C3B Consulting's

INSIGHT-SPECTRA

Applying common sense to using technology intelligently

www.insight-spectra.com

Volume 23 Report 3; May 2010

Progress progresses

Charles Brett, C3B Consulting

Management introduction

Progress Software has possessed an impressive suite of product capabilities for many years, from the original OpenEdge database that was for so long the core of the company through acquisitions that added messaging (Sonic), enterprise/SOA support (Actional), business event processing (Apama) and more. At the same time, Progress was noted for its client approach, which was above all 'helpful'.

Indeed, many times IT customers have told C3B Consulting (C3BC) that the reason Progress had won their business, against competition as formidable as IBM, Oracle and others, was that it consistently went the extra mile when selling and then delivering and was different in that its people were agreeable to do business with.

Much changes at Progress

The past year has seen considerable change at the company, both in its senior management ranks and in the way that its products are developed and delivered. Founder Joe Alsop has retired (though he retains his Progress investment) and under the leadership of new CEO Rick Reidy there have been major switches of direction. The executive team has been revamped, with some notable heavyweight additions. Gary Conway (previously of PeopleSoft) is now the Chief Marketing officer; Dave Benson (from News Corporation) is CIO and Rob Levy (from BEA and before that CA) has responsibility for all product development.

Yet it is in product orientation where possibly the greatest differences are being made. In the past Progress went to market with separate product silos – OpenEdge, Sonic, Actional, Apama, etc. which each tended to have its own marketing, sales and development. The result was as you might expect: any customer might 'enjoy' multiple visits from different Progress salespeople promoting separate products.

No longer. Today, sales, marketing and product development are organized and operate across the company, rather than in their previous silos.

A consequence of this product aggregation is that Progress has been able to make a significant push into what it refers to as Responsive Process Management (RPM). While we consider this a rather ugly confection of words it enables Progress to bring together:

- the Savvion BPM acquisition (made earlier in 2010)
- the Apama business event processing capabilities
- Actional's business transaction management

as a suite of complementary offerings that can be purchased together or which can be used separately to complement other tools that customers already possesses. Thus a customer that already has bought (say) an event processing engine that was not Apama could still exploit Savvion BPM and/or Actional within their portfolio.

This approach is being repeated, albeit with different manifestations, across the rest of the Progress product portfolio. In so doing the company is simplifying execution, from sales through to development. As significantly, Progress can now cross-fertilize, using its assets in different ways in different places and thereby reducing duplication as well as increasing the potential for complementary exploitation — by customers as well as internally.

Analysis

Rather than try to consider the implications across all of Progress, this analysis will focus on the company's RPM suite which makes sense to us despite the name. Business event processing can stand on its own — as the many implementations of Apama's complex event processing have already proved (see also the case study on page 7, with Royal Dirkzwager's use of Apama to create new business opportunities). Similarly, the SOA and service transaction management capabilities that the Actional products delivered have stood alone while Savvion had established itself as a BPM player. When, however, these are brought together, as has now happened under the new prod-

uct management structure, there is a deeper and greater flexibility that will appeal to many enterprise customers.

In essence, the new Progress RPM suite provides functionality which seeks continuing business improvement by combining three capabilities. RPM:

- provides business users, via a human centric-platform, with real-time visibility into business processes and systems (including legacy processes), whether inside or outside the business
- includes business process modeling and automation; business users can continuously improve existing business processes (or rapidly implement new ones) using modeling as a foundation
- monitors millions of business events and/or identifies patterns as well as anticipating opportunities and threats to which the business can immediately respond (either by triggering an automated response, adjusting an already running business process, launching a new business process or alerting a user to take action).

Does RPM achieve all this? It is still a little early to know precisely. What can be said is that the fundamentals are there — in terms of proven product underpinnings and of management approach.

In addition, Progress has anticipated that complexity could become a constraint. To address this it recently announced its Control Tower product. This is a unified environment that displays real-time alerts, interactive interfaces and tools within RPM, providing users a view into what is happening within their business. To C3BC, Control Tower may yet be the key to success for RPM because its interactive framework can be configured to define and deliver performance indicators as well as business information. Operating within the RPM modeling environment (which means that it can work with both Progress and non-Progress products), Control Tower has the potential to create, model, monitor, control and dynamically improve what is happening within an enterprise. If this is fully achieved, it will be a value booster.

There is, however, another important point to make here. The 'new' Progress is firmly, though not exclusively, oriented towards enterprise size customers. One consequence is that, rather than being able to ignore (and largely be ignored by) the 'big guys' (CA, IBM, Microsoft, Oracle, etc.) it must now directly compete with them and show that it is at the least as good and, in the short term, probably better at addressing what enterprises need.

Management conclusion

So far, so good. C3BC likes what it is hearing from Progress as it moves deeper into the 'enterprise space'. But we also have one reservation. One reason the company won business against heavyweight competition was that it went that extra mile. Joe Alsop is also a gentleman, and Progress as a company ran in his image — its people reflected this quality being 'gentle-people'.

This differentiated Progress and in the quest to grow and to compete with 'the other big boys in the software market' it would — in C3BC's analysis — be harmful if the new management's changes lost its founder's graciousness, and that very specific differentiation.

Charles Brett
C3B Consulting
www.c3bconsulting.com

Is there an ostrich in the house?

Peter Bye, Bye Associates

Management introduction

Organizations of any size increasingly depend on IT services to conduct their business. In some cases it is absolutely critical; banks, airlines and on-line operations such as Google and Amazon cannot operate without (IT). While IT may be less critical in other organizations, the chances are that business functions rapidly become impaired if services are lost.

It therefore seems obvious — common sense even — that organizations would plan for interruptions of IT services, up to and including complete loss of the systems running the business. But do they? Peter Bye sees many ostriches with their heads deep in the sand.

Unplanned service interruptions

The consequences of unplanned service interruption vary from seriously inconvenient to catastrophic, with potentially large sums of money and possibly life at risk, depending on the nature of the business. Such a disaster might be caused by loss of a data center due to power failure or some more dramatic reason, such as a serious fire or natural disaster.

