The Role of a Master Data Governance Hub
TABLE OF CONTENTS

Introduction 3
Evolution of MDM Architecture 4
  Initial Deployment 4
  Evolution to Pub / Sub 4
  Real-Time Service Calls 5
  Data Governance Hub 6
Comparing DG-Hub and Operational MDM Hub 7
Summary 8
About Gaine Solutions 9
Introduction

Over the past decade, Master Data Management (MDM) technology has evolved from its experimental beginnings into an accepted component within enterprise architecture. Whereas early MDM deployments played a supporting role to BI or CRM, more mature MDM deployments have become mission-critical operational systems in their own right – we describe this evolved state as an Operational MDM Hub.

An Operational MDM Hub (Op-Hub) typically has a real-time or near real-time role in business processes such as order entry, billing, commissions or help desk support. As the Op-Hub becomes embedded in more business process, there is a correlated limit to the flexibility of the Hub and an increased emphasis on performance. In many ways an Op-Hub no longer has the characteristics that early MDM Hubs provided to the enterprise, such as central cleansing, matching and validation, or to be a neutral staging area for new enterprise data.

In this paper we describe how a Data Governance Hub (DG-Hub) can augment operational MDM to resurrect the flexibility that has been sacrificed in many MDM deployments.
Evolution of MDM Architecture

Here we look at the most common stages of MDM evolution and some of the associated characteristics of each stage.

**Initial Deployment**

Many initial deployments begin with a “disconnected hub” that takes batch files from contributing systems to produce either a “golden record” (repository type hub) or a central cross-reference (registry type hub).

The end-consumer of this simple deployment was often a BI database or a single system such as CRM. The consuming system would typically receive an extract from the hub and the processing of this extract was totally “disconnected” from any hub processing.

In these deployments, operational systems are shielded from performance characteristics of the hub. Extracts are often generated on-demand with little or no impact on other systems linked to the hub.

**Evolution to Pub / Sub**

The logical next step in the evolution of the hub architecture was a move to a publish and subscribe architecture where operational systems may either publish data to the hub or subscribe for updates from the hub. Through this architecture, systems can be contributors or consumers of information or both. Internal systems may become dependent upon updates from the hub as they “listen” for updates generated by other business processes.

Within this architecture subscribing systems may require a certain turnaround time from the hub to deliver subscription files to meet the demands of a particular process. System availability and performance become critical factors of the hub and consequently the hub is more sensitive to configuration changes and the addition of new data sources.

The hub has now lost its ability to act as a neutral staging area for new data files and has become more of a “hardened” operational system.

---

**Figure 1 - Typical Initial MDM Deployment**

**Figure 2 - Evolution to Pub / Sub Architecture**
Real-Time Service Calls
As the hub becomes established within the architecture as a trusted source of enterprise information there will be systems that want to access this data in real-time. Most modern MDM hub technology supports web service calls to both provide and accept information.

Within this service oriented architecture (SOA), performance and availability become even more critical. The real-time hub must be judicious in the type and number of transformations and validation applied to inbound data. Most real-time hubs apply very light processing to inbound data to maximize the performance characteristics of the hub.

This real-time hub architecture represents the current state of mature MDM deployments. While the hub has maximized its value to operational processes, it has also become less tolerant of change and more sensitive to performance considerations. These factors restrict the Op-Hub from being used as an effective Data Governance Hub where flexibility and extensive validation processing is required.

Figure 3 - Service Oriented Architecture
Data Governance Hub

The introduction of a Data Governance Hub addresses the limitations inherent in a real-time MDM hub deployment.

The DG-Hub is typically loosely coupled to other enterprise applications through a pub / sub architecture. It is unlikely that the operational MDM hub will contain all of the data required for data governance processing and it is common that the DG-Hub will require additional source information to apply a broader scope of data validations.

The DG-Hub is able to apply extensive validations and policy checks on the master data because it operates as a background process and no downstream systems are reliant upon a real-time response or subscription file from the DG-Hub. Outputs from the DG-Hub are data governance workflows and inputs gathered from these workflows. Because the DG-Hub is not providing operation systems with enterprise keys or a golden record, it is more suitable for loading new data files without the same concerns for performance impacts or changes to the master files.

The flexibility of the DG-Hub is a natural extension of an enterprise MDM strategy. The DG-Hub simplifies the role of the MDM Hub and provides additional capabilities without compromising performance or system availability.
### Comparing DG-Hub and Operational MDM Hub

<table>
<thead>
<tr>
<th>DG-Hub</th>
<th>MDM Operational Hub</th>
<th>Benefits of the DG-Hub</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contains very rich source data with more attribution that a typical MDM Hub.</td>
<td>Contains a limited set of “enterprise” attributes.</td>
<td>Allows more extensive and complex validations on the master data.</td>
</tr>
<tr>
<td>Extensive workflows for multiple data governance processes.</td>
<td>Simple workflows for the most common data steward activities such as resolve suspect match or conflicting attribution.</td>
<td>A single record loaded into the DG-Hub may generate multiple independent workflows. As no system is waiting for a response, these workflows can be long-running and inter-dependent.</td>
</tr>
<tr>
<td>Manages an organizational structure for the entire data governance organization.</td>
<td>Manages only the data steward roles and data access roles.</td>
<td>Data governance involves multiple stakeholders not just data stewards and councils. The DG-Hub allows the extended organization to be modeled and used in workflow processing.</td>
</tr>
<tr>
<td>Manages data policy and specifications.</td>
<td>Manages rules for the MDM attributes under management.</td>
<td>Data policies typically extend beyond the attributes within an MDM Hub. The DG Hub provides an extended data specification repository to manage more extensive data policies.</td>
</tr>
<tr>
<td>Provides a central repository to monitor the efficacy of the data governance organization.</td>
<td>Provides a central place to manage the master data content.</td>
<td>Improving master data over time requires that the data governance processes and policies are subject to a continuous improvement. The DG-Hub provides a view of these processes and outcomes.</td>
</tr>
<tr>
<td>Focus is on integration with the desktop of the business users.</td>
<td>Focus is on integration with enterprise systems architecture.</td>
<td>DG stakeholders span the entire organization and may have a role to play in the management of this enterprise data. A familiar desktop experience is critical to successful data governance adoption.</td>
</tr>
</tbody>
</table>
Summary

A Data Governance Hub provides a new level of flexibility to a mature MDM deployment. The DG Hub relieves many of the major restrictions inherent in an Operational MDM Hub deployment and restores the advantages of a disconnected system without compromising the benefits of MDM to operational processes.

- A Data Governance Hub is simpler to deploy and support because it isolates validations and workflows from operational processes.
- New data can be loaded into the DG Hub without concerns for impacting the operational environment.
- Manage data policy and specifications for a broader scope of master data.
- Manage a data governance organizational structure for use in workflow processing.
- Provides a central repository of data governance processes and user interactions.
About Gaine Solutions

Gaine is an Enterprise Data Management specialist, creating value for its clients through a range of business services delivered in an on-demand commercial model. The unique Gaine approach accelerates time-to-value and minimizes the time, cost and risk inherent in data intensive initiatives. Gaine is the developer of the widely deployed Master Data eXchange (MDX) platform and provides EDM services to some of the world’s largest and most respected Global 2000 organizations through its offices in San Francisco, Dallas, New York and Cape Town.

For more information please visit www.gainesolutions.com