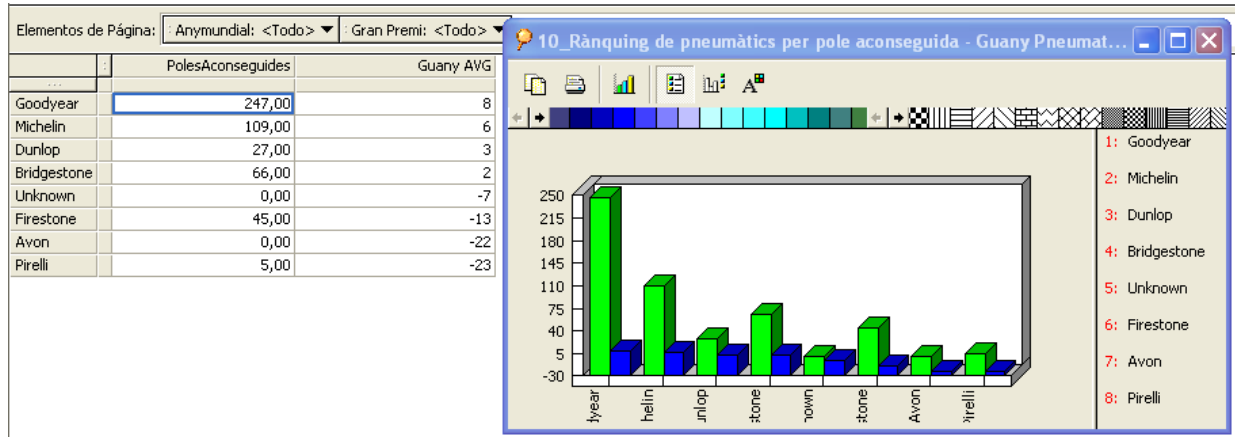


Figura 43 - Rànquing pneumàtics per pole aconseguida

En aquesta captura es pot observar el resultat obtingut per al gran premi de Mònaco, en aquest cas el resultat canvia i el tipus de pneumàtic que es presenta com a millor en cursa és Michelin.

Elementos de Página: Anymundial: <Todo> Nom: GP Monaco					
	Poles Aconseguides	Utilització en Cursa	Porcentaje UtilitzacioEnCursa	Porcentaje Poles Aconseguides	Guany AVG
Michelin	8	104	15%	24%	20
Goodyear	17	337	48%	50%	5
Dunlop	3	48	7%	9%	4
Bridgestone	3	82	12%	9%	-3
Unknown	0	4	1%	0%	-6
Firestone	3	83	12%	9%	-18
Avon	0	4	1%	0%	-20
Pirelli	0	40	6%	0%	-25

Figura 44 - Rànquing de poles per pneumàtic en Mònaco

En funció del mundial els resultats poden canviar també força. A continuació es mostra la captura dels resultats obtinguts en el mundial del 1984.

Elementos de Página: Anymundial: 1984 Nom: <Todo>					
	Poles Aconseguides	Utilització en Cursa	Porcentaje UtilitzacioEnCursa	Porcentaje Poles Aconseguides	Guany AVG
Michelin	13	153	37%	81%	43
Pirelli	0	81	20%	0%	-20
Goodyear	3	179	43%	19%	-24

Figura 45 - Rànquing de poles per pneumàtic en el mundial del 1984

Conclusions i línies de futur

Els objectius plantejats al inici del projecte han estat assolits. S'ha proveït a l'ICEM d'un sistema àgil i eficient per a realitzar els seus informes i a més facilitar l'anàlisi de les seves dades. Amb la creació del magatzem i l'ús de Discoverer els usuaris disposen d'una eina amb una sèrie d'informes predefinitos que a compleixen els requeriments inicials. Aquests informes, a més, es poden personalitzar segons les pròpies necessitats, amb la possibilitat de crear-ne de nous d'una manera senzilla. La millora en la forma de treballar respecte a les dades originals és evident.