The plans to minimize disruption should, therefore, be commensurate with the financial and other consequences of a lack of IT services. There is no point in spending more on provision for loss of services than it costs to be without them. Emergency services and financial institutions are more 'publicly' exposed than, for example, a medium-sized engineering company.

But, whatever the size of the organization or the nature of its business, some provision should be made. And, depending on the jurisdiction, a number of critical commercial or public sector businesses are required by legislation or regulation to provide disaster recovery (DR) facilities.

Common sense too readily goes out the window

However, for reasons I do not really understand, common sense sometimes seems to go out of the window when it comes to planning service continuity. A surprising number of organizations do not make adequate, or indeed any, provisions for loss of their systems. This is in spite of the fact that officers of commercial companies, or senior management of other organizations, may be personally open to civil or criminal proceedings for the consequences of failure to provide cover in the case of service interruption. Furthermore, damages can be substantial if an organization goes out of business because of loss of its IT capability.

Worryingly, there are, in my experience, far too many organizations facing this level of risk who underestimate what may happen. Threats as dire and diverse as terrorism and natural disasters such as hurricanes, volcanos or floods loom ever larger in the public's consciousness.

Failure to provide for adequate service continuity varies in its level of apparent negligence. In some cases, senior management appears collectively to prefer to leave its collective head in the sand, assuming naïvely that 'it won't happen to us'.

This produces a 'makes no provision at all' approach. While such an attitude is, thankfully, not that common, I have encountered it in a number of cases. The reasons advanced for the lack of provision include excuses such as 'the systems are been very reliable, so we don't need any DR' through to 'we assume our supplier would provide a replacement immediately'. I have even heard both of these comments emerge from a single company (which shall remain anonymous).

The former statement ignores the fact that the systems, while indeed reliable, are not fire- or water-proof, or sadly today, bomb-proof. The latter was made in spite of there being no written agreement between the supplier and its client.

Nor were there any other provisions in place, for example for network switching or for where the

'replacement systems' were to be installed. In fact, the company concerned did not have very stringent requirements for DR as it presumed it could operate for two or three days without its systems. Yet DR provision would have been quite easy to make and its need was made more acute by the fact that if any outage was for longer than two or three days, recovery would have seriously compromised the primary business. (Who would have thought 6 months ago that a volcano in Iceland could bring 80% of Europe's air traffic to a halt for several days?)

A more common fault

Perhaps a more common fault is to make provision for DR but fail to ensure that it would work effectively. A DR center might be set up to contain back-up systems for the primary data center, and perhaps some live production as well. Procedures might be established for recovery in the event of various contingencies up to and including the complete loss of the primary data center. Yet when disaster — or lesser problems — strike, lack of familiarity and practice too often means that much does not work out as originally planned.

An example is the case of the failure of an HBOS (Halifax and Bank of Scotland) data center on Saturday 14th November 2009. I was listening to BBC Radio 4's Money Box programme when the failure made the headlines (see http://news.bbc.co.uk/1/hi/uk/8360313.stm). There it was reported that a power failure at one of the HBOS data centers early in the morning had caused a major interruption of IT services: no ATMs, no over the counter services were available; even appointments could not be handled or canceled (as the required client information was in one of the lost systems). The story continued over the weekend and was widely reported; recovery took hours.

It is not clear why it took so long to restore services. Evidently, uninterruptible power supply equipment did not work immediately or systems would have been recovered more rapidly in situ. The data center itself was not damaged as it would have been by a fire, for instance. And, given that HBOS has more than one data center:

- why were services not restored within one of those other (DR) centers?
- was there critical equipment only in one data center?
- were the procedures for handling DR sufficiently documented and automated?

Whatever the reason, it is clear that improvements in DR handling were needed.

Outsourcing does not solve everything

A data center hosting outsourced systems for a number of clients is another example of a catastrophe in waiting. Happily in this case, as I will discuss, it was resolved before disaster struck.

A DR center had been established some hundreds of kilometers away from the live center, a provision that was important given that the primary data center was in an earthquake zone. The primary data center had a variety of systems supporting multiple clients (and bought from different suppliers and using different technologies). The number of systems, and the fact that different customers had different service-level agreements, meant that DR for this outsourcer was complex.

A DR process to manage a switch to the DR center had been established and documented. However, the process was complicated -- taking over 100 pages to describe and requiring 2-3 days to execute by a small, skilled team. In effect, the process was useless and almost guaranteed not to work. The level of complexity was such that there was no chance of executing it without errors. And, given the two to three days required to execute, practising DR was impossible as the data center could never be 'off air' for long enough.

Fortunately for this outsourcer's clients, someone understood what could happen. A project to automate the process reduced the time to execute to less than 30 minutes, with just one operator required. It is now practised regularly.

Management conclusion

I want to finish with three recommendations about DR, which seem to me to be about applying common sense. They are:

- *make a provision for DR commensurate with the exposure if systems are unavailable*
- *apply a high level of automation to the DR process*

- *practise the process regularly to make sure it works.*

The first point seems blindingly obvious, although not apparently to everyone. However reliable systems are they can still go wrong. Much more likely, the environment they live in might fail, for reasons totally outside the control of the data center's owner; floods and hurricanes are examples. It therefore makes sense to avoid putting data centers in known exposed places and also to ensure that two centers should not be affected by the same disaster: don't put them on the same flood plain or earthquake fault line, for instance.

The second point is less obvious but is nevertheless critical. The need to execute a DR process does not often arise -- but when it does, it is serious. Even in relatively straightforward environments it can be complicated, never mind the data centers in the examples discussed above. Complexity and lack of frequency of use means that operators executing manual procedures will make mistakes; they are doing something rare and under pressure. (I know of one case where the operators started the script on page 2.) Automation is essential to minimize the risk of error. Ideally, all that should be required is to click on the icon labelled 'Execute DR' (perhaps with an 'Are you sure?' check-prompt); the rest of the process should then be automated.

It is more difficult to automate the decision to execute DR. Anything requiring human intervention requires some time, which must be added to the execution time when determining the total duration of service interruption. In critical systems, the decision can be automated. In really sophisticated environments, an application can be distributed over more than one system in different centers, raising availability to close to 100% (Figure 1 shows a schematic of such a configuration). Deciding on the level of sophistication required is a business decision, based on the potential financial and other exposure caused by system loss.

