White Paper

Bloor

Managing data growth on System z

A White Paper by Bloor Research
Author : Philip Howard
Publish date : October 2008

IBM Optim provides facilities that ensure that the archived data won't break and it offers a variety of mechanisms for accessing that data

Philip Howard

Executive summary

It hardly needs repeating that data volumes are increasing. Indeed, they are growing at a prodigious rate: estimates suggest that data volumes are doubling every eighteen months. This is partly because there is simply more data that we can usefully collect and use, and also for governance and compliance reasons where it is important to retain information for longer periods than might previously have been the case. At the same time, while disk capacities are getting bigger, disk drives themselves are not getting much faster. This means that, all other things being equal, performance will deteriorate in an unacceptable fashion. Moreover, service level agreements are becoming ever more stringent, thereby requiring performance improvements.

One remedy to this problem is to throw more hardware at it. However, at the same time that data volumes are growing and performance demands are increasing, we have overflowing data centres with escalating power and cooling costs, not to mention data centres where there is simply no more power available. So, while more processing power is an answer it is not a very good one except in extremis. More practical and more cost-effective is to get more out of the installation that you already have. There are several complementary ways to do this. For example, enhanced performance monitoring of the database, network and so forth is a good way to isolate and remove bottlenecks within the system. However, this only caters to the performance part of the problem and does not also address the sheer scale of the data volumes that have to be stored and managed. One way of doing this is to use data archival software.

Data archival as a concept has been around for a long time and yet there are relatively few companies that have implemented archiving policies either in their own right or within a broader data governance context. Even when they have done so, it tends to be on a do-it-yourself basis rather than through the use of automated tools. For example, in a z/OS environment DIY archival will involve sysprogs, DBAs, developers and testers, which is expensive in terms of both skills and resources.

This therefore raises the question: why do so few companies adopt formal approaches to archiving?

The short answers are that it is easy for users to say that they want continued access to everything, and that people are worried that if they archive the data then it might break in some way. So any solution that is going to be widely acceptable must enable continued access to the data after it is archived and must be able to provide guarantees that the archival process won't 'break' the data. We will discuss both of these issues but suffice it so say that they are linked: if you break the data you can't access it.

IBM Optim provides facilities that ensure that the data won't break and it offers a variety of mechanisms for accessing the data. In the sections that follow we will discuss breakage and access in more detail, and then briefly consider how IBM Optim works, particularly with respect to System z, though Optim's capabilities are also generally applicable to distributed systems.

Archiving data

Breaking the data

Traditional approaches to data archiving are relatively simplistic: you identify what you want to archive (either by date or by defining business rules or through some other mechanism) and then you archive the data by extracting it from the relevant tables (or their equivalent if you are not using a relational database) onto near-line storage or off-line storage of some sort.

Herein lies the rub. If you are simply working at the table level, how do you ensure the relationships that exist between entries in different tables remain intact? This is important, not only on an a priori basis but also because it is likely that one of the ways that users will want to be able to access data after it has been archived will be through their existing application software, which will expect the same database structure as exists in the live system and this, in turn, depends on the maintenance of data relationships across tables. In other words there is a strong possibility that the archiving process will break if relationships are not ensured.

Thus the main reason underlying the failure to implement data archiving policies is not just the fear of breakage for fear's sake but fear based on experience that traditional approaches to archiving genuinely do cause breakages. Thus what is required is a solution that maintains relationships and, by implication, does not work at the table level. As we shall see, this is exactly what IBM Optim provides.

Accessing archived data

There are a variety of ways in which users might want to access archived data. For example, they might want to query the data using a business intelligence tool; they might want to access it for compliance reasons (say), in which case access via SQL or something other programming language might be appropriate; or they might want to access the data by means of the originating application as if it was still online in the mainframe. Indeed, with reference to this last option it is also likely to be a requirement that you can reconstitute the archived data back into the live system if necessary.

These various requirements mandate that a variety of archival options are provided. Support for near-line storage (where the data is archived onto larger and slower disks but remain part of the main system) is one requirement. Another useful option is to be able to archive the data into a database on a distributed system where that may be more cost-effective but can still support access from the original application. Another potential method is to convert the data into XML format and compress it (highly), allowing access via XQuery or extended versions of SQL that support XML access. Finally, of course, you want the option to be able to archive data onto offline storage media such as tape.

