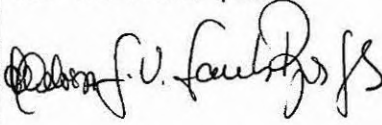

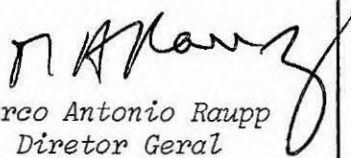


1. Publicação nº <i>INPE-3701-PRE/842</i>	2. Versão	3. Data <i>Novembro 1985</i>	5. Distribuição <input type="checkbox"/> Interna <input checked="" type="checkbox"/> Externa <input type="checkbox"/> Restrita
4. Origem <i>DIN/DCS</i>	Programa <i>DESQFT</i>		
6. Palavras chaves - selecionadas pelo(s) autor(es) <i>BANCO DE DADOS DETERMINAÇÃO DE REQUISITOS</i> <i>METODOLOGIA DE PROJETO PROCESSAMENTO HUMANO DE INFORMAÇÃO</i>			
7. C.D.U.: <i>681.3.016</i>			
8. Título <i>INPE-3701-PRE/842</i>  <i>METODOLOGIAS PARA O PROJETO DE BANCO DE DADOS</i> <i>ASPECTOS PSICOLÓGICOS DA DETERMINAÇÃO</i> <i>DE REQUISITOS</i>		10. Páginas: <i>08</i>	
		11. Última página: <i>06</i>	
9. Autoria <i>Petrônio Osório Costa</i>		12. Revisada por  <i>Heloisa G.V.S. Borges</i>	
Assinatura responsável 		13. Autorizada por  <i>Marco Antonio Raupp</i> Diretor Geral	
14. Resumo/Notas  <p><i>Banco de Dados são sistemas integrados e complexos de "hardware" e "software", e o projeto de suas estruturas de armazenamento e estratégias de acesso também é uma tarefa complexa. Este estudo apresenta algumas metodologias que estão sendo atualmente utilizadas para o projeto de Banco de Dados de grande porte e demonstra que ele não é um trabalho fácil, mesmo com metodologias altamente estruturadas. Durante este processo, são evidenciados os problemas de comunicação entre as pessoas envolvidas, o que pode ser devido ao poder das técnicas e ferramentas de projeto disponíveis e sua ainda pequena utilização por parte dos projetistas.</i></p>			
15. Observações <i>Este trabalho foi aceito para apresentação no VI Seminário ADUNESP-Guaratinguetá, que realizar-se-á de 04 a 09 de novembro de 1985, em Guaratinguetá-SP.</i>			

#### ABSTRACT

*Data bases are hardware and software integrated and complex systems. This fact makes their data structures and access methods design a difficult task too, even with some highly structured methodologies used nowadays for the Data Bases design. As shown in this study, during the design process it becomes evident the communication problems between the involved people, the little power of the available technics and design tools, and their low utilization by the designers.*

# METODOLOGIAS PARA O PROJETO DE BANCO DE DADOS - ASPECTOS PSICOLÓGICOS DA DETERMINAÇÃO DE REQUISITOS

COSTA, PETRÔNIO OSÓRIO - INPE

## SUMÁRIO

Bancos de Dados são sistemas integrados e complexos de "hardware" e "software", e o projeto de suas estruturas de armazenamento e estratégias de acesso também é uma tarefa complexa. Este estudo apresenta algumas metodologias que estão sendo atualmente utilizadas para o projeto de Banco de Dados de grande porte e demonstra que ele não é um trabalho fácil, mesmo com metodologias altamente estruturadas. Durante este processo, são evidenciados os problemas de comunicação entre as pessoas envolvidas, o pequeno poder das técnicas e ferramentas de projeto disponíveis e a sua ainda pequena utilização por parte dos projetistas.

## INTRODUÇÃO

Numa sociedade orientada pela informação, os dados (a informação) são cada vez mais considerados como um recurso vital que deve ser organizado (para maximizar o seu valor) e, neste sentido, além dos sistemas de bancos de dados da própria organização, está havendo uma demanda significativa de serviços de bancos de dados, os quais coletam, organizam e vendem dados.

Qualquer que seja a sua finalidade, o projetista de um Banco de Dados (BD) é colocado frente a um complexo leque de alternativas, e quando mais alternativas ele puder considerar de forma racional, mais chances ele tem de produzir um projeto ótimo. Grande parte dos projetos ineficientes de sistemas de BD (e existem muitos) resulta do fato de o projetista ter considerado somente alguns fatores, ou por ter um horizonte limitado de conhecimentos, ou por ser entusiasta de técnicas particulares que ele conhece bem, com exclusão de outras que poderiam até ser melhores.

