

## **Cientistas Portugueses Participam no Desenvolvimento de Gigantesco Sistema de Computação GRID**

A Computação GRID (a que alguns também chamam Computação em “Grelha”) é uma tecnologia de computação distribuída que replica para o processamento computacional a filosofia e os princípios de funcionamento da *World Wide Web (WWW)* para a disponibilização de informação à escala mundial.

Na verdade, a *WWW* tornou possível disponibilizar de forma distribuída e a partir de milhares de computadores usuais uma quantidade gigantesca de informação que seria impossível disponibilizar com um pequeno número de computadores, mesmo que estes tivessem capacidades extremamente elevadas. De forma análoga, a Computação GRID consegue disponibilizar elevadas capacidades de processamento computacional à custa de distribuir tarefas por vários computadores de forma coordenada e eficiente.

Na Europa, o desenvolvimento da Computação GRID apareceu muito ligado ao CERN, o laboratório internacional de física das partículas situado em Genebra, onde em 1993 foi inventada por Tim-Berners Lee a *WWW* que levou à explosão da Internet.

Em Dezembro de 1994, o Conselho do CERN aprovou a construção de uma nova máquina para permitir um salto significativo na física das altas energias: o *LHC – Large Hadron Collider*. Os trabalhos de construção do LHC começaram em 1998 estando previsto o seu início de operação no primeiro semestre de 2008, com experiências que têm como objectivo esclarecer questões como: O que dá massa à matéria? De que é feita a matéria invisível que compõe 96% do Universo? Porque é que a natureza prefere a matéria à anti-matéria? Como é que a matéria evoluiu nos primeiros instantes de existência do Universo?

A enorme quantidade de processamento computacional necessário para analisar os dados experimentais que serão produzidos pelo LHC, estimados em 15 Petabytes, ou seja 15 milhões de Gigabytes, por ano, durante 15 anos, cedo levou os cientistas do CERN a preocuparem-se com como poderia ser assegurada a capacidade computacional necessária. A Computação GRID apareceu como a tecnologia apropriada.

Por esta razão, foi iniciado em 2001, com financiamento do Programa Quadro Europeu de Investigação e coordenação do CERN, o projecto *European DataGrid* que deu lugar ao projecto *EGEE – Enabling Grids for E-sciencE in Europe*, no qual Portugal participa através do LIP – Laboratório de Instrumentação e Física Experimental de Partículas (Lisboa e Coimbra), das universidades do Porto e do Minho, do Centro de Física de Plasmas do Instituto Superior Técnico, do Instituto de Engenharia Electrónica e Telemática (IEETA) da Universidade de Aveiro, e da Universidade Lusíada (Famalicão).

O projecto EGEE junta cientistas e engenheiros de mais de 240 instituições científicas em 45 países para fornecer uma infraestrutura permanente de Computação GRID que consiste em 41.000 CPUs e 5 Petabytes de disco (5 milhões de Gigabytes), e mantém o funcionamento concorrente de 100.000 processos computacionais. O projecto começou por se dirigir a problemas da física de altas energias e das ciências da vida, e agora integra aplicações a várias outras áreas, como geologia, química computacional e processamento de imagem.

Dada a decisiva importância estratégica da Computação GRID, foi lançada em Portugal no dia 29 de Abril de 2006, pelo Ministério da Ciência, Tecnologia e Ensino Superior, a Iniciativa Nacional GRID (INGRID), prevista no programa Ligar Portugal de Julho de 2005, com o objectivo de recuperar rapidamente o atraso que se fazia sentir no desenvolvimento desta tecnologia no nosso país.

A Iniciativa Nacional GRID foi planeada e é acompanhada e parcialmente financiada pela UMIC – Agência para a Sociedade do Conhecimento, e é concretizada pela FCT – Fundação para a Ciência e a Tecnologia no que respeita à promoção de projectos de I&D e que, logo em 2006, abriu um concurso público para projectos. Neste concurso, foram recebidas 37 candidaturas,

dos quais foram aprovadas 15 com um financiamento total de cerca de 1,7 milhões de euros. Estes projectos são dirigidos a aplicações da Computação GRID a áreas que vão de simulação e análise de dados de física de altas energias como os que serão produzidos pelo LHC e pela física de plasmas e fusão nuclear, a previsão da evolução da costa marítima, simulação de fogos florestais, mapeamento de poluição atmosférica, simulação da estrutura de proteínas, repositórios para aplicações médicas, imagiologia cerebral.

Presentemente, a infraestrutura GRID portuguesa integra 240 CPUs. Em 2007 foi decidido criar com financiamento da UMIC e do Programa Operacional Sociedade do Conhecimento (POSC) um nó principal para a infraestrutura GRID na FCCN – Fundação para a Computação Científica Nacional que envolve a criação de um grande *datacenter* para GRID, com 400 m<sup>2</sup>, e o alargamento da infraestrutura Grid portuguesa para cerca de 650 CPUs nos próximos dois meses e para mais de 1.200 CPUs até ao final de 2008. Atinge-se, assim, uma dimensão em Portugal significativa no plano Europeu.

