

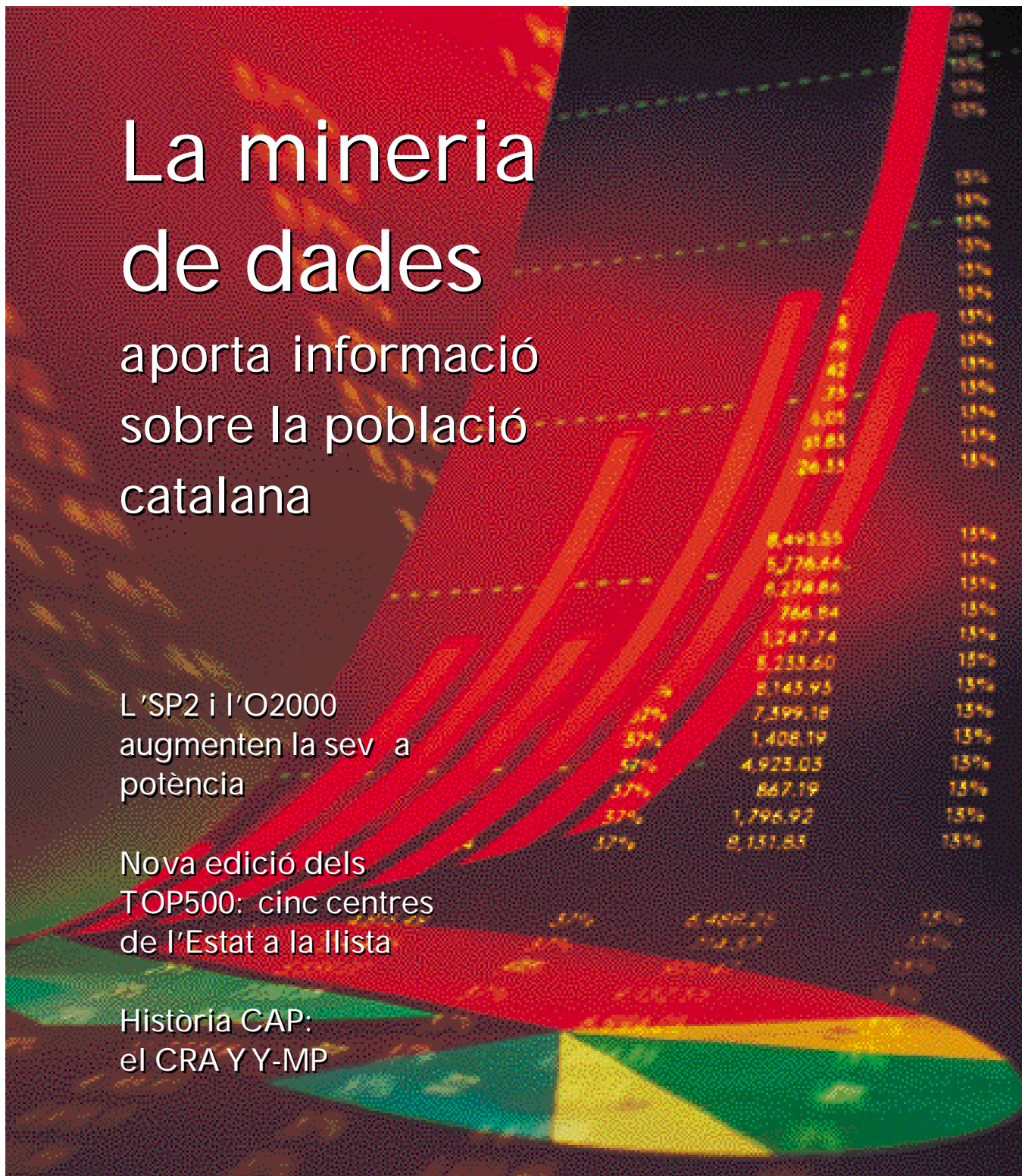
La mineria de dades

aporta informació sobre la població catalana

L'SP2 i l'O2000 augmenten la seva potència

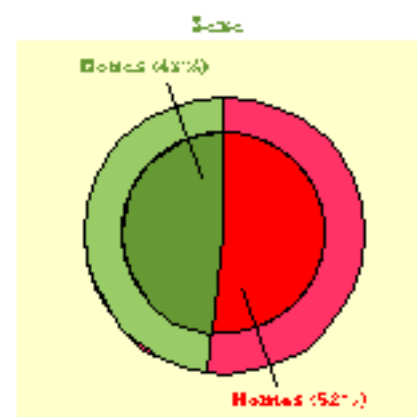
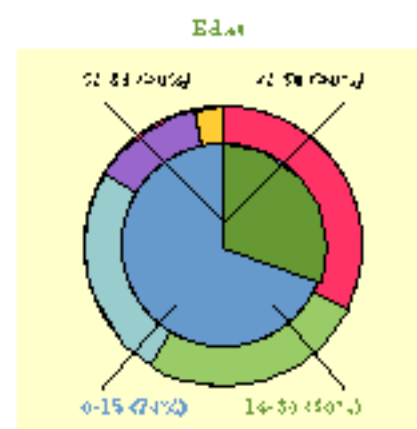
Nova edició dels TOP500: cinc centres de l'Estat a la llista

Història CAP: el CRA YY-MP



La mineria de dades aporta informació sobre la població catalana

Representació de les variables edat, sexe i nombre de persones per llar corresponents al primer cluster dels matrimonis o parelles amb fills.



Les dades estadístiques sobre la població sovint poden donar-nos sorpreses reveladores i poden posar de relleu relacions inimaginables entre variables que aparentment no tenien res a veure. Establir vincles d'aquest tipus és el que ha fet l'Institut d'Estadística de Catalunya amb els resultats obtinguts de l'anàlisi del Cens de Població i Habitatge de 1991 utilitzant la metodologia del data mining (minería de dades), en un projecte fet en col·laboració amb IBM i el CESCA.

Les tres institucions van signar un conveni de col·laboració el 18 de juliol de l'any passat per tal que l'Institut d'Estadística de Catalunya (Idescat), institució responsable de la difusió de tots els resultats estadístics oficials que es realitzen a Catalunya, pogués ampliar els plans de tabulació del Cens de Població i Habitatge de Catalunya de 1991. El projecte analitza sis milions de registres i fins a 150 variables per a cada registre. Els resultats d'aquest treball proporcionaran informació que l'estadística tradicional no pot abordar.

"La supercomputació s'ha expandit més enllà del mercat tradicional de càlcul científicotècnic i per això el CESCA ha impulsat aquesta experimentació, per demostrar la utilitat del nostre maquinari fora del nostre consum habitual", explica Miquel Huguet, director del CESCA. IBM ha debat utilitzar la seva eina de minería de dades per aplicar-la en arxius amb molta informació i l'Idescat es va sumar al projecte en creure que seria una bona oportunitat per trobar relacions poc conegudes entre dades censals. Es va triar aquest cens perquè és el que té més informació. Per analitzar totes aquestes dades calien instruments potents. El 1997 es va instal·lar, en l'ordinador SP2 del CESCA, el gestor de bases de dades DB2 en dos nodes i en un altre node es va instal·lar la versió 1.2 de l'Intelligent Miner. El DB2 va ser el magatzem a través del qual l'intelli-

gant Miner accedia a les dades.

