

RESUMO

Analisa-se a evolução dos Sistemas de Gestão de Bases de Dados (SGBDS) quanto a normalização, apresentando e analisando o trabalho desenvolvido pelo grupo ANSI/X3/SPARC.

Compara-se a arquitectura dos SGDBs actualmente comercializados com a proposta pelo ANSI/X3/SPARC

Finalmente, detectam-se alguns vectores da evolução da informática nos nossos dias, e de como influenciarão os SGDBs de amanhã.

ÍNDICE

a - ÂMBITO DO TRABALHO DO GRUPO ANSI/x3/SPARC

3 – APRESENTAÇÃO DOS PRINCIPAIS TÓPICOS TRATADOS

4 - ANÁLISE DO RELATÓRIO DO ANSI/x3/SPARC

4.1 - Uma Arquitectura Geral para um SGBD

4.1.1 - Nível Externo

4.1.2 - Nível Conceptual

4.1.3 - Nível Interno

4.2 - Especificação de Funções

4.2.1 - Funções de Administração

4.2.1.1 - Administração do Sistema (AS)

4.2.1.7 - Administração das Aplicações (AA)

4.2.1.3 - Administrador da Base de Dados (ABD)

4.2.2 - Programador de Aplicações (PA)

4.3 - Interfaces

5 - DE COMO OS SGBDS DE HOJE SEGUEM AS ESPECIFICAÇÕES DO ANSI/x3/SPARC.

6 - CONCLUSÕES

1 - Introdução

Novas necessidades ser tidas como complexidade crescente das aplicações informáticas vem exigindo

modificações na filosofia tradicional de concepção das aplicações e dos ficheiros por elas manipulados.

Assim, enquanto não houve necessidade de aceder às mesmas informações simultaneamente por mais de uma aplicação, foi frequente os ficheiros serem desenhados de forma a otimizar o processamento de uma dada aplicação. A utilização das mesmas informações por outra aplicação exigia normalmente a reestruturação dos ficheiros em causa. Na época, a filosofia de estruturação das informações é aquilo a que poderemos chamar construção de ficheiros como suporte de aplicações em oposição à tendência actual que será construção de bases de dados como suporte de sistemas de informação.

A filosofia de criação de ficheiros como suporte de aplicações embora útil e adequada em determinados processamentos torna-se impraticável noutros. Hoje em dia temos cada vez mais necessidade de uma filosofia integrada de desenho de sistemas de informação integrados, devido a múltiplos factores, entre os quais é de assinalar:

- Necessidade de que vários programas utilizem simultaneamente a mesma informação;
- Necessidade de controle de coerência entre várias informações;
- etc.

Como resultado das necessidades atrás apontadas, surgiu na década de 60 um novo e importante conceito: bases de dados (BD) e com ele outros que lhe estão indissoluvelmente associados, tais como Sistema de Gestão de Bases de Dados, Sistema de Base de Dados, etc. Estes conceitos estão definidos e posicionados num outro trabalho apresentado a este 1º Congresso Português de informática (Lucas, Murais, 1980). Nele se desenvolve o conceito de Sistema Integrado de Informação (SII), dedicando especial atenção aos diferentes aspectos referentes às Bases de Dados, como o seu posicionamento e interligações face aos SII, uma metodologia para sua construção e a definição de perfis de competência. Quer a metodologia de construção de uma BD, quer os perfis de competência são compatíveis com as especificações do relatório ANSI/X3/SPARC que nos propomos analisar na presente comunicação. Apesar do conceito de BD ter nascido, como se disse, no início da década 60, só em 1971 foi dado um primeiro passo no sentido de se encontrarem especificações precisas para um SGBD. Concretamente, o CODASYL Data Base Task Group (CODASYL1971) elaborou um importante relatório onde são feitas propostas para:

- Linguagem de descrição de uma Base de Dados (Data Description Language);
 - Linguagem de descrição da parte de uma Base de Dados, conhecida por uma dada aplicação;
- Linguagem de manipulação de uma Base de Dados, apresentada como extensão do COBOL (Data Manipulator Language).

Embora as especificações dadas no referido documento, tenham sido fonte de muito debate e controvérsia, não restam dúvidas da sua importância na evolução da tecnologia associada às Bases de Dados.