La creació d'aquesta plataforma, a banda de ser un salt qualitatiu en la gestió de les dades, pot funcionar coma estímul en aquesta entitat per a la creació de nous magatzems de dades departamentals o fins i tot un magatzem de dades corporatiu. En aquest cas, però, s'hauria de tornar a avaluar l'arquitectura de maquinari i programari, ja que el que s'ha presentat en aquest projecte tot i ser adient per a l'estructura actual estaria limitat en creixement: la versió d'Oracle Express està limitada a 4G d'espai de dades i per aconseguir un sistema escalable es podria pensar en fer servir una estructura de base de dades en cabina amb cluster.

Els objectius personals del projecte també han estat assolits. S'han acomplert totes les entregues en les fites indicades, seguint la planificació marcada i aprofundint en cada fase en els coneixements de creació d'un magatzem de dades. A més s'han pogut aplicar part dels coneixements que s'han anat adquirint al llarg de la carrera, principalment els de les assignatures de bases de dades.

L'entrega d'aquest projecte, però no implica una solució tancada que no es pugui ampliar i millorar. Tal com ja s'ha comentat aquest projecte pot representar el punt de partida en la creació de nous magatzems de dades departamentals d'altres dades de l'ICEM, on es podrien realitzar una feina similar a la feta.

Partint, però de la plataforma implementada, una de les primeres millores que es podria implementar és la de fer els informes accessibles des d'un navegador Web. Actualment tothom està acostumat a accedir a la informació a través del seu navegador i el fet de necessitar de la instal·lació d'un programari específic per veure la informació pot generar un cert rebuig per part del usuari, a banda, del manteniment que implica. La suite de Discoverer permet penjar els informes en un servidor d'aplicacions, d'aquesta manera la informació es troba disponible des de qualsevol ubicació que tingui un navegador i accés.

Una altra de les línies de futur en les que es podria treballar seria la de integrar terceres fonts d'informació, ja sigui per afegir nova informació o per complementar les dades actuals. En les dades originals disponibles s'han trobat una sèrie d'errades que no han permès carregar part de la informació de les curses (faltava part de la informació o la que es donava era errònia) i a més, existeix una part de la informació que pot variar cada any com ara el sistema de puntuació aplicat o els circuits on es corren grans premis que no són del país (Ex: Gran Premi d'Europa). Aquesta problemàtica obliga a fer un repàs manual de les càrregues per mantenir el magatzem actualitzat. Incorporar noves fonts de dades milloraria el procés de càrrega i permetria donar més informació per a obtenir anàlisis estadístiques més bones.

Finalment, canviar la versió utilitzada d'Oracle per una de les seves versions comercials permetria no només una major escalabilitat sinó la possibilitat de fer servir tot una sèrie de característiques com ara estructures MOLAP o índex de tipus bitmap que podrien millorar el seu rendiment.

Glossari

Dimensió: Punt de vista utilitzat per analitzar un fet.

Discoverer: Conjunt d'eines desenvolupades per la empresa Oracle utilitzades per a l'anàlisi i explotació de dades.

Drill-Accross: Operació multidimensional que canvia d'un tema d'anàlisi a un altre donat un mateix espai.

Estructura en floc de neu: Estructura de taules on la taula de fets es troba envoltada de taules de dimensions i aquestes a la seva vegada es ramifiquen en nivells.

Estructura en estrella: Estructura de taules on la taula de fets es troba envoltada de les taules de dimensions.

ETL: Sigles amb que es coneix l'extracció, transformació i càrrega de dades en un magatzem. De l'anglès *Extract, Transform and Load*.

Factoria d'informació corporativa: Arquitectura lògica amb el propòsit d'ajudar en l'anàlisi de dades per a la presa de decisions.

Fet: Objecte d'anàlisi.

FIC: Vegeu Factoria d'informació corporativa.

HOLAP: Eina OLAP que combina característiques ROLAP i MOLAP. De l'anglès *hybrid OLAP*.

Indicadors: mesura que permeten identificar un fet que es vol analitzar.

Magatzem de dades: Bases de dades que integren diferents fonts de dades amb informació històrica i d'interès per la empresa utilitzades per al suport de presa de decisions.

Magatzem de dades corporatiu: Magatzem de dades on es guarden integrades totes les dades històriques de la empresa.

Magatzem de dades departamental: Magatzem de dades que resol les necessitats d'anàlisi d'un determinat grup d'usuaris.

MOLAP: Eina OLAP que utilitza matrius n-dimensionals, en comptes de taules relacionals, per a emmagatzemar les dades.