Un dels aspectes més destacats d'aquest estudi és que, com apunta Joaquim Cepellades, subdirector de Producció d'Estadístiques Demogràfiques i Socials de l'Idescat, "no hi ha definició prèvia de les hipòtesis de recerca" i no es pretén trobar cap resultat en concret. "Aquesta és la gràcia d'aquestes tecnologies, que són de descobriment, no de verificació. Has de saber què estàs buscant però no hi ha plantejament previ de què vols obtenir", explica Francesc Subirada, d'IBM.

L'Intelligent Miner, a més a més de les funcions estadístiques clàssiques, utilitza les associacions, els patrons seqüencials, les sèries temporals similars (aquestes dues darreres no s'han pogut usar perquè no es disposava de dades al llarg del temps, és a dir, s'ha treballat només amb informació d'un any), les classificacions, les prediccions de valors, l'anàlisi de les desviacions (tampoc utilitzada) i les segmentacions. Els resultats publicables són, segons l'Institut, els obtinguts a partir d'associacions i a partir de segmentacions, que seran les dades que es faran públiques en la presentació oficial de resultats durant el mes de setembre. Les dades estaran a disposició dels investigadors i de tots aquells que les necessitin.

Un altre aspecte per a destacar de l'Intelligent Miner és que amb aquest estudi s'ha demostrat que "un dels camps d'aplicació importants és el món de les ciències socials i, concreta-



ment, el cens de població", com explica Francese Subirada. La importància rau en el fet que abans aquestes eines s'usaven exclusivament per a recerca i desenvolupament, en el món del màrqueting i el comerç. "Aquest és el primer cop que amb aquest programa es traspasa l'àmbit privat per aplicar-ho a dades de la població i les llars de Catalunya segons el cens del 1991", afirma Joaquim Capallades. "Ham pogut identificar relacions noves entre les dades i, a més, ens ha permès fer una quantificació d'informació que potser ja s'intuïa. Les eines usades mostren la intensitat entre relacions que ja se suposen", afegirà.

Conclusions curioses

El cens de població de 1991 referit a Catalunya suposa sis milions de registres de persones i dos milions de registres de llars. En la prova pilot que s'ha dut a terme al CESCO, s'han analitzat variables demogràfiques, socials, econòmiques, del cens lingüístic i referides a l'habitatge. S'han fet aplicacions a diferents tipus de llars: unipersonals, matrimonis o parelles sense fills, matrimonis o parelles amb fills. A partir de la informació obtinguda, els experts de l'Idescat han pogut interpretar les dades. Una dada rellevant és que en cap moment s'ha treballat amb mostres, sinó que s'ha treballat amb el total de la població.

Analitzant les dades s'ha observat, per exemple, fixant-nos en matrimonis o parelles amb fills, que hi ha un conjunt de persones d'entre 16 i 30 anys que presenten un perfil i uns trets

característics que tradicionalment s'atorgarien a joves menors de 16 anys. Les variables que determinen el perfil són, entre d'altres: l'edat, els estudis en curs, la relació amb l'activitat i l'estat civil. I en aquest cas s'ha observat que, menys l'edat, les dues franges de població comparteixen la resta de variables.

També s'ha pogut observar que un conjunt de joves (d'entre 16 i 30 anys) es troben formant part de llars de

matrimonis amb fills, com a fills, quan, per les característiques que presentaven aquests joves, es podria suposar que formaven un altre tipus de llar. Les variables que determinen aquest grup són, entre d'altres: edat, situació professional, sector d'activitat, estat civil, titulació i relació amb l'activitat. Aquests dos exemples s'han obtingut aplicant tècniques de segmentació.

En un altre resultat, en aquest cas obtingut per associació, s'ha pogut veure que en el conjunt de matrimonis amb fills, s'ha detectat entre les variables una relació amb l'activitat i situació professional. Així, per exemple, s'ha vist que en el 94% dels casos quan hi ha una persona ocupada (assalariada fixa) a la llar, hi ha una persona inactiva, i això afecta al 61,8% de les llars de matrimoni amb fills. De la mateixa manera s'ha vist que en el 90% dels casos quan hi ha un ocupat (assalariat fix) i un parat a la llar, hi ha també una persona inactiva, i això afecta al 3,7% de les llars. És molt important veure que en els exemples es mostra també la intensitat de les relacions.

Joaquim Capallades va voler deixar clara la confidencialitat de les dades que s'han usat per a l'estudi: "l'aplicació d'aquestes metodologies potents no vol dir que hi hagi un trencament de la confidencialitat de les dades estadístiques. Aquestes dades són anònimes: individualitzades però no personalitzades. Tot i que permet avançar en el coneixement científic, no és en cap cas un mecanisme de control".

Totes tres parts es mostren satisfetes de la col·laboració conjunta

en l'ús d'aquestes noves eines de mineria de dades en aquest projecte que no s'acaba amb la presentació de les dades. "L'interès principal no és des del punt de vista immediat. El més important es veurà després de portar un temps de reflexió. De moment ha proporcionat elements perquè els investigadors vagin fent associacions", explica el subdirector de l'Institut d'Estadística de Catalunya.

PROGRAMES DE MOBILITAT

Klaus Bernhard, de l'**INTES GmbH** (Alemanya), ha vingut convidat per Jecú Labarta, del Departament d'Arquitectura de Computadors de la UPC, per desenvolupar el projecte *Parmaf*. Bernhard va estar entre nosaltres del 21 de maig al 4 de juny.

Konnie Corradi, de la **Università degli Studi di Bari** (Itàlia), ha estat convidat per Ferran Sanz, de l'Institut Municipal d'Investigació Mèdica (IMIM, UAB), per treballar en el projecte *Computer-Aided Molecular Modelling of G-Protein Coupled Receptors*, del 15 de juny al 31 d'agost.

Günter Dengel, de l'**Institut für Statistik und Dynamik der Luft- und Raumfahrtforschung (ISD)** (Alemanya), ha estat convidat per Enric Oñate, del CIMNE (UPC), per treballar en el projecte *Computer Analysis of Ships Motion under Different Sea Conditions*. Dengel va arribar el 15 de juliol i estarà a Barcelona fins al 15 de gener de 1999.

Rupert K. G. Croft, de l'**Oxford University** (Anglaterra), desenvolupa el projecte *Large Scale N-Body Simulations in Cosmology*, del 20 de juliol al 24 d'agost, convidat per Enric Cervera, de l'Institut d'Estudis Especials de Catalunya (IEEC).

Klaus-Dieter Höpfe, de la **Heinrich Heine-Universität Düsseldorf** (Alemanya), desenvoluparà el projecte *Advisoring on the Current Research in Computer-Assisted Analysis of Biomolecules* de l'1 d'agost al 30 de setembre, convidat per Ferran Sanz, de l'Institut Municipal d'Investigació Mèdica (IMIM, UAB).

És el primer cop que aquest programa va més enllà de l'àmbit empresarial i s'aplica a dades de la població per a l'estudi sociodemogràfic



LA NOVA ANELLA I L'EXEMPLAR V2250 ES VAN PRESENTAR PÚBLICAMENT EL 29 DE MAIG

Jornada d'inauguracions

La nova Anella Científica i la darrera adquisició de maquinari, l'Exemplar V2250, que es van presentar oficialment el passat 29 de maig, han estat les accions més importants que ha dut a terme el CESCA aquest any.



