

Damia: Data Mashups for Intranet Applications

Pradeep Gangadharaiyah
pg29@buffalo.edu

October 20, 2008

Overview

Damia is a web style data integration platform being developed to address the data problem presented by situational applications by harnessing many of the Web 2.0 paradigms and technologies that have spurred the innovation in assembly manifested by the mashup phenomenon.

Damia feed server is comprised of an integration engine, a directory services component, a storage services component, and a rich client browser interface. The directory services component provides capabilities to search, tag, rate, execute, and otherwise manage information assets which include data mashups, public spreadsheets, and URLs to interesting external feeds. The storage services component handles the storage and retrieval of data and metadata needed by other Damia components. The client browser interface provides facilities for composing, editing, and debugging data mashups graphically.

Damia provides a powerful collection of set-oriented feed manipulation operators namely Import-Feed, Filter-Feed, Merge-Feeds, Group-Feed, Augment-Feed, Transform-Feed, Sort-Feed, Union-Feed and Publish-Feed for importing, filtering, merging, grouping, and otherwise manipulating feeds. The feed-oriented data model that forms the basis for such manipulation can easily represent standard feed formats like RSS, ATOM or more general XML data formats.

Data mashups are built by end users and applications in terms of the feed-oriented data model and feed manipulation operators; however, these are logical constructs that have no direct physical implementation. Feed manipulation operators are compiled into augmentation operators, which are lower level data manipulation primitives that execute within a bind-in context providing values for variables that are referenced in arguments are actually executed by the integration engine. The augmentation-level data model, which forms the basis for data manipulation by augmentation operators, is a derivative of the Xquery data model (XDM).

The integration engine is comprised of a flow compiler, metadata services, and an augmentation engine. The flow compiler receives an XML specification of a data mashup and translates it into an augmentation flow. Metadata services stores the augmentation flow and the original XML representation of the data mashup. The augmentation engine, which executes a data mashup, is conceptually divided into an ingestion layer, augmentation layer, and publication layer. The ingestion layer is responsible for importing data and for mapping imported data into an instance of the augmentation-level data model. It also contains a resource cache from which it can serve resources in order to avoid accessing data sources. The augmentation layer is responsible for manipulating instances of the augmentation-level data model in order to produce the data mashup result. It is comprised of augmentation operators namely

Import, Iterate, Filter, Extract, Expression, Construct, Group, Sort, Fuse, Hsjoin and Union that can evaluate xpath expressions; perform sorts, joins, grouping, construction, and low level manipulations. The publication layer is responsible for transforming an instance of the augmentation-level data model to a specific format like RSS, ATOM, or JSON, which it then serializes for consumption by web applications.

Detailed Comments

The whole concept of data integration and mashup, though not totally a new concept to the tech world when we are aware of so many data integration and mashup tools in the market that are already in use and in large by the corporate world, but, yet I feel the authors of this paper have made a sincere and a noteworthy attempt in explaining the already known concept of data mashup by comprehensively presenting the detailed architecture of Damia and its components. The paper clearly explains the motivation for developing such a tool using two detailed real world scenarios and tries to explain how Damia can be used to easily and quickly solve these problems though some among the audience who put themselves in the users position found it difficult to understand how exactly the Damia works.

The paper is technically sound and comprehensive as far as explaining the architecture of the Damia server is concerned but the paper fails to clearly explain how Damia wins over his competitors in delivering data mashups efficiently and its not clear from the paper as which class of corporate world would potentially form the prime customers for such a product when already there are so many tools in the market.

The flow of ideas in the paper is somewhat confusing as few concepts and new terms like augmentation data model were referred in the beginning of the paper but were indeed explained at the end and hence I could understand the full paper only when I went over it multiple times. Apart from this, I should say the diagrams were flawless and the way each component is explained in connection with various other components of the architecture it is linked is praiseworthy.

At the presentation, some among the audience who were mostly researchers or students of Data Integration found it strange as to why Damia's augmentation model being a derivative of Xquery Data Model does not use Xquery engine to execute the mashup and how is Damia's integration engine better or faster than Xquery engine. I felt the paper lacked the information as to clearly explain this.

Some among the audience felt that Damia was just a scale down of the traditional ETL tools and is confined only to the web and mashup. While I believe that Damia was intended for much wider applications than just mashup and the authors also try to assert this fact, however, the paper fails to present this distinction in a comprehensive way. In fact, the paper gives an overview of the major differences between Damia and other related tools but fails to provide how Damia is better than those tools and in what way.

The paper without a doubt can be used as an initial ice breaker by students and researchers who are new to the concept of data mashups over the web and is helpful in depth understanding of the architecture of Damia which in turn can be related to other similar tools to easily understand their architecture.