simplity

DATA WAREHOUSE MODERNIZATION

A FIELD REPORT

simplity

„The DWH is dead, long live the DWH" – so what? Many organizations deal with the question of whether their data warehouse is still up-to-date or has already become obsolete because of Hadoop, the cloud and other new technologies. We will discuss this question based on several recently conducted modernization projects.

INTRODUCTION

There are many reasons to attempt a modernization project: high costs and a cumbersome legacy environment, lack of user satisfaction, but also new technologies promising higher benefits at lower costs.

Which of the promises can one believe? Which objections and concerns do you have to take seriously? And most of all: when is it really time to undertake a comprehensive modernization of the data warehouse landscape?

DEFINITION

However, before we address the possible motivations for a modernization and the available courses of action that are best suited, a definition must first be made: when is it really a DWH modernization and how can it be differentiated from minor changes such as a database upgrade?

Following common definitions, modernization can be interpreted in a variety of ways. The term stands for simple software updates as well as for the complete reorientation of the DWH architecture. In addition, new model-driven development methods and agile project management approaches are also part of the modernization concept.

From our point of view, all measures that enable an organization to keep its data warehouse environment relevant and competitive and to align it with new business and technological requirements can be summarized as modernization. Simple software updates that do not meet new requirements would not be enough to be considered a modernization.

REASONS

Just as no doctor would make a diagnosis or therapy suggestion without first establishing the patient's medical history, one must first deal with the symptoms when modernizing the data warehouse: if, for example, the DWH does not contain the data required by business, then a migration to a new platform will not contribute to increased user satisfaction.

Without any claim to theoretical completeness, the practice shows that the following causes must be distinguished, even if in "severe cases" several of these causes, if not all, may be co-occurring:

LACK OF PERFORMANCE

Of course, complaining about long response times will often be due to an outdated or undersized platform, which is relatively easy to fix with a hardware upgrade. Often, however, such problems are due to architectural weaknesses that require more extensive remedial action.

BUSINESS COVERAGE

Another reason for the desire for modernization as well as the emergence of the dreaded "shadow IT" in the business departments is the absence of important data areas in the DWH. While an additional source system can often be integrated relatively easily, major deficits may require significant changes to the data model and the ETL processes, which in effect can amount to almost as much effort as building a new DWH.

LACK OF TRUST

When asking business users for the reasons for their dissatisfaction or even the non-use of the DWH, often inadequate data quality or lack of knowledge about the data available in the DWH and its reliability are cited. Various projects have proven that appropriate documentation such as a business information model or business glossary often deliver value much faster and with less effort than technical changes.
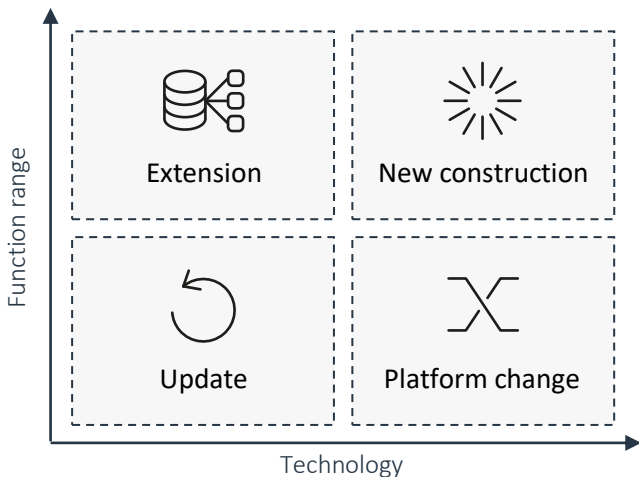
ANALYTICAL CAPABILITIES

Sometimes the desire for modernization is driven by the shortcomings of the analytics and reporting tools used. If this can be identified as the only or at least massively predominant cause, a "refurbishment" can be achieved with a new BI tool and a redesign of the reporting landscape without significant changes to the DWH itself.

As mentioned above, the reality we face at customers often shows a mixture of causes. Nevertheless, the outline may serve as an indication for which measure to take.