Joan Albaigés talla la cinta que envoltava la nova màquina.

La nova Anella (TERAFLOP 33) i la compra de l'Exemplar V2250 (TERAFLOP 33) es van inaugurar coincidint amb la reunió del Consell de Govern, al matí del mateix dia 29 de maig. A una de les sales de l'Edifici Annexus es va instal·lar una estació de treball amb set finestres d'Íbbonc obertes. Les finestres permetien la comunicació amb la FCR, la UAB, la UPC, la UdL, la UOC, la URV i la UdG (aquestes dues darreres, només amb connexió d'àudio). El Comissionat per a Universitats i Recerca, Joan Albaigés va conversar amb els representants de les universitats a través de la videoconferència i va inaugurar, d'aquesta manera, la nova Anella Científica. Els vicerectors també van tenir ocasió de xerrar amb la gent de les seves respectives universitats utilitzant l'Íbbonc: van quedar satisfets de l'experiència.

La presentació oficial de la nova màquina del CESCA, l'Exemplar V2250 també va anar a càrrec de Joan Albaigés. El comissionat va ser l'encarregat de tallar la cinta amb els colors de la bandera catalana que envoltava la màquina com si fos un regal.

Entre la vintena de convidats as-

sistents hi havia Joan Albaigés, comissionat per a Universitats i Recerca; Antoni Giró, director general d'Universitats; Josep A. Pla, director de la Fundació Catalana per a la Recerca; David Serrat, vicepresident de la CIRIT; el rector de la UOC, Gabriel Ferrater, i vicerectors de la resta d'universitats del Consorci; Josep Garriga, director territorial d'empreses de Telefónica, com a membre del patronat de la FCR; Xavier Peiró, director del Centre de Telecomunicacions de la Generalitat de Catalunya; i un representant d'IBM i de Convex, empreses a què pertany el maquinari del CESCA.



Gabriel Ferrater parla amb membres de la UOC a través de l'Íbbonc.

B R E U S

Nou comissionat per a la Societat de la Informació

Des de principis de juny la Generalitat de Catalunya té un nou comissionat que depèn directament del Departament de Presidència: el comissionat per a la Societat de la Informació. Aquest nou lloc l'ocuparà l'actual secretari general d'Indústria, Comerç i Turisme, Miquel Puig (Tarragona, 8-5-1954). Des d'aquest càrrec es vol aglutinar les diverses tasques de telecomunicacions, informàtica i d'altres sectors audiovisuals que fins ara estaven assignades a diferents departaments de la Generalitat.

Vall d'Hebron deixa l'Anella

L'Hospital General de la Vall d'Hebron s'ha donat de baixa com a institució connectada a l'Anella Científica des de mitjan de maig. La prioritat de l'hospital era tenir una connexió d'alta velocitat amb els Estats Units que no podia obtenir amb la proporcionada per RedRIS, i ha optat per passar a un proveïdor privat. Després d'aquest canvi, hi ha 49 institucions connectades a l'Anella.

Canvis en la connexió IBERNET

Les dues línies de 2 Mbps d'IBERNET que connecten RedRIS i el punt neutre han estat substituïdes per un circuit virtual permanent ATM de 3200 Kbps. El canvi de routing es va fer el passat dimarts 27 de maig, a les 8.30 del matí i va suposar una desconnexió d'uns pocs minuts entre RedRIS i el punt neutre.

ELS BENCHMARKS ESTANDARD I ELS DE GAUSSIAN S'EXECUTEN UN 30% MÉS RÀPID ALS NOUS PROCESSADORS

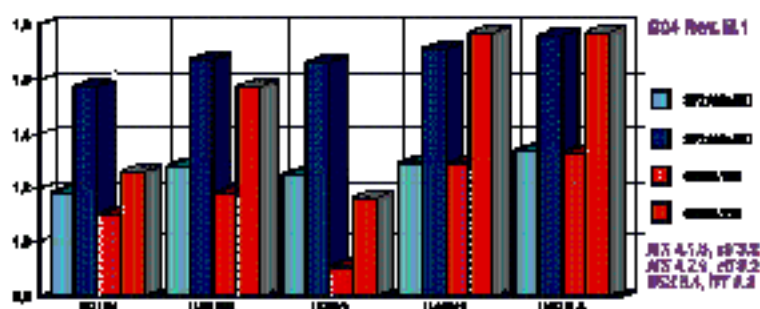
L'O2000 i l'SP2 s'actualitzen

L'Origin 2000 del CEPBA s'ha substituït per una nova màquina de les mateixes característiques però amb els processadors R10000 a 500 Mflop/s de rendiment punta en comptes dels de 392 Mflop/s anteriors. El nou superordinador es troba al CEPBA des del 27 de maig. Per la seva banda, el maquinari del CESCA també ha patit variacions: els 32 processadors thin120 de 480 Mflop/s de l'SP2 van ser substituïts per thin160 de 640 Mflop/s, el passat 26 de juny.

Característiques tècniques i rendiment relatiu al processador wide

	wide	thin120	thin160	PA8200	R10000	
Freqüència (MHz)	66	120	160	240	195	250
Amplada bus	256	256	256	64	128	
Cache dades (KB)	256	128	128	2.048	4.096	
R. punta (Mflop/s)	(266)	1,80	2,41	3,61	1,47	1,88
LINPACK TPP	(236)	1,72	2,25	3,15	1,46	1,91
LINPACK 100X100	(130)	1,79	2,42	1,56	0,88	1,28
SPECint95	(3,8)	1,48	2,26	4,31	2,49	3,87
SPECfp95	(12,4)	1,34	2,06	2,00	1,53	1,97
CESCA/QT	—	1,27	1,67	N/d	1,15	1,48

Speedup respecte al processador wide/1GB



USAN: 8 nodes HP4-R10000 oct-direct 164 base functions
 UASJL: 16 nodes thin120 oct-direct freq est 199 base functions
 UASJL: 16 nodes thin160 oct-direct 148 base functions
 USJL: 32 nodes HP4-R10000 oct-direct opt geom/indop 304 base functions
 UASJL: 32 nodes thin120 oct-direct opt 135 base functions

0:01:12
 3:57:44
 21:44:18
 0:38:21
 4:38:24

Com a conseqüència dels problemes experimentats a l'Origin 2000 des de Setmana Santa, Sibcon ha pres la decisió de reemplaçar totalment la màquina. El rendiment punta de la nova és de 32 Gflop/s, un 28% addicional. El rendiment màxim en resoldre un sistema d'equacions lineal (Rmax) és de 25,30 per a un Nmax de 43.200 elements.

Els 32 processadors thin120 que es ven instal·lar el 5 de maig de 1997 s'han substituït per thin160. El rendiment punta ha passat de 22,29 Gflop/s a 27,41, un 23% addicional respecte al que havíem anunciat anteriorment, i un 48% respecte a la potència disponible l'any passat. L'Rmax estimat ha augmentat de 16,17 a 19,72 Gflop/s.

