

Um Roteiro para Construção de Cubos e Consultas OLAP

Iris Fabiana de Barcelos Tronto, Nilson Sant'Anna
Ministério da Ciência e Tecnologia – MCT
Instituto Nacional de Pesquisas Espaciais – INPE
Laboratório Associado de Computação e Matemática Aplicada – LAC/CAP
E-mail: iris_barcelos@lac.inpe.br

Resumo

A gestão de projeto de software é uma atividade de intenso conhecimento onde os gerentes usam suas habilidades e experiências para tomar decisões. Uma vez que este domínio de negócio depende da experiência e conhecimento dos gerentes é necessário que se tenha técnicas e ferramentas para apoiar o gerenciamento deste conhecimento. Este trabalho apresenta uma abordagem que apóia a análise de informação a partir de uma base de dados de gestão de projetos de software.

1. Introdução

No desenvolvimento de software, geralmente, os gerentes experientes obtêm mais sucesso que os gerentes novatos em termos de atingir metas de cronograma, custos e funcionalidade. Isso é devido à experiência acumulada ao longo de situações ocorridas em projetos no passado e ao conhecimento derivado destas experiências.

Neste contexto, é também importante o uso de métricas, que trazem benefícios expressivos, principalmente nas estimativas de tempo e custo [12]. Os dados resultantes do uso de métricas são coletados e constituem uma base de dados de experiência que pode auxiliar a tomada de decisão no gerenciamento de projetos [3]; [4]; [5]; [7]; [8]; [9]; [11].

Porém, uma grande quantidade de informação útil pode permanecer oculta nestas bases de dados caso não sejam aplicadas técnicas adequadas de extração de conhecimento. Técnicas de Inteligência Artificial têm sido utilizadas para apoiar o reuso de artefatos de um projeto para o desenvolvimento de outro projeto [2]. Mendonça & Sunderhaft aplicam técnicas de *mineração de dados* para explorar dados de engenharia de software [6]. Abordagens como OLAP (*On-Line Analytical Processing*) [1]; [10] e *Mineração de Dados* [6]; [13] parecem adequadas, pois elas permitem selecionar grandes quantidades de dados automaticamente e descobrem informações importantes nos dados.

O objetivo deste trabalho é apresentar um roteiro para criação de cubos OLAP a fim de permitir uma análise mais eficiente de informações contidas em bases de dados de gestão de projetos, tal que o conhecimento obtido dessa análise possa servir de suporte a decisão à gerente de projeto para a melhoria deste processo.

Este trabalho está organizado em cinco seções. A Seção 2 apresenta a tecnologia OLAP. Na Seção 3 é apresentado um conjunto de passos que podem ser utilizados para a geração de cubos e consultas OLAP. Finalmente, a Seção 4 apresenta as conclusões deste trabalho.

2. Tecnologia OLAP

O termo OLAP (*On-Line Analytical Processing*), hoje muito difundido, traduzido para *Processamento Analítico On-Line*, representa a possibilidade de se trabalhar os dados, com operadores dimensionais, possibilitando uma forma múltipla e combinada de análise [1]. A característica principal dos sistemas OLAP é permitir uma visão conceitual multidimensional dos dados armazenados através de cubos. Um cubo é uma estrutura composta de *dimensões* e de uma *tabela fato*. As *dimensões* representam os eixos do cubo e correspondem a atributos do domínio analisado. A *tabela fato* é constituída de medidas (dados) de uma tabela de fatos coletados dentro do domínio analisado. As medidas são relacionadas às dimensões.

A visão multidimensional é mais útil para os analistas do que a visão tabular tradicional utilizada nos sistemas de processamento de transação. Ela é mais natural, fácil e intuitiva, permitindo uma visão das informações em diferentes perspectivas. Isso auxilia efetivamente o tomador de decisão, pois ele pode verificar tendências nos dados, utilizando dados resumidos, trocando as dimensões de lugar e navegando através de suas hierarquias. Dessa forma, os usuários podem testar suas hipóteses e pensar sobre questões que não haviam ainda sido levadas em consideração.

Uma grande novidade no mundo OLAP é o produto MS SQL-Server Analysis Services, que é utilizado pela Ferramenta MS Project 2003, da Microsoft [10], uma conceituada ferramenta de apoio à gestão de projetos.

3. Roteiro para a Construção de Cubos OLAP

Aqui é apresentado um estudo de caso, que consiste da elaboração e aplicação de um roteiro para a geração de cubos e consultas OLAP a partir de uma base de dados de gestão de projeto de software. As Seções 3.1 e 3.2 apresentam os passos que compõem este roteiro.

3.1. Planejamento

A construção de cubos e consultas OLAP é realizada tomando-se uma base de dados do domínio do negócio como ponto de partida. Como neste estudo de caso não se dispunha de uma base de dados pronta, foi necessário desenvolver as atividades de 1 a 5 abaixo apresentadas.