MODERNIZATION STRATEGIES

As soon as there is clarity and general agreement on the causes, it is necessary to opt for one of the modernization strategies available. Although sometimes radical steps may be necessary, be warned against "false prophets" (or evangelists). The DWH is not dead and nor is Hadoop the ultimate solution to every problem out there.

In general, a distinction is made between two dimensions of modernization, namely the technology and the range of functions:

UPDATE

Minimally invasive modernization e.g. using a new database version to benefit from new functionalities or even a slight change of the data model or other components to increase performance.

EXTENSION

Quite often the existing technology is still sufficient for the needs of an organization, but the data available in the DWH environment doesn't meet the analytical expectations of the business users.
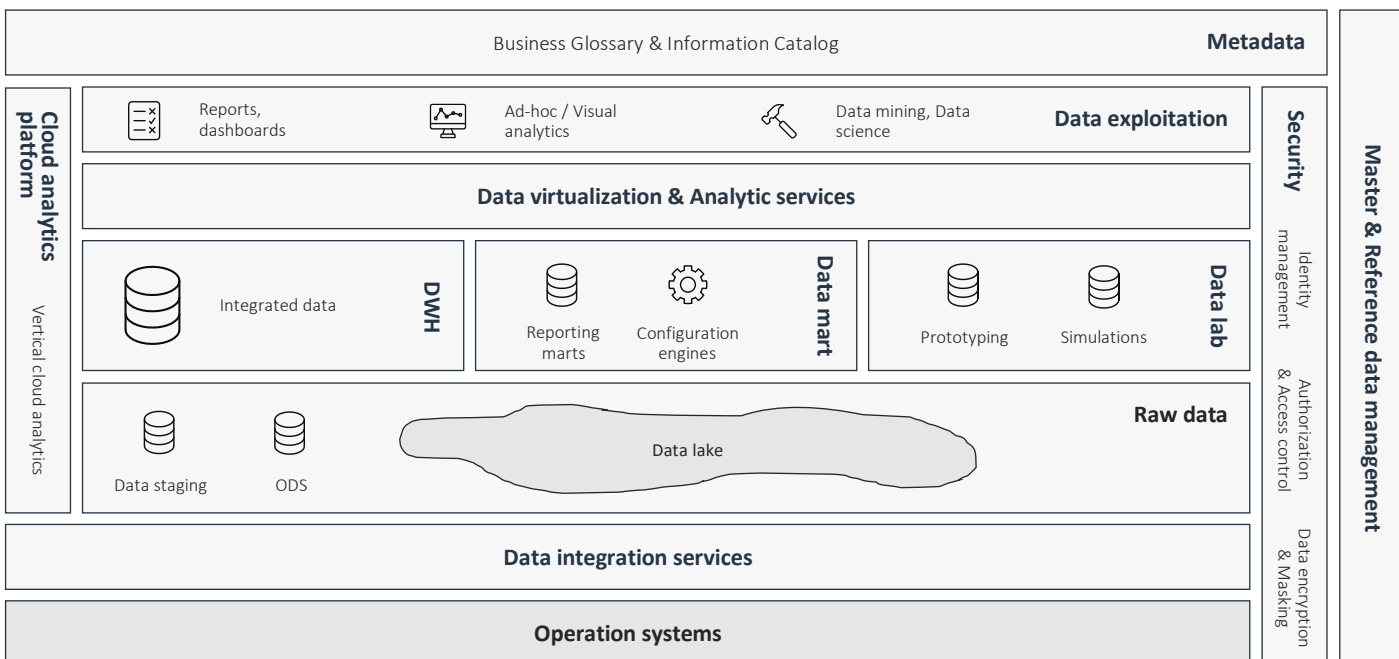
PLATFORM CHANGE

The opposite case of extension, where the data set essentially meets the requirements in terms of quantity and quality, but the technology in use has reached the end of its service life.

NEW CONSTRUCTION

The most radical form of renewal. In addition to a change in technology, the data set as well as the scope of services will be expanded. Often, one of the two dimensions will tip the scales but the other is tackled as well.

As with the causes, it is also true here that in practice certain hybrid forms are to be found. Approaches that not only allow, but actively promote a "peaceful coexistence" of different technologies and methodologies for different purposes have proven particularly promising. For more information, please refer to our recent article on logical DWH and multimodal governance.

