

This special issue brings to bear recent advances in the field of cloud computing that are applicable to data management, data retrieval, data intensive applications and data analysis applications.

Cloud computing represents an important step towards realizing McCarthy's dream that all aspects of computation may some day be organized as a public utility service. It embraces concepts such as *software as a service* and *platform as a service*, which incorporate services for workflow facilities for application design and development, deployment and hosting services, data integration, and network-based access to and management of software. Customers of clouds, much like customers of utility companies, can subscribe to different services at different service levels to guarantee the desired quality of service.

Search and electronic commerce companies such as Google, Microsoft, Amazon and Yahoo have adopted cloud computing technology in a major way as have Fortune 500 companies such as IBM, General Electric and Procter and Gamble. Academic units and government agencies are also key players, and several joint efforts among them are at the forefront of cloud computing technology.

In this special issue, we hope to offer readers a brief glimpse into this exciting new technology, specifically from the perspective of data management and data intensive applications. The articles in this issue cover a wide array of issues on topics that range from specific instantiations of cloud computing technology to the broader perspective of the opportunities and possibilities that the technology open up.

Abadi reviews and discusses the limitations of the cloud computing paradigm, and offers a perspective on the potential opportunities for data management and analysis applications within the paradigm.

Abounaga *et al.* discuss the challenges in deploying database appliances on Infrastructure as a Service clouds, as well as the tools and techniques for addressing the issues.

Bertino *et al.* highlight the important issue of privacy preservation in cloud computing systems and describe an approach to privacy preservation in such systems while at the same time enhancing interoperability across domains and simplifying existing identity verification policies.

Beyer *et al.* describe the key features of an enterprise content analysis platform being developed at IBM Almaden which targets the analysis of semi-structured content.

Cooper *et al.* highlight the significant challenges with building a commercial cloud computing system at Yahoo that emphasizes data storage, processing and querying capabilities.

Grossman and Gu provide a nice overview of cloud computing and what differentiates it from past work. They also distinguish among different types of cloud technology in existence today, and conclude with a discussion on open research problems in the arena.

Paton *et al.* describe an autonomic utility-based approach to adaptive workload execution (scheduling), and illustrate the benefits of their approach on workloads comprising workflows and queries.

Peng, Cui and Li discuss lessons learned from constructing a cloud computing platform in an academic environment, and discuss potential improvements to facilitate massive data processing and enhanced system throughput in the context of a specific web and text mining application domain.

Tsangaris *et al.* describe an ongoing system project called Athena Distributed Processing (ADP), its key components and its challenges within the context of supporting user defined operators and enabling efficient dataflow processing and optimization on grid and cloud infrastructures.

Wu and Wu propose a new indexing framework for cloud computing systems based on the Peer-to-Peer structured overlay network concept that supports efficient dynamic network expansion and shrinkage, and demonstrate its viability on the Amazon EC2 cloud.

We would like to thank Shirish Tatikonda and Sai Wu for their help in assembling this issue. We hope you enjoy reading it.