L'estructura final de l'SP2 serà de 42 processadors thin160 i dos wide, agrupats en dues màquines

Per contra del que s'havia anunciat, les dues màquines no s'ajuntaran. L'estructura final de l'SP2 serà de 42 processadors thin160 i dos wide, agrupats en dues màquines, una d'elles amb 32 thin160 i la xarxa d'interconnexió d'alta velocitat (120 MB/s), i l'altra, amb la resta de processadors connectats per Ethernet (10 Mbps).

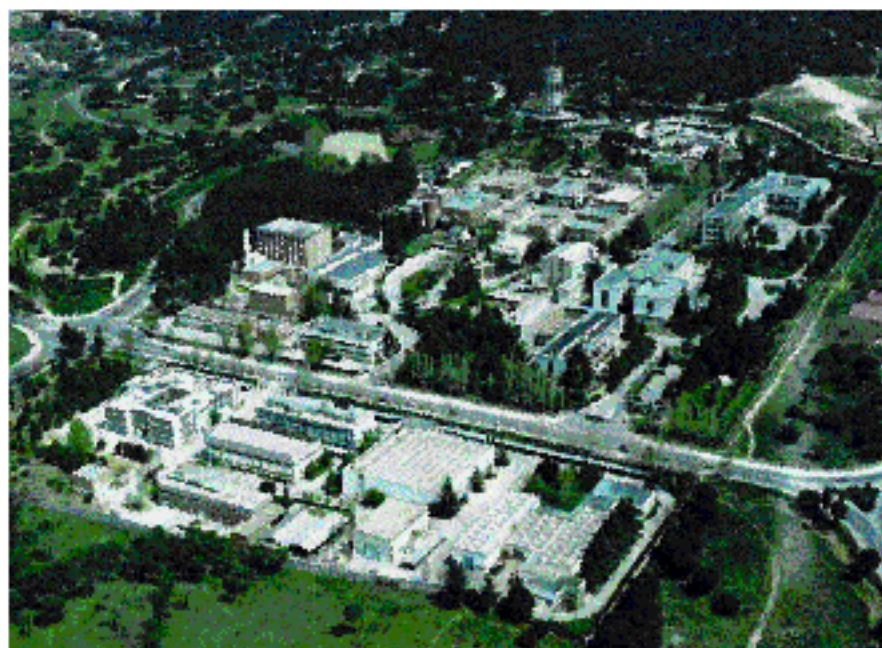
De mitjana, tant els benchmarks estàndard com els nostres benchmarks de Gaussian s'executen un 30% més ràpid als nous processadors de l'SP2 i de l'O2000. Aquest augment és proporcional a l'increment de la freqüència del processador. Els resultats dels nostres benchmarks a l'V2250 no estan disponibles ja que estem pendents de rebre els compiladors de C i Fortran per a la versió HP-UX 11.0. Per aquest motiu la nova màquina encara no està al vostre servei.



EL CENTRE DISPOSA D'UN CRAY J90 I UN CRAY T3E

Ciemat: de la fitxa perfor

El Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (Ciemat) té el seu origen en l'antiga Junta de Energía Nuclear (JEN), que es va crear el 1951 amb l'objectiu d'investigar en el camp de la tecnologia nuclear a Espanya. Des d'aquell any i fins ara el centre s'ha anat transformant i el seu maquinari ha anat evolucionant. Ara per ara el Ciemat té disponibles dos superordinadors de la marca CRAY: un J90 i un T3E.



Vista general de les instal·lacions del Ciemat a Madrid.

El 1982 el centre va passar a dir-se Ciemat i va diversificar les seves activitats en el camp energètic. El 1986 es va reestructurar amb l'objectiu de promoure i dur a terme projectes d'investigació, innovació i desenvolupament tecnològic i per tal d'aconseguir la incorporació de les tecnologies energètiques a la societat. Les seves principals activitats se centren a:

a) Demostrar el paper de la fusió nuclear en les seves vessants científica i tecnològica, i la participació en programes supranacionals com JET o NET-JTER.

b) Progressar en la viabilitat de l'energia de fusió amb activitats que augmentin la disponibilitat i seguretat de

les plantes en operació, desenvolupar projectes d'I+D referents a la gestió de residus radiactius, i vigilar la descontaminació i el desmantellament d'instal·lacions mitjançant estudis.

c) Millorar l'eficiència i la qualitat dels combustibles fòssils amb les activitats per a l'adaptació de les instal·lacions nacionals a les normatives europees, mitjançant el desenvolupament i la implantació dels sistemes avançats de combustió i gasificació dirigits a una generació energètica més eficaç i neta.

d) Incrementar la competitivitat de les energies renovables amb la realització de projectes d'I+D en les àrees de l'energia solar, eòlica i biomassa.

e) Contribuir a la implantació d'objectius de qualitat ambiental amb activitats d'investigació i suport tècnic.

El Ciemat té tres centres d'investigació: un a Madrid, un a Almeria i un a Sòria. El Centre de la Moncloa (Madrid) és la seu central de l'organisme: allotja la majoria dels recursos humans i materials. La Plataforma Solar d'Almeria (PSA), operada conjuntament amb l'institut alemany DLR, és un dels centres europeus d'energia solar tèrmica més importants. El Centre de Desenvolupament de Energies Renovables (Ceder) de Sòria té per activitat principal l'aprofitament energètic de la biomassa.

Per les seves activitats d'investigació i desenvolupament, el Ciemat sempre ha disposat de computadors per al càlcul científic. El primer va ser als anys 60, un UCT, una autèntica "ordenadora" de paraules, sense llenguatges de programació. En els 70 es va incorporar l'Univac 1100, amb el sistema operatiu Exec8, compilador de Fortran II i fitxes perforades. Cap a l'any 78, apareixen els primers terminals i permeten el que avui és trivial: el diàleg directe amb les màquines. El 1979, s'instal·la la base de dades SIMBAD i el gestor documental UNIDAS cosa que suposa un salt qualitatiu en la informàtica de gestió. El 1985, apareix l'arquitectura de xarxa local, amb màquines Digital VAX i IBM 3090 i 4381, i el Ciemat es converteix en el node principal de la xarxa nacional FAENET de Física de Altes Energies.

Als anys 90 es produeix el salt cap a sistemes oberts, basats en Unix:

La potència de càlcul del CIEMAT és de 39,2 Gflop/s



a a la supercomputació

l'estàndard de comunicacions TCP/IP. Es multipliquen els ordinadors personals, es transforma la manera de "fer informàtica" i l'arquitectura dels sistemes d'informació, que passen a client/servidor. El canvi més recent es produeix el 1991 amb la introducció de la supercomputació que es consolida el 1995 i 1996 i dona lloc a l'actual centre de computació d'altas prestacions. Actualment l'arquitectura informàtica del Cimat està basada en una estructura client-servidor i es disposa de servidors especialitzats en càlcul científic:

• **Computador Cray J90.** És una màquina vectorial amb 16 processadors, velocitat de rellotge de 10 ns i un rendiment punta total de 3,2 Gflop/s. Disposa d'una memòria principal de 4 GB compartida i d'una capacitat d'emmagatzematge en disc de 100 GB. Treballa amb el sistema operatiu UNICOS i disposa de compiladors adients al càlcul científic (FORTRAN 77, FORTRAN 90 i C), de llibreries matemàtiques

(IMSL i NAG), gràfiques (DISSPLA i NCAR), i de programari genèric (GAUSSIAN 94). S'usa generalment per a aplicacions de fusió termonuclear, modelització de cèl·lules fotoelèctriques, modelització atmosfèrica, càlcul de resuspensió, càlculs de blindatges, anàlisis d'estructures químiques i simulacions numèriques en general.