OLAP: Sigles que fan referència a les eines d'anàlisi, generalment multidimensional. Aquest tipus de programari permet un millor coneixement de les dades mitjançant un accés ràpid, consistent i interactiu a una sèrie de vistes de les dades operacionals transformades i integrades per a reflectir la dimensionalitat de l'empresa tal com la veu l'usuari. De l'anglès on-line analytical processing.

Oracle: Sistema de gestió de bases de dades relacionals produït per la empresa Oracle Corporation.

PL/SQL: Llenguatge de programació d'Oracle que permet la creació de funcions i procediments incorporant sentències SQL emmagatzemant-los en la mateixa base de dades.

ROLAP: Eina OLAP implementada sobre una sistema de base de dades relacional.

SGBD: *Vegeu* Sistema de gestió de bases de dades.

Sistema de gestió de bases de dades: Programari de gestió de bases de dades. Principalment s'encarrega d'abstraure als usuaris de l'emmagatzemament físic, mantenir la integritat i gestionar l'accés concurrent d'usuaris.

SOA: Arquitectura de programari que defineix la utilització de serveis per donar suport als requeriments de negoci i facilita les activitats de consolidació i integració de serveis. De l'anglès *Service-oriented Architecture*.

SQL*Loader: Eina proporcionada en la mateixa instal·lació de la base de dades Oracle que permet fer la importació de dades emmagatzemades en fitxers de text a la base de dades d'una manera senzilla.

Taula de fets: Taula que agrupa el conjunt de fets associats a un mateix espai de dimensions.

Bibliografia

Llibres de text

Casert, Joe; Kimball, Ralph (2005). *The Data Warehouse ETL Toolkit*.

Inmon, W.H. (2002). *Building the Data Warehouse (3rd Edition)*.

Kimball, Ralph; Ross, Margy (2002). *The Data Warehouse Toolkit. (2nd Edition)*.

Rainardi, Vincent (2008). *Building a Data Warehouse*.

Rius, Angels; Serra, Montse (2003). *Magatzems de dades i models multidimensionals*.

Enllaços d'interès

Corporate Information Factory (CIF) Resources by Bill Inmon, Inmon Data Systems
<http://www.inmoncif.com>

Data Warehousing: A Look at Business Intelligence and Data Warehouse
<http://www.1keydata.com/datawarehousing/datawarehouse.html>

Data Warehousing Review - Data, Data Everywhere!!
<http://www.dwreview.com>

Fórmula 1 - Viquipèdia
http://ca.wikipedia.org/wiki/Fórmula_1

Intranet Journal: Feature: Finding Your Way Around E-commerce
<http://www.intranetjournal.com/features/datawarehousing.html>

Kimball Group: Data Warehouse Training, Consulting, and Kimball University
<http://www.rkimball.com>

Annex

Script de transformació de full Excel a fitxer csv

```
Dim oFSO
Dim oShell, oExcel, oFile, oSheet
Set oFSO = CreateObject("Scripting.FileSystemObject")
Set oShell = CreateObject("WScript.Shell")
Set oExcel = CreateObject("Excel.Application")
oExcel.DisplayAlerts = False

For Each oFile In oFSO.GetFolder("C:\dades\").Files
  If LCase(oFSO.GetExtensionName(oFile)) = "xls" Then
    With oExcel.Workbooks.Open(oFile, 0, True, , , , True, , , ,
False, , False)
      For Each oSheet In .Worksheets
        oSheet.SaveAs "C:\dades\" & oSheet.Name & ".csv", 6
      Next
      .Close False, , False
    End With
  End If
Next
oExcel.Quit
```

En aquest script es suposa que els documents sempre es deixaran en el mateix directori (en aquest cas el directori dades de la c) i que el nom de les fulles sempre serà el mateix per a poder automatitzar el següent pas de càrrega en base de dades.

La crida del script és: `csript excel2csv.vbs`

Cal que la màquina on s'executi l'script tingui instal·lat l'Activex d'Excel, sinó no es podran generar els fitxers.

