

Publication: Data Management Review

Column Title: Data Warehouse Delivery

Author: Douglas Hackney

Headline: Project Status Report Cards

Issue: December '96

[begin copy]

Data warehouse projects, by their very nature, are very large, complex projects. Even a simple, single business unit, single subject area data mart is a non-trivial responsibility. Effectively communicating, much less understanding, where your data warehouse project is, and where the looming crises are, is in itself, a daunting task. Most teams rely on massive project plans and copious proposals and project reporting documents to monitor and communicate project scope, process, progress and requirements.

Unfortunately, this is a bit of overkill when all you need to do is answer a LOB (Line of Business) manager's simple question of "how's the data warehouse coming?". Plopping down a 95 page project plan and 150 page project definition document on them is akin to lighting a candle with a blow torch. Certainly effective, but a bit overtaxing on the available resources. I've found that a much more effective approach is to develop a project status report card. These report cards consist of statuses and descriptions for the major elements of the entire system that makes up the data warehouse. Due to its hierarchical nature, it is simple to collapse the detail out of a report card and instantly have a simple, high level status document ready for discussion or distribution to all parties. They make great discussion documents for team meetings, with trouble areas easily highlighted, and also function as excellent forcing functions for recalcitrant suppliers and teams, i.e. a networking group that hasn't gotten that FDDI backbone in place yet.

So what does a data warehouse project report card look like, and how do you put one together? What you're looking for is a document in outline form that you can flesh out as you move forward with the project. In the beginning there will be many topics that contain only "we don't know" as an entry. Moving forward, you and the team will fill in each of these with increasing levels of detail and empirical facts. At any time in the life of the project you'll have a handy guide to measure progress and provide quick status reporting to upper management.

Due to space limitations, I can't give you a complete picture, but these categories will serve as a good starting point:

- Mission Statement: This is one you can fill in from the beginning, "To become the system of record for decision support in the enterprise."
- Subject Area: Start with one, don't try multiples. It will be determined by balancing three main factors: pain, politics, and utilization.
- Source Systems, internal and external: determined by subject area and user need
- Architecture & Topology: server(s), server(s) locations, network pipes & sizes, network topology, connection to desktops, etc.
- Physical Device Information: server manufacturer, model, configuration, desktop client minimum configurations, etc.
- Process Flow: where things are going to be done and when, i.e. extract, scrub, transfer, load, summarization, access, batch window, etc.
- Data Flow: where the data goes, i.e. DB2 temp files, output queues, temp load tables, transformation engines, etc.

- Content: high level table descriptions, as the project progresses, this should be accompanied by an appendix that details fields & descriptions. Make sure to highlight known exceptions and missing data points requested by users.
- Process and Task Ownership: who owns JCL, who owns RDBMS loads, who owns triggers & stored procedures, who owns user change requests, etc.
- Roles and Responsibilities: data warehouse team roster, steering committee or executive sponsorship committee members, individual sponsors, end user contacts, consultants, vendors, etc.
- Meta Data: description, content, extent, population, maintenance, and availability with various tools.
- Known Caveats: missing data points, performance issues, tool limitations, etc.
- Schedule and Resource Requirements for Missing Elements: Inevitably, the product you deliver will be short of the grand vision that the team had when beginning the project. In this section include the time and resources required to fulfill the vision.
- Data Marts: description, content, replication strategy, subject areas, etc.
- Schedule and Resource Requirements for Data Marts: If you are constructing an enterprise, or “galactic” data warehouse, and then extracting data marts from it, don’t over commit. Make sure to allow some recovery time after the “galactic” roll out. There will be problems to fix prior to being able to stamp out data marts on a regular basis.
- Q/A process and level: overall level of Q/A requirements/commitments at current and future milestones, Q/A process for various deliverables.
- RDBMS Response Times: i.e. for a summary query to a single aggregate table X seconds on average, for a detail query to 2 dimensions and the fact table Y minutes.
- Service Level Agreements: internal, i.e. Unix Administrators to DW Team and external, i.e. DW Team to Users
- Security: strategy, tools, levels, access methods, administration, etc.
- Scheduling & Distribution: tools, strategies, architectures, etc. used for the scheduling and delivery of end user reports and data sets.
- Tools: RDBMS, monitoring, design, mapping, extract, transformation, management, data access, data visualization, etc. Capabilities and their known limitations, i.e. maximum number of rows to return to a desktop “LOWLAP” tool and still get results in a reasonable time.

You can use this project status report card in many ways. I prefer to not get teams bogged down in “analysis paralysis” driving down the atomic levels of detail in each category. I find them much more useful if the team spends no more than a couple of minutes on each point in each review session. Drill down until you hit a wall, put in “we don’t know yet”, then move on to the next point. The “we don’t know yet” points can then be assigned to off-line research or task groups. In this way, you can use the report card to drive status meetings that are an hour in length rather than an afternoon of endless churn.

If you’ve got a project planned or under way, take fifteen minutes and build an outline of these categories, and the others specific to your site you’ll think of along the way. I think you’ll find it a valuable and efficient document for your team and for external reporting to the business.

[end copy]