1. Definição da Área do Negócio. Neste caso, é a gestão de projeto de software
2. Definição de um Processo de Software. Aqui, foi escolhido o Cascata, mas poderia ser qualquer outro.
3. Definição de um modelo lógico – Diagrama Entidade e Relacionamento – DER.
4. Criação de um banco de dados relacional. A partir do DER, cria-se o banco de dados (com suas tabelas, colunas, restrições, etc).
5. Povoamento do banco de dados. Aqui, o banco foi povoado com dados fictícios no domínio em questão.

Se já houvesse uma base de dados disponível que atendesse as necessidades da organização, poderia partir para a implementação do cubo que se refere à Seção 3.2, precedida (em alguns casos) das tarefas de entendimento da base de dados e preparação dos dados.

3.2. Criação de um Cubo

Um cubo é criado tendo-se em mente o tipo de consulta que um gerente de projeto pode querer fazer. Ao projetar um cubo é importante levar em conta as prováveis consultas que poderão ser feitas no cubo, porém não se deve perder de vista que existem muitas consultas potenciais que podem surgir. Portanto, o projeto de um cubo deve priorizar a escolha e a organização das dimensões que podem influenciar nas medidas relacionadas à decisão que o usuário deve tomar. A escolha das dimensões é fundamental no projeto de um cubo. Os passos pra criar um cubo são apresentados a seguir.

3.2.1. Defini-se o Tipo de Consulta Base

O primeiro passo da criação de um cubo é descobrir qual o domínio da informação que o usuário deseja analisar. Por exemplo, para a construção do cubo denominado “Atrasos gerais” tomou-se como base o fato de que o gerente de software precisava decidir sobre a adoção de uma ou outra metodologia em um projeto a ser iniciado, bem como decidir sobre as pessoas que deveriam ser alocadas para o projeto dependendo da metodologia escolhida. Neste caso, a visualização de algumas medidas é significativa para a tomada de decisão. Dentre elas, tem-se: número de defeitos, custo (custo de investimento+ horas trabalhadas * valor da hora) e atraso total do projeto. Além disso, ficou evidente a necessidade das dimensões: pessoas, metodologia projeto.

3.2.2 Cria-se o Projeto de Alto Nível

Cria-se um projeto conceitual e lógico (Projeto de Alto Nível) do cubo, ou seja, são definidos as medidas da tabela fato, as dimensões e os níveis de cada dimensão. A Tabela 1 apresenta as medidas, dimensões e níveis de cada dimensão para o cubo “ATRASOSGERAIS”.

TABELA 1 – Projeto de Alto Nível do Cubo

| Dimensões | Níveis |
|--|-------------------------------|
| Projeto | Perfil e Nome do Projeto |
| Pessoa | Escolaridade e Nome da Pessoa |
| Metodologia | Ferramenta e Nome |
| Medidas: Duração Prevista, Duração real, atraso | |

3.2.3. Confronta o Projeto de Alto Nível com a Base de Dados

Este passo consiste de um processo de verificação das seguintes partes:

Tabela fato: É preciso verificar se existe no banco de dados uma tabela que contenha todas as medidas. Se não existe, podem ser investigadas duas opções: i) Criar uma tabela fato no banco, que contenha todas as medidas, relacionando as tabelas (através de chaves) que possuem cada uma das medidas desejadas ou; ii) Criar um cubo para cada medida que esteja em tabela fato diferente e que se quer analisar. Esta última opção, é considerada a mais viável.

Dimensões: É preciso encontrar uma relação entre a tabela fato e as tabelas que definem as dimensões. Além disso, é preciso definir quais atributos das tabelas que representarão os níveis das dimensões.

Feito isso pode implementar o projeto de alto nível utilizando a ferramenta para construção de cubos.

3.2.4. Implementação do Cubo

Implementa-se o cubo utilizando a ferramenta OLAP desejada. Neste estudo de caso o cubo “ATRASOSGERAIS” foi implementado seguindo o conjunto de passos descritos no manual da ferramenta utilizada, que é a MS SQL-Server Analysis Services.

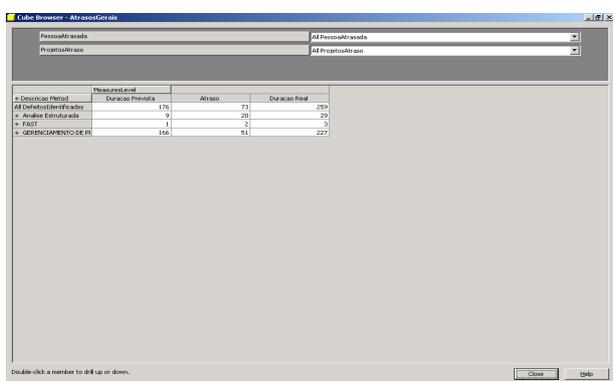
3.2.5 Projeto Físico

Define-se o tipo de armazenamento dos dados do cubo em duas etapas: primeiro define se o modelo será multidimensional, relacional ou híbrido. Segundo, deve-se definir um ponto desejado entre desempenho e capacidade de armazenamento. Pois, se todas as operações forem

realizadas e armazenadas haverá uma grande ocupação de espaço, por outro lado se algumas operações não são realizadas então o tempo de resposta das consultas pode ser maior.