Fitxers de control i crides des de la línia de comandes

En cada crida es fa servir un fitxer de control diferent on se l'indica la taula i els camps de destí, a més cada càrrega genera un fitxer de log diferent. El paràmetre `skip = 1` elimina el primer registre del fitxer on aniria la capçalera amb els noms de les columnes. Per poder realitzar aquesta crida també s'ha de configurar prèviament el `tnsnames` amb la instància XE que es fa servir.

Resultats.csv

```
load data
infile 'c:\dades\Resultats.csv'
into table Resultats
fields terminated by "," optionally enclosed by '"' trailing nullcols
(circuit_data, pilot, equip, pneumatic, posicio nullif posicio=BLANKS, voltes,
temps, velocitat, motiu)
```

Poles.csv

```
load data
infile 'c:\dades\Poles.csv'
into table Poles
fields terminated by "," optionally enclosed by '"' trailing nullcols
(circuit,pais,datac,pilot,temps)
```

Pilots.csv

```
load data
infile 'c:\dades\Pilots.csv'
into table Pilots
fields terminated by "," optionally enclosed by '"' trailing nullcols
(pais,pilot,anys)
```

Les crides realitzades amb el SQL *Loader, 1 per cada fitxer, han estat:

```
sqlldr LOAD_F1@XE/Lf1_1234 control=control1.ctl log=log1.log skip=1
sqlldr LOAD_F1@XE/Lf1_1234 control=control2.ctl log=log2.log skip=1
sqlldr LOAD_F1@XE/Lf1_1234 control=control3.ctl log=log3.log skip=1
```

Programació del treball d'actualització en la base de dades

```
BEGIN
DBMS_SCHEDULER.CREATE_JOB (
job_name => 'F1_DW.JOB_ActualitzaMagatzem',
job_type => 'STORED_PROCEDURE',
job_action => 'F1_DW.ActualitzaMagatzem',
start_date => sysdate,
repeat_interval => 'FREQ=WEEKLY;BYDAY=MON;');
END;

BEGIN
DBMS_SCHEDULER.ENABLE ('F1_DW.JOB_ActualitzaMagatzem');
END;
```

Registres pre-carregats en el magatzem de dades

```
INSERT INTO Pneumatic (nom)
VALUES ('Unknown');

INSERT INTO Equip (nom)
VALUES ('Unknown');

INSERT INTO Motiu (descripcio)
VALUES ('No Retirement');
INSERT INTO Motiu (descripcio)
VALUES ('Unknown');

INSERT INTO GranPremi (nom)
```



```
VALUES('GP East USA');
INSERT INTO GranPremi(nom)
VALUES('GP Las Vegas');
INSERT INTO GranPremi(nom)
VALUES('GP West USA');
INSERT INTO GranPremi(nom)
VALUES('GP San Marino');
INSERT INTO GranPremi(nom)
VALUES('GP Belgium');
INSERT INTO GranPremi(nom)
VALUES('GP Switzerland');
INSERT INTO GranPremi(nom)
VALUES('GP Pacific');
INSERT INTO GranPremi(nom)
VALUES('GP Luxembourg');
INSERT INTO GranPremi(nom)
VALUES('GP Europe');
```

Procediments emmagatzemats

F1_DW.ActualitzaMagatzem

Procediment per actualitzar tot el magatzem de dades:

```
CREATE OR REPLACE
PROCEDURE ActualitzaMagatzem IS
BEGIN

    LOAD_F1.netejaRegistres;
    actualitzaCircuit;
    actualitzaEquip;
    actualitzaGP;
    actualitzaMotiu;
    actualitzaMundial;
    actualitzaPilot;
    actualitzaPneumatic;
    actualitzaResultatCursa;
    actualitzaResultatsMundial;
    LOAD_F1.esborraTaules;

END;
```

F1_DW.actualitzaCircuit

Procediment que actualitza les dades de la taula Circuit:

```
CREATE OR REPLACE
PROCEDURE actualitzaCircuit IS
BEGIN
    MERGE INTO Circuit c
    USING (SELECT DISTINCT TRIM(circuit) nom, TRIM(pais) pais
    FROM LOAD_F1.Poles
    WHERE TRIM(circuit) IS NOT NULL AND TRIM(pais) IS NOT NULL) o
    ON (c.nom = o.nom)
    WHEN NOT MATCHED THEN
```

```
INSERT (nom, pais)
VALUES o.nom, o.pais;
END;
```

