

## Manuel Delfino, director del Port d'Informació Científica

12/2007 - **Telecomunicacions, Electrònica i Informàtica.**

"La tecnologia GRID ja ha demostrat la seva utilitat als investigadors, malgrat no estar encara plenament consolidada." Manuel Delfino és director del Port d'Informació Científica (PIC), un centre tecnològic participat per la UAB, pel Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), per la Generalitat de Catalunya i per l'Institut de Física d'Altes Energies (IFAE), ubicat al campus de la UAB. En la entrevista següent, ens explica què és la tecnologia GRID i quins avantatges implica per a la recerca de moltes disciplines que requereixen emmagatzemar un elevat volum de dades.



Manuel Delfino és llicenciant en Matemàtica Aplicada, Enginyeria Mecànica i Física per la Universitat de Wisconsin (EUA) (1977), Màster en Física (1979) i Doctorat en Física amb especialització en Informàtica (1985). Amb la seva tesi doctoral, realitzada utilitzant l'Accelerador Linial de Stanford, va trobar la primera evidència de la interacció dèbil entre l'electró i el positró. Com a part de la tesi, també va contribuir a la tècnica d'instrumentació Calorimetria Gasosa per a detecció de partícules.

Especialitzat en computació científica aplicada, ha participat en els projectes IFAE-FALCON (1987-1992) i CERN-RD-47 (1993-1995), pioners en la utilització de "granges" d'ordinadors per processament de dades científiques.

Pel que fa a la seva trajectòria professional, ha estat investigador del Supercomputer Computations Research Institute (EUA) (1993) i director de Tecnologies de la Informació del CERN (1999). Actualment, és director del Port d'Informació Científica i catedràtic de Física de la UAB.

Manuel Delfino dirigeix el Port d'Informació Científica (PIC) des de l'any 2003, data en què es va crear, estretament vinculat al projecte Large Hadron Collider (LHC), l'accelerador de partícules més gran del món, ubicat al Laboratori Europeu per a la Física de Partícules (CERN), a Suïssa. Aquest accelerador té per objectiu reproduir condicions similars a les que es van produir al Big Bang per estudiar els orígens de la matèria. És previst que entri en funcionament a mitjans del 2008, però ja han començat a realitzar-se les primeres simulacions.

Paral·lelament a la creació del LHC i del PIC, es va anar desenvolupant la tecnologia GRID, que permet repartir l'emmagatzematge d'aquest volum ingent de dades emeses pel LHC per tot el món i, alhora, que puguin ser accessibles per als investigadors des de qualsevol punt del planeta. Es tracta d'una tecnologia que ha mostrat també la seva utilitat en altres disciplines, com ara la biomedicina o la cosmologia, i amb un futur molt prometedor per a la recerca.

El PIC forma part de la primera corona d'onze centres -europeus, nordamericans i asiàtics- encarregats de l'emmagatzematge i el processament de les dades emeses pel LHC i de la seva posterior distribució a una segona corona de centres, repartits pels cinc continents que, a la vegada, proporcionen aquestes dades als milers d'investigadors de la física de partícules de tot el món. En aquest moment, la capacitat d'emmagatzematge del PIC és d'uns 300 terabites, però és previst que superi els 1.000 terabites en els propers mesos.

### - Quin és el principal avantatge de la tecnologia GRID?

- L'estructura GRID l'entendem com una plataforma, que té la seva similitud amb la xarxa elèctrica mundial, que parteix d'uns centres principals emissors de dades i que, en forma de raïm, acaba arribant als ordenadors personals dels investigadors. El

sistema GRID ens permet emmagatzemar un volum de dades molt important en diversos centres repartits per tot el món, fer-ne el seu processament i, posteriorment, donar, de manera global, a cada investigador, les dades que requereix per a la seva recerca. D'aquesta manera, s'agilita enormement el treball de l'investigador.

#### **- Ens pot posar un exemple de funcionament del sistema GRID?**

- L'accelerador de partícules LHC, per exemple, genera primer unes dades que són processades. És a dir, els detectors del LHC ens donen punts de l'espai per on passen les partícules, però a l'investigador físic l'interessen especialment les trajectòries per on passen aquestes partícules. Per tant, es produeix un procés de reconstrucció per traçar la trajectòria d'aquests punts, que poden tenir milers de trajectòries. Els ordinadors tracten aquesta informació bruta del detector de partícules i la transformen en informació útil per a l'anàlisi científica. En una segona part, els mateixos investigadors seleccionen quines són les dades que volen analitzar mitjançant un procés de filtrat del total de dades disponibles, per establir les seves hipòtesis de recerca. Aquest procés no és únic per a la física de partícules. Una màquina de ressonància magnètica, per exemple, produeix unes mides en punts per representar l'espai dins el nostre cervell i el procés de reconstrucció transforma aquests punts en una sèrie de dades útils per als radiòlegs. Si s'acumulen moltes ressonàncies magnètiques de molts pacients diferents, un radiòleg pot triar quin és el criteri que vol establir per filtrar el total de ressonàncies disponibles i obtenir només les que s'ajustin a uns determinats paràmetres. Això li permetrà establir teories i hipòtesis sobre diferents alteracions cerebrals.