Mais tarde, em 1972, o "Standards Planning and Requirements Committee" (SPARC) do "American National Standards Committee and Computer and Information Processing" (ANSI/x3) respondendo à necessidade de formalização e clarificação do campo abrangido pelos SGBD criou um grupo de estudos ad hoc cuja finalidade era determinar que áreas da tecnologia associada às Bases de Dados eram possíveis do desenvolvimento de normas, e produzir um conjunto de recomendações para cada uma dessas áreas.

Este grupo definiu entre outras coisas uma arquitectura para um Sistema de Bases de Dados (SBD) e produziu em 1975 um importante relatório interno (ANSI/X3/SPARC, 1975).

Pensamos que o aparecimento dos SGBDs e de toda a disciplina que lhes está associada, tendo a sistematizar e melhorar consideravelmente a concepção aos Sistemas de Informação (SI). Esta melhoria tem dois resultados palpáveis:

- As soluções encontradas tendem a atingir cada vez mais os objectivos fixados;
- O natural crescimento de um SI começa neste momento a ser possível sem grandes acidentes.

O trabalho do Grupo de Estudo (GE) ANSI/X3/SPARC vem corroborar esta nossa ideia, pois ocupou-se fundamentalmente de:

1) Definir para os SGBD uma arquitectura em camadas que comunicam através de lotes -fases. Esta arquitectura impõe totalmente, como se verá adiante, uma grande sistematização na concepção de BDS.

2) Especificar as funções necessárias à concepção, implementação e manutenção de BDs (estritamente ligadas às camadas especificadas).

Os interfaces e enunciados em 1 são o Único elemento que o grupo considera candidato a normalização

Porque pensamos que o relatório produzido por este grupo de trabalho contem matéria extremamente importante e novadora, propomo-nos nesta comunicação efectuar a sua análise critica, onde tentaremos fazer sobressair os pontos mais importantes.

2 - ÂMBITO DO TRABALHO DO GRUPO ANSI/x3/SPARC

A primeira tarefa do grupo ANSI/X3/SPARC foi determinar o seu próprio âmbito de Trabalho, concretamente definir que elementos deviam constituir aquilo a que chamaram "Gestão de Bases de Dados" Database Management". Com este objectivo começaram por definir os componentes básicos de um Sistema de Informação (SI) tendo seguidamente analisado quais dentre eles cairiam no domínio da "Gestão de Bases de Dados

Foram assim encontrados 5 componentes básicos para es SI, a saber:

1. Mensagens
2. Registos
3. procedures

4 - Recursos físicos (Resources)

5 - Processos (Processes)

Analisando estas componentes, o Grupo de estudos decidiu que "todos os input/output, concretamente, leitura e perfuração de cartões, listagens, input / output via terminal e ainda transferencia de informação entre processos deviam ser considerados mensagens e como tal caiam na disciplina de Gestão de Mensagens, não fazendo parte da Gestão de Base de Dados.

Decidiu ainda que todas as actividades relacionadas com a preparação, compilação, teste e catalogação de um programa até que esteja disponivel para ser executado cairiam na disciplina de Gestão de Procedures, não fazendo também parte da Gestão de Bases de Dados. Os problemas de dispatching e assignação de drives de disco banda têm a ver com os recursos físicos do computador e caem na disciplina de Gestão de Recursos, também não fazendo parte da Gestão de Base de Dados. O Grupo de Estudos decidiu ainda que todos os objectos relacionados com variáveis locais, áreas de trabalho, contadores de instruções, etc, tem a ver com o estado de um processo caem em consequência na disciplina de Gestão de Processos" que também não faz parte da Gestão de Base de Dados.

Finalmente tomou Registos, campos, ficheiros, etc e as suas descrições e tudo o que diga respeito a índices, técnicas de mapping, métodos de acesso, organizações de ficheiros e linguagens orientados utilizador-final (end-user-languages) como elementos constitutivos da disciplina de Gestão de Base de Dados e consequentemente os únicos a tratar no trabalho que foram incumbidos de realizar.

