The Responsibilities of the Enterprise Data Architect

The role of the data architect is to manage data as it moves through the enterprise, coordinating the efforts of the various teams involved.

In "Enterprise Data Architecture: Why, What and How," we described the need for data architecture. Integrating fragmented data, streamlining the data architecture and improving data quality involve a large number of groups from the IS organization and business units. The enterprise data architect (EDA) must coordinate these efforts so that the enterprise can best leverage this valuable asset.

What Does the Enterprise Data Architect Do?

The EDA is charged with improving the consistency, timeliness, quality, security and delivery of data. In addition, the EDA must streamline data flows and remove unnecessary costs from the complex, fragmented, heterogeneous data architecture that is found in most large enterprises.

Because of the enormous scope of enterprise data architecture efforts, work must be prioritized to produce results. One way is to select a business process and focus on the key data that this process uses. The selection should be done in consultation with business executives (see "Business Processes: A Compass for Architecture").

Several groups have responsibility for aspects of data in the enterprise, including application teams, data warehouse managers and the integration competency center, as well as business people. The EDA focuses on data that moves through the enterprise, and must coordinate work with these groups and other stakeholders. Various teams in the IS organization and the business will execute the more-granular tasks; the EDA will provide overall leadership for driving the enterprise toward an overarching data architecture vision. The EDA's responsibilities include:
Mapping data sources: An understanding of where data is stored and maintained is essential to data architecture. The data map should include descriptions of the business meaning of the data, its uses, its quality, the applications that maintain it and the database technology in which it is stored. Documentation of a data source must describe the semantics of the data so that the occasional subtle differences in meaning are understood. Because business processes increasingly involve business partners, some data sources will be outside the enterprise. These data sources should not be ignored.

Although the application teams are responsible for documenting the details of application data, the EDA must know where data is maintained and accessed throughout the enterprise. Application enhancements that result in the maintenance of additional data subject areas are of particular interest, because these have the potential to increase data fragmentation and redundancy.

As described in "The Architecture Engagement Process," a process must be in place to evaluate the impact of project proposals on enterprise architecture. This process ensures that the EDA assesses the impact of these changes on the architecture.

Documenting interfaces and data movement: Having mapped its sources, the next step in understanding enterprise data is to record how it is moved around the "virtual" enterprise. This includes the frequency of movement, the source and destination of each step, how the data is transformed as it moves, and any aggregation or calculations. The application team, the integration competency center and the data warehouse team play vital roles here, with the EDA coordinating their efforts and fostering consistency.

Designing the movement of data through the enterprise: Having documented the status quo — the sources of data and how the data is moved around — the EDA can then look at how this movement can be improved. Some changes might be obvious, such as eliminating unnecessary movement. Others might be more complex, involving brokering agreements between business units about sharing interfaces. Again, groups such as the integration competency center and the data warehouse team do much of the detailed work, with the EDA identifying and coordinating reuse opportunities.

Defining integrative views of data: These views will draw together data from across the enterprise. Some views will use a database of extracted data — others will bring together data in "near real time." The EDA works with business people and application designers to identify and model these integrative views and
determine the quality of service requirements — data currency, availability, response times and data volumes.

Designing canonical data views: Because the format and the semantics of data differ from application to application, data must be transformed as it moves from its source to its destination. Without a canonical view of data, each interface will perform a unique point-to-point transformation, with the number of these transformations proliferating exponentially and becoming an enormous burden. A shared, canonical view makes these transformations more manageable, with the data from the source being transformed once to the canonical view, and once more from the canonical view to the target view. If an integration competency center does not exist, the EDA must define the canonical views. If it does exist, the EDA must work with this group to ensure that the canonical views assist as broad a range of transformations as possible.

Defining technical standards and guidelines: Data architecture standards and guidelines should cover when and how to use the architected databases (such as the data warehouse and the operational data store), the technologies to be used for various purposes (for example, when to use extraction, transformation and loading tools, or an integration broker), and models of selected entities, objects and processes. The guidelines should encourage reuse of existing data stores, as well as address issues of security, timeliness and quality.

The EDA must ensure that standards are up to date by investigating emerging technologies and new releases. This means working in conjunction with technology architects (where they exist). Participation in proof-of-concept projects and other projects that are early users of new technologies is required from time to time.

Leveraging existing data assets: Often, developers create new sources of data, rather than reuse those already in existence. This increases fragmentation. The EDA must provide guidelines for reuse and ensure that the architecture management processes include the assessment of all proposals to create new data stores.

Managing related metadata: For the integrated views of data to be used correctly, metadata must be available. The metadata should include business descriptions of the data, details of any calculations or summaries, descriptions of the sources of the data, and indications of data quality and currency.

Communicating the data architecture: Like all enterprise architects, the EDA has the responsibility of communicating the
benefits of the enterprise data architecture. This must be directed to a range of stakeholders — the development community, IT operations, IT management at all levels, and business people. The EDA must acquire information from these groups to focus the architecture to meet business and IT needs. The EDA also must also help developers use the architecture.

Ensuring a focus on data quality: The EDA must be concerned not only with the location, definition and use of data, but its quality as well. However, because data quality is a strategic business issue, rather than a technical IT issue, the EDA must educate the business on its importance, and involve and facilitate the work of business constituents on improvement programs. Data stewards are a critical component of such programs (see "Data Stewardship: Critical Component of Data Architecture"). Equipped with a knowledge of data across the enterprise, the EDA can work effectively with data stewards so they can understand data semantics and identify opportunities for improving data quality.

Key Facts:

- The EDA must coordinate all enterprise efforts that affect data, with a focus on data that moves through the enterprise.

- Because the scope of enterprise data architecture is enormous, the work must be prioritized to enable data delivery to increase business effectiveness.

- The EDA is unable to do all that needs to be done, and must work closely with others, including application teams, data warehouse personnel and the integration competency center.

Bottom Line: Enterprises should establish a data architect role to coordinate the movement, enhancement, integration, quality and availability of data throughout the virtual enterprise. The enterprise data architect (EDA) must interact and exert influence with all areas of IT, including application teams, data warehouse teams, and the integration competency center. In addition, the EDA must work closely with the business so that the architecture is aligned with and capable of supporting the business strategy.