The objective is to consider the various concepts such as classic DWH, Data Lake and Cloud or other forms of virtualization not as competitors, but rather as complementary solution components, whereby metadata (e.g. a business glossary or the business information model approach as advocated by us) serve as a logical grouping, which ensures the cohesion of the individual components.

CHALLENGES AND RISKS

Comparable to the concealment of side effects of a prescribed drug, it would be highly negligent not to address the challenges and risks of such modernization plans. Of course, most of them apply to major projects in general, but we want to highlight the DWH-specific aspects.

BUSINESS SUPPORT

Although modernization projects are often initiated by a well-meaning IT organization, it is important not to forget the early involvement of the business departments. Even if, which is by no means guaranteed, the problems are understood correctly, the acceptance of the desired solution depends decisively on the involvement of the "customers". This aspect is particularly important when it comes to redeploying departmental solutions ("Shadow IT") back into the control of the IT department.

REQUIREMENTS MANAGEMENT

As described earlier, extensive modernization measures are required above all when the business users do not have the necessary data at their disposal. Those who spontaneously start sprints out of misunderstood agility without having meticulously collected and really understood the data requirements and analytical needs will often end up in a dead end. However, this does not change the fact that multi-year, monolithic large-scale projects today - rightly - no longer find acceptance. Our approach to business information modeling offers a way out of this dilemma.

BIG BETS

Although the promises are often tempting, be warned not to put "all eggs into one basket" in terms of technology. Most organizations have diverse and often divergent analytic needs; rarely is a single technology the right answer to all these questions. Rather, it's about orchestrating an ecosystem of multiple components. What is often neglected is the lack of experience in dealing with the new technologies. This has often led to new hardware becoming a "dust catcher" in the truest sense of the word, as the necessary resources are neither available in the organization nor in the labor market.

SUCCESS FACTORS

Now that you have decided on the DWH modernization, how should you tackle such a project? Despite numerous successful projects, we have not yet succeeded in identifying a single "silver bullet". However, there are numerous solution components that have proven themselves repeatedly and in the most diverse customer situations.

DATA-DRIVEN PROJECT PLANNING

Despite initial skepticism, it often turns out that the overlap of data requirements across business units is much larger than initially thought or acknowledged. While differences in KPIs and metrics are inherent in nature, the same basic data is usually needed to compute them.
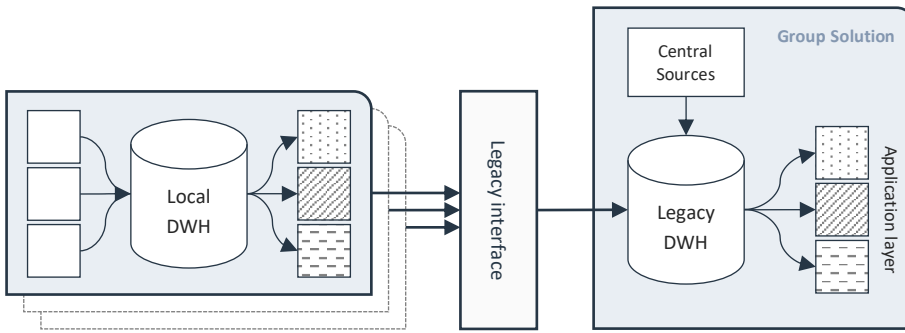
One of the undeniable advantages of business data models and other formalized methods for collecting data requirements is a detailed view of the overlaps between the various business units, which can be an invaluable tool in project planning.

