Paper 147-27

Data Warehousing in the Modern World: A Case Study Revisited
Mary K. Tucker, Royal & SunAlliance Insurance Company, Charlotte, NC

ABSTRACT
Exploiting the SAS® System's end-to-end solutions for delivering integrated data reporting and data mining tools offers many challenges and rewards. In 1999, after developing mainframe applications for many years, I began a new journey: data warehousing in a SAS Client Server environment.

Now some three years later, I will share selected theories and practical applications as a data systems integrator in the real world. SAS-users and non-users alike will find something of interest as we explore the answers to the following questions. What does computing in the modern world look like? What does it take to survive the digital revolution? (And is surviving enough?) What are the key principles for building and delivering successful data warehousing applications?

INTRODUCTION
The goal of data warehousing is simple … to provide business users fingertip access to enterprise information across business segments, data sources, and platforms. Okay, it's simply stated, but perhaps not so simple.

I became interested in data warehousing because I needed a way to systematically access and integrate data from many places, for many people with many diverse, varied business needs. The legacy processes I developed and used before were not simple either, but neither were they effective nor timely.

Making the switch from mainframe processing to a client/server data warehousing environment is not easy. It's worth it, but it's not easy.

COMPUTING IN THE MODERN WORLD
Mergers and acquisitions, downsizing, rightsizing, outsourcing and other business initiatives complicate business rules and contribute to an ever-growing complexity of data sources and platforms and lead to what I call the "Vicious Cycle of Data Mining."

There's data, data everywhere! But how much of it is routinely transformed into useful information? It reminds me of a saying about the ocean to someone stranded on a life boat, "water, water everywhere and not a drop to drink."

The "Vicious" Cycle of Data Mining

Users are frustrated, IS departments are overwhelmed and customers are more savvy and demanding everyday. There's one word to describe computing in the modern world: chaos.

CHAOS
Interestingly, I began to use the word chaos to describe the current state of data warehousing and modern computing before I discovered it is not just a word, but actually a relatively new field of mathematical theory. Chaos theory "attempts to explain the fact that complex and unpredictable results can and will occur in systems that are sensitive to their initial conditions. " (Ross, homepage).

A common example of chaos is the "Butterfly Effect" which states, only partly jokingly, that a butterfly stirring the air in Peking can affect the weather in New York a month later. (Downes, 8).

As noted on the web site of the Chaos Group at the University of Maryland at College Park, "the idea that many simple nonlinear deterministic systems can behave in an apparently unpredictable and chaotic manner was first noted by the great French Mathematician Henri Poincare' <in the early 1900's> … but its importance was not fully appreciated until the widespread availability of digital computers for numerical simulations and demonstrations of chaos in various physical systems. This realization has broad implications for many fields of science and it is only in the past decade or so that the field has undergone explosive growth." (Chaos Group, homepage).

> "Chaos breaks across the lines that separate scientific disciplines. Because it is a science of the global nature of systems, it has brought together thinkers from fields that had been widely separated. The most passionate of advocates even go so far as to say that 20th century science will be remembered for three things: relativity, quantum physics, and chaos." (Gleick 5).

"IN A WORLD OF ORDER, CHAOS REIGNS"
So is data warehousing chaotic? First, consider the three defining features of chaotic systems (Ross, The Mathematics of Chaos) and their relation to data warehousing.

1) *Chaotic systems are deterministic. This means they have something determining their behavior. (Ross).*
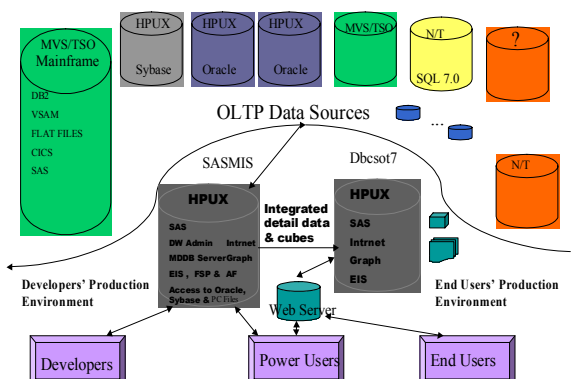
In a truly integrated data warehousing environment there are several determining factors for the behavior of the system which are often well out of the control of data warehouse implementers:
- Source systems
- Source data
- Business rules
- End users

Generally source systems are OLTP (on-line transactional processing) in nature and are designed to capture data in a specific way in order to complete some type of business transaction. While there is a close relationship between source systems and source data, they are not the same. Source data stored by source systems is much more likely based on design parameters for the source system, not the data warehouse. The originating source systems' design and usage predefine the quality, quantity, and accessibility of source data feeding the warehouse.
Likewise, business rules and business users have a similar relationship. Transforming systems source data into business intelligence requires definition and application of business rules

ranging from the simplest to the most complex algorithms that can be designed. End users expect digitized business rules to behave in the way they observe the business rules behaving in the real world. And they notice when this is not the case!