DR processes must be practised to ensure they work as intended. Regulations may require DR to be exercised periodically but, even if there are no rules or legal requirements, all organization should still go through the process regularly. Such rehearsals should also exercise the return to the primary system as well, for often this can be as or even more difficult.

Net, net: do not suffer ostriches lightly. If their heads are in the sand, it may cost you and your organization dear.

Peter Bye
Bye Associates
peter.g.bye@btinternet.com

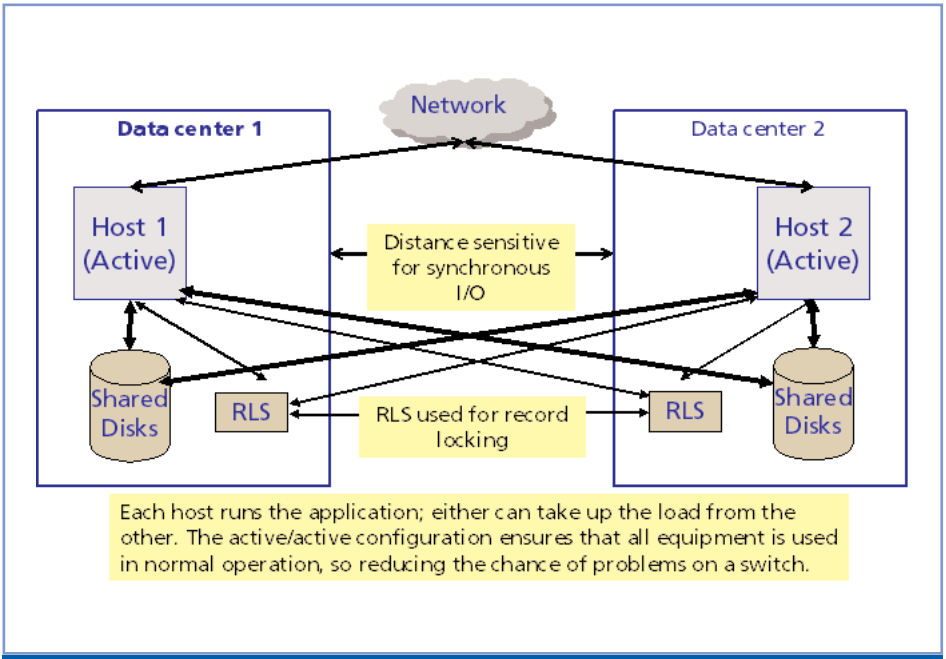

Figure 1: Simplified schematic for near 100% availability

SAP transaction losers — rejoice!

Charles Brett, C3B Consulting

Management introduction

Last summer C3B Consulting commented about the difficulties of finding 'lost transactions' within SAP's hugely successful enterprise application software. These are transactions that have gone AWOL in that they cannot be found but are 'somewhere' in an SAP system. If only one knew where to look.

This analysis considers one solution, from Holland. InterfaceHandler is the product and its source is Frent IT B.V. SAP users facing lost (or AWOL) transactions should definitely look deeper.

The problem

Whether for a bank or a manufacturing company or a government agency, when transactions go missing there is a common cost when trying to find what is 'missing' that comes in three forms:

- the immediate cost associated with (say) an order not going through
- the lost opportunity cost that can come with an unhappy customer
- the direct cost of paying for the skilled assistance to solve the problem.

The scale of the challenge is straightforward to describe in actual terms. A division at one major international SAP user has some 30 SAP instances that require 50+ SAP-skilled people to keep everything operating: the price of retaining such expertise — which is what is needed to find transactions when they go 'missing' — costs that organization nearly $10M/year. Last summer we commented that we found it curious that, with tens of thousands of SAP implementations installed worldwide, such an opportunity had been neither addressed by SAP itself (whose tools in this area remain rudimentary) nor by a third party (given the potential market size).

At the time we also said we expected some measure of relief to arrive via a startup called Conpertus which was launching a product called 'Periscope for SAP' . Unfortunately, the bad news for SAP users was that a threat of legal proceedings against one individual participant meant that, within weeks, Periscope for SAP was no longer available from Conpertus (and Conpertus disappeared).

The good news for SAP users today is that the technology still exists, and had never been transferred to Conpertus. Indeed, after the demise of Conpertus (which resolved all the legal issues), the developers took the opportunity to develop further their technology and to make key improvements.

Enter InterfaceHandler

From February (2010) InterfaceHandler (q.v. www.interfacehandler.com) became available from Frent IT BV, based in the Netherlands. The CEO is Fokko Jan Reinders and in discussions with Fokko Jan he made the following points:

- the time since last summer was used to extend InterfaceHandler, particularly to support Windows 7 in both 32 and 64 bit implementations
- MQSeries and TIBCO messaging support further extend InterfaceHandler's connectivity capabilities
- there is support for both SAP XI and R/3
- much work has been completed in improving the documentation
- installation is simpler and faster.

He also mentioned that InterfaceHandler is already in use at several test customers and that their responses are positive. In particular, InterfaceHandler's ability to free up highly skilled SAP staff, by simplifying the discovery of the 'missing transactions', is regarded as a significant plus — because such people are so valuable.

Analysis

One of the key attractions to C3BC of InterfaceHandler is that it is non-intrusive. Essentially you install it on a PC attached to your SAP 'network' and then, using techniques derived from automated discovery, it identifies the path of a transaction through one or

many SAP instances. From this discovery the application builds a map for the selected business transaction.

Once you have the map, then you know where to look for 'errant transactions' which, once found, may be and repaired and recovered. Transactions (and messages) can be found by IDOC / RFC number or XI/PI GUID. InterfaceHandler enables the details of every step of the transaction message flow to become visible.

InterfaceHandler achieves this by working from the bottom-up rather than from the more traditional top-down associated with work flow monitoring. Using Frent IT's developers' deep knowledge of SAP, InterfaceHandler examines the metadata and linkages associated with each designated transaction to build a picture through SAP's application modules of what happens to the data in the transaction as it traverses SAP's myriad of rules. In so doing it identifies the interfaces used and captures the status of all the components. It can even show the transaction flow performance and measure this to see if the throughput thresholds satisfy service level agreements.

