



Merging customer-referenced data into one database is no simple task. Could there be an easier way? Robert Buckley finds out.

## THE VIRTUAL SHORTCUT

**T**he single customer view is a vital step on the way to effective marketing. A cleaned, deduped view of all customer data and transactions provides benefits to both the marketer and the customer: better targeting and less waste during marketing campaigns; better analysis of customer behaviour and trends; and better customer service.

### Real or virtual?

But achieving a single view is by no means easy. Data is often spread throughout a company in various databases, flat files and other sources. Some way has to be found to aggregate the data from these sources. Most companies opt for the traditional approach of ETL, extracting data at regular intervals from the databases, transforming it using a set of pre-established rules, before loading the cleaned data into a central repository where it is merged with the other sources.

Yet some prefer a “virtual” single customer view, extracting information from the different data sets, transforming it and merging it on the fly, arguing

there are many advantages to it, including a real-time view of customers that is completely up to date. So is the single virtual view a viable option for marketing?

Martin Doyle, managing director of DQ Global, says while he doesn't think one approach is categorically better than the other, “if you can leave the data where it is – and there are very good reasons for doing that – it reduces the time necessary to deploy a solution and get marketing up and running. If you can do that, it's very good news.”

It's a viewpoint Steve Clarke, client services director of CDMS, backs up. By taking advantage of the virtual view approach of the Unica system that his company resells, he claims to be able to get companies putting together campaigns within eight weeks. “A lot of that is meetings, working out matching rules and making sure you provide exactly what the client wants,” he says. “We can get it done in days.”

Done correctly, a virtual view should be identical to a single integrated database in its capabilities. Marketers will be able to extract data from the view

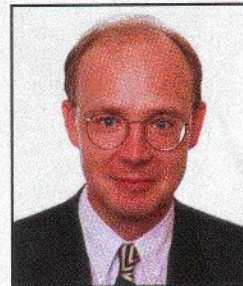
for analysis and campaigns, since single-view systems invariably can export into XML and other standard data formats. So for certain applications, virtual and consolidated approaches will produce similar results.

In fact, there will be some situations where the virtual view might be the only one possible. In many organisations, data intended for use in a single customer-referenced view is often contained in legacy systems or operational systems, such as accounts, call centres and ERP systems. Often it will be impossible to centralise the data from these systems; this might be because of technological shortcomings in the various systems, internal politics or legal issues, such as data protection restrictions. It may even be a deliberate security policy designed to stop a hacker that breaks into one database from having access to all customer data instantaneously.

Tim Pottinger, CRM divisional managing director of Identex, highlights for example, the problems of large merged companies. "They can combine data sets, provided they ask the right question, but often they haven't got control over all their data sets, their partners and legacy systems. It's a big project which may

need more structural types of changes."

Combining the different data sets into a virtual view is not a trivial operation. To link the databases and present this view will require a unique key that allows the systems to match the records in the different databases. This may be a customer number or similar ID. But where data sets refer to different products or have been acquired through merger and acquisition, for example, there may be no such ID.



The individual database owners don't want you mucking around with their records

Terry Hiles, managing director, Capscan.

In these cases, some kind of matching algorithm or tool will be needed to do the work. The matching will typically be based on name and address, or whatever data happens to be available in the source databases.

### Prudential takes its time for single customer views

As a result of mergers, acquisitions, business consolidations and a historically product-centric approach to sales, Prudential UK found itself with 26 policy management systems for its range of pensions and life assurance policies. As well as making marketing difficult, the lack of a single customer view also affected call centre efficiency. Agents often had to keep their customers on hold while they switched between systems searching for the right policies. Customers with more than one policy even had to make separate calls to different units.

Prudential decided to create a hybrid single view, one that would give access to some information from all the databases through a central database put together using ETL, while allowing drill downs into the policy information in particular databases when needed. According to Ally Thomson, lead application architect at PruTech, Prudential's IT services division, one of the company's biggest concerns was performance. "Attempting to do it on the fly, with the variations in formats of the data on the different back-end systems wouldn't have been feasible because of the overheads involved. This way we get the high performance from the static data and use middleware to drill down into the more volatile data. It also fits in quite well with the business processes we want to support as well."

Prudential settled on creating a main database containing customer names, addresses, date of birth, data protection markers, NI numbers and similar data. It used Informatica ETL to extract the data from the disparate databases and ran the Trillium Software System in parallel to standardise the data format, correcting and enhancing data using PAF. Initially, the company processed 15.5 million policy records and 13 million customer records from four key source systems that held policies for pensions, insurance and bonds, going on later to process records from further systems as well. It then compiled

8.8 million unique customer records for loading into its new Oracle-based single view database.

In addition to the standardised data, each record contains policy numbers that act as keys to back-end databases. Whenever a call centre representative needs to drill down for further information, the system provides access to the back-end policy information stored outside the single view.

Prudential's call centres add and update 16,000 records a day, with agents making changes to the single view database directly, which in turn updates customer policies in all the other databases. In addition, each night, the system extracts customer information from the other databases to update the single-view database, repeating the same de-duping and cleaning processes used for the initial database.

Prudential now plans to take its marketing activities back in-house, using the single customer view database as a platform. "Our campaigns are still very much driven from a legacy system we're looking to replace," says Thomson. "The database that drives the marketing campaign is only a fraction of the customer base we want to get. The system we use at the moment is also very inefficient and the marketing department has struggled over the past year."