2) *Chaotic systems are very sensitive to initial conditions. A very slight change at the starting point can lead to enormously different outcomes. This makes the system fairly unpredictable. (Ross).*

Data warehousing is an iterative on-going process. With the introduction of each new data source, subject area, and end-user interface, a new set of initial conditions are added. Much work for the sake of data warehousing is done attempting to do what has never been done before. Some factors dictating initial conditions include:

- Hardware environment
- Software choices
- Network environment
- Business environment

Hardware scalability and availability are key to a successful data warehousing initiative. Under-sizing the hardware environment will diminish the functionality of well-built applications and disappoint its users. Over-sized hardware is only misused capital.

Choices about software applications used in the various stages of data warehousing development are crucial as well. Of course, the best thing to do is make all the right choices, up front, all the time. And equally so, this it not possible. Decisions have to be made along the way about when and what to build and when and what to buy and when to stay the course and when to stop and start over. Sometimes the choices are extremely difficult.

With appropriate hardware and software environments in place, network connectivity and performance are paramount to end-users. Users want fingertip access … and fast!

Ultimately, the only purpose for building a data warehouse is to support the business environment for which it was designed and implemented. Enterprise computing requires enterprise support and commitment. A successful data warehouse will survive changes in business strategy, management and organization if it demonstrates the ability to respond dynamically to the business environment in which it exists.

3) *Chaotic systems appear to be disorderly, even random. But they are not. Beneath the random behavior is a sense of order and pattern. Truly random systems are not chaotic. (Ross).*

Imagine an organization with twenty-one data sources, seven platforms, five hundred users and ten subject areas. Okay, make that two hundred data sources, all known platforms, five thousand users, an unknown number of subject areas and no straight-forward, apparent way to mix and match the appropriate rows and columns across systems and sources for delivery to the right users at the right times. Mind boggling? Perhaps, but not so unrealistic.

In a customized data warehousing environment, the seeming disorder of the above scenarios can be narrowed down to only four specific activities (or levels of abstraction):
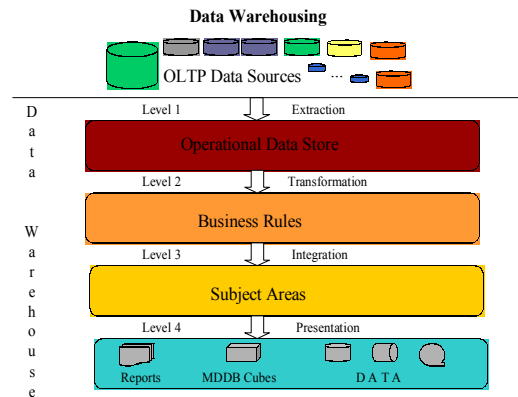
- Level 1 – Data Extraction
- Level 2 – Data Transformation
- Level 3 – Data Integration
- Level 4 – Data Presentation

The goals for data extraction are 1) to identify systems source data entities needed for the warehouse, 2) to identify the access methods and tools needed for replication, 3) to capture and store detailed source data (in it's original format) for future processing and, 4) to restate source data in a single, sharable, common platform. Some call this an operational data store.

Data transformation is where business rules are applied. This is the hard part where design and implementation standards set the stage for the remaining levels, requiring one set of metadata and business rules bridging all source data and subject areas.

If the data transformation step is done properly, data integration is easy. It's a matter of defining a particular subject area, choosing which systems source data tables contain the needed columns and rows of data, and what rules apply for merging and appending data from different sources. Data transformation is done in order to provide cleansed, formatted data for the last level, data presentation.

Data presentation is for users. Front-end access to a data warehousing applications should be intuitive, dependable, accurate and timely.

CONTROL AND TARGETING
So accepting the premise that data warehousing is chaotic, how does that help?

For developers:

Two fundamental aspects of chaos are the following:

1. The exponential sensitivity of orbits to small perturbations.

2. A very complex orbit structure making possible many very different motions in the same system.

   "Actually these two are not independent; they may be viewed as two sides of the same "chaos coin". The first property leads to the difficulty in the prediction of sufficiently distant future states, while the second property leads to the fact that chaotic attractors often have embedded within them an infinite, dense set of unstable periodic orbits. (Chaos Group, Controlling Chaos)."

   "While chaos is commonly viewed as undesirable, these properties can also be of great benefit in certain situations where one wishes to control the system dynamics. In particular, if the system is sufficiently well characterized and noise is not too large, then the first property implies that relatively large changes in the eventual location of orbit points can be made using only small changes in a control variable,

while the second property implies that the dynamical changes so produced may be very diverse, leading to a situation with a great deal of flexibility." (Chaos Group, Controlling Chaos).

