

LOGICAL DATA WAREHOUSE AND MULTIMODAL GOVERNANCE

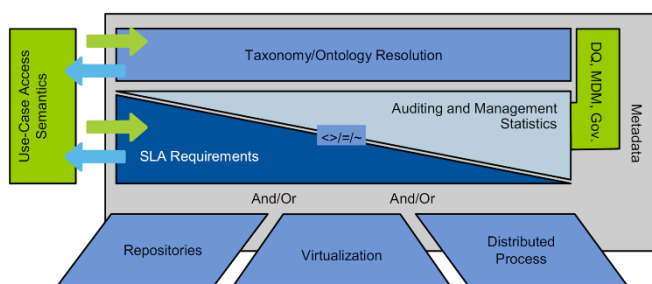
CAUGHT BETWEEN COMPLIANCE, AGILITY AND
FLEXIBILITY

WHITEPAPER

LOGICAL DATA WAREHOUSE ACCORDING TO GARTNER

The logical data warehouse concept was introduced by Gartner in 2012¹ and has been continuously developed and promoted since then. It provides recommendations on how organizations can build a demand-driven, modern data management capability for analytical applications. The logical data warehouse is food for thought, not a recipe. Companies need to translate the concept into a suitable architecture according to their requirements and capacities.

Now, what does a logical data warehouse consist of? The original visualization of Gartner's concept is shown in figure 1.



DQ = data quality; Gov. = governance; MDM = master data management

Figure 1. Logical data warehouse according to Gartner

Without wanting to interpret the diagram in detail, there are some aspects worth mentioning. The gray block starts in the middle of the three slanted blue blocks at the bottom. This suggests that the logical DWH is more about the upper part, which emphasizes use, access, and meaning of the information rather than data storage and transport technologies.

So, a logical data warehouse is not the same as a virtual data warehouse? Remarkably, "logical" is often equated with "virtualized". However, the picture clearly shows that a logical data warehouse can use a mixture of technologies. "Repositories" means using a data warehouse, data marts, or other databases for integration, while "virtualization" and "distributed processing" use middleware. One or two out of three, or all three together make up the infrastructure base

HIGHLIGHTS

- The logical data warehouse concept is food for thought, not a recipe. It must be adapted to each company's situation.
- Classical data warehouses, data lakes and virtualization solutions are not competing approaches, but rather cooperating components of a logical data warehouse architecture.
- Different data quality, maintainability, flexibility and agility requirements not only require different technologies, but also different governance approaches.

for decomposing data warehouse functions into logical components: an ecosystem of collaborative solutions.

The term "taxonomy/ontology resolution" used in the upper part of the illustration appears abundantly abstract and intangible. Essentially, it is about giving the data meaning and context. In the following sections, we will discuss the according Simply approach.

Various applications, from financial reporting to campaign management, place different demands on the same information. Some of these requirements may match with the features and SLAs of a certain IT system, while others may be in conflict. Solutions are needed that help represent data in a form that exists primarily at a logical level but is implemented through a combination of different physical databases and data transport components depending on the needs at hand.

In our opinion, this is the core of the logical data warehouse concept. It must always be questioned how to deal with the sometimes-contradictory accessibility, availability, maintainability and data quality requirements in a heterogeneous data landscape.

¹ Mark A. Beyer and Roxane Edjlali: Understanding the Logical Data Warehouse: The Emerging Practice. ID G00234996, Gartner, June 2012.