Uma consideração particularmente importante no projeto de um BD é armazenar dados de forma que possam ser usados por uma grande variedade de aplicações e de maneiras diferentes, o que era extremamente difícil nas instalações de computadores anteriores à era do BD.

Para atingir a flexibilidade no uso dos dados, que é essencial em muitas situações, dois aspectos do projeto de BD são importantes:

- 1º) os dados devem ser independentes dos programas que os utilizam, tal que os novos dados possam ser adicionados ou reestruturados sem que os programas sejam alterados;

2º) deve ser possível interrogar e pesquisar o BD sem a enfadonha tarefa de escrever programas em linguagens convencionais de programação. São Usadas as chamadas *Linguagens de Consulta*.

O trabalho de projetar um BD está cada vez mais difícil, especialmente se é desejada eficiência ótima. O "software", em geral, está cada vez mais elaborado e suas capacidades são frequentemente mal entendidas ou não usadas eficientemente. Existem muitas maneiras diferentes de estruturar os dados, cada uma com suas vantagens e desvantagens. Além disto, dados diferentes têm características diferentes que afetam a sua organização, e usuários diferentes fundamentalmente tem também necessidades diferentes.

As necessidades são suficientemente diversas e, em geral, nenhuma organização de dados pode satisfazer a todas as necessidades - pelo menos com o "hardware" disponível. Portanto, o trabalho do projetista é uma análise cuidadosa do delicado equilíbrio entre todos estes fatores (Martin, 1977).

#### O CONCEITO DE UMA METODOLOGIA DE PROJETO

Em termos de BD, uma "boa" metodologia de projeto deve (Costa, 1985a):

- 1) produzir uma estrutura de BD que satisfaça os objetivos do usuário em um intervalo de tempo e com uma quantidade de esforço razoáveis;
- 2) ter generalidade e flexibilidade suficientes para que possa ser usada não apenas por pessoas com diferentes níveis de experiência em projetos, mas também por aquelas limitadas por diferentes modelos de dados ou "software" de gerenciamento de BD;
- 3) ser reprodutível, de forma que duas pessoas aplicando a metodologia ao mesmo problema produzam mesmo (ou aproximadamente o mesmo) resultado.

Para atingir estes objetivos, uma metodologia de projeto de BD deve ter os seguintes componentes básicos (Teorey and Fry, 1982):

- 1) um processo de projeto (uma série de etapas, cada uma das quais requerendo uma escolha entre as alternativas de projeto e produzindo um resultado bem definido);
- 2) técnicas de projeto (para realizar a enumeração das alternativas requeridas no processo de projeto);
- 3) critérios de avaliação (para selecionar uma alternativa em cada etapa do processo);
- 4) requisitos de informação (para servir de entrada ao processo de projeto e a cada uma das suas etapas);
- 5) um mecanismo descritivo (para representar a entrada de informação e os resultados de cada etapa do processo).

A evolução das metodologias de projeto de BD é um processo dinâmico e não há um acordo geral sobre qual seja o "melhor" enfoque. As metodologias utilizadas na prática têm se fundamentado em princípios geralmente aceitos, não dependendo de pontos de vista ou suposições.

Entre elas estão as metodologias de Bubenko, Kahn, Sheppard, Tozer, Chen e Gerritsen; todas elas propõem quatro etapas que devem ser consideradas como guias flexíveis para o administrador ou para o projetista do BD:

Fase 1 - Formulação e Análise de Requisitos.

Fase 2 - Projeto Conceitual (ou Projeto de Visões e Integração).

Fase 3 - Projeto de Implementação (ou Projeto Lógico).

Fase 4 - Projeto Físico.

#### FASE 1 - FORMULAÇÃO E ANÁLISE DE REQUISITOS

Requisitos de informação corretos e completos são ingredientes chave no planejamento de sistemas de informação como Bancos de Dados, bem como na implementação das aplicações baseadas nestes sistemas.

Por enquanto, tem havido relativamente pouca pesquisa na determinação de requisitos e existem poucos procedimentos práticos e bem formulados para este fim. Alguns métodos para obtenção e documentação de requisitos têm sido propostos, mas eles tendem a apresentar soluções gerais em vez de métodos alternativos para implementar uma estratégia específica, escolhida para a determinação de requisitos.