3 - APRESENTAÇÃO DOS PRINCIPAIS TÓPICOS TRATADOS

O Grupo de Estudos considerou que, uma vez definido o âmbito do seu trabalho (ver Cap. 2), o passo seguinte seria elaborar um conjunto de especificações a que deviam obedecer os SGBDs.

Foi assim desenvolvida uma arquitectura geral para um SGBD, para cuja definição começaram por analisar o funcionamento dos SGBDs existentes. Chegaram a conclusão que, genericamente, estes sistemas estavam organizados em 2 níveis: a informação tal como é vista pelo sistema e a informação tal como é vista pelo programador.

Atribuíram a esses dois níveis a designação genérica de interno e externo. Observaram ainda a necessidade funcional de um terceiro nível, a que chamaram de conceptual. O nível conceptual representa a descrição global da parte do mundo real modelizada na DO. Esta descrição segundo o próprio grupo, corresponde aquilo que é informalmente invocado, quando numa fase adiantada do projecto, o utilizador e o analista não se entender quanto aquilo a que chamam de especificações do sistema.

Surge assim um modelo rudimentar da arquitectura proposta pelo ANSI/X3/SPARC (ver Fig. 3).

De notar, que mesmo actualmente (1980) nenhum SGBD comercializado satisfaz completamente estas especificações, que devem no nosso entender ser encaradas como um objectivo a atingir nos próximos tempos.

Paralelamente a definição de uma arquitectura para um SGBD foram definidas as funções para a concepção e implementação de uma BD compatíveis com os níveis de arquitectura definidos. Estas funções vêm substituir, em nosso entender, no que diz respeito à criação de BD, as normalmente definidas como analista funcional, analista orgânico e programador.

Foram assim definidas as seguintes quatro funções, que serão melhor analisados no Cap. 4 (ver Fig. 3)

- Administrador do sistema (AS)

Grosso modo define o modelo conceptual

- Administrador da Base de Dados (ABD)

Define o modelo interno

- Administrador da Aplicação (AA)

Define o(s) modelo(s) externo(s)

- Programador de Aplicação (PA)

Programa uma ou mais aplicações correspondentes a um ou mais modelos externos

De notar que um mesmo indivíduo pode executar diferentes funções, podendo uma função ser executada por mais de um indivíduo.

A comunicação entre os diferentes níveis da arquitectura proposta e efectuada através de interfaces, sendo através deles que passa a informação entre os diferentes módulos.

O Grupo de Estudos considerou que a normalização no domínio de "Gestão de Bases de Dados", no caso de vir a ser feita deveria incidir entre os diferentes componentes da arquitectura, e não propriamente na definição do conteúdo funcional de cada um deles (filosofia esta que também foi seguida na normalização de outros domínios da informática). Em consequência designou os interfaces, sua caracterização e definição da informação que os atravessa, como os únicos elementos passíveis, no momento, de desenvolvimento de normas.

4 - ANALISE DO RELATÓRIO DO ANSI/X3/SPARC

O Grupo de Estudos esquivou-se a definição de BD e SGBD limitando-se a dar algumas características

de cada um deles. Assim uma BD contém informação acerca de uma determinada parte do mundo real. A esta "parte do mundo real chamam empresa (enterprise). Representa normalmente informação acerca de um grupo de pessoas, artefactos, ideias, acontecimentos e processos organizados para suportar objectivos comuns".

- Quanto a um SGBD deve contribuir para o aumento da utilização da Informação e para a melhor integração das funções de processamento e acesso a Informação no seio da organização. Deve facilitar o acesso de informação por múltiplas classes de utilizadores. Deve melhorar a confidencialidade e integridade. A integração da informação proporcionada por um SGBD deve assegurar maior consistência entre as informações e controle de redundância. Finalmente um SGBD deve permitir evolução da BD proporcionando data independence. Data Independence permite aos utilizadores a criação de novas aplicações com os objectivos de seguir a evolução da empresa, corrigir erros de concepção ou melhorar a performance e a resposta do sistema".

4.1 - Uma arquitectura geral para um SGBD

A arquitectura proposta pelo Grupo de Estudos (GE) para um SGBD exige claramente a definição de BDs através de metodologias bem definidas e por pessoas com funções bem determinadas. Esta assim orientada para a construção de sistemas bem estruturados e flexíveis, capazes de por um lado evoluir com a própria evolução de empresa para que foram definidos e por outro com o hardware e software disponíveis.

