Hilbert®

White Paper

Early vs. Late Data Binding in Business Intelligence

Version 1.0

Hilbert®

Contents

Introduction

Suppose that a company wants its phone system and cost analyzed per month, per quarter and per year for the last 5 years: The data volume is about 1,000,000 calls per year. Such an analysis would conventionally performed by reading through all phone records and at least a summarizing of the calls per day or per week before this data is loaded into the reporting database or even the data warehouse. To make reporting easier and faster such weekly, monthly and yearly summary records are stored in the reporting database too, to get a fast response when doing reporting. Thus the data have been summarized before being stored into the reporting database – let us call this *Early Data Binding* because the data has been early bound with the solution process (the report or graph etc.). Another way is to store all data into the reporting database and do the summarizing step just at the time the analysis as a report or graph etc. is done – let's call this *Late Data Binding* because the data binding process is done at a very late stage of the solution process.

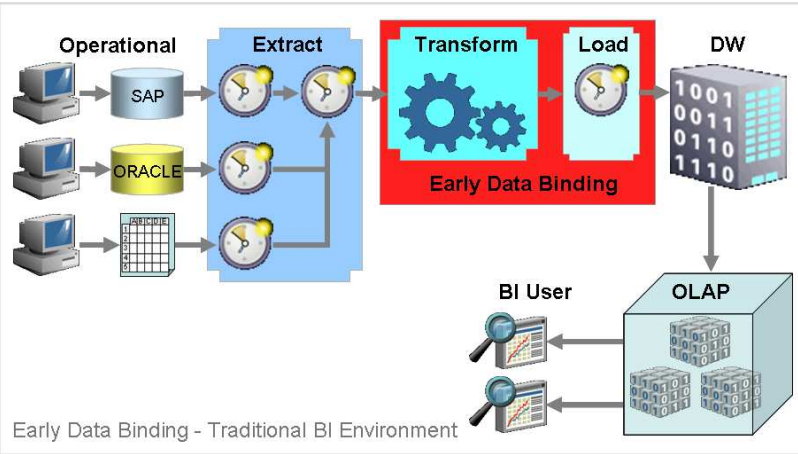
Why do Early or Late Data Binding?

In the past *Early Data Binding* was needed because current Data Analyzing Solutions do show performance weaknesses when they have to collect and calculate large amount of data at run-time and if such data is not prepared with fast search indexes such processes take even longer. Ad hoc queries often cannot be performed in acceptable time because the indexes are not there and re-indexing cannot be done on the fly.

Thus current Business Intelligence solutions have to prepare data for performance reasons with indexes, summary records built and other calculations made much in advance of the solution process and based on the reporting requirements – *Early Data Binding*. If speed is not compromised by a requirement for indexing, or by data volume, type or location of data, all data can be loaded into the reporting database and all reporting can be done without any upfront preparation in real-time and even on an ad-hoc basis – *Late Data Binding*.

Disadvantages of Early Data Binding

For emphasis: *Early Data Binding* does things to the data that <u>cannot be undone</u> once the data reside in the reporting database. *Late Data Binding* in contrast connects the data already in the reporting database to a solution set just before execution of the analysis. Parsing and format conversion are two examples of common *Early Data Binding* activities. The *T* of the ETL (Extract Transform Load) process, required in current Business Intelligence Solutions, is one kind of *Early Data Binding*.



Early Data Binding - Traditional BI Environment

Early Data Binding has important disadvantages because it removes information from the data and inserts assumptions before the solution system gets the data for solution. It may hide vital clues you need for the solution. After all, only the data actually in the database can be used for analysis.

In our example you may want to report on a weekly basis or you may want to highlight the top 10 callers or an analysis by LOB is needed with weekly buckets – all this cannot be performed because your data was initially prepared in an ETL process based on requirements defined at that time and information which is now important has been removed.

