Infuse Your Data Hub Strategy With Data and Application Integration

FOUNDATIONAL **Refreshed:** 20 May 2019 | **Published:** 22 December 2017 **ID:** G00343327

Analyst(s): Eric Thoo, Ted Friedman

Integration is fundamental to executing a data hub strategy. Data and analytics leaders must determine the role of both data integration and application integration technology for connecting endpoint systems to their data hubs.

FOUNDATIONAL DOCUMENT

This research is reviewed periodically for accuracy. Last reviewed on **20 May 2019**.

Key Challenges

- Data hubs typically require a range of integration styles, but most organizations have done little to coordinate the various integration tools and teams across their enterprise.
- Organizations often fail to derive maximum value from their integration technology investments because they use them only in one-off application-specific and use-case-specific deployments.
- Adoption of a data hub strategy compels you to reach broad agreement on data formats and semantics, and to find ways to use data integration tools in synergy with application integration technology.

Recommendations

As a data and analytics leader focused on modernizing your data management strategy, you should:

- Broaden your data hub's applicability by providing a range of data integration and application integration capabilities for connecting endpoints to the hub.
- Maximize the reuse and value of your existing integration technology investments by broadening their applicability in the context of a data hub.

- Foster reuse, semantic consistency and transparency of data flow by adopting common approaches, standards and metadata sharing across all the styles of integration that support your data hub architecture.