Este é um ponto extremamente importante, em nosso entender, pois pensamos que da clara definição de metodologia e funções na construção de uma BD, depende fortemente, senão completamente, o êxito ou falhanço da sua futura utilização.

A arquitectura proposta é composta por 3 níveis

- Externo
- Conceptual
- interno

4.1.1 - Nível externo

ao nível externo que são definidos os vários não modelos externos (external views) da BD, sendo cada um constituído por uma colecção de objectos representando entidades, propriedades e relações de interesse para a aplicação correspondente. Uma aplicação é definida segundo o GE como "uma parte de uma empresa organizada para atingir determinado objectivo, integrado no objectivo global da empresa". É dado como exemplo uma empresa comercial que terá entre as suas aplicações salários, marketing e investigação e desenvolvimento. A cada modelo externo da CD, está associado um esquema externo que não é mais do que a sua descrição numa linguagem compreensível pela máquina.

4.1.2 - Nível conceptual

ao nível conceptual é definido o modelo conceptual da CD, constituído por "uma colecção de objectos representando entidades, propriedades, relações de Interesse para a empresa" A descrição destes objectos numa linguagem entendível pela máquina chama-se esquema conceptual.

Segundo o próprio GE, o modelo conceptual surgiu, como uma indirectão necessária entre os modelos interno e externo(s) de forma a permitir "data independence" necessária entre os modelos interno e externo(s), de forma a permitir "data Independence"

Deve entre outras coisas:

- Permitir a definição de novos esquemas externos ou a modificação dos existentes sem modificações no esquema inteiro;
- Permitir que modificações no esquema interno sejam Invisíveis ao nível externo;

Numa arquitectura de SGBD a dois níveis (a tradicional - ver fig. 5) qualquer modificação no esquema interno (por ex. para melhorar a performance) tem repercussões imediatas nos esquemas externos, obrigando à modificação dos programas que sobre eles trabalham.

Na arquitectura a três níveis proposta pelo ANSI/SPARC (ver fig. 6) a _ mesma modificação no esquema interno, apenas afecta o "mapping" Interno/conceptual não tendo qualquer repercussão nos esquemas externos, e conseqüentemente nas aplicações que sobre eles trabalham.

4.1.3 - Nível interno

É ao nível Interno que e definido o modelo Interno da BD. "A este nível a BD e um conjunto de objectos relacionados com os objectos respectivos do modelo conceptual. É descrito por um esquema interno. O modelo Interno e desenhado com vista a urna utilização mais eficiente dos recursos disponíveis (hardware e software) compatível com as necessidades de processamento da empresa.

Segundo o GE "os objectos conceptuais, que representam a organização do mundo real, devem ter uma vida mais' longa do que as tecnologias segundo as quais o hardware e o software são baseados. Assim o modelo Interno deve estar apto a mudar para reflectir a mudança das tecnologias". Dá-se assim uma grande flexibilidade ao sistema.

O

As correspondências entre objectos aos níveis externo, conceptual e Interno são efectuadas através de "mapping".

Na Fig. 7 e apresentada uma arquitectura geral para um SGBD1 tal como foi definido pelo GE.

4.2 - Especificação de funções

A arquitectura atras definida não estaria completa, se nao fossem também indicadas as funções necessárias a definição de uma BD (quem vai definir o modelo conceptual, modelos externos e modelo Interno?) e sua implementação (quem vai escrever os programas que a utilizarão?).

4.2.1 - Funções de administração

4.2.1.1 - Administrador do sistema (AS)

O administrador do sistema que define o esquema conceptual, sempre que possível numa linguagem directamente compreensível pela maquina.

Mas que quer efectivamente dizer definir o esquema conceptual?

Primeiro quer dizer que deve "identificar as necessidades globais da empresa, para as aplicações existentes, propostas ou potenciais".

Assim o AS determina

- 1) "A totalidade da Informação necessária a empresa;
- 2) A melhor maneira de recolher a informação e a melhor maneira de a modelizar;
- 3) As regras de integridade e segurança sobre a globalidade da BD".

