A data mart is a subject-oriented database that supports the requirements of a particular business department. Peer Connect members share their approaches for developing efficient data marts in organizations, from the ground up.

Overview

A Peer Connect member seeks opinions from fellow members on creating an application architecture for his organization that sits in a number of disparate systems. The peer is also keen to know the methods used by other organizations in establishing a data mart/data dictionary and platform, as his organization is at ground zero concerning data maturity. Peer members provide advice on multiple phases to consider for building data marts in an organization.

Peer Strategies

Define the Organization’s Core Data and Its Purpose

Responding to the thread initiator’s query, a peer advises that figuring a data issue from ground zero usually requires changes across multiple systems (and the organization) rather than just a fix in the data project. The Peer Connect member, who has experience working with a not-for-profit (NFP) and is familiar with budget constraints, adds, “If you are not able to do that, then adding a layer to pull the data together is a good starting point. This can be a ‘classic’ data lake, or if the issues are real-time, some form of data virtualization. You can get a relatively inexpensive start using a cloud data lake to tidy up one domain.”
Another peer member highlights that the most important thing with data, to be useful as information for the organization, is to understand its meaning for the entire organization as well as its quality.

Many peers stress examining the key organizational data concepts, based on the business domains that define the organization. A Peer Connect member suggests, “You do this by developing a glossary of terms with key relevance to your organization (for example, in an insurance organization, it could be ‘what is a claim?’). This glossary must be developed by your main business domain visionaries or ensure that it at least reflects how they would define those terms in regards to the organization’s purpose and communication among the organization.” Per the peer, one needs to document which visionaries/stakeholders within the organization would be responsible for owning a particular data concept. Also, identify who would be the custodian within the organization for any data related to that concept (for example, a CFO for accounting data.) The peer adds, “This process is commonly referred to as developing the organization’s data glossary. This ensures you do not go down the path where applications provide little or reduced value to your organization because of meaningless/poor quality data.”

**Steps to Create a Robust Data Model**

Stressing on the thread initiator’s organization’s data maturity, which is “at ground zero,” a peer emphasizes focusing on gaining an understanding of the organization's data. He proposes a common concept of “3 Vs” for the same:

- **Volume**
- **Variety**
- **Velocity**

The peer shares that this concept will help in making critical educated decisions, irrespective of whether you build, buy and operate on-premises or on the cloud.

**Develop the Application’s Data Dictionary**
According to the peers, the next step will be to define the data your organization’s application(s) process ingress/egress. A peer instructs, “Start with a definition of what it means to the business — the data glossary will help. Document where it comes from, where it goes and where it is stored. Also, identify data constraints as well as the data type and classification, and quality issues (for example, common input errors).” The peer emphasizes that the process is commonly referred to as the “application’s data dictionary.”

Responding to the thread initiator’s query on creating a data mart, a data and analytics (D&A) professional suggests establishing what the organization already has, before deciding what needs to be purchased. He suggests getting answer to the following:

“Is there an accepted data model (conceptual or logical), a data glossary of the vocabulary used within the NFP or any degree of data governance in place?”

The peer adds, “A data mart can take several different forms depending on the end users. With a clear idea of who the intended end users would be and the information velocity through the organization, you will be in a better situation to start looking at solutions for bringing the disparate systems together.”

He recommends to combine a data model and glossary to start building a data dictionary. The peer opines, “Focus first on the ‘core data’ within the NFP and take a step-by-step approach to the dictionary rather than going for any ‘big bang’ approach.”

**Consider the Following Measures While Choosing a Data Solution**

A peer contemplates, “A healthy data mart is built by establishing a set of firm business and IT requirements (a clear vision) upfront, modular-based and documented.” He informs that a data mart is not the same as a data warehouse or an operational data store. The peer further opines, “You may want or need to consider a strong data integration platform considering the ‘3 V’s’ of your organization’s data.”

Regarding data dictionaries, the peer recommends to consider:

- Buying before building.
- The degree to which you need to capture for data lineage. Do you need to go back to the moment the data enter your domain or your data mart?
The peer member further advises, “With any platform consideration, take into account the resource capacity and skill sets. Buying involves upfront costs and usually fewer resource requirements but typically a mountain of support and out-of-the-box capabilities. Whereas, building comes with lower upfront costs and more flexibility, but requires more resources.”

**Engage With Relevant Stakeholders to Understand Your Data**

Stressing the importance of understanding the organization’s data, a Peer Connect member says, “Before you think about what the structure and architecture of your organization's data needs are, analyze:”

- “What do you want to do with the data?”
- “What information products are holding you back as an organization?”

While sharing his experience of working in an NFP environment, the peer adds, “I often find that it is a harder sell to the rest of the organization because there is no direct correlation between initiatives and bottom-line profit/sales.”

Also, for every technical endeavor that the organization intends to launch (warehouse/lake/catalog and so on), the peer recommends launching another one aimed to address skill gaps with regard to data in the organization. Per the peer, “It is important to educate relevant stakeholders to ensure more ‘data savvy’ professionals across the organization.”