• **Computador Cray T3E.** És un ordinador amb arquitectura paral·lela MPP de 40 processadors, cicla de rellotge de 2,2 ns, un rendiment punta total de 36 Gflop/s i un R_{max} estimat de 24,46 Gflop/s. Té una memòria principal de 5 GB, distribuïda en 128 MB per processador, i una capacitat d'emmagatzematge de 206 GB. Treballa amb sistema operatiu UNICOS/mk i disposa de compiladors FORTRAN 90 i C++, així com de llibreries matemàtiques. Els processadors són DEC Alpha 21164. S'usa generalment en fusió termonuclear, dinàmica molecular, simulacions en física de partícules, disseny de lentos electromagnètics, tecnologia nuclear i càlculs de resuspensió.

A més a més, el Cimat té un cluster d'estacions de treball, constituït per 4 computadors HP 9000/C110 configurats en cluster mitjançant una



El Cray J90.

connexió d'alta velocitat. Finalment, també hi ha disponibles servidors per a la gestió administrativa i tècnica i servidors de propòsit general, entre els quals destaca una O2000, basada en el processador MIPS R10000 a 180 MHz, amb una memòria de 256 MB i connectada a un robot de cintes magnètiques STK Wolfcrack 9360 de 25 TB de capacitat. Aquesta màquina actua com a servidor d'emmagatzematge massiu del Cimat.



El Cray T3E, amb un rendiment punta de 36 Gflop/s.

Taula comparativa de les característiques tècniques dels processadors dels dos Cray

	Cray J90	Cray T3E
Freqüència (MHz)	100	450
Amplada de banda (GB/s)	1,6	4,8
Cache de dades (KB)	n/d	8
Cache d'instruccions (KB)	n/d	96
Rendiment punta per processador (Mflop/s)	200	900
LINPACK 100x100	106	n/d
LINPACK TPP	193	n/d

L'INEM i el CIEMAT s'incorporen als TOP500

La darrera edició de la llista dels TOP500 inclou cinc centres de l'Estat espanyol: el CIEMAT, l'INEM, el CEPBA, la Universitat de València i el CESCA (per ordre d'aparició). En l'edició del mes de juny, s'observa que Sun és la marca que ha escalat més en la llista pel que fa al nombre de màquines.

En la llista publicada el 18 de juny s'inclouen cinc centres de l'Estat (un centre més que a l'edició de novembre) amb maquinari de gran potència de càlcul: el CIEMAT, l'INEM, el CEPBA, la Universitat de València i el CESCA. Els dos primers és la primera vegada que apareixen a la classificació, mentre que la resta es mantenen. En aquesta edició, en canvi, desapareix de la llista Puertos del Estado

(EPPE). El CIEMAT ocupa la posició 333 amb un Cray T3E de 36 processadors, amb un R_{max} de 22,04 Gflop/s. L'INEM se situa només una posició per darrera amb un Sun HPC 10000 de 52 processadors de 21,68 Gflop/s d' R_{max} . El CEPBA està a la posició 359 amb el seu O2000 i un R_{max} de 20,10, el mateix que la Universitat de València que, amb la mateixa màquina, ocupa el lloc número 300. L'O2000 del CEPBA i la

UV han augmentat el seu R_{max} de 19,23 Gflop/s a 20,10 Gflop/s i han passat de tenir un N_{max} de 32.000 a 40.000. El CESCA, finalment, està a la posició 358 amb un R_{max} de 15,56. Cal apuntar que les recents actualitzacions del maquinari del CESCA i del CEPBA no han estat incorporades en aquesta llista.

Pel que fa al maquinari, Sun continua la seva escalada dins de la llista: ha passat de tenir 85 màquines a tenir-ne 111; i de 1,44 Tflop/s de rendiment punta de càlcul a 2,56. "Estem en la computació d'altres prestacions per ser els número 1", explica Clark Masters vicepresident del data center/HPC product group, de Sun. Els altres fabricants es mantenen en el seu nivell de màquines presents en la llista o bé baixen de lloc. Així, IBM té 75 màquines: HP, 25. SGI, per contra, passa de les 216 màquines a la llista a les 200, tot que continua essent el fabricant amb més supercomputadors en el rànquing.

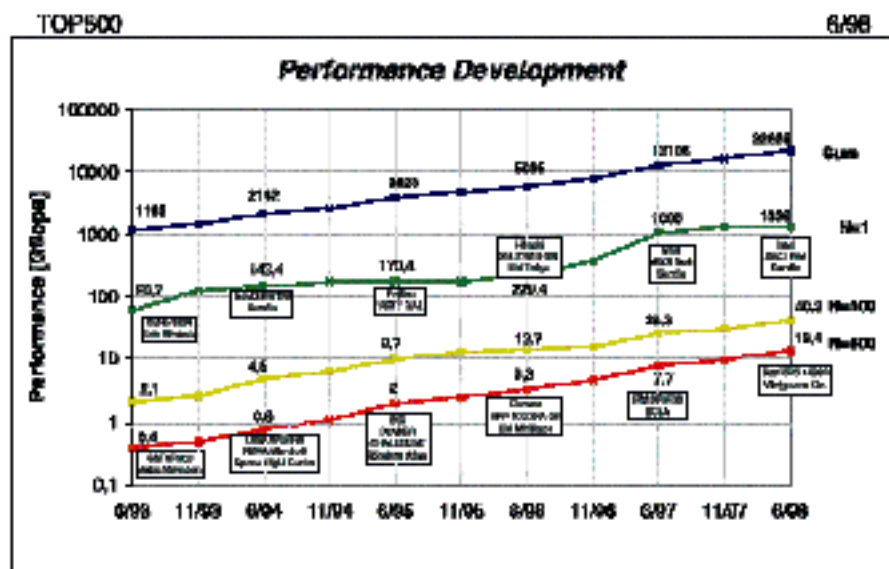
Una dada curiosa és que l'ASCI Red d'Intel amb 9152 processadors: un R_{max} de 1,34 Tflop/s, instal·lat a Sandia National Laboratories, es manté per tercera vegada consecutiva a capdavant de la llista. Hans Meuer que va presentar l'edició de juny dels TOP500, estima que aquesta màquina és tan potent que estarà present en la llista fins més enllà de l'any 2005.

Què són els TOP 500?

Els TOP500 és una llista que elaboren Jack Dongarra i Erich Strohmaier, de la Universitat de Tennessee (EUA), i Hans Meuer, de la Universitat de Mannheim (Alemanya), i que classifica els 500 ordinadors més potents del món segons la màxima potència de càlcul obtinguda (R_{max} o rendiment punta) en resoldre un sistema d'equacions lineals (LINPACK).

La gràfica ens aporta diferents dades.

- La línia de color blau representa la suma total dels R_{max} de les màquines dels TOP500. En un any la potència total ha crescut un factor d'1,9, l'equivalent a l'augment que ha tingut el CESCA.
- La línia de color verd indica quina és la màquina més potent en cada edició. En les tres darreres edicions Intel ha dominat aquesta classificació.
- El color grec indica l' R_{max} de la màquina número 100 de la llista.
- La línia de color vermell indica l' R_{max} de l'última màquina de la llista.