3.3. Execução de Consultas OLAP

Após ter construído e armazenado o cubo, este estará pronto para ser consultado. Utilizando a ferramenta MS SQL-Server Analysis Services, basta selecionar, na tela do Analysis Manager, o nome do cubo que se deseja consultar. Automaticamente será apresentada a tela que exibe todas as possibilidades (combinações de medidas e dimensões) de consultas a partir daquele cubo. A Figura 1 apresenta uma consulta ao cubo ATRASOSGERAIS



| Dimensional | Duration Planned | At Risk | Duration Real |
|------------------------|------------------|---------|---------------|
| a. Descrição Projeto | 176 | 71 | 271 |
| b. Analise Estruturada | 9 | 20 | 29 |
| c. FASE | 1 | 4 | 5 |
| d. GERENCIAMENTO DE PR | 166 | 51 | 227 |

FIGURA 1 – Consulta ao Cubo ATRASOSGERAIS

4. Conclusões

No contexto da gestão de projetos de software, um grande conjunto de dados é coletado, armazenado e posteriormente analisados para dar apoio aos gerentes de projeto na tomada de decisão. Porém, para atender a esse objetivo é necessário dispor de técnicas que dêem suporte a esse processo.

Várias abordagens, métodos e técnicas têm sido utilizadas a fim de superar algumas das dificuldades inerentes à gestão de projeto de software, dentre elas a tecnologia OLAP. Com as ferramentas OLAP os usuários de um banco de dados de gestão de projeto de software podem ter visões multidimensionais dos dados, ao contrário da visão tabular tradicional. Essas visões multidimensionais auxiliam efetivamente ao tomador de decisão, pois ele pode verificar tendências nos dados, utilizando dados resumidos, trocando as dimensões de lugar e navegando através de suas hierarquias. Dessa forma, os usuários podem testar suas hipóteses e pensar sobre questões que não haviam ainda sido consideradas.

O fato de não ter utilizado uma base de dados operacional representa um fator limitante no processo de realização de consultas e interpretação das informações. Porém, espera-se que esta limitação seja superada em uma próxima etapa.

Referências Bibliográficas

- [1] C.Barbieri, “ BI- Business Intelligence: Modelagem & Tecnologia”, Ed. Axcel Books, Rio de Janeiro, 2001.
- [2] C.Liu, “Planning Support to Software Process Evolution”, In International Journal of Software Engineering and Knowledge Engineering, Novembro de 2003.
- [3] CMMI Product Team. “CMMI® for Systems Engineering/Software Engineering/Integrated Product and Process Development/Supplier Sourcing, Version 1.1, Continuous Representation (CMMI-SE/SW/IPPD/SS, V1.1, Continuous)”, (CMU/SEI-2002-TR-011, ADA339818). Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University,2002.
- [4] L. F.Borrego Filho, N.Sant’Anna, M.G.Cereja Junior, L Luque, B.H.Casilo, “Uma Abordagem para Processos de Gerenciamento do Tempo e dos Recursos Humanos Apoiados por Ambientes Integrados de Engenharia de Software”, Recife, SIMPROS Simpósio Internacional de Melhoria de Processos de Software, 2002.
- [5] M.C.Paulk, B.Curtis, M.B Chrissis, C. V. Weber, “Capability Maturity Model for Software, Version 1.1”, Software Engineering Institute – Carnegie Mellon University, Pittsburgh, CMU/SEI-93-TR-024, February 1993.
- [6] M.G.Mendonça, N.L.Sunderhaft, “Mining Software Engineering Data: Um Survey”, A DACS State-of-the-Art Report, Data e Analysis Center for Software, 1999, At. www.dacs.dtic.mil/techs/datamining/datamining.pdf
- [7] PMI -Project Management Institute. A Guide to the Project Management Body of knowledge: PMBOK Guide 2000 edition. Pennsylvania, Project Management Institute, 2000.
- [8] R. C. Vasques, “Gestão de projetos SW-Capability Maturity Model e Capability Maturity Model Integrated”, Curitiba: CITS – Conferência Internacional de Tecnologia de Software, 2002.
- [9] R.S.Pressman, Engenharia de Software, São Paulo, McGraw-Hill, 2002.
- [10] R.V.Vargas, “Microsoft Project 2002: Professional & Server”, Rio de Janeiro, Brásport Livros e Multimídia Ltda, 2002.
- [11] SPICE Software Process Improvement and Capability Determination. Software Process Assessment Part 2: a model for process management version 1.00, 1995.
- [12] T.DeMarco, “Controle de Projetos de Software: Gerenciamento, Avaliação, Estimativa”, Rio de Janeiro, Campus, 1989.
- [13] Z.Chen, “Data Mining and Uncertain Reasoning: An Integrated Approach”, John Wiley & Sons, Canada, 2001.