O AS deve utilizar para a definição do esquema conceptual o "conceptual Database Schema Processor" de uma forma interactiva (ver.Fig. 7).

4.2.1.2 - Administrador das Aplicações (AA)

O AA define o esquema externo para uma ou mais aplicações, devendo "Identificar as necessidades de informação existente(s) na(s) aplicações em causa.

Define assim os objectos externos : podendo um objecto externo ser definido de varias formas: identificando explicitamente os seus componentes. extraindo-o de objecto(s) externo(s) já definido(s), ou extraindo-o da definição do esquema conceptual".

"O AA exerce controle sobre as aplicações que gere, definindo regras específicas para os seus objectos externos, estruturas e descritores para esses mesmos objectos".

O AA define ainda a ligação dos seus objectos externos aos objectos do modelo conceptual, devendo utilizar o "External Database Schema Processor" (ver fig. 7) para definir o(s) seu(s) esquema(s) externo(s) de uma forma interactiva

4.2.1.3 - Administrador da Base de Dados (ABO)

"O ABD em coordenação com o AS define as necessidades globais de Informação interna de forma a satisfazer as aplicações existentes, propostas e potenciais".

É ele o responsável pela definição do esquema interno da BD. Começando por tomar conhecimento dos objectos existentes ao nível conceptual e das necessidades de utilização da informação na empresa, o ABD define os objectos Internos, a sua representação, Organização, armazenamento e acesso. Deve depois definir os "mappings" entre objectos conceptuais e objectos internos. Encarrega-se da reorganização do modelo interno da BD, de forma a obter a melhor performance, de acordo com as prioridades definidas pelo AS. Utilizará utilitários fornecidos pelo SGBD para controlar e examinar a utilização da informação de forma a assegurar-se de que a política adoptada, rotinas e controles são adequados a obtenção dos objectivos da empresa

Como resumo das diferentes funções de administração, diz o relatório 'Pondo o controle político nas mãos do administrador do sistema, assegura-se que a BD desenvolvida reflecte as necessidades globais da empresa e não beneficia apenas alguns interesses particulares.

Isolando a definição, optimização e gestão da informação Interna nas mãos do ABD assegura-se a competência técnica apropriada a realização dessas tarefas".

4.2.2.- Programador de Aplicações (PA)

O programador de aplicações codifica os programas. Ele só deve conhecer o(s) esquema(s) conceptual(ais) correspondente(s) à(s) aplicação (oes) que desenvolve. Estes esquemas externos são formatados numa forma compreensível por um modulo existente no "external schema processor" (ver fig. 7).

4.3 - Interfaces

Como Já vimos atrás, foi definido um conjunto de interfaces que fazem a ligação entre os diferentes módulos do sistema. Remetemos para o relatório do ANSI/ SPARC a descrição de todos eles, de uma forma precisa, limitando-nos aqui a analisar alguns que nos parecem mais significativos (ver Fig. 7).

-1

Interface 1 - Descrição do Esquema Conceptual - "Source Format"

É através deste interface que o AS define e modifica para o resto do SGBD o esquema conceptual da BD. O "Conceptual Schema Processor" executa a sua validação.

4

e Interface 2 - Descrição do Esquema Conceptual - "Object Format"

É a interface através do qual e comunicada a descrição do esquema conceptual ao resto do SGBD via o dicionário de dados. -

Interface 3 - Descrição do Esquema Conceptual - "Display Format"

É através dele que a descrição do esquema conceptual e formatada para utilização pelos AA e ABD.

Os interfaces 13, 14 e 15 tratam da descrição do esquema interno.

Através da interface 13 o ABD descreve o esquema interno tal como o "mapping" conceptual-interno; do 14 essa descrição é comunicada ao resto do SGBD, e do 15 é formatada para utilização por pessoas para isso autorizadas.

Através dos interfaces 4, 5 e 6. são descritos os vários esquemas externos. Os AAs descrevem os seus esquemas externos através do Interface 4. Definem também os "mappings" entre os seus esquemas externos e o esquema conceptual ou ainda entre os seus esquemas externos e outros esquemas externos. Através do interface 5 a descrição do esquema externo tal como dos "mappings" externo-conceptual e eventualmente externo/externo, são comunicados ao resto do SGBD e através do interface 6 "External Schema Formatter" torna os esquemas externos acessíveis aos programas dos utilizadores. Existe um "External Schema Formatter" para cada "host language" p.ex. COBOL, FORTRAN, PL/1, etc.

