

## Article

## Managing performance in a BI environment



[Philip Howard](#), *Director of Research - Technology*, Bloor Research

Published: 18th April 2007

Copyright © 2007. All rights reserved.



There are lots of tools available for managing and monitoring the performance of front-office applications. However, there are not that many that do a similar thing for business intelligence and data warehousing. Moreover, those tools that are available tend to be over-specific: for example, they can monitor the performance of the BI environment but not the data warehouse, or vice versa, or they can monitor how Oracle applications (say) interact with the BI environment but not how this inter-relates to data warehouse usage. In other words, you cannot generally gain a holistic view of the usage of your BI environment, taken as a whole.

One exception to this rule is Appfluent, a US-based company now starting its activities in the UK and Europe. What Appfluent does is to monitor all traffic going into or out of your data warehouse. However, unlike the likes of Symantec and Embarcadero, which do this purely from the perspective of compliance and security, Appfluent also monitors performance: for example, the number of rows retrieved, the sorts of joins employed, the indexes used (and when non-indexed data is retrieved), response time, use of materialised views and so on. Moreover, you can see all of this detail for individual users and, where required, you can drill down to the actual SQL involved.

However, note that this is the SQL coming into the data warehouse, not necessarily what is actually executed, since the database optimiser may re-write any badly constructed SQL. While you might want to look at the latter it is actually important that Appfluent can't do this. The big issue in data warehousing is, of course, performance and one of the major points about Appfluent is that it captures and analyses all of your workload without adding any overhead to production systems because the SQL collection is done at the network packet level. This means Appfluent does not require database or application logs or system resources such as main-memory sampling: it is therefore non-intrusive and does not impact on performance.

In addition to the metrics that Appfluent gathers it also collects details about the originating application so that you can see the application from which queries are being generated, such as Business Objects, Cognos, SAS Enterprise Miner, PeopleSoft (Oracle) applications, IBM AlphaBlox and so on, as well as ETL (extract, transform and load) applications from the likes of IBM and Informatica. However, at the database level the product only supports DB2, Oracle, SQL Server and Sybase IQ at present. The company is working on SAP BW and it is likely that support for other data warehouse products will be introduced in due course.

Based on the metrics gathered, Appfluent automatically generates some 50 standard reports (and you can customise your own) which can be presented using either Business Objects or Cognos, with a third vendor to be announced shortly. For users of other BI tools there is an ad hoc reporting module that can be tailored to work with your preferred query system/dashboard.

Apart from being able to monitor an individual's queries for compliance purposes (both to see who is accessing what in general, and to check for inappropriate usage) the big advantage that Appfluent provides is that it lets you identify where performance problems are occurring, not in isolation in a single part of the environment but across the whole environment. In principle, this means that you can identify where you should concentrate your tuning efforts, what data you have that is not being accessed, where it might be useful to implement indexes and materialised views (or remove those that already exist) and so on. You can also identify where you have redundant queries, the best use of ETL processes and so forth.

I strongly recommend that significant data warehouse users have a look at Appfluent. However, be prepared to be surprised. As an example, one of Appfluent's customers found that its query load was more than twice what it thought it was. The potential benefits of having this sort of better understanding of your requirement should be obvious. To give a more specific example, Pfizer, which uses Appfluent along with its HR data warehouse, was able to halve the size of that warehouse, reduce the number of user IDs it maintained by an order of magnitude, and cut its ETL load times by nearly three quarters. The implementation paid for itself in just 5 months. 'Nuff said.