Other considerations

There is, to C3BC, another compelling advantage to Frent IT's non-invasive approach — SAP Certification. Many SAP users value this certification, which is a process that SAP uses to ensure that a third party offering does not prejudice the company's own applications. As might be expected, SAP Certification can be expensive and is difficult to achieve because of the testing required. However, because Interface Handler is non-invasive it does not change anything in any part of SAP's applications. According to FrentIT, InterfaceHandler's SAP Certification is imminent.

Furthermore, to be expert today in software as complex as SAP's offerings requires expensive skills. Increasingly SAP user organizations need automated tools to find out ('discover') what is happening within and across their SAP applications. In C3BC's analysis there will be two stages in the adoption of a tool like InterfaceHandler:

- the first will focus on immediate problem solving — locating 'missing transactions; this will produce significant day-to-day savings (whether measured in costs avoided or people skills released or both)
- the second will prove be even more valuable — it will rest in building an automated knowledge base of how business transactions function within SAP; this will have longer tern benefits.

Management conclusion

While Frent IT has applied its knowledge of SAP to a specific difficulty, C3B Consulting expects tools similar to InterfaceHandler to emerge — because enterprises simply cannot afford the costs of missing (but perhaps not lost) transactions or events.

Meanwhile, from February 2010, SAP users who are 'losing' transactions should rejoice — at the prospect of relief.

Charles Brett
C3B Consulting
www.c3bconsulting.com

Value-based case study #7

Combining technologies opens up whole new business opportunities at Royal Dirkzwager

Management introduction

Royal Dirkzwager, now more than 130 years old, serves the Port of Rotterdam. It provides maritime information services to all those associated with shipping (owners, charterers, agents, harbor masters, etc.), particularly about when ships will arrive, where they will berth and when they will leave, plus information about these ships (cargo types, tonnage, positions, ETAs and ETDs, etc.). Rotterdam is the largest port in Europe, with a location open not only to the North Sea and the Channel/Atlantic but also inland, via the River Maas, to the Rhine basin in Germany and beyond. Being able to schedule and thereby control shipping movements is critical: otherwise vessels have to anchor and wait for a berth — which costs everyone in time as well as money.

Paul Wieland is the Manager of the Maritime and Logistics ICT Department in Royal Dirkzwager. He joined some 2 years ago (before he had worked for 17 years in ICT in the aviation industry). As he describes in this case study, Royal Dirkzwager is taking advantage of 3 key technologies – AIS, an enterprise service bus (ESB) and event processing. In so doing Royal Dirkzwager is aiming dramatically to improve the efficiency of the Port of Rotterdam while simultaneously extending Royal Dirkzwager's information reach from being local (focused on Rotterdam) to becoming global.

Indeed, what Mr. Wieland describes is nothing less than establishing Royal Dirkzwager as a new type of information utility. Albeit founded on its long established Rotterdam-based business, it is now adding new capabilities and opportunities because it selected technologies to enable this.

The business issues

Until 10 years ago Royal Dirkzwager was locally oriented, around what was needed for the operation the Port of Rotterdam. The processes we used were immensely manual and time consuming, with little automation or interconnection. They involved large numbers of telexes, telephone calls and even voice radio traffic between agents, ships, the Port and our selves. Although shipping movements were far fewer in number than today, the people involvement dimension was enormous.

Since then we have seen two major technological changes. We started to interconnect our systems with other companies (like the shipping companies, for example) so that data can flow system to system without manual intervention. In addition, the mandatory introduction of the Automatic Identification System (AIS) for ships now produces position data for each ship every few seconds

Initially AIS, and its use of the Global Positioning System or GPS, was based on transponders on each ship using VHF (with a limited range) to communicate data. This made sense when the original purpose of AIS was for safety — collision avoidance: ships broadcast their positional data (including speed and course) which other ships within VHF range received and, using this AIS data, the ships could compute courses so that collisions did not occur.

This data was, however, also useful for ports like Rotterdam: if they could receive the same positional information within VHF range (usually up to about 50Kms) and on what courses, with speeds, the ships were proceeding they could anticipate arrivals earlier and more accurately than before. Estimated times of arrival could now automatically be calculated at a greater distance from Rotterdam, without the immense manual effort that had been necessary before.

Royal Dirkzwager went further. It realized that this positional information would be even more useful if the data could be acquired still further away from Rotterdam. It proceeded to add receiver stations around the coasts of Western Europe (for example at the entrance to the Channel) so that these receiver stations would notify Royal Dirkzwager's systems as a soon as a ship entered European waters. In effect we, via these stations, could now assemble a 'picture' of all ships in European waters within VHF range of one of our receiver stations, including where they were, at what speed were they going, on what course, etc.

Add to this that each AIS transponder is associated with details of 'its' host ship (tonnage, vessel type – container, tanker, bulk cargo, liner, cruise ship, etc – plus other relevant data like the presence of dangerous cargos) and Royal Dirkzwager could generate as complete a view of shipping around Europe as was practical given the VHF range limitations.

Satellites add the global dimension

This worked well and was integrated into the Royal Dirkzwager systems, which in turn could now supply improved information to agents, owners, harbor masters and other customers. But it was still restricted to European waters within VHF range.

Yet our customers, who buy our information, wanted more. They wanted to extend their planning horizons.

Though 'coastal AIS' was a great improvement over what had been available previously on the primarily manual basis, customers wanted to 'see' further out. If they possessed this information they could improve utilization and thereby profitability (if a berth is going to be occupied, it often makes more sense to reduce a ship's speed so that she arrives when the berth is available rather than to steam normally and then have to anchor). The pressure was on us to extend what we had been doing.

With the arrival of satellite-based AIS, AIS — which uses satellites to pick up the AIS signals and relay the data from wherever in the world — more changes became possible. While Royal Dirkzwager had partnered with other coast based receiver stations around the world to share data, this only worked if ships were within VHF range of a coast-based receiver station. With satellite relay, a ship's movements can be captured anywhere in the world that AIS-receiving satellites cover, even in mid-ocean.