#### **- Quines àrees de coneixement se'n poden beneficiar?**

- Encara que el primer objectiu del PIC va ser el tractament de les dades del CERN i del LHC, es va tenir també present que aquesta infraestructura pogués ser útil a altres àrees de la ciència. Actualment estem treballant amb enginyers aeronàutics, cosmòlegs, astrofísics i metges, no només per reutilitzar la capacitat de tractament de dades, sinó també la filosofia que implica la tecnologia GRID, de treballar en grans projectes internacionals.

Per exemple, per estudiar la malaltia de Huntington trobem que hi ha investigadors a Alemanya, al Regne Unit, a Itàlia i a Espanya, -a l'Hospital de la Santa Creu i Sant Pau-, que tenen tots moltes dades de les seves recerques, però que calen ser posades en comú per compartir i treballar globalment, pe tal d'obtenir millors resultats en investigació. Aquí té una utilitat real el sistema GRID perquè permet treballar a un conjunt de científics que estan dispersos globalment, però que es conceptualitzen ells mateixos com un únic grup o un únic projecte. Mitjançant serveis de software i acords d'identificació d'usuaris, com ara un DNI electrònic per a investigació acceptat mundialment, es connecten a centres discrets de processament de dades, que són els centres que donen servei al GRID.

Aquesta infraestructura permet que un centre, com ara el PIC, no serveixi només a Espanya o a Catalunya, o a un sol projecte. El mateix centre és part del sistema GRID i els equips d'investigació són organitzacions virtuals donades d'alta al GRID, que poden fer servir les dades emmagatzemades a tots els centres que estiguin en xarxa per a un projecte en concret, encara que estiguin a l'altra punta del món.

#### **- Com estan acollint els investigadors aquesta nova tecnologia?**

- La prova de foc del sistema GRID encara ha d'arribar perquè tota tecnologia innovadora requereix un període d'adaptació per part dels humans, que està calculat que sigui d'uns deu anys, i que permetrà que aquest projecte, que va començar amb caràcter estacional i temporal, esdevingui una realitat constant i permanent. L'any 2009 o 2010 pensem que la GRID europea serà una realitat consolidada plenament.

D'altra banda, hem de tenir en compte que el científic només utilitza quelcom, en aquest cas una tecnologia, si li serveix per fer millor la seva recerca. Jo considero que la tecnologia GRID ja està demostrant la seva utilitat als investigadors, però s'haurà de donar encara un marge de temps per a la seva acceptació entre la comunitat investigadora.

#### **- Qui pot ser usuari del PIC?**

- Triem els usuaris en funció del criteris següents: excel·lència en la recerca, que aquesta recerca es faci en un context internacional i que es generin dades en quantitat suficient com per no poder ser gestionada pels mitjans estàndards.

#### **- I els investigadors han de pagar pels serveis oferts?**

- Els investigadors usuaris del sistema GRID poden accedir a les dades per a la seva recerca des de diferents centres d'emmagatzematge. El sistema GRID els permet saber el cost que els suposa obtenir les dades de cada centre on estiguin disponibles i poden triar el que els suposi el cost mínim. Al PIC, per exemple, hi ha investigadors d'altres centres que ens demanen dades. I, també, els investigadors de Catalunya demanen accés per utilitzar els altres centres de dades a canvi. Existeix una comptabilitat global on cada mes anem mesurant, de manera estadística, el gruix de serveis que donem a cada investigador per tal d'ajustar els costos dels centres i crear una mena de mercat de processament de dades. La nostra no és una plataforma per investigar sobre GRID, és una GRID de producció. Si aturem la GRID, el Large Hadron Collider i altres projectes deixaran de funcionar. Per això, amb finançament europeu, hem organitzat Europa en un seguit de regions i, en cada indret, vertebrem un centre d'operacions que assegura que aquesta infraestructura es mantingui coherent dins la regió i es comunica amb els altres centres de control de les altres regions. Aquest projecte, denominat EGEE (Enabling Grids for E-sciencE), implementa tots els serveis de software que oferim.

