Instructor Name	Masaru Kitsuregawa, professor	Laboratory Location	Institute of Industrial Science	Research Area	Data Engineering

Advanced Technologies for Ultra-large-scale Data: Ultra-high Performance and Green Database Technologies, Cyberspace Mining through Web archive, Global Environment Digital Library, and Cloud Computing

☞ http://www.tkl.iis.u-tokyo.ac.jp/

Outline of research

The volume of data created by human, sensors and computers has explosively increased in recent years. The Kitsuregawa Lab has wider focus on researches based on database engineering, which is the basis of the technology for handling large quantities of data, including system software, advanced applications, hardware and algorithms.

Main research themes

1. Development of World-leading High Performance Database Engine for the Large-scale Database Era and Research on Cyber-physical Systems: Our research group is developing "World-leading High Performance Database Engine" based on a novel out-of-order execution principle. The execution principle has benefits of (1) improving the utilization efficiency of storage system resources significantly by issuing massive amount of asynchronous inputs/outputs, (2) improving throughput remarkably by processing input/output data in a dynamically determined order instead of an programmed input/output order, and (3) utilizing efficiently multi-core processors, each having many core processors integrated in a chip. The engine is projected to achieve x1000 boosting for substantial range of analytical queries. Our group also plans to construct an experimental system of next-generation strategic social services (cyber physical services) that are enabled by the new database engine and to clarify effectiveness of the new database engine.

Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST)
Cabinet Offfice / JSPS

2. Ultra-large-scale Web Archive and Cyber Space Analysis System: The Kitsuregawa Lab has continuously collected Japanese Web pages since 1999 and has constructed a Web archive system including about 30 billion pages. Such Web archive enables us to analyze long term historical changes in cyber space. We are trying to develop an analysis system based on the Web archive. One of our contribution result is Socio-sense that provides functions to figure out structure and evolution process of the cyber space based on long term Web archive, graph mining algorithms, and natural language processing techniques. We have then constructed an interactive visualization system to show a cyber map that is a result from structural and temporal analyses on a large display wall (Fig. 1). We have been trying to analyze topic propagation, temporal changes in word usage, and structure of Web spam. Moreover, we are now promoting research on multimedia Web analysis framework towards development of social analysis software.

3. Petabyte-scale Global Environment Information Fusion System: It is necessary to access usable information on the environment to deepen our understanding of the earth environment, to improve our predictive ability, and to make sound decisions on risk and resources management through the best possible use of earth observation data. We have been developing a large scale global environmental information fusion system for data integration and analysis that includes the supporting functions of life cycle data management, data search, information exploration, scientific analysis, and partial data down-loading (Fig. 1).

Earth Observation Data Integration & Fusion Research Initiative
(http://www.editoria.u-tokyo.ac.jp/)

4. Reliable and Power-saving Network Control Technologies for Cloud Computing: Cloud computing -- large-scale on-demand distributed computing environment -- has become very popular. We are trying to develop a new technology to construct cloud applications with dynamic load balancing and fault resilience. Large-scale search problems have some difficulty to apply these techniques,therefore we're focusing on distributed computing platform to develop such applications with ease.

✉ kitsure@tkl.iis.u-tokyo.ac.jp

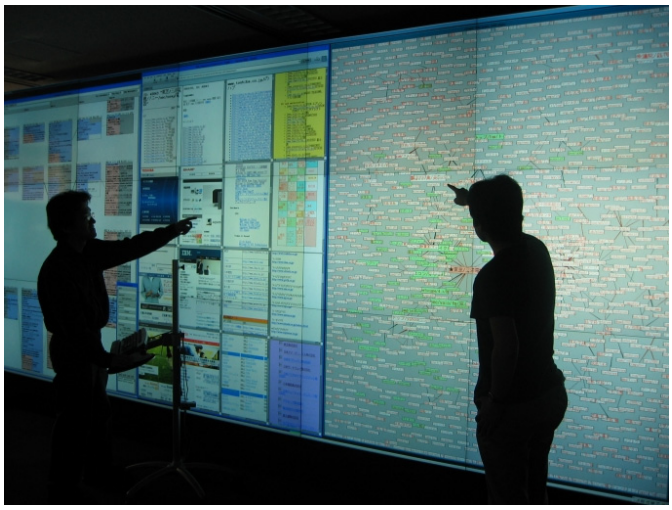

Fig. 1 The cyber map visualization system on the display wall

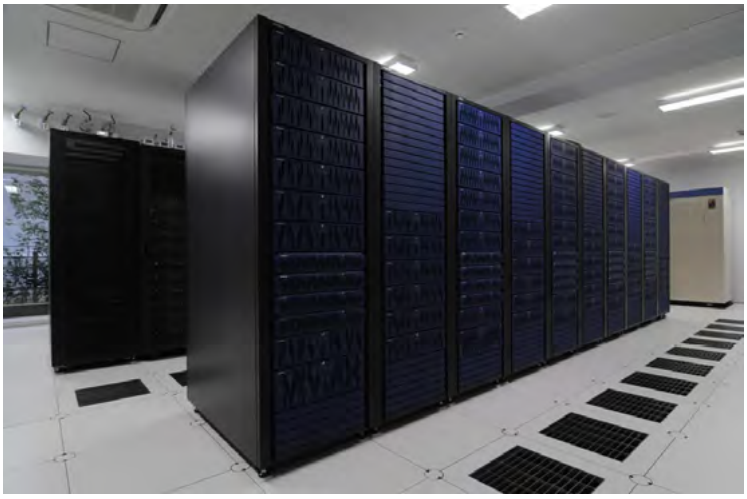

Fig. 2 Petabyte-scale global environment information storage