White Paper

Top 5 Data Architecture Challenges

By Ron Huizenga

April 2015

Table of Contents

Introduction

Data architects face many challenges on a day-to-day basis. This paper will highlight five major challenge areas, and provide insights regarding how to address them with data modeling:

- Evolution of methodologies & culture
- Adapting to changing architecture
- Complex data environments
- Data quality
- Business focus

Evolution of Methodologies and Culture

When we review the evolution of new methodologies, along with the corresponding changes in corporate culture, we can see that there have been numerous approaches over the years. In the earlier days of traditional / waterfall processes for data modeling, there was a more rigid organizational structure with data modelers, programmers, and system analysts. Projects had rigid schedules with specific activities, delivering solutions in a linear, time-consuming fashion. Coping with changes also proved to be difficult, extending timelines even further.

Agile methodologies emerged in an attempt to address the shortcomings of traditional practices. The focus changed to iterative delivery from self-organizing teams that eliminated traditional bureaucracy. This challenge to the status quo was very difficult for many organizations and individuals to accept, as they perceived their worlds to be turned upside down. At the other extreme, some interpreted this to justify a lack of discipline, which quickly turns into a free-for-all. However, those that embraced it sensibly achieved tremendous results. Continual feedback for improvement is a fundamental tenet of agile practices. Ensuring that there are deliverables for each time-boxed iteration achieves this, as well as continuous collaboration between business and technical stakeholders.

Many organizations have successfully adapted to a hybrid approach, leveraging agile practices for operational execution, within a larger enterprise architecture and project delivery framework. This team approach to solution

delivery emphasizes the importance of principles of the methodology along with the technology framework and architecture toward achieving the goals. Having the agility to show progress toward deliverables without waiting for 100% completion enables more milestones to be achieved toward the release goals.

From a data architecture perspective, a modeling tool that allows granular check-out and check-in of specific objects gives a data architect the flexibility to work on a subset of the model for a specific task or milestone without negatively impacting the rest of the project. Advanced compare-and-merge capabilities allow updates to be quickly and easily integrated into the core model when the task is complete. These capabilities enable data professionals to streamline the model enhancements.

Adapting to Changing Architecture

Technology has been evolving at a much faster pace than methodologies, presenting even greater challenges to organizations that are trying to leverage them. The underlying architecture of databases and modeling tools has also changed. The rapid proliferation of unstructured platforms, also called schema-less or 'big data', needs to be understood and properly managed as part of an enterprise portfolio. This also demands enhanced integration capabilities. Otherwise, organizations will simply repeat the mistakes of the past, such as application silos, but with different technology.

We must extract and consolidate the metadata into models to promote comprehension, consistency and reuse. Powerful data modeling capabilities give us the ability to do so. We need to reverse engineer from the various diverse platforms into relevant data model constructs and metadata. This gives us the ability to represent business objects and data constructs or consistently across platforms, while providing visual maps of how the data components fit together.

For platforms like MongoDB, each document in a collection can have a different schema, so we can't just query the system tables. We must query a representative sample of the actual documents in the collections. Just because we can change big data schemas easily, doesn't mean we should do so without the proper controls and documentation. We can design the changes within the data models, properly connected to business glossaries and terms for comprehension. Those changes can then be generated from the models for deployment.

Complex Data Environments

Corporate data environments are also evolving and becoming extremely complex. Part of this is driven by mergers and acquisitions in which the companies have invariably been using different platforms and applications. It is also becoming standard practice for organizations to purposely buy and integrate a number of solutions, often combined with some internally developed solutions as well. This is often complicated further by not decommissioning obsolete systems, adding even more clutter. This proliferation of disparate systems needs to be reined in and proactively managed.

To combat this, an enterprise class data modeling tool can provide a multi-level hierarchy for models and glossaries that corresponds to the functional decomposition of the enterprise. Metadata can be extended to catalog and categorize data assets. Naming standards and business glossaries provide a basis for common nomenclature and meaning. The data models and sub-models themselves provide a map of the data landscape. Universal mappings show how manifestations of the various entities are linked back to the concepts, across models and platforms. Business process models can reference the data model constructs, giving context to the use of data in the organization.

Data Quality

According to the Data Management Body of Knowledge, aspects of data quality include accuracy, timeliness, completeness, consistency, relevance, and fitness of use. Knowing that your data is current, correct, present, and usable is key to making good business decisions. Some estimates indicate that poor data quality costs a typical company the equivalent of 15-20% of revenue, and significantly impacts corporate efficiency.

How do we maintain a high level of data quality, and avoid problems with 'dirty' data? All too often, companies don't take steps to deal with data quality until they have a major breach or disaster. A continuous improvement philosophy can address data quality at the source. The business leadership must establish a data culture and enforce accountability at the points of data creation. Modeling and metadata management can help to measure, control, and improve the quality of the data.

Business Focus

Data is not just a technology issue - it's imperative to the business. Without data, most businesses would function very poorly, or not at all. It must be clearly understood that data is owned by the business, and as such, the data architecture should be business-driven. Business and IT teams must work together on the data strategy. Emerging roles such as that of the Chief Data Officer are critical. The business leadership has to drive the data culture in the organization. Data architects need to be the champions for data value and data quality, and ensure that everyone in the organization understands what the data is and be able to explain and rationalize it in business terms. A collaborative environment for the definition and utilization of key components, such as business glossaries, will encourage participation and alignment between teams, and help to eliminate inaccuracies and siloes.

Summary

The challenges described above have made data modeling and metadata management more important than ever. The models and associated metadata are the only means by which complex data environments can truly be understood and managed. Without comprehension, it is impossible to manage data quality. A well-defined data architecture makes it possible to address all of the described challenges and is a foundation to improve data quality, master data management and data governance in general.

With enterprise scale capabilities such as business glossaries, data dictionaries, reverse engineering, forward engineering and cross-organizational collaboration, Embarcadero® ER/Studio® offers a comprehensive suite of data modeling tools to address the challenges of data architecture not only for today, but also the future.

Learn more about ER/Studio and try ER/Studio for free.

About the Author

Ron Huizenga is the Senior Product Manager for ER/Studio at Embarcadero Technologies. He has over 30 years of experience as an IT executive and consultant in Enterprise Data Architecture, Governance, Business Process Reengineering and Improvement, Program/Project Management, Software Development and Business Management. Ron's experience spans multiple industries including manufacturing, supply chain, pipelines, natural resources, retail, health care, insurance, and transportation.

About Embarcadero Technologies

Embarcadero Technologies, Inc. is a leading provider of award-winning tools for application developers and database professionals so they can design systems right, build them faster and run them better, regardless of their platform or programming language. Ninety of the Fortune 100 and an active community of more than three million users worldwide rely on Embarcadero products to increase productivity, reduce costs, simplify change management and compliance, and accelerate innovation. Founded in 1993, Embarcadero is headquartered in San Francisco, with offices located around the world. To learn more, please visit www.embarcadero.com.

Download a Free Trial at **www.embarcadero.com**

Corporate Headquarters | Embarcadero Technologies | 275 Battery Street, Suite 1000 | San Francisco, CA 94111

www.embarcadero.com | sales@embarcadero.com