In practice, this usually assumes one of the following forms: either data areas are prioritized according to the degree of overlap between the departments for data integration, or a data area that is only of interest to a small user group can be identified for pilot purposes.
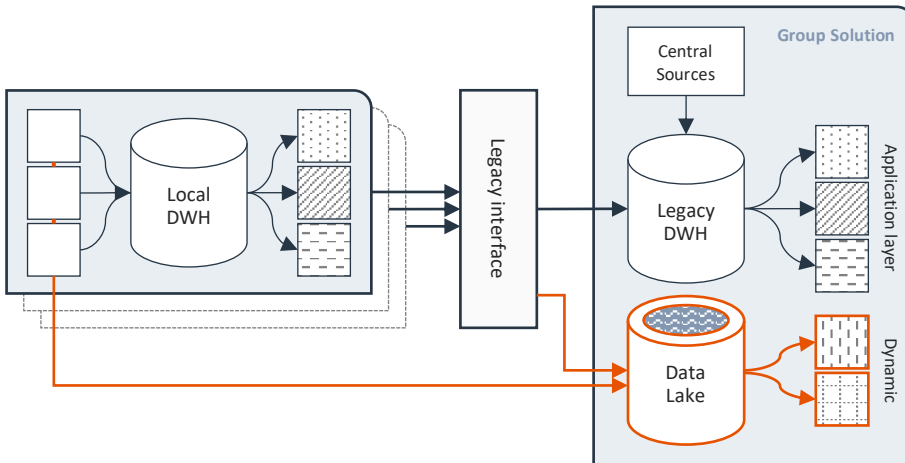
TRANSITION ARCHITECTURE

While the current and the target architecture are usually sufficiently planned and documented, the necessary intermediate steps from one to the other often receive too little attention. Especially in large and complex organizations, e.g. those that need to consider the needs of subsidiaries or foreign branches, it may often be necessary to define several intermediate steps.
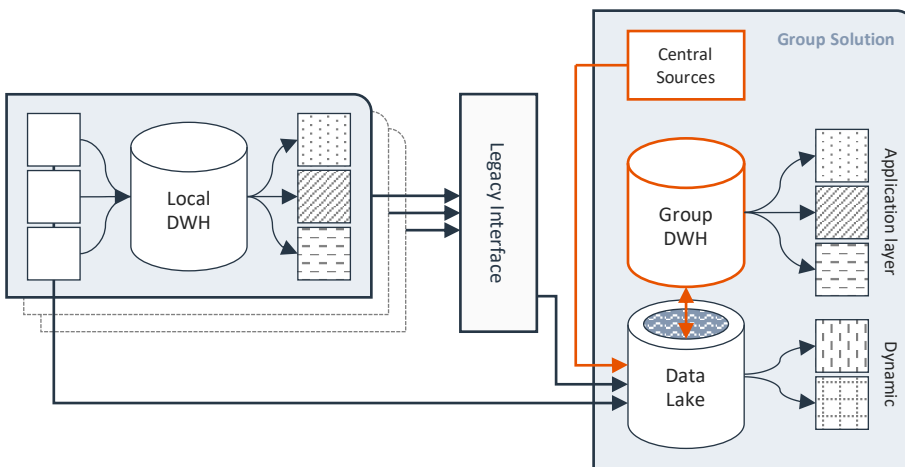
The following illustration shows this idea schematically simplified with the example of one of our clients, an international banking group:
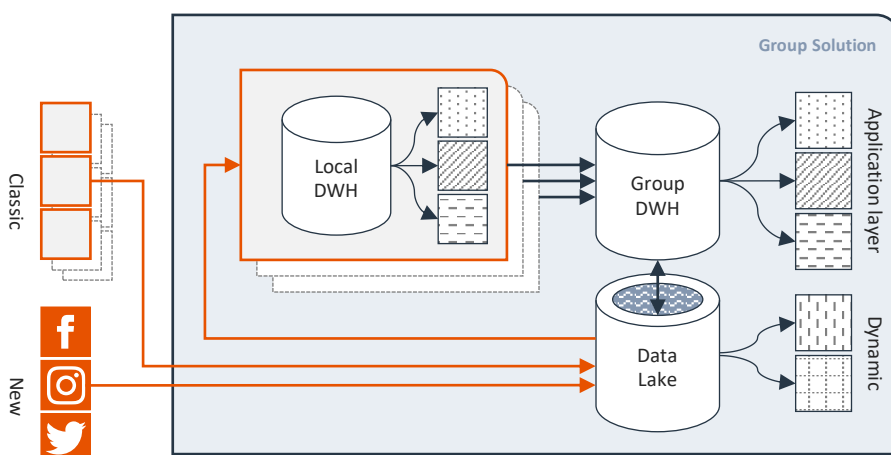
Step 1: represents the situation before the project start, in which the subsidiaries deliver via a standard interface to the group DWH, which also holds the data of the central source systems.