Table of Contents

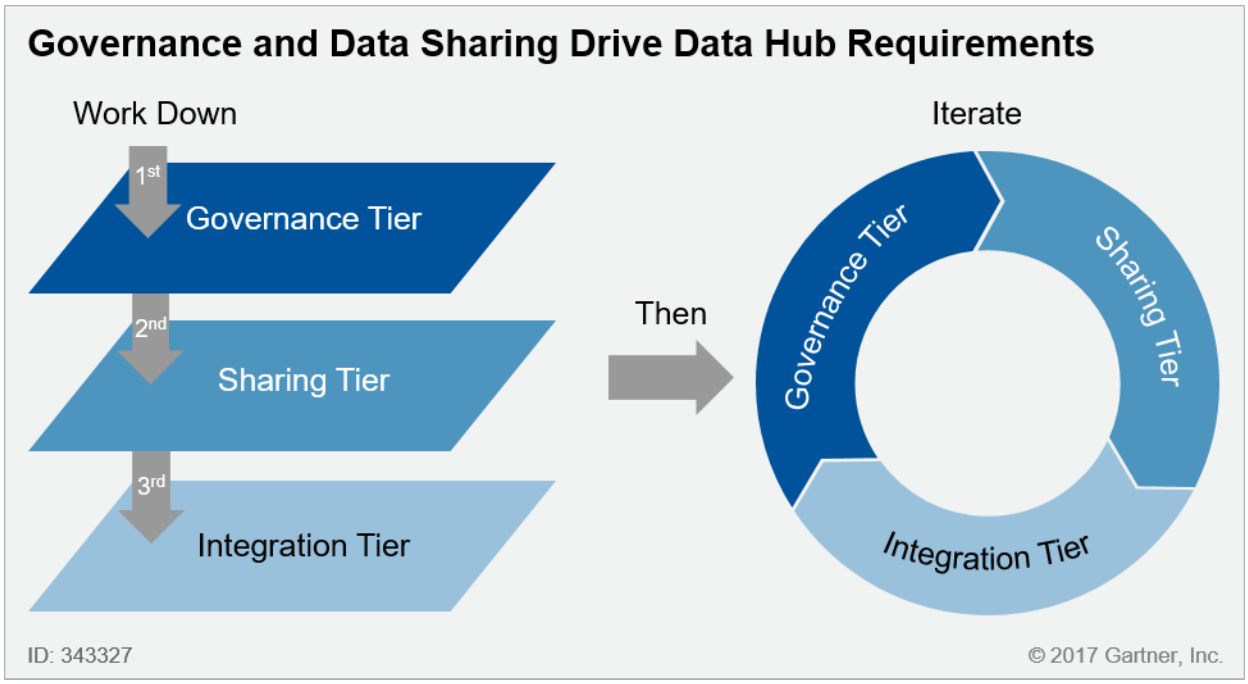
List of Figures

Introduction

Integration is fundamental for providing data hubs with efficient, frictionless ways to share data in order to meet consumption requirements across an enterprise. The diversity of data resources scattered across applications and data structures, and the complexity of data flowing between them, means that the integration requirements of a data hub strategy are extensive.

A data hub strategy is one that recognizes three important, related, dependent but distinct efforts (see Figure 1).

Figure 1. Governance and Data Sharing Drive Data Hub Requirements



Source: Gartner (December 2017)

The most important effort, but the least pursued and the hardest so far, is to determine the appropriate form and structure for the governance of data across an enterprise. This "governance tier," when established, has spawned new solutions, including master data management (MDM) and information stewardship solutions.

A second effort, to establish a "sharing tier," explores the business requirements for access to shared data. This effort follows on from the governance effort, but is distinct from it, since access to data does not imply automatic involvement in governance — many organizations focus on sharing data but ignore governance concerns.

Once governance and sharing requirements are established, the work of establishing an "integration tier" — the third effort — can be informed more effectively. In a complementary manner, integration approaches and capabilities may inform the work of governance, so this is not always a top-down approach.

This is where we pick up the story in this report. (For more details, see "Use a Data Hub Strategy to Meet Your Data and Analytics Governance and Sharing Requirements.")

Data hubs that enable the sharing of data between diverse sources and consumers of data entail a lot of integration complexity. This is because the data being shared and the producing and consuming systems can have wildly varying characteristics. For example, some data is produced in bulk at a single point in time, whereas other data exists in always-on event streams. Some data is highly structured, and some far less structured. Also, some consumption points require data on-

demand, while others need to receive data on a scheduled basis. The various combinations of these (and other) characteristics mean that a single style of integration technology cannot address all requirements for data flow and accessibility in a typical data hub deployment.

Data and analytics leaders pursuing data hub projects must address data and application integration requirements as a whole, not in isolation. This involves managing associated competencies from an enterprise perspective. Multiple styles of integration, coordinated and applied in a controlled and flexible manner, are at the core of a data hub architecture.