F1_DW.actualitzaEquip

Procediment que actualitza les dades de la taula Equip:

```
CREATE OR REPLACE
PROCEDURE actualitzaEquip IS
BEGIN
    MERGE INTO Equip e
    USING (SELECT DISTINCT TRIM(equip) nom FROM LOAD_F1.resultats
    WHERE TRIM(equip) IS NOT NULL) o
    ON (e.nom = o.nom)
    WHEN NOT MATCHED THEN
    INSERT (nom)
    VALUES (o.nom);
END;
```

F1_DW.actualitzGP

Procediment que actualitza les dades de la taula GranPremi:

```
CREATE OR REPLACE
PROCEDURE actualitzaGP IS
BEGIN
    MERGE INTO GranPremi gp
    USING (
    SELECT DISTINCT 'GP '||TRIM(Pais) nom FROM LOAD_F1.Poles
    WHERE TRIM(Pais) IS NOT NULL) o
    ON (gp.nom = o.nom)
    WHEN NOT MATCHED THEN
    INSERT (nom)
    VALUES (o.nom);
END;
```

F1_DW.actualitzaMotiu

Procediment que actualitza les dades de la taula Motiu:

```
CREATE OR REPLACE
PROCEDURE actualitzaMotiu IS
BEGIN
    MERGE INTO Motiu m
    USING (SELECT DISTINCT INITCAP (TRIM (SUBSTR (motiu, 1, DECODE (INSTR
    (LOWER (motiu), 'with '), 0, LENGTH (motiu), INSTR (LOWER (motiu), 'with ')-
    1)))) descripcio
    FROM LOAD_F1.resultats
    WHERE TRIM(motiu) IS NOT NULL AND TRIM(motiu) <> '-') o
    ON (m.descripcio = o.descripcio)
    WHEN NOT MATCHED THEN
    INSERT (descripcio)
    VALUES (o.descripcio);
END;
```

F1_DW.actualitzaMundial

Procediment que actualitza la taula Mundial:

```
CREATE OR REPLACE
PROCEDURE actualitzaMundial IS
BEGIN

    MERGE INTO Mundial m
    USING (SELECT DISTINCT TO_CHAR (TO_DATE (TRIM
    (SUBSTR (circuit_data ,
    INSTR (circuit_data , '/' , INSTR (circuit_data , '/')+1)+1)),
    ' Day, Month dd, yyyy', 'NLS_DATE_LANGUAGE=English'), 'yyyy') anyMundial
    FROM LOAD_F1.Resultats
    WHERE INSTR (circuit_data , '/') > 0) o
    ON (m.anyMundial = o.anyMundial)
    WHEN NOT MATCHED THEN
    INSERT (anyMundial)
    VALUES (o.anyMundial);

END;
```

F1_DW.actualitzaPilot

Procediment que actualitza la taula Pilot fent servir les taules Resultats i Pilots:

```
CREATE OR REPLACE
PROCEDURE actualitzaPilot IS
BEGIN

    MERGE INTO Pilot p
    USING (SELECT DISTINCT TRIM(TRIM (SUBSTR (pilot, INSTR(pilot, ',') +1))
    || ' ' || TRIM (SUBSTR (pilot, 1, INSTR (pilot, ',') -1))) nom ,
    TRIM(pais) pais
    FROM LOAD_F1.pilots) o
    ON (p.nom = o.nom)
    WHEN NOT MATCHED THEN
    INSERT (nom, pais)
    VALUES (o.nom, o.pais);

    MERGE INTO Pilot p
    USING (SELECT DISTINCT TRIM(Pilot) nom, 'Unknown'
    FROM LOAD_F1.Resultats pais) o
    ON (p.nom = o.nom)
    WHEN NOT MATCHED THEN
    INSERT (nom, pais)
    VALUES (o.nom, o.pais);

END;
```

F1_DW.actualitzaPneumatic

Procediment que actualitza la taula Pneumatic:

```
CREATE OR REPLACE
PROCEDURE actualitzaPneumatic IS
BEGIN

    MERGE INTO Pneumatic p
    USING (SELECT DISTINCT TRIM(pneumatic) nom FROM LOAD_F1.resultats
    WHERE TRIM(pneumatic) IS NOT NULL) o ON (p.nom = o.nom)
    WHEN NOT MATCHED THEN
    INSERT (nom) VALUES (o.nom);

END;
```