Seres humanos aparentam ser tão versáteis com respeito ao uso da informação que as restrições ou limitações humanas são frequentemente ignoradas. Isto significa que "questionar" usuários sobre os seus requisitos de informação não leva, necessariamente, a um conjunto completo e correto de requisitos. Estas limitações são (Davis, 1982):

- A memória humana de longo período tem capacidade essencialmente ilimitada e requer apenas alguns milissegundos para ser lida, mas o tempo necessário para o registro (nela) é muito grande.
- A memória humana de curto período é muito rápida, mas de pequena capacidade (esta capacidade tem sido caracterizada como " $7 \pm 2$ ", isto é, entre 5 e 9 peças de informação podem ser retidas nesta memória durante uma operação (Miller, 1956); esta memória é utilizada (no processamento humano de informação) para operações do tipo "calcule" e "compare". Estas limitações podem afetar a habilidade humana em definir requisitos (um usuário sendo entrevistado não pode manter nesta memória um grande número de itens para discussão ou análise e, assim, suas respostas são limitadas. Isto também afeta o número de requisitos que o usuário considera importantes). Estas limitações podem ser significativamente reduzidas pelo uso de memória externa (como uma folha de papel) e de metodologias que sistematicamente elegem e registram pequenas quantidades de peças de informação.
- Existem evidências substanciais que mostram que os seres humanos não estão livres de interferência na sua seleção e utilização de dados. O efeito sobre a determinação de requisitos é uma interferência significativa naqueles requisitos baseados em procedimentos atuais, informação correntemente disponível, eventos recentes e inferências a partir de pequenas amostras de eventos. O analista e o usuário que

compreendem estas interferências podem compensá-las. Entre as características humanas que interferem na seleção e uso de dados estão:

- a) "Ancoragem e aproximação": existe uma tendência para fazer julgamentos com base em algum ponto tomado como fundamental (ou âncora) e para fazer ajustes e aproximações em torno dele;
- b) "Concreteza": pessoas que tomam decisões tendem a considerar apenas a informação disponível, na forma como ela é apresentada, e não pesquisar dados nem manipular aqueles que lhes são apresentados. No caso da determinação de requisitos, significa que os usuários sofrerão interferência da informação que eles já tinham a respeito dos seus requisitos;
- c) "Recência": as pessoas são mais influenciadas pelos eventos recentes do que pelos do passado. Novamente, a determinação de requisitos será prejudicada por aqueles acontecimentos que ocorreram mais recentemente, ao passo que necessidades originadas por acontecimentos mais antigos serão relegados a segundo plano;
- d) "Análise estatística intuitiva": seres humanos não compreendem intuitivamente o efeito do tamanho da amostra na variância e, assim, tiram conclusões duvidosas a partir de pequenas amostras ou de um pequeno número de ocorrências. Isto é importante porque muitos fenômenos de uma organização ocorrem a uma taxa muito pequena. Além disto, existe a tendência de identificar causalidade com ocorrência conjunta e assinalar causa onde não existe causa. Tudo isto pode resultar em julgamento errado da necessidade de informação.

Baseada nestas limitações, uma metodologia de determinação de requisitos deve:

- 1) auxiliar o analista a limitar e estruturar o espaço-problema (a maneira como um indivíduo que toma decisões representa o problema com o qual vai trabalhar), pois estima-se que os analistas gastem 75% do seu tempo nesta atividade (Davis, 1982);
- 2) auxiliar a pesquisa eficiente dentro do espaço-problema (deve complementar a definição, descobrindo requisitos que não foram obtidos pelo princípio da "Ancoragem e aproximação" e que estão limitados pela memória humana de curto período);
- 3) auxiliar a supressão de interferências, tais como "recência", "concreteza" e pequenas amostras;
- 4) prover garantias de que os requisitos sejam corretos e completos.

Metodologias diferem no detalhamento da estrutura provida. Algumas provêm estrutura conceitual, mas pequena estrutura de documentação e de processos; outras dão estruturas detalhadas para todas as tarefas e toda a documentação. A importância de estrutura detalhada pode variar conforme as circunstâncias. Por exemplo, analistas e usuários com pouca experiência podem achar uma estrutura detalhada muito útil, ao passo que aqueles (analistas e usuários) experientes nas suas áreas de aplicação e aptos a definir requisitos podem achar uma metodologia com estrutura muito detalhada um tanto inibidora e frustrante.

A estratégia de determinação de requisitos varia conforme o problema ou o sistema que está sendo projetado, mas a sua escolha pode ser sumariada como mostra a Figura 1, onde a estratégia escolhida é função da incerteza na informação (ou conhecimento) que o usuário tem do sistema.

Incerteza	Estratégia
Baixa ↑ ↓ Alta	Questionamento/entrevistas. Derivação a partir de um sistema existente. Síntese a partir das características da utilização do sistema. Descoberta a partir de experimentação.