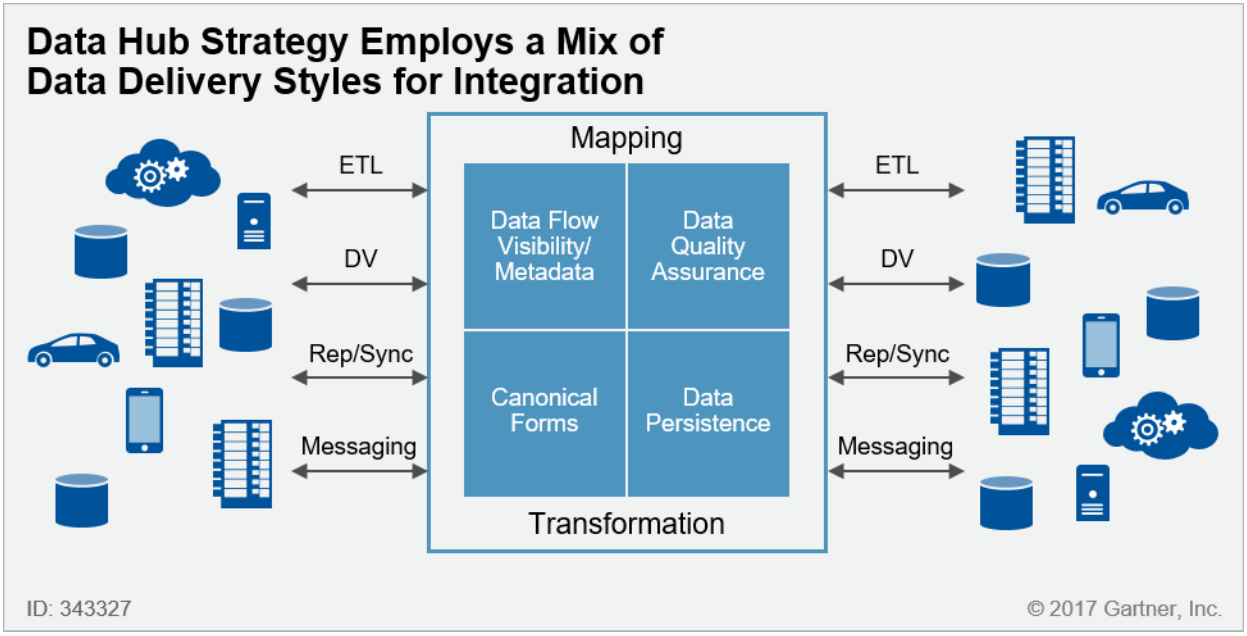
Analysis

Broaden Your Data Hub's Applicability

Data hubs require a balanced mix of data delivery styles for integration (see "Modernize Your Data Integration Architecture for Digital Business by Combining Data Delivery Styles"). Efficient data integration strategies across applications, Internet of Things networks, enterprises and ecosystems are crucial to the pursuit of a data hub architecture that helps determine effective mediation of semantics (for governance and sharing). Integration capabilities are central to a data hub because they are needed for ingesting, integrating and provisioning data for a range of producing and consuming applications and data stores. A wide range of technologies may be used to implement the key components of a data hub architecture, including various integration technologies (see "Implementing the Data Hub: Architecture and Technology Choices").

Central to a hub architecture must be technology to implement data flow — that is, to get data into, or connect data to, the hub from producing applications and other data sources, and then to provide it to consumers. The flow of data into, through and from the hub can be supported by a range of data integration capabilities. In fact, all styles of data delivery for integration can be relevant, and most hub deployments will use various styles to support the varying latency, granularity and persistence needs of the different types of data being shared (see Figure 2).

Figure 2. A Data Hub Strategy Must Employ a Mix of Data Delivery Styles for Integration



DV = data virtualization; ETL = extraction, transformation and loading; Rep = replication; Sync = synchronization

Source: Gartner (December 2017)