The two methods that are useful in utilizing chaos in an effective ways are:

Control

Feedback stabilization of selected unstable orbits embedded in the attractor. (Chaos Group).

"Instead of trying to increase signal strength to drown out more and more noise, engineers should settle for a modest signal, accept the inevitability of errors and use a strategy of redundancy to catch and correct them. (Gleick, 92).

Targeting

The rapid steering of a chaotic orbit to the vicinity of a desired point phase space. (Chaos Group).

"The boundary between two or more attractors in a dynamical system serve as the thresh hold of a kind that seems to govern so many ordinary processes, from the breaking of materials to the making of decisions. Each attractor in such a system has its basin, as a river has a watershed basin that drains into it. Each basin has a boundary… This branch of dynamics concerns itself not with describing the final, stable behavior of a system but with the way a system chooses between competing options. (Gleick 233)."

For Users:

"The sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work." Joh Von Neumann. (Gleick, 273).

They just want it to work!

THE DIGITAL REVOLUTION

According an entry in the Factasia web site, what people normally refer to as the digital revolution, *the information superhighway,* is just the last phase of the process. The digital revolution includes the establishment of global broadband digital networks for consumers, the transition of all kinds of information including wideband broadcast media into digital formats, and the emergence of a rapidly moving information economy in this new digital marketplace. (Jones).

The digital revolution is really about artificial intelligence. Somewhere between 1623 and 1952 (apparently it's disputable) the digital computer was invented, presenting the potential to truly revolutionize the automation of intellectual processes. The design of the digital computer depends on a formal method of notation (a machine code) to provide instructions to the computer while at the same time notating results easy to read by people. This presents the need for automatic translation and well-defined syntax and semantics … leading to the "Evolution of Data Representation." (Jones).

"The need to extend delivered functionality forces continued evolution in the notations used to represent information, both for man/machine interaction and for use internal to or between computers. From the earliest days of computing, digital communications begin to emerge. Gradually the importance of networking of computers grows until this becomes a dominant theme in computing. As the *information superhighway* becomes a global obsession,

formal notations have to be perceived as suitable for mediation of interaction between networked computers." (Jones).

In the world of digital technology two controlling variables, storage capacity and band with, can not be ignored. Two laws, Moore's Law and Metcalfe's Law, put these variables in perspective.

"The unrelenting, exponential improvements in semiconductor speed, size and cost that have operated since the 1960's follow Moore's Law, a prediction by Intel founder, Gordon Moore, that every eighteen months in the foreseeable future, chip density (and hence computing power) would double while cost remained constant, creating ever more powerful computing devices without raising their price … i.e. faster, cheaper, better." (Downes, 5).

"Less well known than Moore's Law is the observation made by Robert Metcalfe, founder of 3-com Corporation, that networks (whether telephones, compilers or people) dramatically increase in value with each additional node or user." (Downes, 5).

Operating in this new digital environment requires change. I am not sure where I first heard the term "killer app" but that has been the phrase I use to describe the type of data warehousing applications needed in today's work place. I was delighted to find a book on the subject , Unleashing the Killer App by Larry Downes and Chunka Mui, last summer. On their web site about the book, www.killer-apps.com, they have this description for why they wrote the book, which was published in 1998:

"Making the change from traditional strategy (or no strategy) to digital strategy is hard work. We know, because we resisted it ourselves. For years, we worked on systems, reengineering, and planning projects that were unrelated not only to each other, but often to the changing dynamics of our clients' marketplaces, changes being wrought, visibly or invisibly, by digital technologies."

"When dramatic developments like e-mail, the World Wide Web, and increasingly powerful desktop computers seemed to come from nowhere, upsetting our careful plans, we started to wonder if there wasn't something fundamentally wrong with our approach. "

"Three years ago, we began working with a few pioneering firms to develop a new approach to planning and strategy, one that was better suited to the world of plummeting technology prices, exploding networks, and "marketspaces" with few transaction costs. We rejected many of our traditional analytic tools and replaced them with experimentation. We studied the obstacles that keep organizations from success, and developed tools and techniques for overcoming them. The result was what we call digital strategy. We believe it is nothing short of revolutionary, and we wrote "Unleashing the Killer App" to be the manifesto for that new revolution ." (Downes, Authors' Statement).

IS SURVIVNG ENOUGH?

"The best way to predict the future is to invent it" Alan Kay (Downes, 57).

Traditional strategy is implemented through what Michael Porter, author of the book Competitive Advantage published in 1980, calls the "value chain".

"The value chain is a set of activities an organization performs to create and distribute its goods and services, including direct activities like procurement and production

and indirect activities like human resources and finance. Each of these activities adds some value to the product, a value Porter refers to as 'margin.' The more competitive the value chain of the organization, the more the overall product's value exceeds the sum of its parts, hence the more margin the firm can realize as profit." (Downes 69).

