Executive Summary

Data Architecture and Beyond:
Strategies for Improving Your Data Ecosphere

by Scott W. Ambler

Enterprise Architecture Executive Summary Vol. 10, No. 10

In the accompanying *Executive Report*, I explore data architecture in detail, putting it into the context of your overall IT architecture strategy and describing the critical aspects that should be addressed by effective data architecture. Data architecture is a concern at both the enterprise and system levels. An enterprise approach to data architecture and management is clearly important, yet at the same time, data is also important for individual systems. The critical issues, for which you'll need both enterprise- and system-level strategies for, are:

- **The data model.** Your data model must address several critical issues. In particular, it should include a domain view that indicates the majority entity types and the relationships between them; a location view that identifies the source(s) of record, if any, for the data; and optionally, an attribute view indicating the details behind the entity types. There are several key architecture concerns when it comes to your data model: usage, accuracy, timeliness, reuse, performance, and evolution.

- **Physical data storage.** Your physical data storage architecture must address basic technical issues such as security, reliability, availability, performance, internationalization, and the storage needs of both online transaction processing (OLTP) and online analytical processing (OLAP) applications. Other concerns are process-related, such as addressing operation and support issues, industry regulations, and storage volume growth.

- **Data security.** Data security issues include authentication, authorization (access control), and encryption/decryption.

- **Data transport.** Put simply, the goal of data transport is to get data from point A to point B. There are two ways to transport data: either electronically via a network or physically on disks, tapes, or SRAM chips. For electronic transport, you'll need strategies for supporting cryptography (encryption/decryption of data), high performance, and high availability.

- **Data access.** There are two fundamental issues when it comes to data access: data access paths and database encapsulation. First, data access paths are an architectural issue because they indicate topology requirements; they bring source of record issues to the surface; and they reveal potential data security, timeliness, accuracy, and reliability requirements. Database encapsulation is the act of hiding the implementation details of your database(s), including their physical schemas, from your business code. In effect, this layer provides your systems with persistence services — the ability to read data from, write data to, and delete data from — data sources. There are three database encapsulation strategies that you should consider using: data access objects (DAOs), persistence frameworks, and services.

CHALLENGES IN THE DATA ECOSPHERE

The "data ecosphere," a subcomponent of the IT ecosphere, is composed of your data architecture, data management process, "data culture," and your data governance approach. It isn't sufficient to have a technically viable data architecture if it doesn't reflect the realities of other aspects of the data ecosphere. Unfortunately, we face serious challenges. A recent data management survey found that in organizations with a data group, two-thirds of respondents indicated that developers will sometimes go around the data

group, implying that the majority of data management groups are failing to execute their mission effectively. Furthermore, some estimates are that data quality problems result in over $600 billion in annual losses in the US alone. Yes, that's $600,000,000,000. A big part of the problem is the cultural impedance mismatch between the development community and the data community.

It started when the object community started adopting new evolutionary development techniques that differed from the traditional techniques favored by the data community. The problem grew worse in 2001 with the publication of the Agile Manifesto and the adoption of highly collaborative and quality-driven strategies. Agile teams are now achieving measurably higher success rates than traditional application development teams and data warehousing projects, calling into question the approaches preferred by the traditional data community. When you step back and think about it, traditional software processes are based on several very questionable assumptions, all of which I believe are false. These assumptions are: you need specialized roles, such as programmers, business analysts, data architects, testers, and so on; you need to model the details up front; you need to write everything down; you need to take a data-driven approach; review and inspections are an effective way to ensure quality; and the data management group needs to govern data.

AN ACTION PLAN FOR IMPROVING DATA ARCHITECTURE

Improving the technical aspects of your data architecture is straightforward, but you must address the challenges within your overall ecosphere. To do this, you need to follow five strategies:

1. **Adopt concrete quality techniques.** Traditional quality techniques are great in theory but just aren't getting the job done in practice, and it's clear that we need to rethink our traditional assumptions and adopt new strategies. The agile community has put significant effort into developing and evolving such techniques, which include evolutionary/agile data modeling, database refactoring, database regression testing, and continuous database integration.

2. **Promote teamwork over politics.** We need to improve teamwork by building cross-functional teams, promoting non-solo development, improved training and education, and other team-building strategies.

3. **Adopt a lean approach to governance.** This strategy is based on the philosophy that the most effective way to govern the actions of intellectual workers is through motivating and enabling them, not by a command-and-control process.

4. **Go beyond data management.** The management of data is only the beginning, you must also manage your intellectual property (IP) and do so in compliance with applicable regulations.

5. **Become as agile as you possibly can.** Agile techniques enjoy a measurably higher rate of success than traditional approaches. This is because they take a collaborative, evolutionary approach that is focused on quality, maximizes ROI, and meets the actual needs of stakeholders. It is important to recognize that you might not ever become fully agile due to constraints in your environment, but that you should strive to be as agile as you possibly can given your situation and try to improve wherever possible.