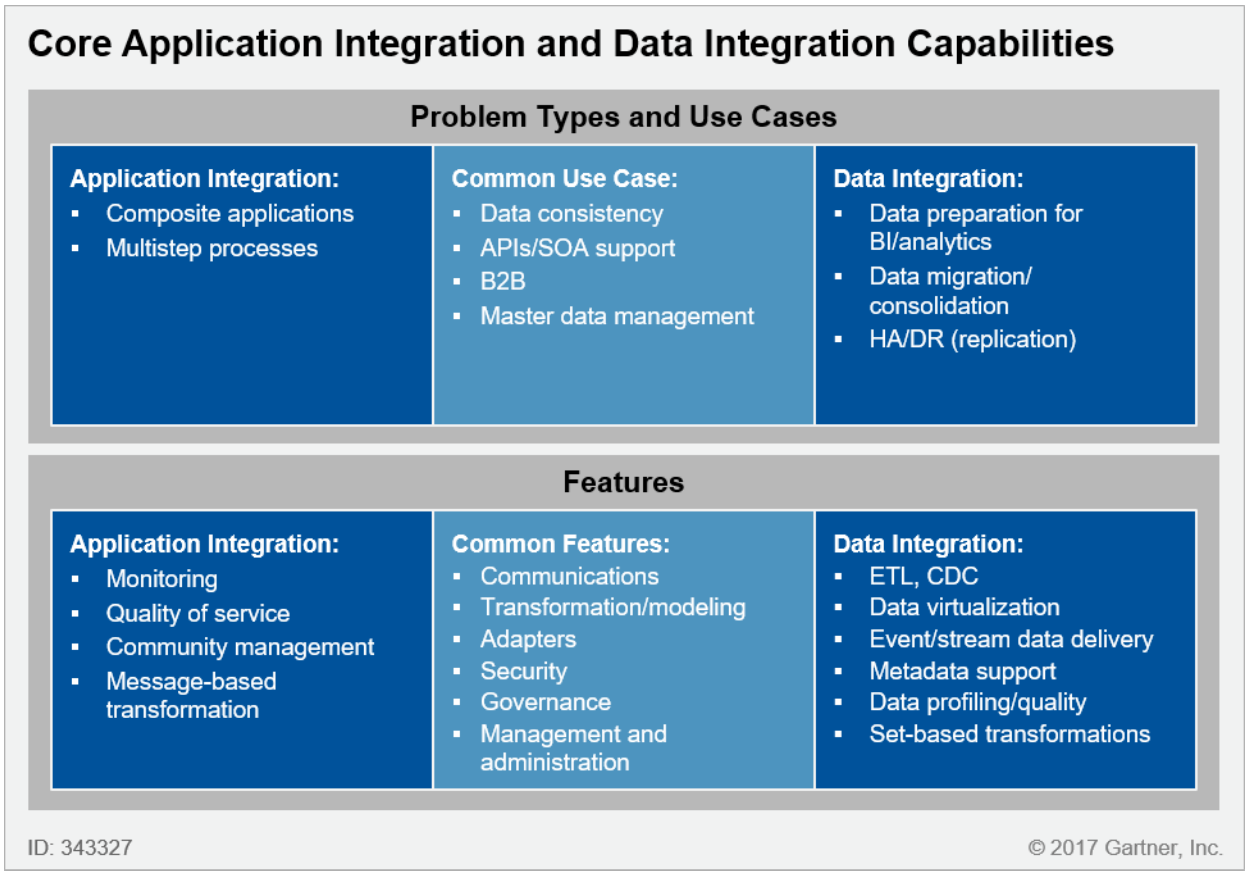
Recommendations:

- Use a mix of data delivery styles as part of a data hub strategy to extend the flexibility and adaptability of your data integration infrastructure.
- Enable broad applicability of your data hub by providing a range of synergistic data integration and application integration capabilities for connecting endpoints to the hub.

Maximize the Reuse and Value of Your Existing Integration Technology Investments

As a data and analytics leader, you should start by identifying common capabilities within data and application integration approaches that can be used to build and manage your data hub. Because integration capabilities for a data hub approach draw on concepts and capabilities from the discipline of application integration, integration efforts should draw on the capabilities that organizations typically already have for bringing data and applications together. To amass data and orchestrate flows in data hub implementations, neither the applications nor the data for integration requirements should be regarded as stand-alone (see Figure 3).

Figure 3. Core Application Integration and Data Integration Capabilities



API = application programming interface; B2B = business-to-business; BI = business intelligence; CDC = change data capture; DR = disaster recovery; ETL = extraction, transformation and loading; HA = high availability; SOA = service-oriented architecture

Source: Gartner (December 2017)

A common set of application and data integration capabilities must be pursued cohesively to ensure the agility of data flows for the varying use cases that arise for a data hub.

A data hub may, for example, synergize data flow by identifying a change (through change data capture) in a customer's buying characteristics, and pass the captured data to message queues to enable a campaign application to offer incentives. This enhances the customer experience via data and application integration.

Both data integration and application integration approaches are necessary to solve common problems with a data hub strategy, to synergistically support the same business strategy and goals. Common goals include support for data consistency, API and service-oriented architecture (SOA) initiatives, B2B integration and MDM. Data hub projects supporting systems of innovation that draw on mobile, web and postmodern ERP applications need a mix of data integration and application integration techniques to gain just-in-time insights and respond to continuously evolving events.

