

SBC - Grandes Desafios em Computação
Ciência em Larga Escala

Marta Mattoso

Guilherme H. Travassos

Programa de Engenharia de Sistemas e Computação
COPPE/UFRJ

Temos assistido ao longo dos anos uma evolução silenciosa e cientificamente apoiada, que, em futuro não muito distante, provocará o desaparecimento do computador. Embora isto não represente, absolutamente, o fim da computação, já que estamos vivenciando neste momento uma transformação extremada onde o que já é grande necessita ser ainda maior (cluster de computadores em *grids* computacionais para o desenvolvimento de ciência em larga escala) e o que já é pequeno se torna cada vez menor (alta integração dos circuitos permitindo a confecção de sensores e dispositivos capazes de ser arquitetados em soluções sistêmicas que podem decidir, estrategicamente, a permanência e sobrevivência de uma nação), representa um ponto de inflexão que merece reflexão e o desenvolvimento de estudos específicos. Atualmente, a indústria de software tem conhecimento limitado sobre como software é projetado, implementado e evoluído ao longo do tempo. Sente-se a necessidade de modelos adequados para analisar processos e produtos de software; sente-se a necessidade de entendimento dos limites da tecnologia em determinados contextos; e sente-se a necessidade de habilidade de análise e de experimentação. Em geral, uma tecnologia que aparentemente resolve um problema genérico em um determinado contexto pode não ser aplicável em outros contextos. Frequentemente, essas limitações são desconhecidas e ignoradas pelos desenvolvedores. A ausência de modelos de processos e produtos adequados, de conhecimento dos limites das tecnologias de software e de análise e estudos experimentais limita o conhecimento e a evolução dos conceitos subjacentes à construção de software.

Mesmo com todo o avanço científico obtido até então, onde soluções conseguem ser fornecidas através da utilização de metáforas e heurísticas computacionais básicas, ainda não conseguimos abordar, de forma precisa, o desafio de construir soluções computacionais para problemas que em breve se tornarão realidade. Software vem, aos poucos, deixando de ser tratado como um produto único e monolítico para assumir o papel de serviço, espalhado por diferentes sensores e dispositivos que podem ser utilizados em conjunto para a solução de problemas efetivamente complexos. Este espalhamento nos leva a uma situação onde cotidianamente a utilização dos serviços computacionais passa a fazer parte do dia a dia das pessoas como algo muito natural, ubíquo, fazendo com que a figura do computador deixe de existir. Até na situação onde um alto poder de processamento de dados se torne necessário, a interação homem-computador passa a ser realizada no nível de dispositivos e sensores, fazendo com que os tradicionais equipamentos se transfigurem em elementos para o efetivo conforto e bem estar do indivíduo.

Apesar dos resultados significativos que alcançamos até então, todas as nossas propostas, dos algoritmos aos processos de software, não levam em consideração esta nova condição. O volume de componentes de softwares disponíveis para a ciência em larga escala, faz com que essa base de serviços computacionais mereça um tratamento sofisticado quanto a sua descrição, disponibilização, interfaces, histórico de uso, etc., de tal forma que o cientista possa compor itens de solução de problemas da ciência em larga escala. Tradicionalmente, a área de pesquisa em bancos de dados vêm se especializando, com sucesso, em adicionar semântica no

armazenamento e gerência de dados, distribuídos ou não. No entanto, tais técnicas não prevêm o uso de semântica na gerência de serviços computacionais, por exemplo, software. Consideramos fundamental a combinação da engenharia de software com as técnicas de bancos de dados no sentido de prover um ambiente capaz de apoiar a ciência em larga escala no que diz respeito à gerência harmônica de serviços computacionais, incluindo componentes de software e dados no contexto da ubiqüidade.

Sem essa combinação e com as tecnologias atuais, não conseguimos responder à pergunta básica: Como deve ser construído, numa perspectiva de engenharia, um produto computacional que contemple esta ubiqüidade? Num País de dimensões continentais como o Brasil, obter resposta para esta pergunta pode representar um diferencial estratégico para o futuro. Por exemplo, O ambiente em que vivemos é sem dúvida um ponto crítico para a sociedade. Em cenário de preocupação crescente com o meio ambiente, monitorar algumas variáveis como poluição do ar, rios e solo, temperatura, nível de chuva, com objetivo de apoiar iniciativas dos órgãos responsáveis por lidar com a qualidade de vida da sociedade é essencial. Alguns exemplos onde soluções computacionais precisarão ser consideradas incluem a bioinformática, a localização para áreas de plantio, investimentos em infra-estrutura (esgoto, distribuição de água, pavimentação, recuperação de rios). Este é um grande desafio que se apoiado de forma consistente pela computação, poderá trazer grandes benefícios para solução de um problema de extrema relevância para o Brasil.

Esse modelo de ubiqüidade e distribuição no desenvolvimento e execução de soluções baseadas em software, dados e sensores, por exemplo, vêm sendo chamado de experimento. Entretanto, essa distribuição não pode se distanciar de um monitoramento integrado. Existe um conhecimento valioso na combinação e uso de recursos computacionais num experimento que não pode ser perdido.

Para alcançarmos este patamar tecnológico, deveremos considerar a estruturação de uma agenda de pesquisa com ênfase em Processos e Técnicas para desenvolvimento e monitoramento de aplicações baseadas em experimentos eletrônicos que contemple os desafios das seguintes áreas, nos novos cenários de ubiqüidade e sensores:

- a) Gerência de Workflow Científico
- b) Ambientes Computacionais em Grid, redes móveis, sensores, etc
- c) Proveniência e curadoria de dados
- d) Mecanismos para Gerenciamento e Manutenção de Conhecimento Científico
- e) Engenharia de Aplicações baseadas em tecnologias Web

Marta Lima de Queirós Mattoso é professora adjunto do Programa de Engenharia de Sistemas e Computação da COPPE/UFRJ. Atua na linha de Banco de Dados. Pesquisadora 1D do CNPq. Coordenadora de diversos projetos de pesquisa nacionais e internacionais. Suas áreas de interesse incluem Distribuição e Paralelismo em Sistemas de Bancos de Dados em ambientes de grid, web e computadores paralelos, Gerência de Experimentos em Bioinformática, Gerência de Dados e Serviços Web. (Lattes- <http://buscatextual.cnpq.br/buscatextual/visualizacv.jsp?id=K4780390J5>)

Guilherme Horta Travassos é professor adjunto do Programa de Engenharia de Sistemas e Computação da COPPE/UFRJ. Atua na linha de Engenharia de Software. Pesquisador 2 do CNPq. Suas áreas de interesse incluem Experimentação em Engenharia de Software, Qualidade de Software (Inspeção e Teste de Software), Ambientes de Desenvolvimento de Software e Desenvolvimento de Aplicações baseadas na Web. Representante brasileiro na ISERN – International Software Engineering Research Network. (Lattes- <http://buscatextual.cnpq.br/buscatextual/visualizacv.jsp?id=K4723122E7>)