#### **- Com es finança el PIC?**

- Tenim un pressupost base, a fons perdut, que és precisament el que genera el valor afegit que l'investigador ve a buscar, és a dir, oferim una excel·lent infraestructura física i una base humana que vetlla pel bon funcionament del sistema. Part del pressupost base es destina també a equipaments i ens permet realitzar projectes pilot per incorporar noves àrees d'investigació.

Participem també en projectes conjunts amb els investigadors i demanem finançament per a totes les necessitats d'emmagatzematge i de registre de dades del projecte que presentem.

La suma de tota aquesta infraestructura beneficia tots els investigadors que utilitzen el PIC i, com a mitjana, cada investigador es beneficia de la inversió realitzada i d'aquest valor afegit de què parlàvem.

#### **- El sistema GRID i la supercomputació competeixen entre si?**

- No, en absolut. Aquest dos sistemes no s'han de veure com a incompatibles, sinó, més aviat, com a complementaris. A més, aquesta tecnologia permet treballar tant als supercomputadors com als centres de súper emmagatzematge, com ara el PIC. El que passa és que mentre un centre de supercomputació té moltes unitats CPU's que es poden comunicar entre elles molt ràpidament, en el sistema GRID aquest tipus de comunicacions ni tan sols es produeixen perquè tots el "diàlegs" ocorren entre els computadors i els magatzems de dades que tenim. El que fa el sistema GRID és tractar processos en què a cadascun dels ordinadors se li assigna una parcel·la de dades i l'ordinador és capaç de processar aquestes dades sense necessitat de "parlar" d'una manera explícita amb la resta d'ordinadors.

#### **- Encara té futur la supercomputació, doncs?**

- Segur. Encara més, ara estem vivint un renaixement de la supercomputació, ja que els progressos tecnològics han permès construir ordinadors cada cop més grans sense augmentar el consum d'energia elèctrica i l'emissió de calor. El supercomputador Mare Nostrum, ubicat al Centre Nacional de Supercomputació, n'és un excel·lent exemple, perquè està construït de components estàndards, que es beneficien del progrés tecnològic i li han permès quadruplicar la seva potència. Aquests supercomputadors permetran unes simulacions molt més completes, per exemple, del canvi climàtic. O, per exemple, on abans es podia simular menys d'1 mm d'ample d'un ala d'un avió, ara s'estan simulant centímetres d'aquestes ales. Això genera una gran quantitat de dades, el que implica que els supercomputadors no són ja tan diferents dels detectors del LHC o del conjunt de màquines de ressonància magnètica i esdevindran una font més de dades que s'hauran d'emmagatzemar. Si aconseguim que l'usuari usi el supercomputador amb un DNI electrònic per a recerca i accedeixi al PIC amb aquest mateix DNI, tot el procés serà molt més àgil i fàcil.

#### **- I quan succeirà això?**

- Estem a punt. Està funcionant però no està desplegat per usar dia a dia. Ara hi ha una tecnologia GRID de centre de processos de dades, on el LHC ha tingut una influència fonamental, i hi ha una GRID de centres de supercomputació. Ambdós sistemes utilitzen estàndards diferents i ara hem d'arribar a un acord per unificar aquests estàndards.

#### **- En el futur entendrem la ciència i la vida humana sense la computació?**

- Jo diria que sense infraestructures digitals ja no s'entén i això creixerà en el futur, quan desaparegui la fractura digital que hi ha a la societat actual degut a que encara hi ha persones de les generacions anteriors al naixement de la World Wide Web que no s'han acabat d'integrar en el nou sistema de comunicació virtual. L'ésser humà ha sofert una transició, en els últims deu anys, cap a una situació que no pot girar cua. Les següents generacions no entendran l'existència sense les xarxes de telecomunicació que hem aconseguit tenir gràcies al progrés tecnològic. Ara, la gran novetat és que tots estem interconnectats i es dona una personalització de la informació. El web original tenia el defecte que només un podia escriure la pàgina i la resta llegia. Avui dia, el concepte d'organitzacions virtuals permet, amb un DNI electrònic, participar activament i no només ser receptor d'informació, sinó també emissor.

#### **- Què li sembla que la UAB organitzi enguany l'Any de la Computació?**

- És una molt bona idea perquè permet concentrar l'atenció sobre un tema de gran actualitat que té molta incidència, tant en l'àmbit de la recerca com en la vida quotidiana. Espero que les iniciatives que estan previstes dur a terme permetin apropar les noves tecnologies i les infraestructures de comunicació a tota la comunitat universitària i ajudin a entendre millor la seva importància.

**Entrevista:** María Jesús Delgado

**Foto:** PIC