A successful shift from traditional strategy to a digital strategy promotes the opportunity to build killer apps.

DESIGNING THE KILLER APP
The authors of Unleashing the Killer App have categorized the steps for designing killer apps in twelve steps grouped in three stages.

The first stage of designing any structure is to identify and evaluate an appropriate site. "In cyberspace the new business environment is largely defined by the organizations and individuals with whom you interact – your business partners." (Downes, 74).

Stage two is concerned with designing the structure, and in particular the face it presents to the outer world. "Interfaces for killer apps are built in software, a material that is both cheap and highly flexible." (Downes, 75.)

The goal in stage three is to rehabilitate the interior structures "including the pre-existing portions, to a new set of uses." (Downes, 76).

> "Rehabbing your organization requires a degree of ingenuity. Old structures must be torn down and new ones constructed, but always within the organizational culture that serves as the foundation." (Downes, 76).

THE TWELVE PRINCIPLES OF KILLER APP DESIGN
Reshaping the Landscape
1) Outsource to the customer.
2) Cannabalize your markets.
3) Treat each customer as a market segment of one.
4) Create communities of value.

Building New Connections
5) Replace rude interfaces with learning interfaces.
6) Ensure continuity for the customer, not yourself.
7) Give away as much information as you can.
8) Structure every transaction as a joint venture.

Redefining the Interior
9) Treat your assets as liabilities.
10) Destroy your value chain.
11) Manage innovation as a portfolio of options.
12) Hire the children.

RESHAPING THE LANDSCAPE

OUTSOURCE TO THE CUSTOMER
By outsourcing to the customer you recognize that the customer is the best customer service representative and the best product developer for their own needs. Incorporating user feedback into the iterative development of data warehousing applications is essential. (Downes, 81-82).

CANNABALIZE YOUR MARKET
Cannibalizing your market means offering new distribution of goods and services to existing clients. To do this you must "understand that the present value of current channels needs to

be balanced against the unrealized power of other information assets to be exploited in cyberspace. Chief among them is brand. New competitors have to start from scratch but you can jump start digital markets with the credibility and goodwill already associated with your existing organization." (Downes, 90).

TREAT EACH CUSTOMER AS A MARKET SEGMENT OF ONE
Customers like personalization, especially when they participate first hand. By treating each customer as a market segment of one you can lower the development costs associated with giving customers exactly what they want (especially since they don't always know until they get what they initially asked for.) Your customers can then become personally vested in the processes, products and services your organization offers. "As Moore's law makes it possible to put intelligence into every device (soon down to every light bulb), Metcalfe's law spreads out data network standards for communicating with these devices." (Downes, 99).

CREATE COMMUNITIES OF VALUE
Communities of value are created by…valuing community, after all cyberspace is ruled by "network economics." (Downes, 101).

Joint Application Design (JAD) is required for enterprise data warehousing and data mining practices and is the basis for the long-term success of data warehousing initiatives.

BUILDING NEW CONNECTIONS

REPLACE RUDE INTERFACES WITH LEARNING ONES
The authors of Unleashing the Killer App pose these questions:

> "Stop for a moment and think about the five most recent direct contacts you've had as a *consumer* with the checker at the grocery store, an airline reservation agent, the customer service department of your credit card company, or even a directory assistance operator. Did you receive superior customer service? Did you receive any service at all? Did you think to yourself that you'd just as well take your business elsewhere if only you weren't convinced it would be just as bad? And even if you did receive superior service, was there any indication that the company collected the information you gave it, so that it could better serve you next time? Did the company give *you* the opportunity to structure the transaction the way you wanted?" (Downes, 115).

ENSURE CONTINUTIY FOR THE CUSTOMER, NOT YOURSELF
Don't worry. The authors also say "it can only get better." From a data reporting perspective, the older the application, the more difficult the interface. The front-end, the face to the user, needs to be re-invented on a regular basis, with users directing improvements while receiving uninterrupted, quality service. "Starting with the customers viewpoint is only part of the solution. Your interface must also bring the customer along as he or she is most comfortable from the old world to the new (Downes 124.)"

GIVE AWAY AS MUCH INFORMATION AS YOU CAN
In the past, an organization's proprietary information has been considered to be a competitive advantage, but as the world wide web and the information super highway continue to grow, deciding how much information should be kept behind the firewall and how much should be given away becomes more important.

> "The age of closed systems is over… the new world is fueled by open systems. In designing killer apps, the rule for the interface is always to make it as open as possible and to give away information rather than hoard it." (Downes, 125).

STRUCTURE EVERY TRANSACTION AS A JOINT VENTURE
On-line, networked applications do not *create* partnerships but rather, provide new opportunities for *managing* business relationships.