Step 2: introduces a Hadoop-based data lake, which is fed by both the standard interface and local source systems. For the time being, processing at the group level will remain unaffected, but the data lake will already be made available to the business users.

Step 3: the data lake becomes the new acquisition layer of the DWH architecture, the group DWH is migrated to the new solution. The central source systems from now on are sourced via the data lake.

Step 4: the local DWHs are integrated into the group solution. In addition to the traditional data sources, new sources such as social media are also being used.

As the example hopefully shows, the changes do not necessarily have to be large or relevant to everyone involved in the project in order to justify an intermediate step. Rather, it is about providing additional data or new analytical capabilities in sometimes small but fast steps and avoiding years of waiting for the "big bang".

A popular remedy in this regard, also in conjunction with the following section, is to provide business users with access to often difficult-to-access raw data in the form of data labs for explorative analysis while simultaneously tackling the more complex data integration.

EARLY VISIBILTY OF RESULTS

The "fruit" is never really as "low hanging" as in the colorful presentations that serve as a basis for management decisions. No vendor has yet discovered the "Holy Grail", which makes it possible quickly and easily, which was previously expensive and time-consuming.

Nevertheless, it pays off to carefully select a pilot topic, e.g. an application that is migrated to the new platform without any structural changes or a small data area for which an end-to-end implementation can be done.

While the benefit is often limited from an IT perspective, one should not underestimate the psychological impact of the visibility of outcomes when it comes to getting and keeping decision makers on board. Here, a pragmatic approach is clearly to be given preference over the pure doctrine.

MODEL-DRIVEN DEVELOPMENT

In recent years, the buzzword of "DWH automation" has been making the rounds: not only highly praised start-ups, but also every well-known vendor of data bases and integration tools offers functionality in this regard. In our opinion, the promises should be carefully examined, customer projects show usually only limited time and effort savings.

One approach that has been successfully adopted by many companies is model-driven development, for which the business information model approach already mentioned above represents an excellent starting point.

In customer projects, we not only derived technical data models from them, but also significantly accelerated the generation of ETL routines. It would be presumptuous to talk about full automation, but a cost reduction of up to 80 percent is possible, depending on the project environment, model quality and tool support.

GOVERNANCE

Since the requirements gathering is necessarily carried out in close cooperation with the business users, modernization projects are the ideal opportunity to review the processes and responsibilities for data. This usually leads to adjustments or even the setup of applicable structures for the first time.

Those of our customers who opted for our Business Information Modeling approach were able to take advantage of this, because a specialized data model is better suited than any other tool to define business responsibilities for the different data areas. This ensures that data governance does not degenerate into a "paper tiger" but benefits the organization.

CONCLUSION

To answer the question asked at the beginning: no, the DWH is not dead; on the contrary, it enjoys good health in many places - or at least is on the road to recovery. New technologies usually provide a valuable complement, but almost never an equivalent replacement for the DWH. So far, we have not met a CFO who wants to prepare financial statements based on data from the Data Lake only.

The modernization of the DWH will succeed so long as it is not seen only as an IT priority, or even exclusively as an IT task, but is done with close involvement of the respective business units. Another advantage of this can be a containment of the uncontrolled IT proliferation within business.

Equally important is a cautious approach, because the big hit rarely succeeds. Rather, a step-by-step approach based on a solid, data-driven foundation leads to success. If, on top of that, you still succeed in defining up-to-date data governance, your DWH will have many more years of fulfilled analytical existence!

ABOUT SIMPLITY

Simplity offers services around data management and analytics and has supported numerous customers in Central Europe as well as internationally in the successful implementation 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 suite 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.

SELECTED CUSTOMERS

THE AUTHORS

STEFAN RUHLAND

Principal Consultant at Simplity and business lead for numerous successful implementation projects. Previously, he worked for various consulting firms and IT service providers.

stefan.ruhland@simplity.eu
+436766486918

STEFAN MARKUS

Senior Consultant at Simplity and lead data architect for customers in various industries. He began his career as a Data Modeler and BI Consultant at Teradata.

stefan.markus@simplity.eu
+436765756950

CONTACT