As AIS became global so Royal Dirkzwager was presented with the opportunity to move from its new Western European waters-orientation to being global. In so doing it could become the shipping information source for all those who wish to subscribe.

Let me give you an example. A ship might leave (say) Valparaiso in Chile. Using satellite AIS we would see when she cleared port. She would then (say) head for the Panama Canal. When she arrived in the Gulf of Panama she would likely have to wait her turn before entering the first locks of the Canal; again we would see this and as she progressed through the Canal. When she eventually emerged at the northern (Caribbean) end, we would see that normal steaming speed had resumed. Now we can calculate and then anticipate when she will enter European waters as well as know to what port she is headed.

Furthermore, using the satellite capability, even when on the high seas (in mid-Atlantic, definitely out of VHF range), we can see her position course and speed. Thus both owner and port can constantly update for when she will arrive, make sure a berth will be available and schedule the unloading. Owners can then advise their customers, or trans-shippers (for a barge going up the Rhine or onto a train going elsewhere) when cargos will be ready for collection.

Volumes increase by at least order of magnitude

Going global has its own ramifications. Before satellite AIS, we tracked around 5000 ships at any one moment. By going global this number rises to 50,000+, a ten-fold increase. This increase meant that we had to do some major rethinking and retooling of our systems if we were to be able take advantage of the opportunities presented.

The first stage in addressing this was to contract with LuxSpace. LuxSpace provides us with an AIS data feed from all the ships that are sending out data. We then have to process the vast amount of resulting data in order to make it useful.

To handle this we had previously implemented an Enterprise Services Bus (ESB) as the mechanism to connect our own systems together and also to make the connections with the systems of our customers. Whenever a message is originated on one system for another one (whether it comes from one of our own applications or from a customer), this message enters our ESB (we had implemented the Sonic ESB from Progress Software) and then this is routed to the appropriate destination according to the logic running on the ESB.

The Sonic ESB was already, therefore, a critical element for interconnecting applications, both inside Royal Dirkzwager and also with customers and data suppliers. We had proven its relevance before acceptance by interconnecting our systems with the Port of Rotterdam systems, ensuring that all worked. Subse-

quently it has become an essential component of our ICT infrastructure because it 'switches' data between services and applications (whether internal or external). It also possessed 'always-on' capabilities; we can even introduce maintenance updates and changes without having to take the ESB down (a vital consideration for us).

The introduction of AIS expanded our needs, especially when the global dimension (using the satellite information) arrived. We contracted with LuxSpace to send us a data feed with the global AIS data. But this needed to be processed — the AIS position 'events' have to be analyzed and systems updated. The ESB was not the place for this.

In talking with Progress we learned about its Apama event processing engine. This can handle huge amounts of data per second (as it does in financial centers). By building rules we decided could use Apama to analyze the AIS data stream and use Apama to create alerts twice a day (or more often nearer to destination ports) as to where a ship is located plus we could define boundaries or areas so that when a vessel crossed a boundary (say a line between the Lizard and Ushant for entry into the Western Channel) a specific type of alert would be issued.

In effect we have the LuxSpace data feed coming into the Apama engine. This runs the rules looking for data that satisfies the rules and when rule conditions are satisfied, an alert is passed to the ESB. The ESB then routes that event, with associated data (like position and speed) to all those applications that need this data.

This combination of the event processing engine and the ESB are what gives us both flexibility and automation. To give some measure of what this means we were handling about 500 event-messages/second (which is a tiny proportion of Apama's throughput capability). We expect to go 5000 event-messages/second as we embrace the global dimension.

Personalizing applications and CruiseSMS

In addition we can 'personalize' the event analysis. For example the owner of a fleet of ships might want to know where each one is and what she is doing on (say) an every six hour basis for the fleet. By writing owner-specific rules we can assemble what a customer wants and, via the ESB, communicate this to that fleet owner. This, in turn, can be used as an input to trigger business processes. We now provide this for the Port of Rotterdam; we track ships destined for Rotterdam.

We have also created a consumer product called CruiseSMS. In the latter case whenever a named cruise ship is approaching (say) Rotterdam, we will generate arrival information that is sent by SMS to those who subscribe. Currently this is aimed at friends or relatives who may want to know when to pick up passengers from a cruise ship or go to the harbor to take photos, etc. But you can imagine how usage could expand around (say) the Caribbean Islands: holiday and tour agents as well as other professionals could subscribe so as to know when they need to be ready for specific cruises. While this is only a modest application at present, we believe it has real potential to generate new revenues for us in the longer term.

There are innumerable other possibilities. Consider ships carrying hazardous cargo. The movements of these are restricted (you may recall that a Chinese ship carrying coal 'took a wrong course' into the restricted area of Australia's Great Barrier Reef). We can define areas and, knowing about hazardous cargos, we can not only alert the authorities that limits may have been breached but even inform the ship's crew that her course is not appropriate and advise them how to remedy their position.

Similarly, safety applications exist. Using Apama we can create a rule that triggers an alert when Apama detects that a ship is 'there' one moment and then stops transmitting for a predetermined period; if this happens we can generate an alert to the search and rescue authorities as well as give a position that might only be minutes (or seconds) old.

I should add that, with the exception of the Port of Rotterdam and CruiseSMS applications, we are only starting to exploit.the capabilities I describe above But our experience so far indicates that all these capabilities are made possible by the combination of technologies embracing satellite AIS, Apama event analysis and alert distribution by an ESB.

Lessons learned and best practices

My first lesson learned — and I would add this is a required and necessary 'best practice — is that it is really important to choose technologies that are close

to the fundamentals of the business you are in or want to be in. Too much separation or difference and you will find yourself 'bending technology to fit your needs'; with software this is almost always possible — but at what cost (in time and money)? With the Sonic ESB first and then Apama's event processing we found technologies that matched our business. In addition, by buying from Progress we obtained a professional software supplier that was able to respond to specific needs and willing to use its experience.