F1_DW.actualitzaResultatCursa

Procediment que actualitza la taula ResultatCursa:

```
CREATE OR REPLACE
PROCEDURE actualitzaResultatCursa IS

    CURSOR c_res IS
    SELECT DISTINCT circuit_data , TRIM (SUBSTR (circuit_data ,INSTR
(circuit_data ,'/')+1 ,INSTR (circuit_data,'/',INSTR (circuit_data
,'/')+1)-2)) circuit, SUBSTR(TRIM(circuit_data),-4) anyMundial,
TRIM(pilot) pilot, NVL(TRIM(Equip),'Unknown') Equip,
NVL(TRIM(Pneumatic),'Unknown') pneumatic, posicio, velocitat,
INITCAP (TRIM (SUBSTR (motiu, 1, DECODE (INSTR (LOWER (motiu),'with '), 0,
LENGTH (motiu), INSTR (LOWER (motiu), 'with ')-1)))) motiu , ROWID
FROM LOAD_F1.Resultats;

    v_pilot number;
    v_equip number;
    v_pneumatic number;
    v_motiu number;
    v_circuit number;
    v_punts number;
    v_abandonament char(1) := '0';
    v_pole char(1) := '0';

BEGIN

    FOR v_res IN c_res LOOP

        BEGIN

            SELECT idPilot INTO v_pilot
            FROM Pilot WHERE v_res.pilot = nom;

            SELECT idEquip INTO v_equip
            FROM Equip WHERE v_res.equip = nom;

            SELECT idPneumatic INTO v_pneumatic
            FROM Pneumatic WHERE v_res.pneumatic = nom;
```

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

TFC – Magatzems de Dades

```
SELECT idCircuit INTO v_circuit
FROM Circuit WHERE v_res.circuit = nom;

SELECT NVL(MAX(punts),0) INTO v_punts
FROM LOAD_F1.PuntuacioMundial
WHERE AnyMundial = v_res.anyMundial
AND posicio = v_res.posicio;

IF (v_res.posicio IS NULL) THEN
v_abandonament := '1';
ELSIF (v_res.motiu <> '-' AND v_res.motiu IS NOT NULL) THEN
v_abandonament := '1';
ELSE
v_abandonament := '0';
END IF;

IF v_abandonament = '1' THEN
SELECT MAX(idMotiu) INTO v_motiu
FROM Motiu WHERE descripcio = v_res.motiu;

IF v_motiu is null THEN
SELECT idMotiu INTO v_motiu
FROM Motiu WHERE descripcio = 'Unknown';
END IF;

ELSE
SELECT idMotiu INTO v_motiu
FROM Motiu WHERE descripcio = 'No Retirement';
END IF;

SELECT to_char(COUNT(*)) INTO v_pole
FROM (SELECT DISTINCT * FROM LOAD_F1.Poles
WHERE TRIM(circuit) = TRIM(v_res.circuit)
AND TRIM(Pilot) = TRIM(v_res.pilot)
AND SUBSTR(TRIM(datac),-2) =
SUBSTR(TRIM(v_res.anyMundial),-2));

INSERT INTO ResultatCursa(AnyMundial, IdGranPremi, idPilot,
idEquip, idPneumatic, idMotiu, idCircuit, posicio, velocitat,
punts, abandonament, pole)
VALUES(v_res.anyMundial ,getGP (substr (v_res.anyMundial, -2),
v_res.circuit,v_circuit), v_pilot,v_equip, v_pneumatic,
v_motiu, v_circuit,v_res.posicio, v_res.velocitat, v_punts,
v_abandonament, v_pole);

EXCEPTION WHEN OTHERS THEN

INSERT INTO LOAD_F1.ERRORRESULTATS(CIRCUIT_DATA, PILOT, EQUIP,
PNEUMATIC, POSICIO, VOLTES, TEMPS, VELOCITAT, MOTIU)
SELECT * FROM LOAD_F1.RESULTATS WHERE ROWID = v_res.ROWID;

DELETE FROM LOAD_F1.RESULTATS WHERE ROWID = v_res.ROWID;

END;

END LOOP;

END;
```