In implementation terms the difficulty may be in determining what access methods users want to employ but this is an organisational issue: in terms of archival support you want a tool that is as flexible as possible in allowing multiple access methods.

IBM Optim

IBM Optim is available for both z/OS and distributed systems. In the case of the former there is an ISPF interface that supports batch JCL while the latter works in a conventional client/server fashion.

Optim works on the basis that what you archive are business entities. To take a simple example, if you wanted to archive customer details then the relevant business entity would include all orders for that customer and all the order details pertaining to each of those orders. The software will automatically recognise all the relationships that are implicit in the database schema (or equivalent, including what is defined in COBOL copybooks) making use of things such as primary and foreign key relationships. However, there may be additional relationships defined at the application level which it cannot recognise so there are also facilities within Optim to define business rules that capture these relationships so that you can define complete business entities. You then deploy the product's capabilities by archiving to your chosen media at the business entity level. The big advantage of this approach is that, provided that business entities are fully defined, the system cannot break when you implement archiving.

Currently Optim archives data on a timed basis: that is, you define schedules for archiving data once it is x weeks or months old. IBM is working to extend these capabilities so that you can choose other archiving policies (for example, you might want to archive data that has not been accessed for some period of time) based on business rules that you define.

In so far as the options available for archiving media are concerned, there are a number of options: you can archive to tape or near-line storage and you can archive to a database (DB2) instance on a distributed system. Note that in this last case this even applies if the original data is not managed by DB2 (if you are using IMS for example) as Optim includes normalisation capabilities. There are also options to store the archived data in a format suitable just for query access via a business intelligence product such as Cognos (IBM).

Conclusion

There is a clear business case for adopting formal archiving policies; lower costs, improved performance, a better ability to meet service level agreements and reduced pressure on the data centre being just some of these. However, only a minority of companies have implemented appropriate strategies. We believe that the main reason for this is because of fear of breakage and concerns over how the data can be accessed and used after it has been archived. IBM Optim, by working with business entities rather than at the table level, should remove concern over breakage. In addition, it offers a range of deployment options than enable the sort of flexibility that users require when accessing data. We believe that users of z Series (amongst others) should seriously consider the implementation of a solution such as IBM Optim for addressing the issues discussed in this paper.

Further Information

Further information about this subject is available from http://www.bloor-research.com/update/981

Bloor Research has spent the last decade developing what is recognised as Europe's leading independent IT research organisation. With its core research activities underpinning a range of services, from research and consulting to events and publishing, Bloor Research is committed to turning knowledge into client value across all of its products and engagements. Our objectives are:

- Save clients' time by providing comparison and analysis that is clear and succinct.

- Update clients' expertise, enabling them to have a clear understanding of IT issues and facts and validate existing technology strategies.

- Bring an independent perspective, minimising the inherent risks of product selection and decision-making.

- Communicate our visionary perspective of the future of IT.

Founded in 1989, Bloor Research is one of the world's leading IT research, analysis and consultancy organisations—distributing research and analysis to IT user and vendor organisations throughout the world via online subscriptions, tailored research services and consultancy projects.

Philip Howard
Research Director - Data

Philip started in the computer industry way back in 1973 and has variously worked as a systems analyst, programmer and salesperson, as well as in marketing and product management, for a variety of companies including GEC Marconi, GPT, Philips Data Systems, Raytheon and NCR.

After a quarter of a century of not being his own boss Philip set up what is now P3ST (Wordsmiths) Ltd in 1992 and his first client was Bloor Research (then ButlerBloor), with Philip working for the company as an associate analyst. His relationship with Bloor Research has continued since that time and he is now Research Director. His practice area encompasses anything to do with data and content and he has five further analysts working with him in this area. While maintaining an overview of the whole space Philip himself specialises in databases, data management, data integration, data quality, data federation, master data management, data governance and data warehousing. He also has an interest in event stream/complex event processing.

In addition to the numerous reports Philip has written on behalf of Bloor Research, Philip also contributes regularly to www.IT-Director.com and www.IT-Analysis.com and was previously the editor of both "Application Development News" and "Operating System News" on behalf of Cambridge Market Intelligence (CMI). He has also contributed to various magazines and published a number of reports published by companies such as CMI and The Financial Times.

Away from work, Philip's primary leisure activities are canal boats, skiing, playing Bridge (at which he is a Life Master) and walking the dog.

Bloor