My second best practice reflect this: look for outside support early if you are not certain. Let me give one instance of the value here by talking about what happened to us. We have to run both the AIS event analysis and the ESB 24x7; we cannot afford to have 2 hours downtime for updates during which we cannot detect vessels' positions. This means that we need to be able to install maintenance and application updates when the systems are running. This is not trivial if you do not know how to do it. We found that using Progress' existing knowledge and experience with its other customers made this much simpler for us than having to work it out for ourselves. The result is stability, obtained much faster than if we had tried to do everything ourselves.

Finally, when you introduce new technology, even if it does seem to fit the business need well, start slowly and carefully. As important, do not attempt to build or deploy mission critical elements until you have proved that the technologies do what you expect (and want). Be prepared to deploy simpler or less vital applications to start with, to build internal and external credibility. This is why we introduced CruiseSMS; it was unlikely to create a business problem if it did not work perfectly at the start (though it did). Starting with (say) looking for environmental infringements would have been higher profile; but what would have happened if we missed the first oil spill, or coal ship entering prohibited waters? Once we had established our credibility we could move onto our more vital business activities with confidence.

Management conclusion

Royal Dirkzwager has always been an information gatherer and provider, primarily to those associated with the Port of Rotterdam. The introduction of the three key technologies — ESB, AIS and event processing — is enabling the company to evolve from being a local, Dutch, information provider to becoming a global information provider. This is a major expansion of focus and capability.

The future for Royal Dirkzwager now offers many more possibilities than the past. All that movement data (5000 message events/second) allied to the data about ships and their cargos will enable Royal Dirkzwager to expand its service offerings. No longer need it focus solely on port movements in Rotterdam (although this will remain a core competence) but it can look to consumer revenues (as in CruiseSMS), enforcement (identifying ships contravening regulations), safety (spotting ships possibly in distress by their pattern of movement) and much more. In effect, new technologies have created a whole new raft of opportunities for Royal Dirkzwager.

'Information fusion' considerations for customers

Nick Denning, Diegesis

Management introduction

It is important for organizations to understand the data that they hold, to exploit that data as an asset yet also be aware of the liabilities and obligations associated with holding that data. Fines for organizations that are subject to regulation are regularly reported in the press when obligations to follow processes have either not been met or the evidence for compliance has not been maintained.

In this first part of a two part analysis, Nick Denning considers Meta Data Management (MDM) as well as whether information fusion can work for, and be relevant to, customers. He starts by looking at what can happen.

There are many ways to fail ...

There are, unfortunately, an ever increasing number of ways in which an organization can find itself in trouble:

- failing to observe regulations and or due processes
- not providing evidence that compliance has been observed
- finding that the process of legally authorized discovery and the associated forensic analysis locates 'embarrassing information'
- falling foul of freedom of information legislation, that requires an organization to be able to locate information that it holds and provide it in response to enquiries; this can open doors that were thought to be closed.

In the UK, recent changes to the Data Protection Act place additional obligations on organizations to ensure that data held is both accurate and secure. This matters when valuable information exists in many databases which can provide a customer profile, better understanding of a customer and the opportunity to identify future customer needs and cross sell products from different parts of the business to that customer. It can also, if not managed appropriately, open access to 'connected data' that might not otherwise be accessible.

What needs to be done ...

Improved customer management should enable organizations to own their customers and protect their business in preference to individual sales people or even agents. Increasingly, organizations must manage and report the key risks to the business and avoid misrepresentation of material facts. To focus on delivery, failures can produce personal liabilities on directors and managers.

Organizations, therefore, face increasing pressures to deploy an effective 'fusion of information' management approach which enables them to:

- manage, locate and exploit for business advantage all information within an organization
- monitor activity in the business to ensure that due processes are being followed; this should also detect deviations from processes as well as document compliance reporting obligations and management actions which address weaknesses
- analyze information so that it can be used to improve risk management and decision making, to support analysts searching and to generate alerts and warnings for managers
- identify information which can legitimately be deleted by information weeding (thereby reducing the risk of exposure to subsequent discovery processes and minimizing the cost of electronic storage of information)
- store additional information that might create new business advantages.

Historically the principle of 'proportionality' enabled organizations to avoid actions which were possible but where the cost was out of proportion to the benefit of interested parties. As product based solutions become available the costs of implementation are dropping. One little commented upon consequence is that many actions previously considered disproportionate (in cost) can now be argued to be both reasonable and proportionate — obliging previously avoidable actions to be undertaken.

Consider then some of the problems in providing the necessary capabilities. These can be considered under the following headings:

- search and analysis
- data warehousing
- Master Data Management (MDM)
- data extraction
- data quality improvement
- Service Oriented Architectures (SOA)
- text analytics
- process records
- security
- discovery.

Search and analysis

Many organizations have large numbers of databases, particularly where the organization has built stove pipe systems to support each functional area of the business or where acquisitions have occurred. Searching for all information across many different databases to support a decision can take significant time and effort.

For example, one customer — with only six databases — reported that two members of staff regularly took three elapsed days to prepare a simple task specification. Major improvements in productivity resulted once users were able to perform a single query and search across all information in the corporation in parallel (subject to security and legal constraints).

A possible approach is to migrate legacy systems into a common database, against which all systems work. For many years the goal of a 'Common Customer Object' has been put forward, though few have made it this far. Recent proposals from many organizations prefer the creation of common web services.

However, consolidating systems to use a central common database may not be ideal. Firstly the effort involved in re-engineering databases to merge data models can be significant. Indeed the data in old systems may not contain the attributes necessary to populate a new one; in addition the data may not be sufficiently clean. Thus a consolidation process can involve substantial cost and effort.

If an organization then sells a business unit, the effort to separate the data and systems from the parent can be still substantial. Furthermore, if the legacy system is servicing existing contracts but is taking no new business, there may be little justification for the potential cost involved.

Data warehousing

The techniques for building a data warehouse are mature. We can use ETL technologies to extract, transform and load data from each source systems into a central data warehouse.

This is not cheap, and needs services effort, specialist products for ETL, warehouse databases and business intelligence. Typically a subset of data is loaded into the warehouse and, where data conflicts, a mechanism is put in place to select the most appropriate value into the warehouse data model, usually by discarding some information.