As a data and analytics leader responsible for data management and integration initiatives, you need to modernize how you provision integration capabilities to power a data hub. One potential option is a hybrid integration platform (HIP) approach (see "Innovation Insight for Hybrid Integration Platforms"), one focused on bringing together data integration and application integration to align disciplines and technologies that can benefit a data hub architecture. A data-hub-related integration effort is one scenario that HIP capabilities increasingly need to support. Data integration tools that operate well with related platforms — such as enterprise service bus (ESB), integration platform as a service (iPaaS), and event stream processing platforms — will become foundational elements of HIP-enabling technologies (see "Converging Data and Application Integration: A Step Toward Pervasive Integration Using a Hybrid Integration Platform").

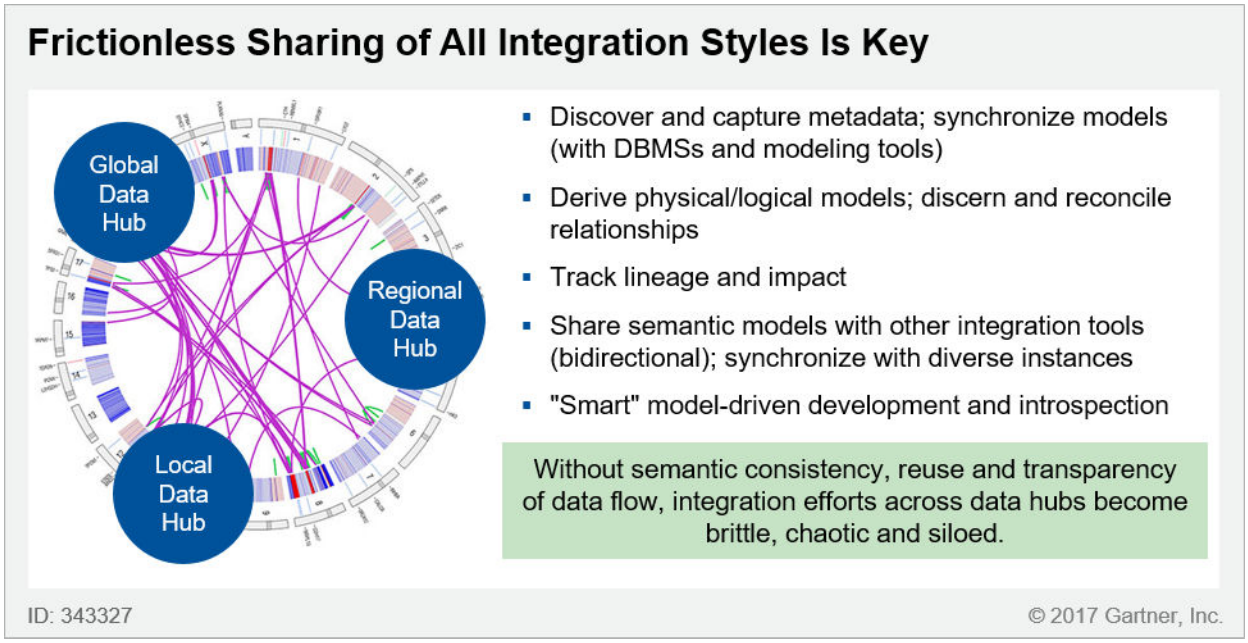
Recommendations:

- Identify an initial set of common capabilities that can be deployed in a reusable manner across integration tools, such as a common vocabulary, approaches to using data integration and application integration technology together, and shared models and metadata.
- Use data integration tools and architecture that are already established in your enterprise, in addition to augmenting existing architecture with new capabilities, where required.
- Apply common capabilities to scenarios relevant to your data hub strategy. These may include scenarios involving key application infrastructure or data management infrastructure requirements, such as to aggregate data and orchestrate business activities in support of API and SOA initiatives and MDM.
- Create opportunities to combine application integration and data integration by using a combination of capabilities, such as ESB, ETL tools, iPaaS and a framework of HIP approaches.