At the moment, Prudential sends the data from the legacy system to a third-party that cleanses the data and extracts mailing files. But using the legacy system, the time to plan a new campaign is about nine weeks. The new system will use the existing single customer view in conjunction with product data that Prudential will then transfer to a data warehouse. Marketing will exploit this database using Chordiant Marketing Director and SPSS applications to plan campaigns and analyse the data, something that should result in a quicker turnover and efficiencies of £100,000 a year, Thomson calculates.

Once the company has matched the equivalent records, tagging them with corresponding IDs for future matching will reduce overheads and matching time. However, it may not be possible to do this, because of write restrictions in some databases, meaning that matching will need to be done on the fly in the virtual view.

### Master and slave

With the records stored in separate databases, deduping is a harder task since deleting a duplicate will require the permission of the data set's owner, something that may not be granted. There will also be the usual issues associated with deleting or merging duplicates in operational data sets: with a single, separate marketing database, it's possible to delete and merge records incorrectly, usually with only minimal ill effects; in an operational database, these kinds of operations need to be done with absolute certainty.

Even cleaning is likely to be harder. Terry Hiles, managing director of Capscan, points out that while a marketer may want an address stored in one format, "the individual database owners don't want you mucking around with their records, might not want an address stored in that format and won't thank you for deciding that you have the authority and wisdom to dump data from their records."

As well as cleaning the databases at the beginning of the operation there needs to be a way to keep all the databases clean. Says CDMS's Clarke, "If you've got different databases and they're being updated by different means by the call centre, secretaries, those kind of groups, you do need a central cleaning place so the cleaning process is the same for all databases." Without consistent maintenance of all the databases, the virtual single view will begin to lose its authority and cohesiveness.



We probably couldn't use a virtual view for compliance reasons

Colm O'Hara, database manager, EuroDirect.

Cleansing on the fly, in which these activities are done as the system pulls together records for the single view, may be possible, but this can be complex, since it adds additional integration layers on top of the integration already required to pull together the databases. "Effectively, you're creating spiders' webs of complexity and information, that logistically are quite difficult to manage," cautions Ed Wrazen, VP International of Trillium Software. "There are also limitations in terms of the technology being able to provide the performance necessary."

The virtual single view will also generally require a master-slave relationship between the databases, with one database – usually the cleanest and most complete – taking precedence in cleaning operations. This will typically also be the database that marketing has most control over and which is in the best format for marketing operations. When in doubt, the customer invoicing system is usually the best master to pick.

The complexities involved in the virtual view mean that many companies eschew the virtues of real-time insight into customers in

favour of the traditional single marketing database with regular ETL updates from other data sources.

Colm O'Hara, database manager at EuroDirect, says that regulatory and operational reasons prevent his firm, which creates and hosts single customer views for a number of companies, from using virtual views. "We've been able to use strong name and address matching capabilities since we don't use a virtual view. We probably couldn't even use a virtual view for compliance reasons: how could you provide evidence for your decisions in a virtual view?"



**If you can leave the data where it is, it reduces the time necessary to deploy a solution**

Martin Doyle, managing director, DQ Global.

Similarly, Shelagh Register, client services director of CACI, says that it's operationally easier to go for a standard integration approach rather than the virtual view. "The issue with having a virtual view is that because of the hit that obviously makes on central processing, it makes it harder to plan operational work since you're making ad hoc use of the 'juice'. Frankly, the tools available for the front end are so good at holding massive data sets, it's just the same as creating a virtual view anyway."

Similarly, Arthur Kay of Synaxis says that few of his clients are interested in virtual view technology once they see what a single integrated database is capable of doing. "The virtual view shouldn't be seen as a way of achieving a single customer view, not if marketing messages are important to you."

### **Fit for purpose?**

While Clarke claims that Unica's universal data interconnect technology can interface with pretty much any database to create a virtual database, oth-

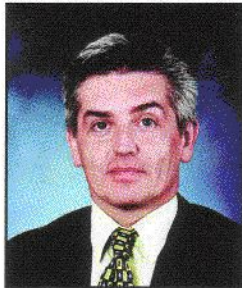
ers are not so sure that a virtual view is so simple.

Phil Good, founder and managing director of Hopewiser, highlights the problems of accessing legacy systems that don't provide standard interfaces and which will require coding to extract data.

"Somewhere like Shell or BP, where the systems have been around 20 years and the guy who programmed them has left, if you ask them to stick a web wrapper round the systems so you can access them, they'll just say, 'I don't think so, mate.'"

Equally, Chris Cuffe, managing director of helpIT Systems, argues that virtual views that pull in data from a number of systems are liable to fail, simply because most big IT implementations fail. This can prove financially catastrophic for many organisations. He also suggests that few marketing applications need real-time data.

"What's the disadvantage to most marketers of having data a week old? If you're looking back three years, what's changed in a week? For the price you have to pay to achieve a virtual view, I can't say it makes a good deal of difference."



**You do need a central cleaning place so the cleaning process is the same for all databases**

Steve Clarke, client services director, CDMS.

The virtual view appears best suited to marketing applications where a small number of systems are involved, when the company needs quick access to reasonably accurate data as quickly as possible and where it's difficult or impossible to create a single integrated database from the various available data sets. For normal marketing activities, its complexity and the increased difficulty of periodic data cleansing looks like it would be a poor choice for many companies. ■