> "Partnership Portfolio Management more than ever, the ability to form relationships and keep them healthy, is a core requirement for any business, whether a large multinational corporation or a self-employed entrepreneur. As the distinctions between you and your business partners become blurred by the breadth, depth and persistence of your connection, managing your relationships becomes even more central." (Downes, 135).

REDEFINING THE INTERIOR

TREAT YOUR ASSETS AS LIABILITIES
Today's market place is changing. The world's economy is moving from operating in a physical environment to that of distributed networking.

> "Information providers, service firms, educational institutions, and even public utilities are struggling to become digital as quickly as possible. Doing so is hard for several reasons, not the least of which is the unsettling effect on your balance sheet. Organizations discover that the true value of their company is its information, but that value is now sunk into the various containers and equipment used to transform it from one good or service to another." (Downes, 141-42).

> "It isn't that the value of today's organization is disappearing so much as its form is changing. You must now rethink not only the value of the company's bits but also the true value of its atoms, some of which you may not need in the near future. It's important to shift your investments to bits, because those new competitors who have *none* of your fixed assets – no real estate, no manufacturing equipment, no distribution network – will suddenly look competitive in the new business environment. Your assets become liabilities." (Downes, 142).

DESTROY YOUR VALUE CHAIN
Destroying the value chain, the traditional way of doing business, requires not one paradigm shift, but many in order to make new rules from the old.

> "Many organizations are preemptively destroying their own value chains. Recognizing that change is coming … they use digital technology to break rules, implicit or explicit, dictating how buying, selling, and producing of goods and services is done, and form new relationships with customers and competitors by automating expensive processes, or giving away proprietary tools others can use." (Downes, 147).

MANAGE INNOVATION AS A PORTFOLIO OF OPTIONS
Killer apps are not built so much as they are discovered.

> "Organizations unleash killer apps when they combine ideas, digital technology and the will to change…In most organizations that we are familiar with, investments in innovation almost always follow a conservative model. Up to 90 percent of IS funds are spent developing and operating current systems, most of which run on older or even outdated technology. Of the 10 percent remaining for research or technology advancement, another 90 percent may be spent studying technologies that are almost mature, like the next release of Microsoft Windows or application suites like SAP. For developments that are more than 18 months from commercial use, there is usually no activity, except for kind of skunkworks investigation that a field

manager might perform. Yet Moore's Law and Metcalfe's Law guarantee that innovations will be arriving and achieving critical mass during that time." (Downes, 154).

> "The problem is a mismatch between investment styles and investment instruments … At best, investments in innovations are justified using rolling five-year plans and the exacting business case approach of return on investment (ROI). If projects cannot guarantee satisfactory returns, they are not funded." (Downes, 154).

> "The key to the deadlock … is taking a small stake in many ventures, keeping the exit price low for investments that fizzle, leaving the possibility for a dramatic upside – killer app upside – for investments that do pay off. By diversifying the innovation investment, the 10 percent of the IS budget set aside for research could be managed not as a series of discrete projects, but as a *portfolio."* (Downes, 154).

HIRE THE CHILDREN
"Young people, especially children, implicitly understand digital technology in a way that the rest of us can only watch and admire." (Downes, 158).

I know this is true when my daughter shows me where to plug the floppy disk drive into my new laptop, or a 'secret' panel in the new printer that catches the paper when it prints (as opposed to the chair I was placing next to it for the same purpose … and I do this stuff for a living!)

The good news is children come in many shapes and sizes. We need not wait for today's children to grow up before we make the leap to designing killer apps, but we can embrace their viewpoints and philosophies and look for innovative leaders willing to take the risks necessary to learn for the future.

DIGITAL STRATEGY IN PRACTICE
Becoming digital is an on going process. Key players from every discipline in the organization need to be, not only involved in the planning and implementation of digital strategy, but also engaged and invested in the project and process.

> "To succeed at digital strategy, your organization must be a learning organization, more focused on ideas and experiments than detailed plans and forecasts. The change is not as dramatic – or weird – as it first sounds. Organizations are at their core learning entities, reinvented over and over again throughout their history. Without a learning capability, the organization would never have come into being and if the need to change is ignored too long, the organization will perish as surely as if it was denied capital, employees and markets." (Downes, 167-68).

> "Organizations that practice digital strategy must learn to communicate with their organization's change center. The language it speaks is the language of ideas, scenarios, options and what-ifs." (Downes, 168).

KEY PRINCIPLES OF DATA WAREHOUSING
Data warehousing initiatives evolve around bridging the gap between the people operating in the business environment and the computer systems and interfaces that process their operations.

