

A Era dos Dados

A grande maioria dos projetos de GIS tem como base levantamentos de dados, sejam eles cadastrais, planimétricos, altimétricos ou imagens

Temos observado grandes evoluções da tecnologia de informática, ao longo dos últimos anos. Os computadores que usamos têm se tornado cada vez mais rápidos e, principalmente, têm capacidades de armazenamento cada vez maiores. Esta constatação parece óbvia, pois trata-se de um fenômeno presente na vida de todos e extensivamente explorado pelos meios de comunicação. Esse fenômeno, somado à crescente presença dos computadores em nosso cotidiano, leva os mais entusiasmados a declarar inaugurada a "Era da Informação", ou mesmo a "Era do Conhecimento". Mas se observarmos mais de perto, podemos perceber que a maioria das atividades informatizadas destina-se a, apenas, armazenar e manipular dados, às vezes de qualidade duvidosa, às vezes organizados de maneira confusa. Daí uma das mais infames máximas da informática: "lixo entra, lixo sai". A capacidade de analisar e trabalhar esses dados eficientemente para produzir *informação* é relativamente rara.

A própria evolução tecnológica tem também tornado evidente a nossa cada vez menor capacidade de lidar, de forma eficiente e competente, com os grandes volumes de dados que temos acumulado. Um recente artigo de Fran Berman, publicado na revista *Communications of*

the ACM, reúne observações sobre esse fenômeno, e observa que a recente evolução das redes de computadores têm uma participação expressiva nele. Afinal de contas, como as redes ainda são bastante lentas, cada vez mais as pessoas produzem cópias e mais cópias locais da informação disponível, por exemplo, na Internet. Mesmo alguns mecanismos que permitem à Internet funcionar com maior agilidade e velocidade, tais como os caches, são essencialmente equipamentos dedicados a copiar dados e deixá-los mais próximos de nós. Além disso, a intensidade do movimento de publicação de informação na Internet está levando a um espetacular aumento do volume de dados acessíveis a partir de qualquer computadorzinho doméstico. Aliás, "computadorzinho doméstico" é uma descrição injusta: qualquer adolescente maníaco por jogos sabe que, hoje em dia, ter em casa uma máquina com 20 GB de espaço em disco não é nada incomum, pelo contrário, é até bastante barato.

Com isso, Berman afirma que a década que acaba de se iniciar é a "década dos dados", um período da história da informática em que os usuários estarão acumulando cada vez mais dados, em uma velocidade superior até mesmo à do vertiginoso crescimento computacional. O crescimento da Internet é uma prova disso.

Em outro exemplo, verifica-se que projetos como os de seqüenciamento de genes não são muito mais do que grandes e complexos esforços de levantamento e catalogação de dados. A cada genoma completado, os cientistas apressam-se em informar à imprensa que trata-se de apenas um passo no sentido da compreensão biológica dos seres pesquisados, que *poderá* levar a algum avanço do conhecimento no futuro, quando as ferramentas de análise se tornarem mais adequadas.

Da mesma forma, bibliotecas virtuais digitais, projetos de imageamento em escala planetária (como o que foi recentemente conduzido pela NASA), sistemas de coleta de dados ambientais e outros são também grandes esforços de coleta de dados. O problema está na nossa capacidade de compreensão e análise de um volume tão fabuloso de dados, para gerar informação e, a partir dela, conhecimento. Uma coisa é conseguirmos obter uma imagem de tomografia computadorizada para facilitar um diagnóstico médico. Outra coisa é conseguir armazenar essa imagem em um banco de dados, torná-lo acessível pela Internet, e a partir daí dar ao médico a possibilidade de conseguir a opinião de um especialista baseado em outra cidade ou outro país, em um projeto de telemedicina. Outra coisa ainda mais diferente é conseguir reunir, pela rede,

1 kilobyte (kB)	2^{10} bytes	1024 bytes
1 Megabyte (MB)	2^{20} bytes	1.048.576 bytes
1 Gigabyte (GB)	2^{30} bytes	1.073.741.824 bytes
1 Terabyte (TB)	2^{40} bytes	1.099.511.627.776 bytes
1 Petabyte (PB)	2^{50} bytes	1.125.899.906.842.624 bytes
1 Exabyte	2^{60} bytes	1.152.921.504.606.846.976 bytes

Referências:

Morrisson, Philip e Morisson, Phyllis. "The Sum of Human Knowledge?". In *Scientific American*, July 1998. Artigo disponível em <http://www.sciam.com/1998/0798issue/0798wonders.html>

Berman, Fran. "From TeraGrid to Knowledge Grid". In *Communications of the ACM* 44(11):27-28, 2001.

dezenas de milhares de imagens de tomografias de centenas de hospitais do mundo inteiro, juntamente com dados médicos de cada paciente, para conduzir uma pesquisa sobre o Mal de Alzheimer. É necessário admitir que estamos ainda engatinhando nesse processo.