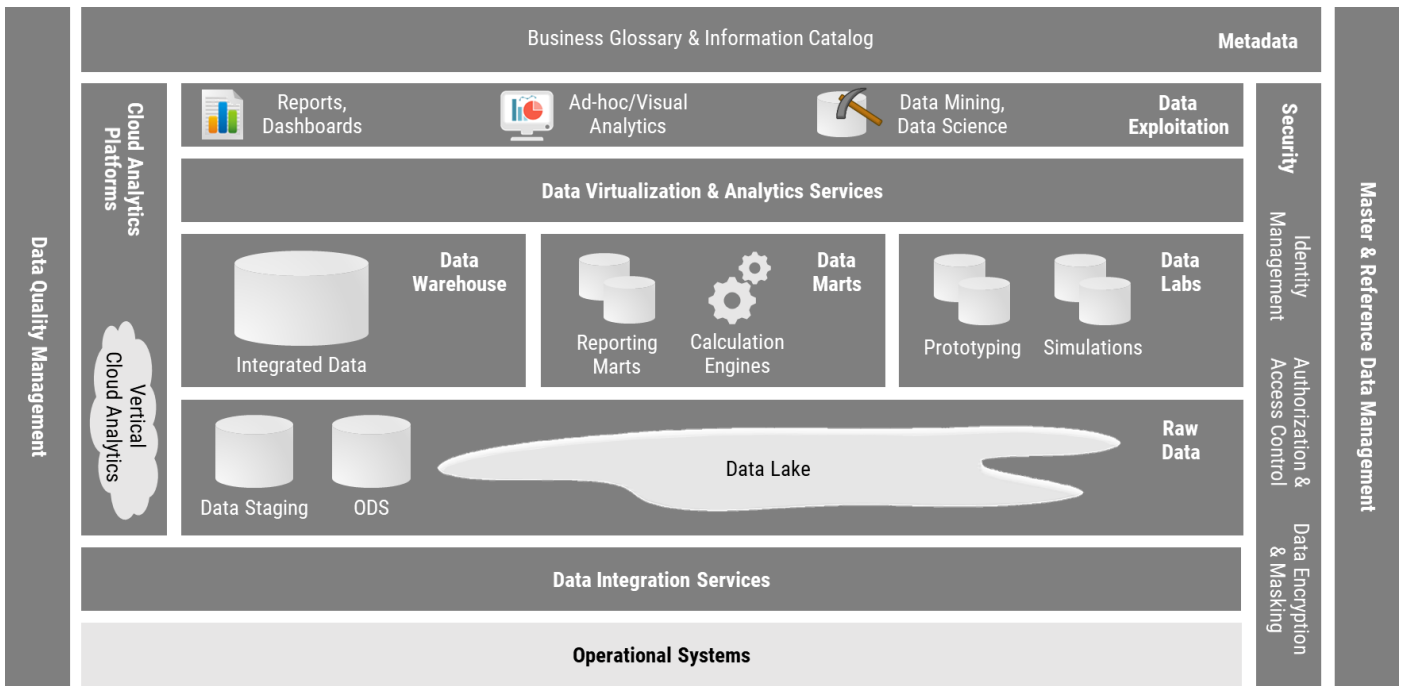


Figure 2. Simplity's data & analytics architecture

SIMPLITY'S DATA & ANALYTICS ARCHITECTURE

To operationalize the presented Gartner concept, Simplity has developed a reference architecture for data management in analytical applications. It is shown in figure 2. As with the general concept the technologies listed are intended as building blocks that can be combined based on the requirements to a tailored solution. Architectural approaches such as classic data warehouses, data lakes and virtualization are not to be understood as competing solutions, but as complementary components of a logical data warehouse architecture.

The architecture components are glued together through metadata shown in the figure as a business glossary and information catalog. The best-practice Simplity approach to this is an integrated business information model that is both understandable enough to serve as a glossary of business terms and formal enough to support an implementation in IT; a more detailed discussion is given in the next section.

Together with the metadata, data quality management, data security as well as master and reference data management form the governance framework of the architecture.

MULTIMODAL DATA GOVERNANCE

Different requirements for maintainability, traceability, flexibility and agility require not only different technologies, but also different data governance approaches. For this purpose, Simplity has defined three archetypes as shown in figure 3. In reality, of course, hybrids are possible and necessary.

Strict Governance. For applications such as regulatory reporting, where data quality and particularly traceability are of the highest priority, a detailed conceptual data model ("business information model") has proven to be successful. This approach allows to unambiguously and precisely

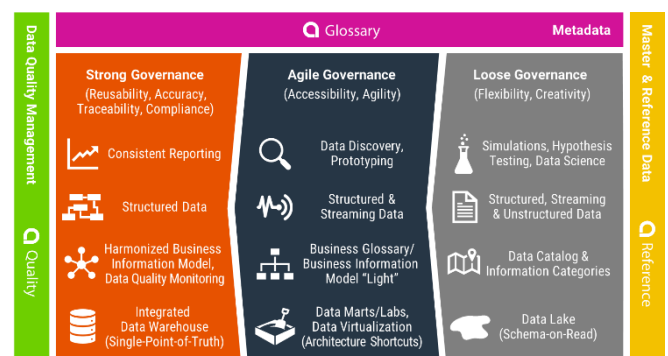


Figure 3. Multimodal data governance

represent actual business objects such as customers, products, or transactions. These objects, their properties as well as relationships are structured in a comprehensive information model that can be aligned and shared across the different business units. KPIs will be broken down to an atomic level, these constituent parts need to be represented in the model as well.

