



Measure Twice, Cut Once: Avoid Common Pitfalls When Building a Data Warehouse

SUMMARY

Taking the path of least resistance is not a particularly sound business rationale, and it is often too easy to forgo the hard work required to follow your data warehouse vision. The result is a shoddy and unstable foundation resulting in delays, unforeseen expenses and failure to meet objectives.

A data warehouse should be like a Ferrari: a well-oiled, high-octane machine. Then again, if you want a data warehouse that just putters along and consumes valuable resources while coughing up toxic exhaust, stop reading. You can't be helped. Otherwise, read on.

Three main themes are often repeated when it comes to poorly built and managed data warehouses: insufficient leadership, cutting corners, and conflict avoidance. The list here within presents ten ideas to combat these oft-repeated themes.

10 GUIDELINES FOR BUILDING A DATA WAREHOUSE

1. **Enlist a Strong Data Visionary** – This is the General of your data warehouse, the person who sets and enforces policies. They don't need a tank or high-tech weaponry to have teeth – though these would be useful job perks – just your company's support. Give it willingly; along with a commitment to get this person the training they'll likely need to deal with, let's say, challenging colleagues. You'll need a strong individual to step on some toes, buck culture and challenge the status quo.
2. **Eliminate Upstream Redundancies** – Often, there are multiple sources for the same type of data. One of the data sources will be better than the others. Use this source and discard the same data from the other sources. Don't worry about hurt feelings . . . it's not like you're choosing one of your kids over another. Data flows are like arteries. Keep 'em clean, and they'll run smoothly for years. Let them clog up with redundant data, and it's only a matter of time before operational issues arise and mushroom.
3. **Load ONLY Clean Data** – Known data inconsistencies are loaded into warehouses with a surprising frequency. The data warehouse isn't a landfill for data (toxic or otherwise), nor is it the island of misfit data. Fixing data integrity issues is less costly and more predictable when done proactively, yet on-the-fly fixes seem to be the solution of choice when deadlines and cost pressures loom. Perception IS reality. If colleagues think that your data warehouse has quality issues, game over. It's tough to dig out of that hole. The key lesson here: find and address data quality issues before loading data. Trust, but verify all inputs.
4. **Enforce a 'Correct-at-the-Source' Mentality** – Don't be afraid to push your data providers to clean up their mistakes. You're one company, not a massive collection of fiefdoms, each with its own misguided set of priorities. No one said this would be easy. Data owners will persist with excuses that threaten delays and obstruction when you try to get them to take ownership of the quality of their data. However, the alternative will cause inconsistencies between the data source and the warehouse, as well as a long list of potential problems. If all else fails, send the head of the company samples of the inconsistent client statements that will inevitably result from having data in the warehouse that doesn't match the source. Tell him there's more coming.
5. **Reconcile Source Data With Warehouse** – You want data integrity and quality, right? Quality checks are critical when dealing with portfolio balances, performance, and other investment attributes. It isn't enough to have robust processes to capture all changes – there will be times when something goes awry. Develop a plan to reconcile data and stick to it. All incoming data needs to be periodically reconciled with the source. A well-designed, fully-automated reconciliation system ensures the Six Sigma quality that financial data warrants.

Measure Twice, Cut Once: Avoid Common Pitfalls When Building a Data Warehouse

10 GUIDELINES FOR BUILDING A DATA WAREHOUSE (continued)

6. **Document Clean Up Activities** – How else will you know the ‘who,’ ‘what,’ ‘when,’ ‘why,’ and ‘how’ of data clean up? Not even Daniel Craig with a cool wrist laser will be able to help you blast your way out of that mess when you’re asked to reconstruct what went wrong. By documenting clean up activities, you can identify recurring themes and resolve to correct these issues so they no longer impair data quality. Be methodical and keep an audit trail. Impose the same controls on anyone whose fingerprints might be found on warehouse data.
7. **Be Careful When Reloading Data** – Update, or replace? A complete purge of old records is usually the cleanest practice, but if you’re feeling adventurous, then go with an update. But, you’ve been warned: problems from faulty data reloads are difficult, if not impossible, to troubleshoot.
8. **Keep Data Warehouse Calculations to a Minimum** – Since you have nothing with which to compare these calculations, this can mask problems with the calculations. Okay, so you must have those customized client report, and these reports require calculations or logic that the source systems just can’t do. Fair enough, but ensure you put some ground rules in place and designate a Sheriff to monitor and validate these calculations. Better yet, since these calculations are often requested by separate groups and for different purposes, pass off clean data to these groups, and let them do their own calculations. Take responsibility for the data in the warehouse and put the onus on these other groups to perform custom manipulations.
9. **Check Licensing and Redistribution Agreements Before Loading Proprietary Third-Party Data** – Sure, you can probably sneak one by a 3rd party data provider; but the use, storage, and distribution of proprietary data is usually restricted. Ignorance is not an excuse and will not be readily accepted by data providers. The cost of negligence can be steep. Know the rules of distribution before you proceed with any external delivery of proprietary data. Take it a step further by including protected data parameters and instructions with the data and suggest that all warehouse clients reference and adhere to these rules.
10. **Not All Data Belongs in the Data Warehouse** – Don’t be afraid to exclude data that could present a challenge to the integrity or capabilities of the data warehouse, including high-frequency data and high-maintenance data. A good data General will never allow it, so make sure you pick a strong, iron-willed person for the job.

Can you build a data warehouse without following the above guidelines? Sure, you can. You can also build a house from the top down, but you won’t have much of a foundation.

Measure twice and cut once.



Laurie J. Hesketh, CIPM, Managing Director, leads strategic engagements and specializes in data management and investment performance projects.



Elizabeth M. Colebrooke, Principal Consultant, brings 20 years of experience in performance, risk, data, and asset allocation derived from working with hedge fund, fund of hedge fund, and mutual fund strategies.

Meradia
119 North High Street
West Chester, PA 19380
Phone: 610-738-7787
Fax: 610-738-7797

Info@Meradia.com
www.Meradia.com

© 2013 Meradia. All rights reserved.

