BearingPoint®

Data Lakes

Comprehensive analysis using Data Lakes

The vast growth in data volumes and the availability of data sources allows companies to gain insights that were unheard of in the past. To date, data has been typically used in traditional data warehouses with predefined use cases that constrain users in what they can find. Data Lakes aims to enable IT affine business analysts to access larger volumes and a broader variety of data, empowering them to dig deeper into available data as well as enrich them with new data sources to find answers to more complex questions.

"Data Lake technologies will enrich traditional analytics based on Data Warehouses."

With massively decreasing costs for storage, and the availability of cloud-enabled Hadoop technologies, raw information of all kinds and sizes can easily and efficiently be stored, processed, analysed and enriched with metadata. Petabytes of information stored in Data Lakes can be integrated into Logical Data Warehouses, allowing for comprehensive analysis using structured and unstructured data. To use the full potential of Data Lakes, there must be descriptive meta data of high quality and the right processes to maintain it.

Several questions arise: Why don't we replace relatively expensive Data Warehouses with Data Lakes (cost ratio 10-100) using appropriate technology? Why don't we store the data in Data Lakes and replace traditional ETL mechanisms by analysis on read access using the latest tools available? Why should we continue using relatively static queries with a fixed structure, if ad-hoc and on-demand queries can be performed on both structured and unstructured data, leveraging the new technology?

Challenges

Data Lake technology vendors promise that their tools are able to extract all relevant metadata from the raw information. So far, none of them can fully satisfy the customers' requirements in respect of query performance, access security, governance, and lifecycle management. To have fit-for-use qualitative information available from Data Lakes it is critical to adapt the functionality and configuration to business needs and to leverage the right tools.

Users must aquire new skills in order to gain flexibility in what they can analyse, and how they can select the appropriate set of tools to extract and structure the data they want to use.

Key Questions

The following questions need to be answered:

- When can we leverage the full potential of Data Lakes and when is a traditional DWH better suited? What about a hybrid approach where both are combined?
- What information in Data Lakes is useful for analytics, and how is it prepared to make it usable?
- How can data governance issues (information lifecycle, accountability, security) be addressed?
- Does it make sense to provide Self-Service analytics as "Data/Analytics as a Service"? In which business (functional) areas does it make most sense?
- Who are the target users for such a system? How is the collaboration model for business and IT set up?

Our Point of View

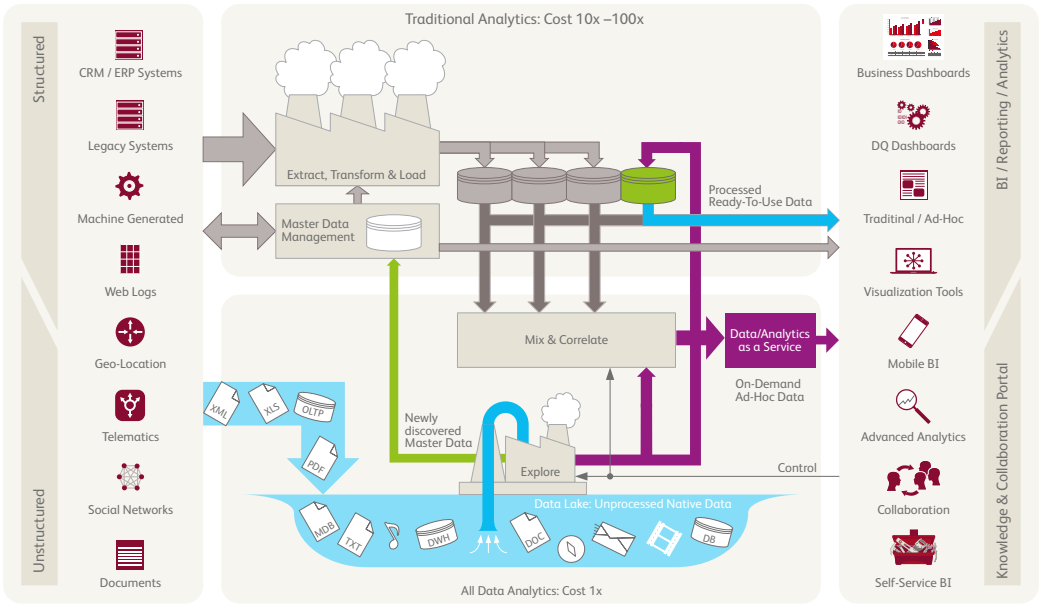
Traditional DWH-based analytics is relatively inflexible and expensive, but very fast and effective in managing data governance. Changes in the data model requires efforts in ETL, which will cause delays between business requests and technical implementations. The traditional DWH-based approach is being challenged due to its high price tag and high-maintenance management. Nevertheless, standard and trusted operational reporting will always be required.

With Data Lake technologies, ad-hoc self-service analysis is possible on structured and unstructured data. To structure data on-the-fly, it is essential to have corresponding and up-to-date metadata, which is derived through appropriate semantic analytics tools. Gaining insight is not limited to data in predefined cubes, where prepared questions are answered. Additional business insight comes from the full context of the raw information, harnessing all types of structured and unstructured data.

How BearingPoint can help

BearingPoint has longstanding project expertise that's cross-industry, and a sound knowledge of the technological developments and trends. Using our well-established toolkit to assess the organizational need for the new Data Lakes technology, we verify the business case for large transformation projects.

Looking at your company's prerequisites and industry benchmarks, we set the optimal state of your future target landscape. Using our industry experience and technical capabilities, we are able to provide you with clear guidance on architectural prerequisites, timing questions and other specific target actions you need to take to define a clear change roadmap.



About BearingPoint

BearingPoint consultants understand that the world of business changes constantly and that the resulting complexities demand intelligent and adaptive solutions. Our clients, whether in commercial or financial industries or in government, experience real results when they work with us. We combine industry, operational and technology skills with relevant proprietary and other assets in order to tailor solutions for each client's individual challenges. This adaptive approach is at the heart of our culture and has led to long-standing relationships with many of the world's leading companies and organizations. Our 3350 people, together with our global consulting network serve clients in more than 70 countries and engage with them for measurable results and long-lasting success.

www.bearingpoint.com

Contact

Matthias Roeser
Partner
matthias.roeser@
bearingpoint.com