Do entanto, as primeiras experiências para o tratamento de problemas dessa escala, denominada por enquanto de "tera-escala", já estão sendo conduzidas. Berman informa que a National Science Foundation, o CNPq deles, está financiando a construção de uma rede denominada TeraGrid, essencialmente uma estrutura de supercomputação distribuída em quatro universidades. Os números são assustadores: meio petabyte de espaço em disco e uma capacidade agregada de transmissão de dados da ordem de 40 gigabits por segundo. Vejamos o que isso

significa: meio petabyte, ou 562.949,953.421.312, bytes, seriam suficientes para armazenar todas as fotos reveladas anualmente por toda a humanidade. O dobro disso, um petabyte, seria suficiente para armazenar todos os filmes produzidos pela indústria do cinema, desde que o cinema foi inventado, segundo cálculos da revista *Scientific American*. Com 40 gigabits por segundo seria possível transmitir um filme de duas horas em DVD a cada segundo. Agora, para armazenar tudo o que a televisão transmite de útil em todo o mundo, sem contar as repetições, precisaríamos de 1000 petabytes, ou quase um exabyte (veja quadro). A idéia do TeraGrid é justamente desenvolver métodos e técnicas de computação que permitam alavancar a geração de informação e de conhecimento a partir de imensos volumes de dados.

Guardando as proporções, nossa área de trabalho não é diferente. A grande maioria dos projetos de implantação de SIG tem como base grandes levantamentos de dados, sejam eles cadastrais, planimétricos, altimétricos, imagens ou outros quaisquer. A acumulação desses dados, como temos feito até hoje, é bem mais fácil de fazer do que o tratamento, a análise e a geração de resultados, mas trata-se de um passo fundamental. É importante observar que a qualidade desses resultados depende da qualidade dos dados, e portanto do processo de coleta e armazenamento. Temos hoje a responsabilidade de contribuir não apenas para a formação de nossos bancos de dados, mas também de preparar o terreno para a geração de conhecimento a partir desses dados. Com isso, talvez seja possível dizer, na próxima década, que a Era da Informação finalmente chegou. ■



Clodoveu Davis é engenheiro civil, doutor em Ciência da Computação e Assessor de Desenvolvimento e Estudos da Prodabel - Empresa de Informática e Informação do Município de Belo Horizonte. E-mail: cdavis@uol.com.br

Sensores Digitais Aerotransportados para profissionais



Alcance a máxima qualidade de dados em suas missões de vôos com os dois sensores digitais da LH Systems - o ADS40 Airborne Digital Sensor e o AL540 Airborne Laser Scanner - a chave para o sucesso dos seus projetos de fotogrametria.

Os nossos sistemas são oferecidos com programas da LH Systems para:

- Planejamento das missões
- Administração e controle durante o vôo
- Processamentos básicos de dados no solo.

Em seguida, os dados filtrados podem ser inserido no programa de fotogrametria SOCET SET, o líder do mercado em fotogrametria digital.

As funcionalidades do sistema SOCET SET são ainda aumentadas com o uso do sistema robusto de aerotriangulação ORIMA e com o sistema de restituição PRO600, que opera em combinação com os produtos MicroStation Geographics da Bentley e com os produtos TerraModeler e TerraScan da Terrasolid.

Soluções para informações precisas de imagens

LH Systems, LLC
2 Inverness Drive East
Suite 106-108
Englewood, CO 80112-5508
USA

Tel: + 1 303 799 9453
Fax: + 1 303 799 4809

Distribuidores autorizados:

- Brasil: Companhia A Importadora Wild Ltda, +55 11 3311 6200
- Argentina: Wild Sistemas S.R.L, +541 1 4342 9398
- Chile: Womers Scientific Co, +1 688 673 0038
- Colômbia: GEDSYS S.A., +56 2 235 0085
- Costa Rica: PROSOL S.A., +57 1 835 7240
- México: Womers Geosystems SA de CV, +52 5563 5011
- Peru: GUINOA SUZA S.A., +51 1 211 4056
- Paraguai: GEDSYS S.R.L., +598 2 915 4535



www.lh-systems.com
e-mail: informes@lh-systems.com