Informació Quàntica a Benasc

José Ignacio Latorre i Pere Pascual, del Departament d'Estructura i Constituents de la Matèria de la Facultat de Física de la Universitat de Barcelona, comenten en el seu article els fonaments teòrics de la computació quàntica. Aquest serà un dels temes en què es farà especial èmfasi a la trobada Centre de Física de Benasc (vegeu contraportada) que es farà del 5 al 25 de juliol en aquesta petita localitat dels Pirineus.

■ Els ordinadors digitals han sofert en el decurs d'aquests darrers anys un gran desenvolupament augmentant considerablement la seva velocitat i disminuint la seva mida, però sempre basant-se en els mateixos principis computacionals. La mecànica quàntica pot canviar això. Un electró pot tenir el seu espín orientat en una direcció arbitrària de l'espai, representem aquest estat amb la notació $|\psi\rangle$, o pot tenir-lo orientat en el sentit contrari, $|\phi\rangle$. La novetat és que l'electró pot també trobar-se en un estat superposició dels dos anteriors,

$$|\psi\rangle = \alpha|\phi\rangle + \beta|\psi\rangle, \quad |\alpha|^2 + |\beta|^2 = 1.$$

Quan mesurem el seu espín en la direcció considerada hi ha una probabilitat $|\alpha|^2$ de trobar-lo en aquesta direcció i $|\beta|^2$ en l'oposada. Aquesta propietat fonamental que està a la base de la mecànica quàntica va portar a D. Deutsch¹ a

definir el *quantum-bit*, dit *qubit*, que contén les idees de computació clàssica permetent superposicions de bits. És aleshores possible tenir un registre físic d'entrada, per exemple, del tipus

$$|\psi\rangle = \alpha_1|11\dots 1\rangle + \alpha_2|11\dots 0\rangle + \dots$$

Aquest input és superposició de molts inputs clàssics. A més, es poden construir portes lògiques que actuen com a transformacions unitàries en aquest sistema. Tenim doncs els conceptes bàsics per a construir computadores quàntiques, m'assivem ent per a les dades que els registres d'entrada poden contenir la superposició d'un gran nombre de registres clàssics.

El recent desenvolupament de la computació quàntica ha estat catalitzat pels nous resultats experimentals de diferents grups arreu del món.

Cal destacar l'aplicació d'aquestes idees per transmetre informació en

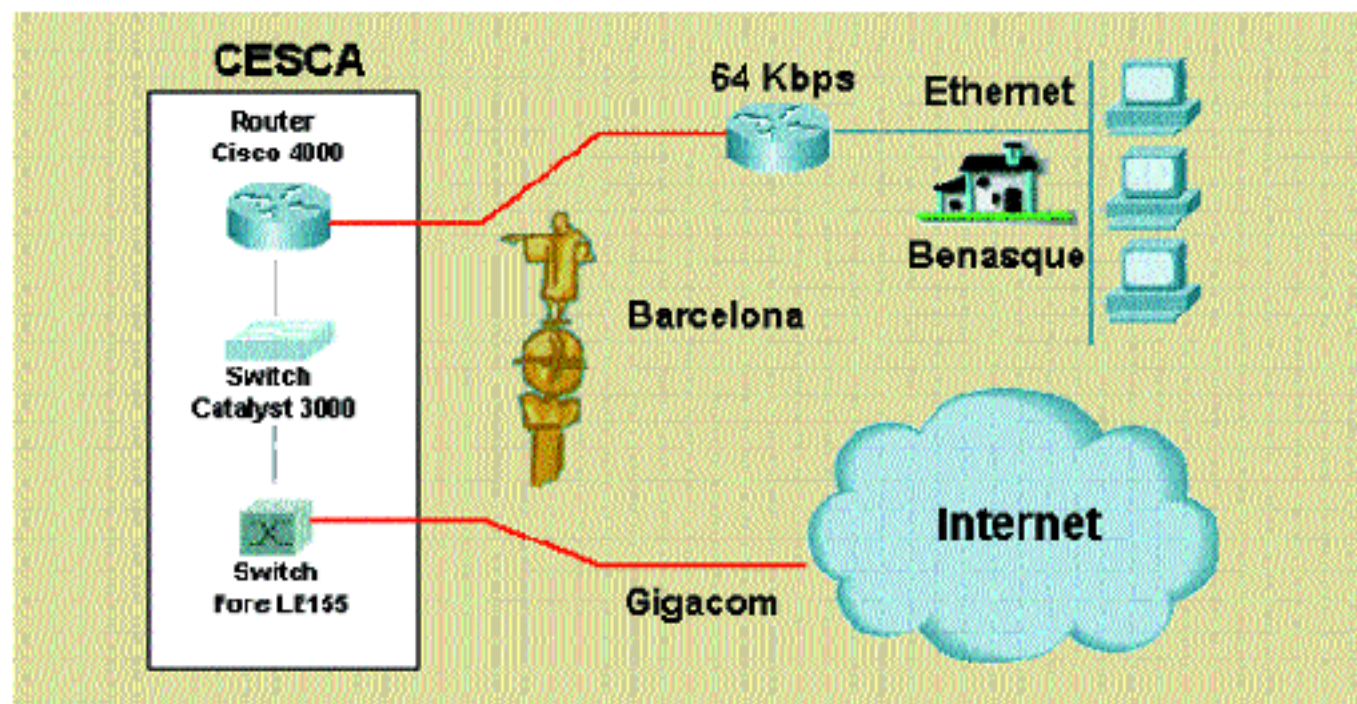
forma segura (encriptació quàntica) com ja s'ha fet en un prototip que opera entre Nyon i Ginebra, finançat per la PTT; l'ús d'entrellaçat quàntic per la teleportació, la construcció de les primeres portes lògiques quàntiques universals, etc. Des d'un punt de vista més teòric s'han descobert nous algorismes quàntics. Per exemple, P. W. Shor² ha proposat un mètode de factorització polinòmic, és a dir, que el temps necessari per factoritzar un nombre no creix exponencialment amb la grandària de l'input d'entrada. Un problema clàssicament no polinòmic queda doncs reduït a un problema polinòmic. Les classes de complexitat són diferents en la lògica quàntica. El lector més interessat amb tots aquests desenvolupaments pot consultar la referència.³

1. D. Deutsch, Proc. R. Soc. London Ser. A400 (1985) 97.

2. P. W. Shor, <<http://xxx.unizar.es/abs/quant-ph/9508027>>.

3. A. Ekert i R. Jozsa, Rev. Mod. Phys. 68 (1996) 733.

4. A. Steane, "Quantum Computing", <<http://xxx.unizar.es/abs/quant-ph/9708022>>.



Cray Y-MP: un record recent

El darrer supercomputador que es va desconnectar al CESCA, el desembre de 1997, va ser el Cray Y-MP. Molts dels nostres usuaris recordaran amb estima aquesta màquina que encara es pot veure a les instal·lacions del CESCA.

Breu cronologia

Data de disseny	1988
Data de fabricació	1990
Data de compra	1992
Data d'instal·lació	1993
Data de desconnexió	1997



JOAN PASTOR

El Cray Y-MP i els seus discos, en la seva configuració final.