Building a warehouse is relatively non-disruptive for an organization, and has a low impact on existing systems. It does not, however, give a complete picture of an organization's data. That lack of completeness can be significant, as I will describe in more detail later in this analysis.

Master Data Management

Master data management products enable the creation of a master record in a central core database along with a mapping between this master record and related records in each underlying database. To deliver this, many of the techniques used by ETL processing (for data cleansing and transformation) are applied to normalize data prior to the fuzzy matching by which records can be grouped and linked to an associated record (later on, groups can be merged or split where subsequent information has identified errors in the grouping process).

The key aspect of the core database and its references is that organizations can now query the core database and then reach back into the original records. In this way they can build a complete profile of understanding of customer — with the key word being 'complete'.

Data extraction

It may be of benefit to use a data weeding capability to remove 'old' data related to completed transactions and old entities. If old data is not weeded out this can adversely impact system performance, extend

backup periods and incur increased data storage costs (thereby also providing a justification for information weeding).

However, while entities (such as customers) may no longer be active in one application or system, they may be active in others. It may well be of value to maintain customer history across all systems in order to be able to understand fully the behavior and profile for marketing purposes and other cross selling opportunities. A customer may be buying recurring products enabling that customer to be targeted with an offer when the current agreement with another supplier terminates (enabling them to be re-acquired).

It also may not be possible to 'discard' data if there are on-going potential liabilities associated with a particular customer or other obligations to maintain financial records. Consequently, while organizations may wish to extract data relating to entities which are currently inactive (often to minimize platform costs) it remains necessary to maintain data in some form of store which remains within analysis capability.

For example, when buying or selling companies, the target company (or its assets) may be acquired with or without historic liabilities. In some cases the acquired company may have been sold without historic liabilities. The capability to maintain a copy of all relevant data, while only transferring to the purchaser that data related to the business when sold may enhance the competitive position and enable risks to be better managed.

Data quality improvement

Once an MDM capability has been implemented it is possible to identify data in underlying systems where values are not consistent across all systems, for instance an address. Having detected such differences the potential is created to implement processes to 'fix up' these fields across the various underlying systems.

Such a process may be automatic but it could also involve contacting the customer. This additional customer contact may provide more sales opportunities, or alternatively may help to verify the MDM rules that have merged the customer details to produce the correct result. The benefits of lower administration costs in the future are significant — provided the grouping of customer records is correctly completed.

Services Oriented Architectures

Earlier I identified some of the dangers in modifying legacy applications to exploit central web services, because of the cost of re-engineering a solution to remove data from the legacy database and replace it with service calls to central services. An MDM capability can provide the basis for common read-only services querying the master data record while the service remains even if the underlying systems are changed. The update logic in each system is unaltered as the MDM capabilities propagate any changes into the central service.

From now, each time there is customer contact, the MDM profile of the customer can be interrogated and any other aspects of customer information which may be in doubt verified. Furthermore, when bringing on a new customer, MDM records can be interrogated to verify customer details and history, speeding up the process and enabling additional risk management based on a customer's history of issues — such as credit rating and other costs previously related to servicing the customer. In addition, if any changes to customer details are propagated to other systems, this further improves data quality across the enterprise.

If a business is to be sold the effort to disconnect the systems that support the business are minimized. The system (or a copy of it) including the MDM capability can be provided with the MDM instance — only now connected to those systems included in the disposal (once records in MDM database related to disconnected systems have been purged).

Text analytics

Many suggest that 80% of corporate data is held not in a structured format (like a database) but in some unstructured form. Text is the most obvious example, but images, voice and other data are relevant.

Specialized search products exist to connect to unstructured data sources — such as files in directories on disk, email systems, instant messaging, content management systems, web pages, pictures, video and voice. Mechanisms for the extraction of raw text, analysis and mining of information from that text to create and store facts are well understood, allowing the location of entities of interest and identification of relationships between entities.

The analytical capabilities are increasing in sophistica-

tion in this area. It is possible to analyze the sentiment of a message, categorize the type of document within which an entity is located, parse text to determine the meaning of sentences and identify more information concerning the context of the relationship between two or more entities. It is also possible, using specialist engines, to analyze a range of other non-text data objects such as:

- photographs to perform face recognition and generate meta data describing a person detected in a photo
- car registration with identification coming from the number plates in video, after which owner details can be retrieved from external database
- speech, which can be analysed and converted for text analysis and even for biometric voice pattern identification.

Using computational linguistics the meaning of ambiguous terms can be established based on the context within which they are used in each case. For example the context can indicate whether the word 'June' means a girl's name or the name of the month.

There is, though, the challenge of uniquely identifying entities by name from within unstructured information. Names are seldom guaranteed to be unique identifiers. An entity name is typically a poor mechanism for unique identification (as ETL and MDM technologies have shown). The same person may be referred to by many different 'descriptions' — full name, forename and surname, forename only, nickname or abbreviation, role or title, alias and so on.

However, if text analytics can extract other attributes of the entities from the context within which the entity is detected then there is a significant chance that, by matching an entity against an overall profile maintained within a structured database or better still an MDM master record, it becomes possible to narrow down the possible candidates for an entity so that even if an entity cannot be uniquely identified then the references can be passed to an analyst to make the assessment or determine the match. Thus, if one can link entities such as people in unstructured information effectively to the people records in master databases, organizations have the capacity to fuse data into a profile for an entity. This potentially produces a greatly enhanced understanding of an organization's interactions.

From this it is then possible to determine the most appropriate and profitable actions for an organization to take in order to exploit its data and deliver business goals.

Process records

If internal processes require that every person participating in an endeavor records the actions that they are taking, then an organization has the ability to analyze internal business processes to ensure that what should be being done is being done. As significantly, this can be used to detect when deviations from processes occur. Records in email or instant messaging can even provide the basis for applying text analytics and thereby process analysis.

This capability can be enhanced still further if more structure can be applied by using ad-hoc work flow products, in which staff define tasks and link information used when carrying out each task in a process. Such capabilities also enable all staff participating in a process to share a common view of the data used in that process, to record historic information from a library of previous processes, improve the corporate memory of the organization and learn from previous experience.