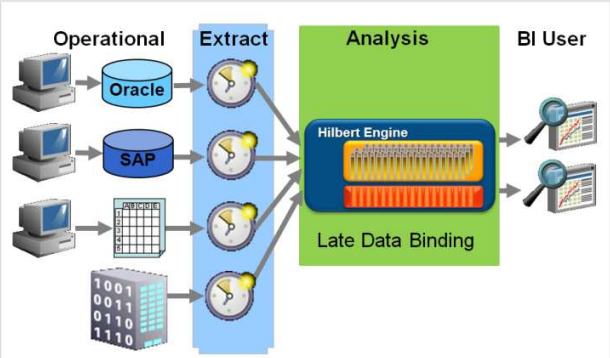
Advantages of Late Data Binding

Late Data Binding has important advantages, because it allows exploring all the data relating to the problem domain in an unrestricted way – at any time and without the need to foresee any upcoming data analysis requirements.

In the given example weekly or LOB reporting as well as the Top-10 callers can easily be examined because all details of the data are available within the reporting database.

Hilbert equals Late Data Binding

Hilbert's philosophy of business intelligence is 100% supporting the *Late Data Binding* approach. When Hilbert connects to a data source like a database or unstructured text information in Word or PDF documents, it does a 100% conversion (called quantification) of all data in all its detail and converts it to numerical representations in Hilbert space. This numeric representation allows complex but highly efficient mathematical operations to handle text information like any other numerical data. The consequences are dramatic. Searching and sorting are performed at the full numeric processing speed of the computer, with no added overhead. Working in a complete numerical environment, the entire body of several hundreds of year's worth of mathematical theorems and processes can now be applied in analyzing all the data. Thus Hilbert can do all data analysis at the time the analysis, the report, the graph, or the scorecard is calculated on the basis of the full data set without any approximation or simplification.



Late Data Binding – Hilbert

Complex Problems Solutions

The more difficult a problem is to solve, considering the tools at hand (in terms of existing solutions, ordinary database environments and query languages such as SQL) the more confusion there tends to be about the approach to take. If the problem is too complex, too onerous, too confusing, too time consuming, and analysts are tempted to use approximations and take shortcuts. Again this is a typical *Early Data Binding* process to trim the problem down to what is to be considered a manageable size. But manageable size could also mean a wrong solution or analysis because important influencing factors may have been deleted in the *Early Data Binding* process. Solutions are driven to this kind of successive complexification because the tools at hand (calculus and differential equations, for instance) tend to become tedious. More pragmatically, for conventional database environments, simplifications are used to get the solution in finite time. Given the proper analytical environment (Hilbert) companies don't need to make simplifying assumptions for the computer. Hilbert's numerical representation of all data in Hilbert Space and the use of the advanced mathematical operations give performance advantages which outperforms any shortcut or workaround required by a traditional, *Early Data Binding* environment.

Conclusion

Enterprises should not attempt to solve any part of data analysis within the ETL process elements of parsing and loading data in a *Early Data Binding* mode. Instead the data should be loaded in its entirety except, of course, to break it into logical pieces and to place it in a format compatible with the database storage format. Just get the data into Hilbert and do ALL of the analysis later in a *Late Data Binding* Mode.

About Hilbert

Pennsylvania based Hilbert Technology Inc. is an international provider of data management and analysis solutions for large and medium size enterprises worldwide. The offering is based on the revolutionary, patented Hilbert Engine technology for the ultra high-speed access, manipulation, storage and analysis of large volumes of structured and unstructured data. The Hilbert solutions are offered as industry or process specific solutions. Organizations in public services, law, government, finance, communications, whole- and retail sales, transportation & tourism and chemical & pharmaceutical can employ Hilbert solutions to gain unparalleled speed in access and analysis over large data volumes.

Hilbert Technology Inc.
P.O. Box 330
Newtown, PA 18940
USA
Web: www.hilbertcompany.com
E-Mail: info@hilbertcompany.com
Tel: +1 (212) 252 1600
Fax: +1 (212) 252 1615