F1_DW.actualitzaResultatsMundial

Procediment que actualitza la taula ResultatsMundial:

```
CREATE OR REPLACE
PROCEDURE actualitzaResultatsMundial IS
BEGIN

    MERGE INTO ResultatsMundial r
    USING (
        SELECT AnyMundial, IdPilot, IdEquip, IdPneumatic,
        DECODE (position, 1, '1', '0') guanyat
        FROM (SELECT c1.AnyMundial, c1.IdPilot, c1.IdEquip, c2.IdPneumatic,
        RANK() OVER (PARTITION BY c1.anyMundial ORDER BY MAX(c1.totalPilot) DESC,
        SUM(c2.punts) DESC) position
        FROM (SELECT anyMundial, IdEquip, IdPilot, SUM(punts) totalPilot
        FROM ResultatCursa
        GROUP BY anyMundial, IdPilot, IdEquip) c1, ResultatCursa c2
        WHERE c1.anyMundial = c2.AnyMundial AND c1.IdPilot = c2.IdPilot
        AND c1.IdEquip = c2.IdEquip
        GROUP BY c1.anyMundial, c1.IdPilot, c1.IdEquip, c2.IdPneumatic)) o
    ON (r.AnyMundial = o.AnyMundial AND r.IdPilot = o.IdPilot
    AND r.IdEquip = o.IdEquip AND r.IdPneumatic = o.IdPneumatic)
    WHEN MATCHED THEN
    UPDATE SET r.mundialGuanyat = o.guanyat
    WHEN NOT MATCHED THEN
    INSERT (AnyMundial, IdPilot, IdEquip, IdPneumatic, mundialGuanyat)
    VALUES (o.AnyMundial, o.IdPilot, o.IdEquip, o.IdPneumatic, o.guanyat);

END;
```

F1_DW.getGP

Funció que retorna l'identificador de Gran Premi per una cursa concreta:

```
CREATE or REPLACE
FUNCTION
getGP(p_any IN VARCHAR2, p_circuit IN VARCHAR2, p_idCircuit IN NUMBER)
RETURN NUMBER AS

    v_total NUMBER;
    v_pais VARCHAR2(100);
    v_nom VARCHAR2(50);
    v_id NUMBER;

BEGIN

    SELECT pais INTO v_pais FROM Circuit WHERE idCircuit = p_idCircuit;
    SELECT COUNT(*) INTO v_total
    FROM LOAD_F1.Poles
    WHERE SUBSTR(TRIM(datac), -2) = p_any AND TRIM(pais) = v_pais;

    IF v_total = 1 THEN
        SELECT idGranPremi INTO v_id
        FROM GranPremi WHERE nom = 'GP ' || v_pais;
    ELSE
        SELECT COUNT(*) INTO v_total
        FROM LOAD_F1.GPExcepcions
```

```
WHERE circuit = p_circuit;

IF v_total = 0 THEN
SELECT idGranPremi INTO v_id
      FROM GranPremi WHERE nom = 'GP '||v_pais;
ELSE
SELECT idGranPremi INTO v_id
      FROM (SELECT nom, pospreferent,
RANK() OVER (PARTITION BY circuit ORDER BY pref, pospreferent)
pref FROM(
SELECT DISTINCT e.nom, e.circuit, pospreferent,
RANK() OVER(PARTITION BY e.nom ORDER BY pospreferent) pref
FROM LOAD_F1.GPExcepcions e, LOAD_F1.Poles p
WHERE TRIM(e.circuit) = TRIM(p.circuit)
AND SUBSTR(trim(datac),-2) = p_any)
WHERE circuit = p_circuit) sel, GranPremi
      WHERE sel.pref = 1 AND GranPremi.nom = sel.nom;
END IF;

END IF;

RETURN v_id;

END getGP;
```