Security

Any mission critical system has to ensure the '-ities' are implemented -- security, reliability, availability and so on. Of these probably the most important is security. Implementing a security model is well understood though not without its challenges and costs.

In the information fusion arena there are particular challenges. Items of information that are by themselves unclassified can, if connected, create classified information that must be protected. For example, the names of each of the US or UK's nuclear deterrent submarines are unclassified. There is also nothing inherently classified about a grid reference or a time in the future. But the result of associating the three together may be highly classified.

Mechanisms for implementing security access control within a relational database are well established with the ability to label items of data, typically records in a database table, to grant or deny access to specific items by specific attributes or using hierarchical levels of security. In contrast, when dealing with unstructured information, it is normal to consider each data

item as an atomic component and hence apply similar mechanisms as those used for relational databases to document management systems.

Using single sign-on and propagating the credentials of the logged-in user through search tools ensures that we can comply with access controls on each type of data. Also important are security mechanisms — such as evaluation of access devices to ensure data retrieved from a repository is protected in transit and cannot be stored in an unprotected environment or accessed by unauthorized individuals.

Discovery

An important aspect of any information management solution is the ability to be able to locate all the computers that hold information. In small organizations that should be straightforward but in large organizations this is more complex.

Software products are available which enable a system management team to scan the network for machines and then identify those which are running software with the characteristics which suggest they are holding data of interest. For instance, RDBMS products often use common standard port addresses. It is more difficult to identify systems holding unstructured information, such as a file server or even desktop.

However, increasingly organizations have been configuring routers to decompose the infrastructure into separate domains to prevent unauthorized access. It may be more difficult now for discovery products to scan the internal infrastructure.

Remind me — why are we considering all this?

The justification for improved information management is that this enables an organization both to:

- exploit for itself the data it possesses
- satisfy legal and regulatory obligations.

Business results are delivered though excellent performance of the right actions. Selecting the best actions to take depends on effective decision making based on locating good information that is accurate, relevant to the context of the decision and complete and provided to trained people who can understand it — people, process and information.

Having made a decision the next need is to execute the process to implement the decision. This will generally require participants to access the information used to take the decision, adding to it as the process progresses. Therefore organizations need to have the tools which will enable them to locate the data required, to process it to generate information, to organize that information to make a decision and then to communicate the 'information package' to those that need to use it.

If a team propose a 'big' decision, the basis for the decision will likely need to be escalated up the decision making hierarchy for authorization. It must also be presented in a form which is relevant to each person who needs to participate in the decision.

Having taken a decision and successfully delivered the outcome an organization will want to exploit this success by being able in the future to recall this success, at least repeat and hopefully improve its performance in the future. It might even wish to incorporate the approach into its processes -- thereby becoming progressively better at doing it, reducing the cost of doing it and generating more profit until competitors catch up reducing the profit of this activity. Meanwhile it may be necessary to protect intellectual property of the approach to maintain a competitive advantage such as by patent or as a trade secret.

Organizations, therefore, benefit from recording the steps taken to achieve success, so that each organization can own the process and train others to exploit and develop this further, while keeping this information confidential. While email might hold this information, subsequent exploitation may be challenging.

Use of process tools which record the steps taken and the information used to take them can provide records which can subsequently be analyzed and potentially re-used. It will also enable a process to be broken down into sub-processes, with each person only having access to those sub-processes where they participate (limiting the number of people that have visibility of the overall business process and hence protecting that organization's intellectual property).

Information objectives

To make sense of all this, organizations need specific and predictable items of information to support well defined processes. A bigger challenge is to identify the information that we require to support the deci-

sion making that underpins change. In information management terms, the key capabilities are to:

- record the thinking process by which an analyst was able to locate, assess, categorize and filter the information
- identify the options considered, the resources needed for each option, the risks associated with each option and the benefits likely to be obtained
- document the justification for the option selected, the contingency plans for addressing risks should they arise and the mitigation effort to minimize the impact of any risks that materialize
- identify future checkpoints in the plan at which provisional decisions will be confirmed and how contingency plans will be implemented if the situation changes during the course of a project
- be able to monitor changes to the information upon which any decision was reached and determine whether such changes impact the completion of the plan
- catalog assets of the organization and in particular those being developed during the plan so that these assets can be exploited effectively (or work on them canceled in the event that the plan changes)
- maintain links between each element of information and to be able to backtrack and adjust the plan exploiting the work to create the original plan, rather than tearing up the plan and restarting.

Management conclusion

As discussed above, there are many tools and approaches now available to create an 'information fusion' picture which in turn creates a vision of conceptual access to information. Search, query, correlation, reporting and presentation tools provide the capability to navigate the information of required, then present it in the most appropriate format for use to be made of what has been determined.

To use the data, the information fusion analyst needs to be able to store the thought process and querying steps which describe how the target data set was created so that the mechanism for generating the product can be audited and reviewed to ensure confidence in the output, can be re-run to re-generate the product periodically and the decisions taken (in generating the data set) and can be reviewed and modified. This means it must be possible to navigate back to the original source data items so that confidence is maintained in the providence of the results generated.

Ultimately, in order to deliver long term benefit, and information fusion, all this needs to be able to be managed under configuration control so that there is a history of the product status. It also needs to be updated automatically so that as the underlying data changes, those changes are immediately alerted to the user if they are material. In this way all participants can monitor the current status and the impact on decisions previously taken.

In the second part, in the next INSIGHT-SPECTRA, Mr. Denning will discuss the impact on the business and 'what is in information fusion me'. This will address examples of business outcomes that can arise, how an ROI can be defined and delivered and the implications for each stake holder.

Nick Denning
Diegesis
www.diegesis.co.uk

INSIGHT-SPECTRA

is published and distributed worldwide by:

C3B Consulting Ltd.
19 St. Michael's Road
Winchester SO23 9JE
UK

Telephone:
+44 787 233 4000
+34 686 116 993

Email:
insight-spectra@
insight-spectra.com
or
charles.brett
@c3bconsulting.com

World Wide Web:
www.insight-spectra.com