El segon Cray que al CESCA va instal·lar el 1993 era un Y-MP 232 provinent del Centre Meteorològic de Finlàndia (2 CPU, 32 MW o 256 MB de memòria central). Quan la màquina va actualitzar-se el 1996 i va convertir-se en Y-MP 464 (4 CPUs, 64 MW o 512 MB de memòria) va venir de Peugeot, a França. El CESCA va ser l'únic centre d'Espanya que va allotjar un Cray Y-MP d'aquest model que era el més potent de tot l'Estat en la seva categoria. El sistema operatiu que s'usava era l'UNICOS, que era molt robust i sofisticat.

Els Cray d'aquest model venen amb els primers que podien refrigerar-se per aire mitjançant ventiladors, ja que els anteriors precisaven d'un sistema de refrigeració per aigua. Quan es va passar de l'X-MP a l'Y-MP es va doblar el nombre de processadors (d'un a dos), el rendiment punta va passar de 332 Mflop/s a 667 Mflop/s, i la memòria i els I/O també eren més ràpids.

A més del programari heretat del Cray X-MP, es van instal·lar dos productes molt peculiars: l'UniChem i l'Ensignht. L'UniChem constava d'una interfície gràfica instal·lada a una estació de treball SGI (de la qual s'aprofi-

tava la potència visualitzadora) i d'uns programes de química computacional que usaven la potència de càlcul del Cray. El cas de l'Ensignht (abans anomenat MPQS) era diferent: només visualitzava, no calculava, però el que havia de visualitzar era tan complex que necessitava que el Cray ajudés en la visualització. Les primeres imatges del projecte Meteo, estudis de la cripta de la colònia Güell, o la simulació del xoc de dues estrelles nenes blanques es van fer amb aquest producte.

Amb el Cray Y-MP bàsicament es van continuar els projectes que ja hi havia el Cray X-MP ja que, a mesura que l'IBM 3090 es va anar fent vell, els usuaris es van passar al Cray. Alguns d'ells van ser SEAT; l'Institut d'Investigació Aplicada a l'Automòbil de la

Generalitat (Idiada), que va fer uns càlculs per idear un mecanisme que evités el refament del sostre d'autobusos en cas de bolcada; SENER, una enginyeria de Bilbao que s'encarregava del càlcul d'estructures d'enginyeria civil, entre d'altres usuaris. També cal destacar el projecte d'Ignasi Fite (CSIC) que va resoldre l'estructura de la molècula més gran que mai abans s'havia resolt per raigs X.

Un dels problemes que presentava el Cray Y-MP era la manca d'espai de disc temporal. Això va fer-se bastant evident quan als químics van començar a fer gran quantitat de càlcul. La màquina era petita per als 350 usuaris que tenia. Hauria hagut de ser 10 vegades més gran (especialment quan es treballava amb 2 processadors) del que era en realitat. La càrrega de CPU estava sempre al 100% i els discos sempre es quedaven petits. Va ser llavors quan va quedar de manifest que s'havia d'apostar pel m'acquirir: comprar un ordinador nou.

El Cray Y-MP encara va duplicar de nou el nombre de processadors i la seva potència (que va arribar al 1,2 Gflop/s) i va ampliar en 30 GB el seu espai de disc (fins a 72 GB) el maig de 1997, però finalment es va desconnectar el 19 de desembre del mateix any perquè els seus costos d'operació eren massa elevats. Tot i així, la màquina encara es pot veure a les instal·lacions del CESCA.

Característiques tècniques del Cray Y-MP

	Cray Y-MP 232	Cray Y-MP 464
Freqüència (MHz)	167	167
Rendiment punta (Gflop/s)	0,67	1,33
LINPACK TPP (Gflop/s)	0,60	1,16
LINPACK 100X100 (Mflop/s)	161	226



Cap
Pedro de Miguel

Integrants

Jesús Carretero, Fernando Pérez i Félix Garcia.

Període

Octubre 1996 - 2000.

Publicacions

- "High Performance Cache Management for Parallel File Systems". *3rd International Meeting on Vector and Parallel Processing (VECPAR'98)*.
- "XMPI: Una implem entación multithread de MPI". *Informe Técnico FIM-109.1-DATSI-98*. Facultad de Informática, UPM, 1998, p. 11
- "Evaluating the ParFiSys Cache Coherence Protocol on an IBM SP2". *Informe Técnico FIM-104.1-DATSI-98*, Facultad de Informática, UPM, 1998, p. 18.
- "Eliminating False Sharing in Cache Coherence Protocols for Parallel and Distributed File Systems: The ParFiSys Approach". *Proceedings of the IASTED International Conference on Parallel and Distributed Systems, Euro-PDS'97*, Barcelona, June 9-11, 1997, p. 31-37.
- "Evaluating ParFiSys: a High-Performance Parallel and Distributed File System". *Journal of Systems Architecture*. Elsevier, vol. 43, 1997, p. 533-542.
- "Cache Coherence in Parallel and Distributed File Systems". *Proceedings of the 5th EUROMICRO Workshop on Parallel and Distributed Processing*, IEEE Computer Society Press, London, January 22-24, 1997, p. 80-85.

Sistemes de fitxers paral·lels per a multicomputadors

"ParFiSys és un sistema de fitxers paral·lels, la principal característica del qual és que utilitza caches coherents distribuïdes pels nodes clients de l'arquitectura". *Pedro de Miguel*

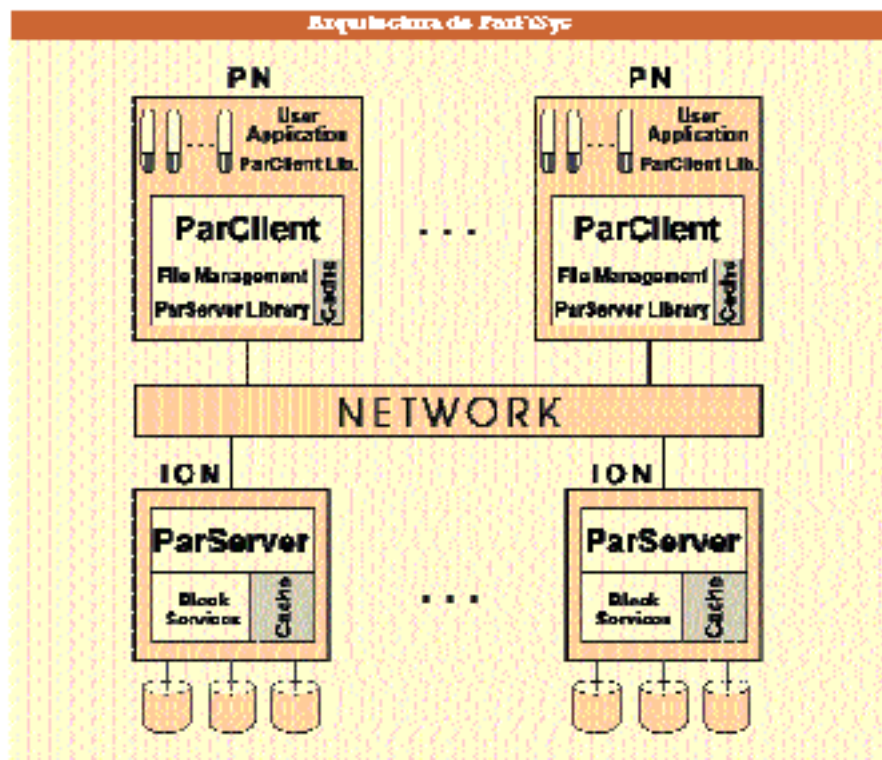
Tot el treball que estem desenvolupant sobre sistemes de fitxers paral·lels té el seu origen en el projecte anomenat ESPRIT GPMIMD (P-5404) de la CEE. L'objectiu d'aquest projecte, que va començar el 1990, era la construcció d'una màquina massivament paral·lela de propòsit general, denominada GPMIMD, juntament amb el programari de sistema i les aplicacions necessàries per a convertir-la en un producte comercial. Entre tot el programari de sistema a desenvolupar per aquesta màquina, es trobava la construcció d'un sistema de fitxers paral·lel per part de la Universidad Politécnica de Madrid (LPM).