Figura 1 - Estratégias para determinação de requisitos.

#### FASE 2 - PROJETO CONCEITUAL (OU PROJETO DE VISÕES E INTEGRAÇÃO)

Nesta fase, os requisitos de informação dos usuários são modelados conforme a necessidade de cada um deles e, a seguir, combinados em um modelo global (também chamado esquema conceitual) que especifica todas as interações mútuas entre os usuários.

Das metodologias existentes para o projeto conceitual do BD, pode-se classificar qualquer uma delas como fundamentada na filosofia "top-down", ou na filosofia "bottom-up". Um exemplo da primeira é a chamada "Análise de Entidades", na qual as visões dos usuários são modeladas, pouco a pouco refinadas e, em seguida, consolidadas. A metodologia conhecida por "Síntese de Atributos", por sua vez, é do tipo "bottom-up" e nela primeiro são identificadas todas as entidades e seus atributos (elementos de dados), conhecidos pelos usuários. Em seguida, procuram-se os relacionamentos entre as entidades e constrói-se um modelo gráfico de toda a informação, o qual é confrontado com as expectativas dos usuários.

Tanto a "Análise de Entidades" como a "Síntese de Atributos" utilizam o modelo Entidade-Relacionamento; embora outros modelos existam, este é simples e bastante conhecido, sendo largamente utilizado no desenvolvimento de BDs centralizados. No caso de sistemas (de bancos de dados) distribuídos, as mesmas metodologias utilizadas para o projeto de sistemas centralizados são usadas com algumas adaptações, não existindo até o momento uma metodologia de projeto que considere explicitamente as características e os objetivos de um sistema distribuído, em todas as fases de seu desenvolvimento. Neste sentido, estudos estão sendo feitos (Costa, 1985b) visando a estruturação de uma metodologia de projeto de BD distribuídos, híbrida (no sentido que utiliza a filosofia "bottom-up" em algumas fases e a "top-down" em outras), que permite antecipar algumas decisões importantes de projeto - inclusive a sua viabilidade técnica e econômica.

### FASE 3 - PROJETO DE IMPLEMENTAÇÃO (OU PROJETO LÓGICO)

Aqui tradicionalmente o modelo (ou esquema) conceitual é traduzido segundo um modelo interno de dados escolhido (relacional ou hierárquico, por exemplo) e descrito conforme as exigências do Sistema Gerenciador de Banco de Dados (SGBD) especificado.

A fronteira do projeto lógico não está bem estabelecida, sendo esta apenas uma fase intermediária entre os projetos conceitual e físico, caracterizando-se tão somente por formular um esquema implementável (processável pelo SGBD), isomórfico ao esquema conceitual, e por permitir uma avaliação da performance da estrutura lógica.

Tradicionalmente, no caso de sistemas de BD distribuídos, é nesta fase que são considerados pela primeira vez alguns parâmetros que orientam a distribuição propriamente dita. Entretanto, esta consideração tardia obriga o projetista a efetuar novas coletas de requisitos junto aos usuários, em busca de informações próprias para a distribuição; estes novos requisitos podem, entretanto, afetar o esquema conceitual, tornando-o obsoleto e invalidando todo o trabalho (ou parte dele) já desenvolvido.

### FASE 4 - PROJETO FÍSICO

Este é o nível mais baixo de abstração de um BD, e seu objetivo é produzir um BD físico que atinja a maior performance com um menor custo. Basicamente, os itens considerados são:

- Projeto de formato de registros e "clustering".
- Projeto de métodos de acesso.
- Mecanismos de manutenção de integridade e segurança.
- Projeto de programas de aplicações.

### REFERÊNCIAS

- COSTA, P.O. Metodologias e ferramentas para o projeto de banco de dados, INPE, São José dos Campos, 1985a (no prelo).
- COSTA, P.O. Proposta de uma metodologia para projeto de banco de dados distribuído, dissertação de mestrado em desenvolvimento, INPE, São José dos Campos, 1985b.
- DAVIS, G.B. Strategies for information requirements determination, *IBM Systems Journal*, Vol. 21, nº 1, 1982.
- MARTIN, J. Computer Data Base Organization 2nd edition, Prentice-Hall Englewood Cliffs, New Jersey, 1977.
- MILLER, G.A. The magical number seven, plus or minus two: some limits on our capacity for processing information, *The Psychological Review*, Vol. 63, nº 2, 1956.
- TEOREY, T.J.; FRY, J.P. Design of Database Structures, Prentice-Hall, Englewood Cliffs, New Jersey, 1982.