“How will your people work with data? Are they capable of adopting analytics platforms and tools, or do they need some help?” enquires the peer. He highlights that it is essential to consider both the professionals and end users during the process. Therefore, the peer adds, “Whatever you put in place, it should be something you can ‘live with’ in the longer term. The governance is crucial as it will determine the decision rights and define the stakeholders.”

The peer insists that the above-mentioned procedures are fundamental considerations at the outset of any D&A initiative. He summarizes the process in the following steps:

- “Develop your strategy around the outcomes you want to see and the vision you want to share.”
- “Look at the personas within the business, determine what moves them on and get them where they need to be. The applications and technology should fit around that, not the other way.”
Obtain a Clear Business Perspective

Next, define the business purpose of your to-be data mart and identify the business users, recommends a peer. He instructs to ensure that the business purpose is narrowly defined to a key business need for information concerning the organization. The peer further states, “You then use your data glossary and existing application data dictionaries to determine which existing data will best fit that need for information.” According to the peer, it is important to plan on having a staging area for the selected application data for it to go logistically for processing before preparing for the data mart.

A peer, who is a D&A director in his organization, emphasizes gaining a clear understanding of the business motivation behind the data request right at the start. He states, “It is always good to start with a clear ‘Why.’” The peer shares asking some questions that should be answered to gain additional details like:

- Getting a consolidated enterprise view of the business
- A quick reporting solution for a particular area of business
- The need for integrating data, and operationally sharing it with internal and external stakeholders
- Replacing the legacy source application

These are not discrete choices, the peer opines. However, according to him, “They will guide you to evaluate your choices and investment priorities. Also, having clear ‘principles and guidelines’ will help your team to understand the boundary lines while looking for a solution. For example, cloud-based or on-premises, federated or centralized and so on.”

Another Peer Connect member directs to start with how you answer questions/ingest information currently in your organization and whether there is a desire to change that or not. She says, “Curated truth is very different from open self-service. The good news is that there are common foundations that account for starting while figuring that part out.” The peer also advises starting with some baseline questions that the leadership and staff would like to obtain easy, real-time answers for — there are often some accessible lifts that can validate/invalidate potential approaches.
She adds that there is no data that lives outside of context; therefore, she suggests to match the organization's plans and deliverables with its culture as it will be the key to continued investment and success. Addressing the current environment around data solutions, the peer informs the thread initiator, “There are many good tools out there these days; hence, you already are in a better spot than some of us with massive legacy stacks to deal with. Ensure to build with staff investment in mind as many vendors will stress on automation, but humans are key to ensuring the data is usable, not just landed.”

Another Peer Connect member, whose organization successfully made considerable progress in the data analytics space, says, “We now have an informatics environment that stores almost all our operational and analytical data. Before I dive more deeply into what you need, I would examine how the data will be used, who will use it and what they want along with their technical appetite for being able to retrieve it.” Further, the peer recommends validating the request at the highest possible level in the executive chain. The consumption requirements will lead directly to the architecture and design.

**Determine Your Data Architecture and Strategy**

A peer member, who is an information management professional in an insurance organization, states, “A year ago, we were in the same boat as you are now. Disparate sources resulted in manual consolidation of business measures.” The peer explains that his organization defined the data architecture and strategy based on the pain points that are needed to be addressed. His organization took an incremental build approach on a modern data platform because they wanted to deliver incremental business value in an iterative fashion.

The peer explains, “The waterfall approach to building has its benefits (and drawbacks) but also creates a waiting period for the business benefits.”

**Define Processes**

Another way to start the work is to define the main processes and their connection points, recommends a peer. According to him, the applications should support processes and the master data should be known in the integration points. He says, “Master data should be updated only in one place (even if it is copied elsewhere). Maybe you need to do this in parts, like one main process at a time.”
Key Take-Aways

- Understand the data requirements for your organization and build a data glossary consisting of all crucial business terms.
- Set up clear IT and business expectations to create a healthy data mart. Ensure to examine the “3 V’s” while selecting a strong data integration platform.
- Educate all stakeholders on the organizational data for better understanding and smooth adoption of data analytics platforms and tools.
- Identify the business users and gain a clear understanding of the business purpose for your data mart. This would guide you in managing your investment priorities and solutions for your organizational data management.

About Peer Connect Content

Peer Connect is a private community in which Gartner clients can exchange insight and advice on their mission-critical priorities.

Peer Connect Perspectives are summaries of opinions and points of views expressed by Peer Connect community members. See the full Peer Connect discussion from which this document was summarized.

Recommended by the Author

Some documents may not be available as part of your current Gartner subscription.

Best Practices for Designing Your Data Lake

Building Data Lakes Successfully — Part 1 — Architecture, Ingestion, Storage and Processing

Building Data Lakes Successfully — Part 2 — Consumption, Governance and Operationalization

Classifying Data Requirements for Agile Processes and for Durable, Adaptable Systems