Key Principles from the Business Perspective
- Be Customer Centric
- Identify Business Opportunities
- Provide Useful Information
- Pay Attention to Detail: Share Metadata

Key Principles from the Technical Perspective
- Develop using Rapid Application Development
- Employ Joint Application Design
- Use Formal Project Management Methods
- Integrate legacy systems and isolated client/server systems
- Deliver Return on Investment

BE CUSTOMER CENTRIC
Giving our users the ability to provide operational feedback is key to the success of data warehousing and data mining initiatives. Users are the best ones to ask about how to best present the end results, how to streamline business processes and what future needs to anticipate. Users come in many shapes and sizes, based on experience level and data needs. For our purposes we can classify users in three types – end-users (business users), power users (analysts) and (application) developers (Berry, 378).

END USERS
End users are, after all, the target audience for the information provided. Their needs drive the development of applications, the architecture of the warehouse, the data it contains and the priorities for implementation. Many choose to interface the data warehouse through printed reports as they always have. Even in this case they will experience improvements from the data warehouse as reports become more accurate, more timely and easier to access. (Berry, 380-81).

More importantly, business users who are willing to use the computers on their desks have the advantage of powerful new tools providing direct access to information they need and know how to manipulate. (Berry, 381.)

POWER USERS
Power users usually want as much data as possible and use special purpose tools for their work. They are often the primary audience for a data warehouse. Usually few in number and technically sophisticated, their work is often highly leveragable. (Berry, 382).

APPLICATION DEVELOPERS
Much of the software that runs on a data warehouse is customized or custom-developed for its particular environment.

These application developers need:

1) an environment where applications in development are shielded from changes in the data warehouse structure and where operational changes have little impact on existing applications.

2) access to valid field values and to know what the values mean (metadata). (Berry 382).

IDENTIFY BUSINESS OPPORTUNITES
Data warehousing done for the sake of its users obviously must be business driven. Other factors do exist such as hardware and software standards and limitations, budgets, and time frames but by allowing your business needs to direct your development you ensure proper prioritization and investment along the way.

PROVIDE USEFUL INFORMATION
Transforming and presenting data as information is the primary goal for data warehousing. The latest evolution in efficiently transforming and presenting data is formal data warehousing practices with browser based front ends.

Identifying the right source data is crucial to the results of an analysis. Using operational data presents many challenges to integrators and analysts such as bad data formats, confusing data fields, lack of functionality, legal ramifications, organizational factors, reluctance to change, and conflicting timelines. Also, the more disparate the input data sources, the more complicated the integration.

Defining a logical view of the analysis process, independent of source data availability, is equally important. Here end-users and developers get to define their ultimate business and technical needs proactively rather than reactively.

PAY ATTENTION TO DETAIL: METADATA
Formal recording of metadata is also crucial to data warehousing. Metadata describes data in terms of entities, attributes and relationships that are meaningful on the business level. It describes product hierarchy, the customer attributes, the relationships between the business and various partners, and other data attributes such as when and where the data is available and what applications use it. Metadata must be flexible to change since a data warehousing environment is not static and must respond repeatedly to changes in the business and systems environment it was built to support. (Berry, 365).

> "Metadata provides the key link between the business users and the data, describing the data in business terms. A good metadata system gives users the ability to browse through the metadata on their desktops, making users more comfortable with the data warehouse and encourages them to take full advantage of it." (Berry 365).

RAPID APPLICATION DEVELOPMENT
RAD is a methodology for compressing the analysis, design, build, and test phases into a series of short, iterative development cycles. This has a number of distinct advantages over the traditional sequential development model. RAD projects are typically staffed with small integrated teams comprised of developers, end users, and IT technical resources. Small teams, combined with short, iterative development cycles optimizes speed, unity of vision and purpose, effective informal communication and simple project management. (Creative Data, Inc).

> "An important, fundamental principle of iterative development is that each iteration delivers a functional version of the final system. It is a properly engineered, fully working portion of the final system and is not the same as a prototype. For example, the first iteration might deliver 100% of 10%, the second iteration 100% of 25%, etc." (Creative Data, Inc).

> "With an iterative methodology, you break your warehousing project into small, manageable chunks, referred to as *projects*. In the iterative approach, you perform the same planning tasks that are required in the big bang approach, but you don't have to evaluate all of your deliverables up front. You must design your overall architecture, but when you enter the planning phase, you need to concentrate only on your first project or iteration. After each project, you review your architecture, your development process, and your business requirements." (SAS Institute).

JOINT APPLICATION DESIGN
Joint Application Design, or JAD, is a process originally developed for designing a computer-based system. It brings together business area people (users) and IT (Information Technology) professionals in a highly focused workshop. The advantages of JAD include a dramatic shortening of the time it takes to complete a project. It also improves the quality of the final product by focusing on the up-front portion of the development lifecycle, thus reducing the likelihood of errors that are expensive to correct later on. (Creative Data, Inc.).

