The field of data management faces new challenges generated by Web-scale data. The amount and kind of data found on the Web, and the need to offer large-scale Web services for email, IM, social networking, photo sharing, etc, requires us to rethink the data models, the processing models, and the platform architectures for supporting large-scale services. Increasingly, more Web "documents" have in fact structure, and are best modeled using database concepts, such as schemas, attributes, and tuples. Web communities generate huge amounts of both structured and unstructured data, such as user profiles, blogs, listings, advertisements. Large-scale, high-availability Web services require a re-design of the data management infrastructure and find new ways to operate it. In this issue of the Data Engineering Bulletin we have invited six papers to illustrate the new challenges faced by Web-scale data. Most of this research is done at the major Web portal companies, which have access to the data, have an infrastructure for search, offer large-scale Web-services, and have the financial resources needed to build the expensive platforms that can cope with Web-scale. But new ideas also emerge from research groups in academia, who are already thinking about the next generation Web-scale data management techniques.

We start with a paper by Bernstein, et al from Microsoft Research that describes very clearly the challenges and tradeoffs faced by large Web services platforms. The authors report from their own experience with Windows Live^{TM}, and discuss design tradeoffs in such large scale platforms. The second paper, from Yahoo! Research, illustrates a different set of issues that arise from managing and combining Web communities, Web services, and advertising services, and highlights the need to support both high-throughput storage/retrieval and analytical processing.

In the next paper Madhavan et al. from Google discuss the structure of data on the Web. It includes a classification of the types of structure, and a discussion on querying both structured and unstructured data. Alba et al. from IBM Research, present in the fourth paper their experience in designing and building the IBM Semantic Super Computing platform, which stores, manages and queries billions of documents.

The last two papers of this issue are representative of the new ideas coming from academic research groups. Agichtein, from Emory University, is proposing to understand the user needs, in order to better guide the information extraction tasks, while Cafarella et al. from the University of Washington discuss novel approaches for expressing queries over both structured and unstructured data on the Web.