

Publication: Data Management Review

Column Title: Data Warehouse Delivery

Author: Douglas Hackney

Headline: Eating the Rat

Issue: January 1998

[begin copy]

A recent survey of CIO's in Information Week revealed that data warehousing ranked along with the intranet as the anticipated two most important market segments in two years, along with the ubiquitous Y2K. What this means is that once the dust settles from the panic surrounding the new millennium, enterprises the world over will still be faced with the requirement to provide an easily accessible integrated, scrubbed source of historical and current data. In fact, you could make a solid case that it is already a competitive requirement in the consumer packaged goods and retail segments, with the prevailing opinion being that if you don't have a data warehouse in these segments, you'll be out of business by the year 2000, or soon after anyway, so there's not much reason to divert resources from warehousing to Y2K storm troopers. Throw in telcom and the newly deregulated utility industries and there are fewer segments where data warehousing is not a necessity than those where it is a required strategic and tactical competitive weapon.

However, while data warehousing is becoming an established science and a burgeoning economic segment, it still lacks the clear view of the perspective of history. Like a 200 year old "historical monument" in the US, warehousing lacks the context of a 1,000 year old church in Germany, or a 2,500 year old palace in China. This lack of perspective prevents practitioners, IT management and business management from fully understanding the evolution required for organizations implementing data warehousing.

Three distinct phases of data warehousing organizations have evolved, each with its own unique distinguishing characteristics. If you have implemented, are in the process of implementing, or are planning a data warehousing effort, it is critical to know where you are along the evolutionary path of data warehousing. One of the largest challenges facing data warehousing efforts is the tendency for business executives to be thinking in terms of advanced Phase 3 capabilities on day one, while it will take IT another one to two years simply to deliver the Phase 1 data resource.

Organizationally, culturally, and politically it will take your organization many years to grow from a Phase 1 data warehousing site to a Phase 3. Gastronomically, you can think of it as similar to a snake eating a rat. If you are in Phase 1, the rat is just going in, Phase 2, the rat is in, but the lump is still moving down the snake's body, and in Phase 3, the lump is gone and the snake is ready to take on another rat. While the snake is involved in eating the rat during Phase 1 and 2, it is relatively immobile, and incapable of taking on a predator, or more food. Therefore, like the snake, it is imperative that for the duration of the data warehousing experience, you have the time and safety required to fully digest the warehouse and its effects on your organization.

To enhance your chances of success, it is imperative that you understand where you are on the evolutionary curve of data warehousing. It is nearly always fatal to let the business cultivate Phase 3 expectations, while you are struggling to merely deliver Phase 1 within your lifetime.

Phase 1:

Phase 1 organizations are consumed by the original mandate of data warehousing: "build it, and they will come." While wise organizations have learned from the previous failures that have taken this maxim to heart, and have built successes by only building to solve specific business

pain, the “buy to build” mindset still pervades this segment. These organizations are primarily trying to “automate the past”, and tend to conduct user interviews based on a data driven scenario, instead of a business process driven scenario. Other defining characteristics are:

- “One Version of the Truth” mission statement
- Build teams consumed with integrating tabular data
- “Buy to Build” mindset, ignorant of, or in denial of the systemic nature of a data warehouse or data mart system
- Still on a very steep and high learning curve
- Building a system to almost exclusively serve a knowledge worker/power user audience
- Limited business scope of the project
- End user access & analysis tools primarily in the Q&R (Query & Reporting) & LowLAP (Low end online Analytical Processing) segments
- Highly likely to be building the system manually
- Limited to no meta data
- Limited to no system monitoring capability
- Little to no mission critical nature of anything that has been built to date
- Little to no “Butler Factor”. This measurement is named for Tim Butler, a marketing manager for GE, who’s critical evaluation factor for any technology is “what will be different on a Tuesday afternoon?” Phase 1 organization’s efforts usually have little impact on the day to day life of the majority of the organization.

Phase 2

Phase 2 organizations have moved beyond their initial efforts simply to integrate the tabular data of their business. They have learned from the early failures of the “build it and they will come” crowd, and have built specific solutions to specific business pain. In doing so, they have constructed a system of architected data resources, both data warehouses and data marts, that provide high level impact throughout their organizations. These organizations have built and cultivated a clear vision of the systemic nature of data warehouse and data mart systems, and have created and sustained all the necessary technical and human processes required for sustenance and growth. Other defining characteristics are:

- “Change the business” mission statement
- Data warehouse and data mart systems are fully integrated into the processes of the business
- Multiple closed loop OLTP/data warehouse systems
- Enterprise class EMT (Extract, Mapping and Transformation) tools
- Very scalable technology and processes
- Highly developed monitoring systems for all aspects of the data warehouse and data mart systems including load and utilization processes
- Multiple, fully integrated, non-OLTP data sources including 3rd party data sets and data islands
- Independent data warehousing team responsible for core competencies for all data warehouse and data mart initiatives
- Sophisticated analysis capabilities, including data mining and statistical analysis
- Data warehouse steering committee fully involved in prioritization and air cover
- Industrial strength OLAP tools
- 100% server centric tools
- Access and analysis tools that are fully capable of resource sharing, scheduling and distribution
- Integrated, open and extensible meta data repository, with a single point of entry for all users
- Almost all members of the enterprise are users, including casual users and information consumers
- Fully functional thin client system access
- Fully mission critical
- Large business scope
- Multiple architected data warehouses and data mart systems
- 100% Butler Factor, i.e. everything is different on a Tuesday afternoon for every user

Phase 3

Phase 3 organizations have moved beyond their own internal audiences and are using their data warehouses to leverage partner and public relationships. They have fully leveraged the power of integrated information resources to fundamentally change not only their internal world, but also those of other stakeholders, such as customers and the public at large. Technically, they have moved beyond mere tabular data and have incorporated multiple data types, much to the delight of the data base vendors who so desperately needed a reason for this technology to exist. Phase 3 organizations operate outside the borders of their own organizations, and use their warehouses to change the lives of everyone. Other defining characteristics are:

- "Manage and share knowledge to change the world of all stakeholders" mission statement
- Internal, partner and external closed loop systems
- Internal, partner and external user audience for all levels of access & analysis, from reporting to industrial strength OLAP
- 100% thin client enabled
- Multiple data types, including text, image, sound, & video
- Data warehouses and data marts that are only one element of the Phase 3 system
- Fully integrated multiple internal and external data warehouse and data mart systems
- Fully replicatable, extensible core competencies
- Networks of build and sustenance teams
- Enterprise business scope for internal and partner organizations
- Mission critical status for internal, partner and external organizations
- 100% Butler factor for all internal, partner and external users.

If you're involved in data warehousing, you are snake with a rat somewhere either in your sights or in your digestive tract. It is imperative to be honest and realistic about where you are along the path. If you are solidly in Phase 1, it is suicidal to take on another subject area, another data mart, or another data warehouse. Just like a snake, you will choke if you try to eat more now. Your organization must digest your first efforts before you try more. It takes a significant amount of time for the organization to realize a reasonable "Butler Factor" from your initial efforts, and until people's lives start changing on a Tuesday afternoon, you are not going to have the widespread political impact you'll need to start or sustain further efforts.

To survive, avoid predators while you've got a big rat stuck in your mouth, eat slowly, don't overeat, don't let your eyes be bigger than your stomach, and allow sufficient time to digest before attempting to eat more.

And don't worry about the rat, it tastes like chicken.

[end copy]