> "JAD centers around a structured workshop session. Everyone gets together in a room and talks it out. Everyone hears what the rest of the group has to say. There's no

delay between question and answer, no "telephone tag" or waiting for memos to come back. JAD eliminates many of the problems with traditional meetings. Meetings are not well regarded as a productive form of work. JAD turns meetings into workshops. They are less frequent, more structured, and more productive. An agenda provides the structure, a facilitator directs the process, visual aids clarify concepts being discussed and the group dynamics, with constant feedback, stimulates creativity (Creative Data, Inc.)"

PROJECT MANAGEMENT
Formal project management practices are required to properly implement RAD and JAD based projects, but must be customized and flexible enough to enable the creativity and speed needed to compete in the digital market place.

"Modern thinking leads systems developers to have a plan of their project, so that various phases can be identified and dealt with in an orderly and effective manner. There has been written criticism in the I.S. literature of planning as stifling innovative project work." (Allan, pm-mthds).

"A contributing factor to the unrest in present attitudes to computer projects stems from the days when computing was considered a highly specialised 'black art' totally the preserve of programmers. It seems incredible to today's computer-based-systems developers that prior to the 1960's there was little discernible project management and hence software appeared to "emerge and just happen" in the hands of computer programmers." (Allan, proj-man).

With Rapid Application Development and Rapid Application Design, end users, analysts and developers participate in all phases of design and implementation allowing more opportunity for deriving

useful feedback from the project management processes and documentation.

"While project management approaches are becoming indispensable to organizations because of rapid product development, better utilization of resources, and increased cross-functional communication, little time is spent on evaluating and learning from individual experiences. A few key questions would yield better results -- Was the investment worthwhile?; Did it go according to plan?; If yes, how? If not, why not?" (Pinto).

SYSTEMS INTEGRATION
Truly integrated systems pave the way to the information super highway. The authors of "The Two Overlooked Aspects of IT Management" discuss the importance of identifying and engaging members from all departments that will be impacted by the release of new systems designed for enterprise computing and reporting. (Campbell).

"Operations integration comprises the body of knowledge and practices needed to be applied to ensure that a new system produces the desired results by a certain time at a specified cost. To succeed, its disciplined approach must be followed by all departments that will use the new system. Among its many required actions are:
-- Clear communication of the corporate goals for the new system.
-- Making it crystal-clear which workers will be expected to use it.
-- Fine-tuning and aligning the company's work processes with the business processes embedded in the new system.
-- Documenting the tuned and aligned processes in the company's official Policies and Work Procedures manual.
-- Modifying workers' roles and job descriptions to account for the changes in the business processes and the workers' use

of the new system.
-- Training users to handle the work processes that use the new system.
-- Creating incentives and disincentives for using the new system." (Campbell).

Data warehousing is all about systems integration, but there is more to it than getting systems to talk to each other. The owners of the technical, organizational and business functions must collaborate in order to ensure the proper assignment and execution of the roles and responsibilities involved in systems integration development.

"To move forward with operations integration The bottom lines are simple:
-- Operations integration must be done in order to mitigate organizational and business risks… the variations on when to begin can run the gamut from the problem-oriented approach, which takes the longest elapsed time, to the parallel approach, which takes the shortest time to full business value. Regardless of which option is chosen, senior management must demonstrate strong leadership and require that operations integration be done from the very start of the project.
-- In addition, top management must require that the operations integration effort not be owned and led by the IT department, but by the highest ranking manager of the user groups that will be using the new system in their day-to-day business.
-- IT leadership's role must be to ensure that top management fully understands that (1) IT will manage technical risk, and (2) only user management can properly mitigate organizational and business risk.

IT managers who do not insist that operations integration be properly owned and led run the very real risk that the IT system will not pay off for the company, and they will be blamed." (Campbell).

"In summary, always keep in mind that risk cannot be avoided. It must be accepted and managed in any project that involves it. But in managing risk, any project will have three risk components to manage successfully -- technical, organizational, and business." (Campbell).

RETURN ON INVESTMENT
What return on investment can be expected from a successful data warehouse? Initial returns include automation of existing processes, on line reporting, and re-engineering and streamlining existing applications. Many opportunities exist to realize other time and cost savings. Bigger returns include improved routine access to data, innovation and creativity for solving business problems and new ways of looking at and analyzing data.

THE TRUE VALUE OF DATA WAREHOUSING
A data warehouse delivers true value when it allows its users to move past data *reporting* and into data *mining*. The role of data mining is to provide tools that inspire creativity based on observations and conclusions drawn on observations in the data.

Data mining is a creative process. Determining what actions to take, based on data mining, depends on what questions are asked, who is asking them and what the results are.