Muito mais haveria para dizer sobre os interfaces, que não cabe no entanto no âmbito desta comunicação.

5 - DE COMO OS SOBOs DE HOJE SEGUEM AS ESPECIFICAÇÕES DO ANSI/X3/SPARC

Urra rápida análise dos SGBD; comercializados mostra-nos que:

- A Importante arquitectura em 3 níveis definida pelo que não existe de todo, ou se existe e apenas de uma forma rudimentar:

- Assim na definição do "esquema" de uma BD os 3 níveis encontram-se perigosamente misturados, tal como as funções de quem os define, o que leva a definição de BDs pouco flexíveis e que não atingem frequentemente sequer os objectivos iniciais;

- Vejamos o que se passa nos SGBDs actual. quanto aos diferentes níveis do esquema:

- Conceptual: não é mais do que a união de todos os esquemas externos (segundo o ANSI/SPARC é qualquer coisa de

qualitativamente diferente, pretendendo modelizar a globalidade da empresa, tendo em conta não só as aplicações existentes como também as propostas e as potenciais);

- Externo e interno - Não há entre eles uma separação nítida, não só do ponto de vista das linguagens disponíveis para a sua definição, como também do ponto de vista dos modelos actualmente existentes em que, como se sabe, excepto no modelo relacional, as estruturas lógicas reflectem quase directamente as estruturas físicas utilizadas para as implementar.

- A definição de funções na concepção das BDs actuais é bastante ambígua. Esta ambiguidade reflecte de uma forma clara a própria arquitectura dos SGBDs actuais.

6 - CONCLUSÕES

A definição de uma arquitectura a três níveis para um SGBD parece-nos ser a questão mais importante tratada pelo ANSI/X3/SPARC. Concretamente a definição do modelo conceptual como modelo global da BD, entre os níveis interno e externo e talvez o ponto mais inovador do seu trabalho.

Resta-nos perguntar:

- Que futuro para as especificações do ANSI/X3/SPARC1 uma vez que não são praticamente seguidas pelos SGBDs actuais?

E com esta pergunta ligam-se estas outras:

- Como serão os SGBDs de amanhã? E como serão definidas as BDs sobre eles implementadas?

Se analisarmos os SGBD actualmente comercializados, verificamos que não se atingiram alguns dos mais

importantes pontos que estavam na sua origem, a saber:

Clara separação entre estruturas lógicas e físicas da informação;

Flexibilidade e capacidade de evolução das BDs sobre elas construídas;

Melhoria de performance no tratamento de grandes volumes de informação.

Atendendo a alguns parâmetros da evolução da informática nos nossos dias, vemos que:

A necessidade de criação de BDs é irreversível;

- Exigem-se cada vez mais melhores tempos de resposta à interrogação dessas BDs, que serão também cada vez mais utilizadas por pessoas sem formação informática;
- Inadaptabilidade das arquitecturas dos computadores actuais ao processamento de BDs. Assim os SGBDs são obrigados a manter por software complicadas estruturas da informação;
- Com a vulgarização das redes de computadores as BDs estarão cada vez mais na ordem do dia.

Preve-se a seguinte evolução para os SGBDs:

- Melhoramento de performance, utilizando novas arquitecturas de computadores (p. ex. processadores associativos paralelos) que executarão por hardware as tarefas mantidas hoje em dia à custa de complicado software;

Melhoramento das arquitecturas actualmente propostas, utilizando um modelo em níveis, muito identificado como o proposto pelo ANSI/X3/SPARC, ou uma sua evolução;

- Melhoramento da "data independence", pela utilização de modelos de dados completamente independentes das estruturas físicas da informação;
- Desenvolvimento de linguagens de interrogação a ser utilizadas por pessoas sem preparação informática,
- Desenvolvimento de Sistemas de Gestão de Bases de Dados Repartidas (SGBDR), a funcionar sobre redes de computadores.