Foster Reuse, Semantic Consistency and Transparency of Data Flow

Tools that help produce synergy between the disciplines of application integration and data integration are crucial if data hubs are to support any data flow. You should emphasize reusable capabilities, rather than narrow support for individual applications or data structures. The capabilities must be flexible enough to switch dynamically between integration patterns, and interact with and share metadata knowledge of the landscape of application and data assets (see Figure 4). This is important because the ways in which data producers and consumers connect to the hub can differ, and may change over time.

Figure 4. Frictionless Sharing of All Styles of Integration Capability Is Key to Enabling Reuse of Data Flow Across Data Hub Projects



DBMS = database management system

Source: Gartner (December 2017)

To support a data hub, the use of potentially different, complementary integration technologies is common and expected. To work coherently with heterogeneous integration tools, and support semantic consistency between tools and across data hub projects, metadata capabilities are crucial in a number of ways:

- Openness to share metadata with diverse integration tools. Example: the use of data virtualization technology to expose federated data flows as inputs to a message-oriented process supported by an ESB.
- Ability to insulate data hub projects from the underlying deployment complexities of physical data structures and specific integration tool choices. Example: "smart" metadata-driven development and introspection of data sources, removing the need to understand the structure of (or interact directly with) the underlying applications and data environments.
- Abilities to monitor, track, analyze and share metadata — a focus of metadata management solutions — enable the inventory of data assets and metadata, which can support design and development of integration flows for diverse data needs (see "Data Catalogs Are the New Black in Data Management and Analytics").
- By coordinating data flow via the hub's enabling technologies, the organization gains clarity about data lineage and relationships between sources and targets. Unlike project-specific approaches in which metadata representing data flow is not visible across projects, the hub

approach must employ a shared metadata approach through which cross-data-hub projects can obtain insights to ensure the semantic consistency of diverse data flow implementations.

Joint practices and collaboration across teams will increase the synergy of the data and application integration skills and capabilities necessary for data hubs. Although many organizations already have a collection of data integration and application integration technologies, they are often supported by differently skilled staff for each tool, in which case differing approaches and capabilities lead to inconsistent delivery and approaches. To optimally exploit a data hub strategy, you must support a full range of data flow service levels and types, and flows of data assets. You should do so by bringing data integration and application integration initiatives and resources into a unified approach for greater leverage, consistency and collaboration. In some cases, a team focusing on integration competency is already in place.

Recommendations:

- Assess how your data and application integration technologies need to work together. Specifically, consider the common connectivity, design environments, administrative tools and metadata, and the aligned skills that will be shared by various types of data hub project.
- Aim for accessibility of metadata and lineage of integration constructs in your data and application integration tools, so that models and runtime metadata can be imported, exported and otherwise accessed beyond the boundaries of individual technologies.

Gartner Recommended Reading

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

“Data Hubs: Understanding the Types, Characteristics and Use Cases”

“Innovation Insight: The Digital Integration Hub Turbocharges Your API Strategy”

“Use a Data Hub Strategy to Meet Your Data and Analytics Governance and Sharing Requirements”

“Implementing the Data Hub: Architecture and Technology Choices”

“Data Hubs, Data Lakes and Data Warehouses: Choosing the Core of Your Digital Platform”

GARTNER HEADQUARTERS

Corporate Headquarters
56 Top Gallant Road
Stamford, CT 06902-7700
USA
+1 203 964 0096

Regional Headquarters
AUSTRALIA
BRAZIL
JAPAN
UNITED KINGDOM

For a complete list of worldwide locations,
visit http://www.gartner.com/technology/about.jsp