**Hardware and Software**
**Engineered to Work Together**

# Information Architecture for Big Data Management Systems

If you're starting on the big-data path, the first few months are critical. Look closely at your organization's current data practices."
Phil Simon, Harvard Business Review Blog

The clamor for Big Data solutions from both IT management and lines of business can be deafening. While insights from Big Data can be game-changing, the integration of Big Data solutions into existing information architectures as well as wider business processes in order to monetize the value of the data under management presents substantial challenge. Planning and development of Big Data solutions using well-defined design patterns can help scope projects as well as simplify the communication of architectural components and purpose. In this solutions brief, we identify a compact flow model for Big Data architectural design patterns and highlight their characteristics.

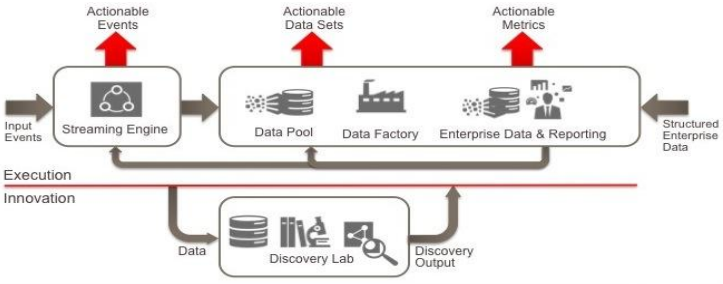


**Figure 1. Flow Based Big Data Model**

### Tips for Pattern-Based Information Architecture

The use of design patterns can simplify the integration of Big Data technologies into existing and new information architectures. When architecting a Big Data solution, consider the following:

- Is this system targeted at execution or enabling innovation?
- Do the outputs of this system need to couple tightly with business-critical processes?
- Are there existing components in the architecture that must be factored into governance and SLAs discussions?

### Extending Information Architectures for Big Data

When defining the role of Big Data components within information architectures, it is helpful to align components and their purposes along the flow of data into the larger data environment. In this compact, flow-based model, we define a set of components, many of which are already present in existing information architectures. These components are as shown in Figure 1 are:

- Streaming Engine: Components which process data in-flight in order to identify actionable events of value to the business and persisting in a durable storage system.
- Data Reservoir: Economical, scale-out storage and parallel processing for data which does not have stringent requirements for formalization or modeling. Typically manifested as a Hadoop cluster.
- Data Factory: Management and orchestration of data into and between the Data Reservoir and Enterprise Information Store as well as the rapid provisioning of data into the Discovery Lab.

- Enterprise Information Store: Large scale formalized and modeled business critical data store, typically manifest by an Enterprise Data Warehouse. When combined with a data pool, these form a *Big Data Management System*.
- Reporting: BI tools and infrastructure components for timely and accurate reporting.
- Discovery Lab: A set of data stores, processing engines, and analysis tools separate from the everyday processing of data in order to facilitate the discovery of new knowledge of value to the business.

The interplay of these components and their assembly into solutions can be further simplified by dividing the flow of data into *execution* -- tasks which support and inform daily operations -- and *innovation* – tasks which drive new insights back to the business. Arranging solutions on either side of this division helps inform system requirements for security, governance, and timeliness.

### Architecture Design Patterns for Big Data Management Systems

Through our recent work in this area, we observe that organizations' priorities will define the scope of the deployed solution, as well as the way in which various components may be combined. From a Big Data management perspective we have been able to characterize three main themes:

- Data Application: where the solution lends itself to Big Data technologies, but as it has a very specific purpose there is no requirement for integration to the broader Information Management estate; e.g. log processing, genome sequencing.
- Information System: a specific solution involving Big Data technologies as well as some limited integration into existing architecture and process; e.g. ETL pre-processing for the Data Warehouse, or the inclusion of more granular data for analysis.
- Information Platform: a broader, bolder and more complete platform based approach that establishes a Big Data Management solution and puts it at the heart of the business. Issues of governance, user skills, procurement, and support are of the utmost importance for this kind of deployment.

### Design Patterns and Strategic Information Architecture

The compact, data-flow based model for Big Data Information Architectures presented here allows for the simple composition and communication of Big Data systems. Use of the resulting design patterns can aid in prioritizing architectural changes and balance strategic aims with system requirements. When applied judiciously, and combined with the corresponding logical architecture perspectives, these patterns serve as a solid foundation for information architectures which maximize the value of Big Data while minimizing risk.

### CONTACT US

Ready to refine your information architecture for Big Data? Find more information at http://www.oracle.com/bigdata.