

Don't Make Bad Data an Excuse

Using new technology to eliminate the shortcomings of traditional warehousing and reporting

A NetSearch White Paper

Created: 11/08/2010

Version: 2.0

Created by: Richard Lewis

Ref: Don't Make Bad Data an Excuse V2.doc

Private, Confidential and Commercially Sensitive

Don't Make Bad Data an Excuse

If you are a manager of a department or organisation, have you ever had to ask?

Why does it take so long to get the answers I am after?

Why can't you get me the reports I want?

Why can't you show the information I need in one report?

Why is it that I can't find this information out myself?

Have you ever had any of these answers back?

"The data doesn't join up properly as the data isn't coded consistently."

"I don't have access to the data that you require as it's too difficult to extract."

"The data is in lots of different systems so I can't bring it together into one report."

"It's too difficult for you to do this yourself. You need to be a SQL expert."

So what do we do about it? We commission a Data Warehouse project to get the data out of the various systems into one big database, fix the data and build a set of business views and cubes and a load of reports.

Several man years later you ask the same questions as before. This time the answers are slightly different but you still don't have the information or access to the data that you really need.

Why does it take so long to get the answers I am after?

There is a lot of data and it's slow to generate the results you need.

We need to update the warehouse first or the results will be out of date.

There are only two of us and we are up to our necks in it at the moment.

Why can't you get me the reports I want?

Only one of us knows how to write a report on that data and they're on holiday.

We'll have to add that information into our cubes as we can't currently get at it.

I am not sure where the data is held for that report.

We don't have enough licenses.

Why can't you show the information I need in one report?

We don't have that data in the data warehouse.

It's too difficult to combine the data into one view.

The tool doesn't allow me to do that.

Private, Confidential and Commercially Sensitive

Why is it that I can't find this information out myself?

You haven't sufficient access to query that data.

You need to understand the schema to get at that information.

You need to be a power user to do that and you haven't been on the course.

Badly coded and inconsistent data, difficulties in accessing data, large volumes and a requirement for "hardcore" technical expertise are all excuses that are used to justify why it is not possible for managers to access the information they need themselves.

Does this have to be the case?

A New Approach

Our view is that the traditional methods of ETL and Warehousing are only ever going to be a "part solution" to the problems of self-service and coping with erroneous and difficult to join up data.

The good news is that technologies do exist that can solve many of the problems highlighted above, however a fresh approach needs to be taken to shortcut the lengthy, costly and technically challenging processes involved in building enterprise warehousing and reporting solutions.

Where an organisation already has a data warehouse, all is not lost! All of the expense and the sweat and tears spent on analysing business requirements, designing load and transformation scripts, data schemas, cubes, universes and what ever other aggregate structures are required to simplify the report creation processes have resulted in the development of valuable information assets. These assets can be leveraged by newer, faster and more scalable technologies.

If you don't have a data warehouse then your Financial Director is going to be your new best friend as there are new ways of providing the structures and assets required for your enterprise reporting needs that can be delivered at a fraction of the price.

So how do we produce the panacea of business intelligence solutions?

Introducing NetSearch

NetSearch has been built specifically as a 'next generation' BI tool. The product utilises the raw power of search technology in order to 'assemble' data for querying/reporting purposes.

NetSearch enables users to answer business queries rapidly with the minimum amount of fuss, in a completely secure environment.

Private, Confidential and Commercially Sensitive

So What?

Search technology is incredibly fast and designed to allow users of any skill level to find information themselves. Search technology is most commonly associated with crawling and searching for documents and web pages, however it can be used to crawl and search databases providing an extremely easy and familiar access point to your corporate data.

If you have a warehouse, NetSearch can crawl and index it and provide a browser based front-end that allows you to run ad-hoc queries, build your own reports with minimal to no training.

If you don't have a warehouse, point NetSearch at your operational data stores and it will create a search engine that you can then use to access your data irrespective of its original location from a single point of access and through your favourite web browser.

As it's a search engine, you can also index other corporate assets such as your file servers, your email systems, and your local intranet. Combine this with data from your operational data stores and you have a single point of access to everything.

Removing Bad Data as an Excuse

So how does NetSearch address the common problems and restrictions inherent in data warehouse solution?

Why does it take so long to get the answers I am after?

Reports and queries will return with sub-second speed.

Data is automatically joined as it indexes and searches, speeding up query times.

Bad data can be merged, combined and fixed to remove duplicates and coding errors.

Users can perform their own ad-hoc queries and build their own reports.

Data can be kept up to date with the original source data for real-time analysis.

Why can't you get me the reports I want?

It no longer matters! Users can self-service their information needs.

The search engine can crawl any data source so you can have unlimited data access.

All of the data is accessible from one place and can be found by "Googling" the data.

We sell on an unlimited user basis so you won't run out of licenses.

Why can't you show the information I need in one report?

As all the data is brought together into a single search engine, you can!

You can keep on adding reports to create multi-panel dashboards yourself.

Private, Confidential and Commercially Sensitive

Reports are calculated in real-time and displayed in seconds over millions of records.

Why is it that I can't find this information out myself?

Security can restrict your access however if you have the rights, you'll find it!

If you can use Google you will have no problems searching for the data.

You" search" for the data and don't need to worry about where the data came from.

Conclusion

Bad data is always going to exist. In fact with the explosion of data we are all experiencing and the increase in the number of ways information is collected and stored there is going to be even more of it.

Traditional Data Warehousing and reporting solutions are expensive to implement, support and license however things are changing.

Many of the excuses we hear about why reporting is difficult are true.

It's difficult to do. When there's lots of data, it's slow. Where the quality of data is poor, it takes even longer to join the data together and eliminate coding errors from the report outputs. Designing schemas and delivering complex reports is a highly skilled job. It takes years to become expert in SQL or OLAP.

It's difficult because the tools that are used to do this job are often outdated and no longer fit for purpose.

NetSearch is a modern and innovative approach to data access and ad-hoc reporting. It simplifies the process of finding information and demystifies the process of ad-hoc analysis and reporting by removing many of the barriers incumbent in making BI tools available to the masses.

Live with your bad data but use technology to prevent this from being an excuse for poor, inaccurate and slow reporting. Search engines are here to stay and can play a big role in empowering people with the know how to handle their own management information requirements.