According to the authors of Data Mining Techniques for Marketing, Sales and Customer Support, "to really achieve its promise, data mining needs to become an essential business process, incorporated into other processes, including marketing, sales, customer support, product design and inventory control." (Berry 18).
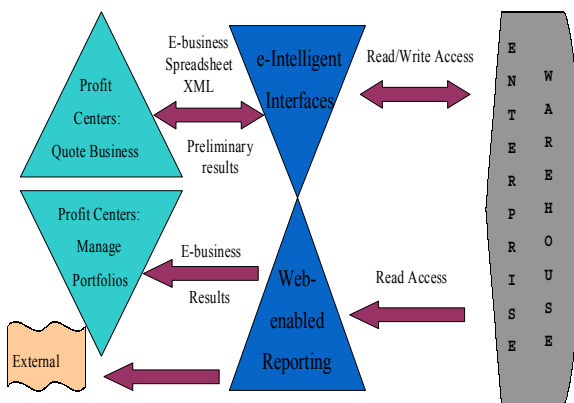
THE VIRTUOUS CYCLE OF DATA MINING

Measurement provides the feedback for the continuous improvement of results. These measurements are what makes `the cycle of data mining' *virtuous.* The specific measurements needed depend on many factors: the business opportunity, the sophistication of the organization, past history and the availability of data. "The 'virtuous cycle' incorporates data mining into the larger context of other business processes. It focuses on action based discovery and not the discovery mechanism itself." (Berry, 18).

> "In the business world, data mining provides a fundamentally new capability, the ability to optimize decision-making using methods learned from past actions. In recent years, technology has converged to enable this capability and in the next few years, more advanced tools, hardware and software, will continue this trend. Parallel processing machines will become more common and less expensive. Software is increasingly available to take advantage of parallel processing and distributed resources for identifying useful patterns in data…Different organizations adapt data mining to their own environment, in their own way. In successful environments, the stages of the virtuous cycle will be in place, suitably transformed to fit into the particular environment." (Berry, 34-35).

In today's world successful data warehousing applications are developed (or "unleashed") by using lessons learned from the study of chaos theory and its controls, understanding the ramifications of Metcalfe's and Moore's Laws as they relate to the digital revolution and placing a new emphasis on customer service, creativity and innovation.

If I've said it once, I've said it a thousand times (and I never exaggerate), data warehousing and systems integration in the real world is not easy. It is not easy, but it *is* worth it and the to begin time is now!

The "Virtuous" Cycle of Data Mining

REFERENCES

Allan, George. Research and Publications: Project Management. University of Portsmouth. January 2002 <(http://www.dis.port.ac.uk/~allangw/proj-man.htm)>

Allan, George. University of Portsmouth, Research and Publications: Project Management Methods. January 2002 <(http://www.dis.port.ac.uk/~allangw/pm-mthds.htm)>

Berry, Michael J. A. & Linoff, Gordon (1997) *Data Mining Techniques For Marketing, Sales and Customer Support.* USA: John Wiley & Sons, Inc.

Campbell, Mike and Dutch Holland (2001) *The Two Overlooked Aspects of IT Management* USA: The McGraw-Hill Companies, Inc. <http://www.pmboulevard.com/pm_community/news/itRisk12_28_01.htm>

Chaos Group, University of Maryland at College Park, homepage January 2002 <www-chaos.umd.edu/intro.html>

Chaos Group, University of Maryland at College Park, Controlling Chaos January 2002 <http://www-chaos.umd.edul/ -controlChaos>

Creative Data, Inc. Development Methodology – RAD and JAD. 1995-2001 (http://www.creativedata.com/research/methodology.html.)"

Downes, Larry and Chunka Mui (2000)*Unleashing the Killer App* USA: Harvard Business School Press

Downes, Larry and Chunka Mui. Authors' Statement: Unleashing the Killer App Homepage, May 1998 <http://www.killer-apps.com>

Gleick, James (1988), *Chaos: Making a New Science* NY, NY: Penguin Books

Jones, R. B. Factasia: Computing: The Digital Revolution, August 1996 <http://www.rbjones.com/rbjpub/cs/dr000.htm>

Pinto, Jeffrey K. and Om P. Kharbanda. *How to Fail in Project Management.* Bibliography. July-August 1996 <http://www.c3i.osd.mil/bpr/bprcd/6084.htm>

Ross, Mike, et all *The Chaos Experience* homepage, 1997 <http://library.thinkquest.org/3120/text_only/index.html>

Ross, Mike, et all *The Chaos Experience* : Chaos History, The Mathematics of Chaos, 1997 < http://library.thinkquest.org/3120>

SAS Institute, Planning and Design, Warehousing Concepts, Methodologies, 2001 <http://www.sas.com/rnd/warehousing/methods.html>

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

    Mary K. Tucker
    Royal & SunAlliance Insurance Company
    9300 Arrowpoint Blvd
    Charlotte, NC 28273
    Work Phone: 704.522.3048
    Fax: 704.522.3413
    Email: Mary_Tucker@rsausa.com