Unidade Curricular: Sistemas de Informação II

Ficha T. Prática n.º 1

Objectivo:
Reflectir sobre a necessidade para o DW, enquadrando a sua génese numa perspectiva histórica dos SIs.

Introdução

Os acetatos iniciais do cap. 1 desta UC apresentam um texto, já discutido durante a primeira aula, servindo aí para mostrar a necessidade do DWarehousing. Importa também abordar o texto introduzindo a visão histórica e comparativa com o modelo das 4 gerações e também introduzindo a perspectiva dual dos SI: vertente operacional e informacional.

Se estão lembrados, percebemos que a evolução histórica dos SIs e a alteração do ambiente de negócio e suas implicações levaram a um contínuo atraso e desajustamento das necessidades face às disponibilidades. Especialmente a questão da desintegração e heterogeneidade dos sistemas e inexistência de dados com suficiente profundidade histórica levaram à desadequação dos SIs entretanto desenvolvidos, em termos da satisfação das necessidades informacionais.

Vamos agora reflectir mais aprofundadamente sobre essa temática, introduzindo novos conceitos e perspectivas.

1.º Texto em Reflexão

Parte 1

In the beginning were applications. Users thought that applications would provide them with information. And insofar as the stated requirements of the applications were concerned, the applications sufficed. But over time the business requirements changed. Keeping the applications in sync with the changing business requirements was a difficult thing to do.

Along the way in trying to keep up with changing requirements, the end user encountered some other limitations to the world of applications. Those limitations were the need for integration, and historical data.

The applications that the corporations had created or otherwise acquired had no notion of integration. One application thought a customer was one thing. Another application has yet another interpretation of what a customer was. When it came to the corporate understanding of data, there was – simply stated – no corporate understanding. From a corporate perspective, the manager could not answer such basic questions as: Who is a customer? What is a product? What is a transaction?

In short, the different applications were never designed to work together in an integrated manner.

Parte 2

The 2nd issue was that applications focused inevitably on very current data. Applications could reveal: 1) how much money a customer has in the bank right now; or 2) where a shipment was right now; 3) what the status of an insurance policy was right now; 4) what the quota of a salesperson was right now; 5) and so on.

The applications were designed to keep track of what is going on right now. But when it came to a sense of need and importance of historical information, the applications treated historical data with no respect at all.

Unfortunately, integration and history were a very important component of information. And applications simply did not measure up.

Parte 3

The end user's first reaction was to rewrite the applications of yesterday. But this idea quickly fell by the wayside. The end user found that – as far as the applications were concerned – the clock could not be turned back. There were simply too many applications, too much undocumented code, too much fragile code, to much complexity to even attempt to roll back the tide of applications.

Parte 4

Thus was born the notion of a data warehouse, an alternative to the dilemma of the end user who needed information but could not impose change on the legacy applications environment.

Like **all radical** and **fresh concepts**, the notion of a data warehouse was derided and scorned by academics and theoreticians. Since the idea of a data warehousing had not risen among their ranks, it could not possibly be a valid concept. Today data warehousing is no longer a theory. It is conventional wisdom, and corporations around the world recognize that the road forward leads to the data warehouse.

Parte 5

Data warehousing forms the center of a wide universe. From the corporate data warehouse, with its granular, corporate integrated data, spring many different kinds of decision support activity. The data warehouse form the basis for such DSS processing as 1) data mart, department processing; 2) simple reporting of corporate information; 3) exploration processing; 4) data mining; 5) operational data stores; 6) project warehousing, and so forth.

But data warehousing did not happen all at once. Like a jigsaw puzzle, data warehousing had been put together a piece at a time. The world of data warehousing has been led by writers and by practitioners who became writers. These leaders have described from their experience what works and what does not.

***W. H. Inmon**, foreword from Objected-Oriented Data Warehouse Design*

2.º Texto em Reflexão

OLTP applications typically automate clerical data processing tasks such as order entry and banking transactions that are the bread-and-butter day-to-day operations of an organization. These tasks are structured and repetitive, and consist of short, atomic, isolated transactions. The transactions require detailed, up-to-date data, and read or update a few (tens of) records accessed typically on their primary keys. Operational databases tend to be hundreds of megabytes to gigabytes in size. Consistency and recoverability of the database are critical, and maximizing transaction throughput is the key performance metric. Consequently, the database is designed to reflect the operational semantics of known applications, and, in particular, to

minimize concurrency conflicts.

Data warehouses, in contrast, are targeted for decision support. Historical, summarized and consolidated data is more important than detailed, individual records. Since data warehouses contain consolidated data, perhaps from several operational databases, over potentially long periods of time, they tend to be orders of magnitude larger than operational databases; enterprise data warehouses are projected to be hundreds of gigabytes to terabytes in size. The workloads are query intensive with mostly ad hoc, complex queries that can access millions of records and perform a lot of scans, joins, and aggregates. Query throughput and response times are more important than transaction throughput.

Retirado de: Chaudhuri, S. & Dayal, U. "An Overview of Data Warehousing and OLAP Technology"

Tarefas propostas:

1. Relativamente à Parte 1 do texto (4 primeiros parágrafos), onde o autor introduz as primeiras questões a reflectir, identifique as características dos SIs descritos e respectivos problemas subjacentes.
2. Enquadre o ambiente dos SIs descrito nesses parágrafos numa perspectiva histórica, usando o modelo das 4 Gerações de Ambientes Computacionais (acetatos 12 a 18).
3. Na segunda parte do texto, o autor identifica outro problema. Refira-o e enquadre-o também no modelo das 4 Gerações de Ambientes Computacionais.
4. A terceira geração de ambientes computacionais, dita "Dispersão" introduz um stress adicional no ambiente de SI anterior. Mostre como se procurou reagir às novas alterações do ambiente (quer de negócio, quer de TI).
5. A terceira parte do texto mostra também essa situação de "revolta" e os problemas com que os utilizadores se debatem no momento. Identifique-os e relacione-os com aqueles identificados na questão anterior.
6. Na parte 4 do texto o autor introduz o *Data Warehousing* referindo que ele surge como uma resposta ao dilema enfrentado pelo utilizador (conhecido, neste contexto, por *knowledge worker*). Identifique o dilema e mostre como o *DWing* lhe deu resposta.
7. Nessa mesma parte do texto o auto refere que o conceito do *DWing* foi "escarnecido e refutado" por académicos e teóricos. É capaz de identificar algumas razões para tal?
8. Mostre o enquadramento do novo ambiente de SI no modelo das 4 Gerações de Ambientes Computacionais.
9. Mostre como as características das transacções (e, consequentemente, dos próprios modelos de dados) levaram à dualidade de sistemas: operacional e informacional apresentado no 2.º texto em análise e nos acetatos 22 a 28.
10. Na parte 5 do texto e no 2.º parágrafo do texto 2 são identificadas algumas utilizações dos dados do DW. Mencione-as e discuta a sua relevância.