The business information model does not only serve as common language across business units, but also as a basis for the implementation, i.e. technical data models will be mapped against it. This is imperative to ensure that model as well as mappings will be kept up-to-date in order to prevent divergence between specification and implementation.

Establishing and maintaining mappings between data requirements and source as well as target systems helps to ensure consistency and traceability. It will eventually lead to improved data quality by reducing semantic ambiguity. The respective business units are not only continuously involved throughout the entire data life cycle, but actually in the “driver seat”. As a result, all stakeholders – be it in business or IT – will have a common understanding of the requirements and data. Furthermore, time-to-market as well as implementation costs are going to be reduced.

The fully-fledged business information model approach provides harmonized data definitions as well as data lineage, but also needs sufficient time, commitment and resources for implementation. The architectural equivalent to the “strict governance” concept is an integrated data warehouse providing a single point of truth.

Agile Governance. In some cases, agility and time-to-market are more important than enterprise-wide alignment. This requires a less rigid governance approach. While the business information model is still the preferred vehicle, the number of business units involved is usually restricted, resulting in “local” attributes and data definitions. For example, Marketing could use strictly governed data for financial reconciliation with Controlling, amended in an agile manner by data that is not used by other business units such as contact history.

In terms of architecture, some shortcuts (such as bypassing the data warehouse core layer) for data with less need for alignment are usually tolerated. While data lineage is still possible, a physical single point of truth no longer exists.

“Loose” Governance. When in the realm of exploration and data science, where you can fail, but need to “fail fast” and the value of a certain data set is not known beforehand, governance must be limited in order not to restrict flexibility and creativity.

Instead of developing a detailed business information model, data sets (e.g. files in a Hadoop-based data lake) are clustered into high-level data categories, which in turn can be linked to strongly governed data areas. These data categories (e.g. “PII according to GDPR” or “risk relevant data”) are linked to physical data sets based on a tagging approach.

While this approach does not provide detailed data lineage, it nonetheless enables further analysis which subsequently can lead to transferring data into a more strongly governed area.

ACCURITY DATA GOVERNANCE SUITE

As shown in figure 3, Simplity’s data governance software suite Accuracy is focused on supporting the various governance concepts laid out in this whitepaper. Centerpiece of the governance suite is Accuracy Glossary for the creation and maintenance of a business glossary (resp. business information model) as well as the mapping against technical source and target data models (information catalog).

The second module is Accuracy Quality to measure and monitor data quality from a business point of view. Tightly integrated with Accuracy Glossary, business-driven data quality rules can be defined based on the business information model, thus ensuring consistency while drastically reducing efforts. Eventually, Accuracy Reference supports all aspects of reference data management, e.g. maintaining product catalogues or reporting hierarchies. In the near future, the software suite will be extended by a module for audit-proof data corrections and adjustments.

HOW CAN WE HELP?

- Simplity offers services around data management and analytics and has supported numerous leading financial institutions in Central Europe in successfully implementing of such initiatives.
- The service portfolio ranges from strategic consulting in the areas of architecture, data governance and data quality management to technical implementations of data warehouses, data lakes and master data management solutions.
- The Accurity software product created by Simplity can be used both as a supporting tool in Simplity projects and as an independent data governance solution for metadata, data quality and reference data management.

AUTHOR AND CONTACT

Dr. Torsten Priebe is CTO at Simplity, where he is responsible in particular for data governance and architecture. He was previously a solution architect at Teradata and headed the Business Intelligence team at Capgemini in Vienna.



Dr. Priebe is co-author of three books, has written various publications and gives presentations at international conferences. He is also a regular guest lecturer at the University of Regensburg, where he received his PhD on the integration of structured and unstructured data.

priebet@simplity.eu | +4369914166193
www.simplity.eu | www.accuracy.eu