A Iniciativa Nacional GRID também previu um reforço da conectividade nacional e internacional com um significativo aumento da largura de banda da rede de investigação e ensino nacional, a RCTS – Rede Ciência, Tecnologia e Sociedade, gerida pela FCCN e financiada pela UMIC. Este aumento de conectividade teve várias vertentes: a instalação de cabo de fibra óptica da FCCN de Lisboa a Braga (em 2005) e a Valença (em 2007) facultando ligações até 10 Gigabits por segundo (Gbps) entre instituições de investigação e universidades neste eixo litoral e permitindo a ligação à rede Europeia de investigação e ensino Geant2 através de Espanha, a ligação da RCTS à rede Geant2 a 2,5 Gbps no final de 2005, a extensão do cabo de fibra óptica da FCCN ao trajecto Lisboa-Setúbal-Évora-Portalegre-Fronteira do Caia que será realizada nos próximos seis meses, também com financiamento da UMIC e do POSC, e assegurará a ligação à rede Geant2 em anel redundante de fibra através de Espanha, a ligação da RCTS à rede Europeia Geant2 a 10 Gbps nos próximos dois meses.

No plano da cooperação internacional, destaca-se a decisão tomada na Cimeira Luso-Espanhola de 2006 de integrar numa rede coerente as infraestruturas de Computação GRID dos dois países – a Rede Ibérica de Computação Grid (IBERGRID) – e a participação de Portugal, através do LIP, nos principais projectos Europeus de Computação GRID com financiamento da União Europeia (UE), nomeadamente nos já referidos *European DataGrid* (2001-2004) e EGEE (fases I, II e III, 2004-2010), e ainda nos projectos CrossGrid (2002-2005), EELA (2006-2007), Int.Eu.Grid (2006-2008). Note-se que o EELA é um projecto de cooperação da UE com a América Latina (Argentina, Brasil, Chile, Cuba, México, Venezuela).

É ainda de destacar a participação no projecto *LCG – LHC Computing Grid* (2003-2023) do CERN que constitui o maior sistema de Computação GRID mundial, no qual o LIP opera um dos cerca de 60 centros mundiais de *Tier2* (o LHC também tem 11 centros de *Tier1* e o centro de *Tier0* operado directamente pelo CERN).

A junção de Portugal e Espanha na IBERGRID permitiu, em conjunto, constituir uma rede conhecida no projecto EGEE por Federação do Sudoeste Europeu, que presentemente já tem mais de 1.500 CPUs e passou a desempenhar um papel significativo na infraestrutura GRID da Europa.

Em 2007, foi iniciado um projecto de dois anos financiado pela UE para preparar a criação e o modelo de organização da Iniciativa GRID Europeia (*EGI – European Grid Initiative*). Este projecto envolve as iniciativas nacionais GRID de 38 países, incluindo os 27 países da UE e ainda Bielorrússia, Croácia, Israel, Moldávia, Montenegro, Noruega, Rússia, Sérvia, Suíça, Turquia e Ucrânia. Portugal é representado pela UMIC e pelo LIP, cujos dirigentes máximos integram o Conselho de Políticas da Iniciativa Europeia GRID, para cujo Presidente foi eleito Professor Gaspar Barreira, Director do LIP.

Em menos de dois anos, Portugal entrou para o mapa Europeu da Computação GRID e assegurou as condições para participar em posição de influência nas decisões sobre o futuro da Iniciativa GRID Europeia e do gigantesco sistema de Computação GRID que está a ser criado.

## O que é a Computação GRID

A Computação GRID (a que alguns também chamam Computação em “Grelha”) é uma tecnologia de computação distribuída que nasceu em 1994-95 no Laboratório Nacional Argonne dos EUA pelas mãos de Ian Foster, Steven Tuecke e Carl Kesselman, reconhecidos como os “pais” da Computação GRID.

A ideia da Computação GRID é replicar para o processamento computacional a filosofia e os princípios de funcionamento da *World Wide Web (WWW)* para a disponibilização de informação à escala mundial. Na verdade, a *WWW* tornou possível disponibilizar de forma distribuída e a partir de milhares de computadores usuais uma quantidade gigantesca de informação que seria impossível disponibilizar com um pequeno número de computadores, mesmo que estes tivessem capacidades extremamente elevadas. De forma análoga, a Computação GRID consegue disponibilizar elevadas capacidades computacionais à custa de distribuir tarefas de processamento por vários computadores de forma coordenada e eficiente.

## Para que serve a Computação GRID

A Computação GRID permite revolucionar a forma de trabalhar em várias áreas da ciência e tecnologia. Numa era em que o trabalho científico de qualquer domínio do conhecimento produz cada vez mais informação, como por exemplo a descodificação do genoma humano, a cura de doenças infecto-contagiosas, imunológicas ou cancerígenas, ou o estudo do comportamento da matéria e da energia em experiências com aceleradores de partículas cujos dados são analisados e partilhados por investigadores dispersos por todo o mundo, a Computação GRID vem oferecer respostas às enormes exigências ao nível da capacidade computacional e de armazenamento que o processamento de grandes quantidades de dados colocam.