LOAD_F1.netejaRegistres

Procediment que esborra els registres de les taules de càrrega inicials que es consideren erronis:

```
CREATE OR REPLACE
PROCEDURE netejaRegistres IS
BEGIN

INSERT INTO ERRORResultats SELECT *
FROM Resultats WHERE INSTR(circuit_data,'/') = 0;
DELETE FROM Resultats WHERE INSTR(circuit_data,'/') = 0;

INSERT INTO ERRORResultats SELECT *
FROM Resultats WHERE Pilot = 'N/A' OR TRIM(Pilot) IS NULL;
DELETE FROM Resultats WHERE Pilot = 'N/A' OR TRIM(Pilot) IS NULL;

INSERT INTO ERRORResultats SELECT *
FROM Resultats WHERE Equip = 'N/A' OR TRIM(Equip) is null;
DELETE FROM Resultats WHERE Equip = 'N/A' OR TRIM(Equip) is null;

INSERT INTO ERRORResultats SELECT *
FROM Resultats WHERE posicio IS NOT NULL AND ISNUMERIC(posicio) ='0';
DELETE FROM Resultats
WHERE posicio IS NOT NULL AND ISNUMERIC(posicio) ='0';

INSERT INTO ERRORResultats SELECT *
FROM Resultats WHERE velocitat IS NOT NULL AND ISNUMERIC(velocitat) ='0';
DELETE FROM Resultats
WHERE velocitat IS NOT NULL AND ISNUMERIC(velocitat) ='0';

INSERT INTO ERRORPilots SELECT *
FROM Pilots WHERE TRIM(Pilot) = 'N/A';
DELETE FROM Pilots WHERE TRIM(Pilot) = 'N/A';
```

```
INSERT INTO ERRORPoles SELECT *
FROM Poles WHERE TRIM(Pais) = 'N/A';
DELETE FROM ERRORPoles WHERE TRIM(Pais) = 'N/A';

INSERT INTO ERRORPoles SELECT *
FROM Poles WHERE TRIM(Pilot) = 'N/A';
DELETE FROM ERRORPoles WHERE TRIM(Pilot) = 'N/A';

INSERT INTO ERRORPoles SELECT *
FROM Poles WHERE TRIM(Datac) = 'N/A';
DELETE FROM ERRORPoles WHERE TRIM(Datac) = 'N/A';

INSERT INTO ERRORPoles SELECT *
FROM Poles WHERE ISNUMERIC(SUBSTR(TRIM(Datac),-2)) = '0';
DELETE FROM ERRORPoles WHERE ISNUMERIC(SUBSTR(TRIM(Datac),-2)) = '0';

END netejaRegistres;
```

F1_DW.getProfit

Funció que calcula el valor de guanyar o perdre una pole en funció del tipus de pneumàtic que es fa servir

```
CREATE OR REPLACE
FUNCTION getProfit
(p_anymundial in number, p_idgranpremi in number, p_idpneumatic in number)
RETURN NUMBER AS
    v_cars NUMBER;
    v_pneumatics NUMBER;
    v_pole NUMBER;
BEGIN
    SELECT COUNT(*) INTO v_cars FROM ResultatCursa
    WHERE AnyMundial = p_anymundial AND Idgranpremi = p_idgranpremi;

    SELECT COUNT(*), MAX(TO_NUMBER(pole)) INTO v_pneumatics, v_pole
    FROM ResultatCursa
    WHERE AnyMundial = p_anymundial AND Idgranpremi = p_idgranpremi
    AND Idpneumatic = p_idpneumatic;

    IF v_cars = 0 THEN
        RETURN 0;
    ELSE
        IF v_pole = 0 THEN
            RETURN -v_pneumatics/v_cars*100;
        ELSE
            RETURN (v_cars-v_pneumatics)/v_cars*100;
        END IF;
    END IF;
END;
```


LOAD_F1. esborraTaules

Procediment per esborrar les taules de càrrega inicials:

```
CREATE or REPLACE
PROCEDURE esborraTaules IS
BEGIN
    DELETE FROM Resultats;
    DELETE FROM Pilots;
    DELETE FROM Poles;
END esborraTaules;
```

LOAD_F1.isNumeric

Funció que retorna 1 o 0 en funció de si el valor que se li passa és numèric o no:

```
CREATE OR REPLACE
FUNCTION isNumeric (param in varchar2) RETURN CHAR AS
    dummy number;
BEGIN
    dummy:=TO_NUMBER(param);
    RETURN('1');
EXCEPTION
    WHEN OTHERS THEN
        RETURN('0');
END;
```