Els resultats del treball de la LPM ven donar lloc, durant l'any 1995, a un primer prototip de sistema de fitxers paral·lel, denominat en l'actualitat PerFiSys. Aquest és un sistema de fitxers paral·lels les principals característi-

ques del qual són: un gran rendiment, capacitat de creixement, distribució de dades i accés paral·lel a fitxers, coherència de dades a nivell global, possibilitat de donar servei de forma concurrent, facilitat per transportar aplicacions des de sistemes ja existents i extensions paral·leles a la interfície del sistema de fitxers, entre les que es troben operacions d'E/S cooperant, multi-fitxers, operacions asíncrones, operacions de preassignació d'espai per a fitxers i serveis per a controlar el funcionament de la cache.

ParFiSys ha estat transportat amb èxit a diferents plataformes: multicomputadors basats en transputers T800 y T9000, xarxes d'estacions de treball i màquines IBM SP2.

El nostre treball futur se centra en el desenvolupament d'un sistema de fitxers paral·lels obert, adient per executar diferents tipus d'aplicacions: científiques, multimèdia, etc.



Quarta edició del Centre de Física de Benasc

La trobada Centre de Física de Benasc* (Osca) arriba aquest any a la seva quarta edició. Les jornades reuneixen del 5 al 25 de juliol uns 70 científics de 16 països per tal de treballar en aspectes teòrics i experimentals de la Teoria de la Informació Quàntica i, en concret, aprofundir en la computació quàntica, la criptografia i la comunicació, i l'òptica quàntica i la física atòmica. Com cada any, el CESCA possibilita la connexió externa, amb una línia punt a punt de 64 Kbps

que posa Telefónica i que connecta Benasc amb el router de l'edifici Annexus. Les màquines les proporciona Silicon Graphics i els físics pensen fer un cluster de 10 estacions de treball. La coordinació científica d'enguany correspon als professors A. Ekert (Oxford Univ.) i I. Cirac (Univ. Innsbruck). Podeu trobar un article sobre el tema principal que es tractarà en la trobada a la pàgina 9 del TERAFL0P.

* <http://sophia.ecm.ub.es>

Edita



AMB EL SUPORT DE



Generalitat de Catalunya



Universitat de Barcelona

Universitat Autònoma de Barcelona

Universitat Politècnica de Catalunya

Universitat Pompeu Fabra

Universitat de Girona

Universitat Rovira i Virgili

Universitat de Lleida

Universitat Oberta de Catalunya

CSIC

TERAFLOP

DIRECTOR

Miquel Huguet

COORDINADORA

Alicia Martínez

REDACCIÓ

Mònica Tudela

COL-LABORADORS

Jesús Casado

Cristina Guisande

Enric Torres

DISSENY I PRODUCCIÓ

Subirà & Associats

CESCA

Gran Capità, 2-4

08034 Barcelona

Tel. 93 205 64 64

Fax: 93 205 69 79

<http://www.cesca.es>

teraflop@cesca.es

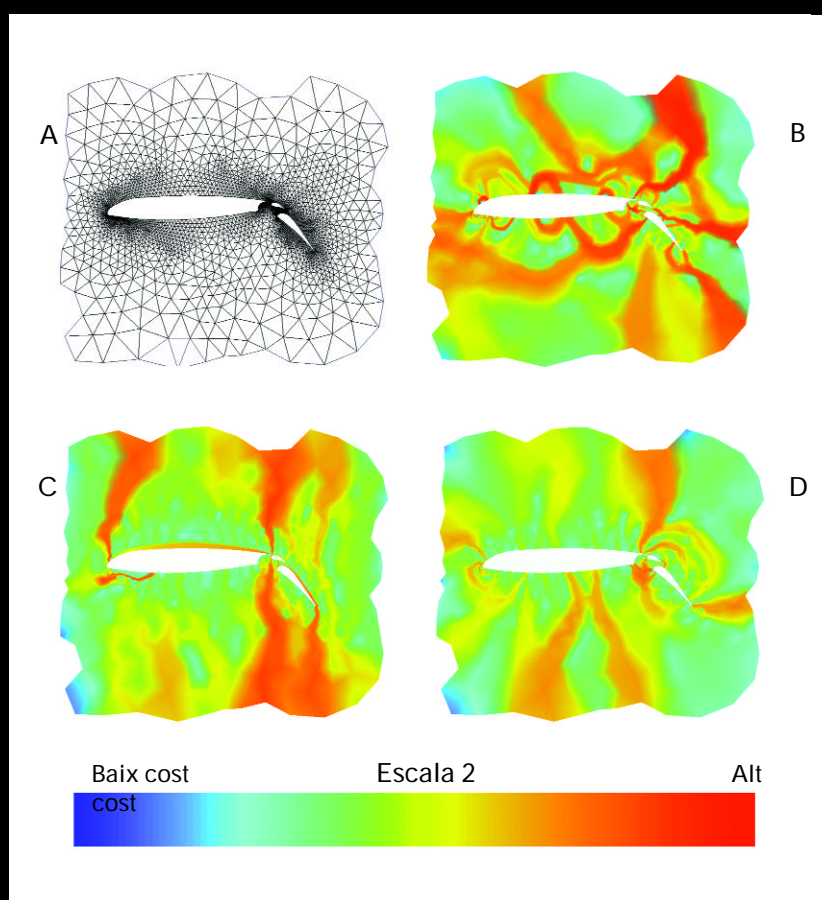
DIPOSI T LEGAL: B-33512-94

ISSN: 1134-6671



Exemplar gratuït

F O T O / N O T Í C I A



L'estudi d'heurístiques d'aproximació per a problemes intractables és un actual i actiu camp de recerca on el paral·lelisme juga un important paper. La figura A mostra una instància del problema de l'Arranjament Lineal Mínim. La resta d'imatges mostren a través de l'escala de colors la qualitat de les solucions generades per diferents heurístiques: Escalada Monòtona (B), Algorisme Voraç (C) i Seqüenciació Espectral amb Refredament Simulat (D). La paral·lelització d'aquesta darrera heurística ha estat desenvolupada i avaluada per Jordi Petit a l'SP2, donant lloc a uns resultats excel·lents, tant pel que fa a la qualitat de la solució com pel que fa al temps d'execució.