Também na simulação computacional de sistemas de grande dimensão, como por exemplo em meteorologia, oceanografia, genómica e proteómica, a Computação GRID permite a conjugação de recursos computacionais distribuídos para a resolução de problemas que requerem elevado desempenho computacional.

## Alguns exemplos de aplicações da Computação GRID

Como exemplos de aplicações da Computação GRID referem-se as que foram desenvolvidas no âmbito do projecto *CrossGrid*, cuja infra-estrutura teve os seus sistemas centrais de gestão instalados em Portugal sob a responsabilidade operacional do LIP – Laboratório de Instrumentação e Física Experimental de Partículas (Lisboa e Coimbra). No âmbito do projecto *CrossGrid* foram consideradas aplicações de controlo, previsão e simulação de cheias em bacias fluviais, de modelação de poluição atmosférica e de serviços de apoio à meteorologia, bem como uma aplicação de apoio à cirurgia vascular. Nesta última aplicação, partindo das imagens de TAC ou Ressonância Magnética de um paciente, o cirurgião pode simular o efeito de um *by-pass* e visualizar o fluxo sanguíneo, decidindo por tentativas sucessivas a melhor solução para a intervenção. Esta aplicação permite encurtar para poucas horas o tempo de intervenção que medeia entre a recolha da imagem e a intervenção cirúrgica otimizada pela simulação prévia.

Um outro exemplo, no âmbito do projecto *EGEE – Enabling Grids for E-science in Europe*, foi simulação computacional de componentes para possíveis remédios para a Gripe Aviária que envolveu a análise de 300.000 componentes de possíveis fármacos contra o vírus da gripe das aves H5N1. Ao utilizar a computação Grid para identificar as ligações mais promissoras para os testes biológicos pode-se acelerar o processo de desenvolvimento de remédios contra o vírus. As simulações foram realizadas na infra-estrutura GRID do projecto EGEE durante um mês,

quando demorariam cerca de um século num único computador. Foram gerados e arquivados, numa base de dados relacional, mais de 60.000 ficheiros com um volume de informação de 600 Gigabytes.

Ainda outro interessante exemplo do projecto EGEE foi a criação de uma das maiores colecções de metadados multimédia a nível mundial. Em apenas 16 semanas, foram processados 37 milhões de imagens, gerando cerca de 112 milhões de objectos de texto e imagem – quase 5 Terabytes (5 milhares de Gigabytes) de dados – contendo no total mais de 150 milhões de características extraídas por análise computacional automática das imagens. Este número é equivalente ao processamento de 300.000 imagens por dia. Este tipo de técnica permite a expansão de motores convencionais de busca de forma a poderem efectuar procuras em imagens e conteúdos audiovisuais, mesmo que não estejam à partida associadas a texto. Permite, também, contribuir para a descrição automática em texto de imagens para cegos ou pessoas que por qualquer razão não possam visualizar as respectivas imagens. Em Portugal, o LIP participou disponibilizando recursos de computação GRID para efectuar a criação de metadados descritivos das imagens a partir da sua análise computacional.

### **O que é a INGRID e quais são os seus objectivos?**

A INGRID é a Iniciativa Nacional GRID portuguesa, lançada no dia 29 de Abril de 2006, em cumprimento do que ficou estabelecido no programa Ligar Portugal, definido em Julho de 2005 com o objectivo de orientar o desenvolvimento da Sociedade da Informação no país.

A INGRID foi planeada, é acompanhada e é parcialmente financiada pela UMIC – Agência para a Sociedade do Conhecimento, e é executada no que respeita a projectos de I&D através de financiamentos concedidos pela FCT – Fundação para a Ciência e a Tecnologia, na sequência de concurso público e avaliação internacional independente.

Estão presentemente em curso 15 projectos de I&D com um financiamento total concedido pela FCT de cerca de 1,7 milhões de euros, os quais foram aprovados num concurso aberto em 2006 que recebeu 37 candidaturas.

Os principais objectivos da INGRID são:

- Reforçar as competências e capacidades nacionais em Computação GRID devido à sua especial importância estratégica.
- Reforçar as competências e capacidades nacionais e prosseguir a integração de Portugal na rede internacional de Computação GRID.
- Melhorar as condições para as actividades científicas e para aplicações de interesse económico e social que envolvem computações complexas ou com elevadas quantidades de dados.
- Reforçar a multidisciplinaridade e a colaboração entre as comunidades de investigadores e utilizadores de meios computacionais de elevado desempenho.
- Reforçar as condições para as empresas encontrarem em Portugal instituições científicas e recursos humanos com conhecimentos e experiência de Computação GRID.