

Fujitsu technology puts big data to use in minutes

April 5 2012

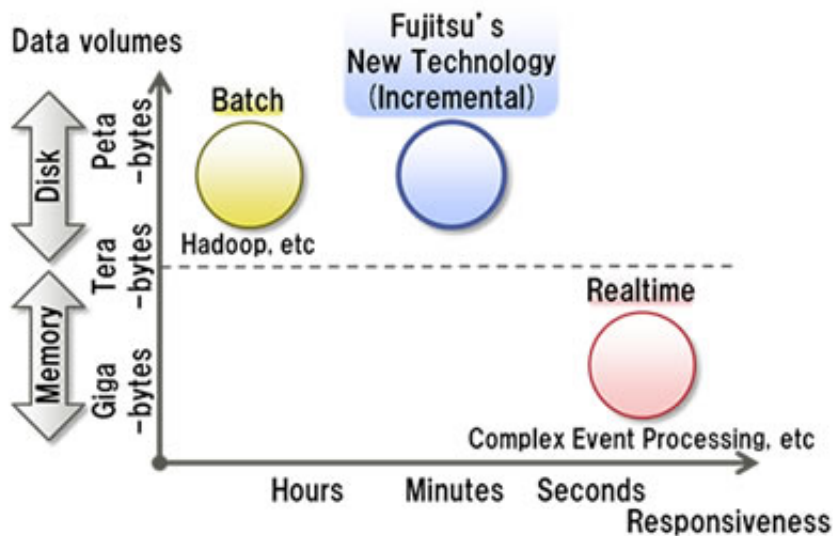


Figure 1: How Fujitsu's new technique compares with existing techniques.

Fujitsu Laboratories today announced that it has developed new parallel distributed data processing technology that enables pools of big data as well as continuous inflows of new data to be efficiently processed and put to use within minutes.

The amount of large-volume, diverse data, such as [sensor data](#) and human location data, continues to grow, and various data processing technologies are being developed to enable these pools and streams of big data to be quickly analyzed and put to use. When the priority is on

high-speed performance, methods that process the data in memory are used, but when dealing with very large volumes of data, disk-based methodologies are typically used as volumes are too large to process in memory. When using disk-based techniques, however, if the objective is to immediately reflect the newly received data in the analytical results, many disk accesses are necessary. This results in the problem that analytical processing cannot keep pace with the volume of data flowing in.

To address this problem, Fujitsu has developed [technology](#) that slashes the number of disk accesses by approximately 90% compared to previous levels by dynamically reallocating data on disks to match trends in data accesses. Whereas producing analytic results of new data could take several hours in the past, with this new technique results are available in minutes. This development excels at both volume and velocity when processing big data, an objective that has been difficult to achieve until now.

This technology will be one of the technologies underpinning human-centric computing, which will provide relevant services for every location.

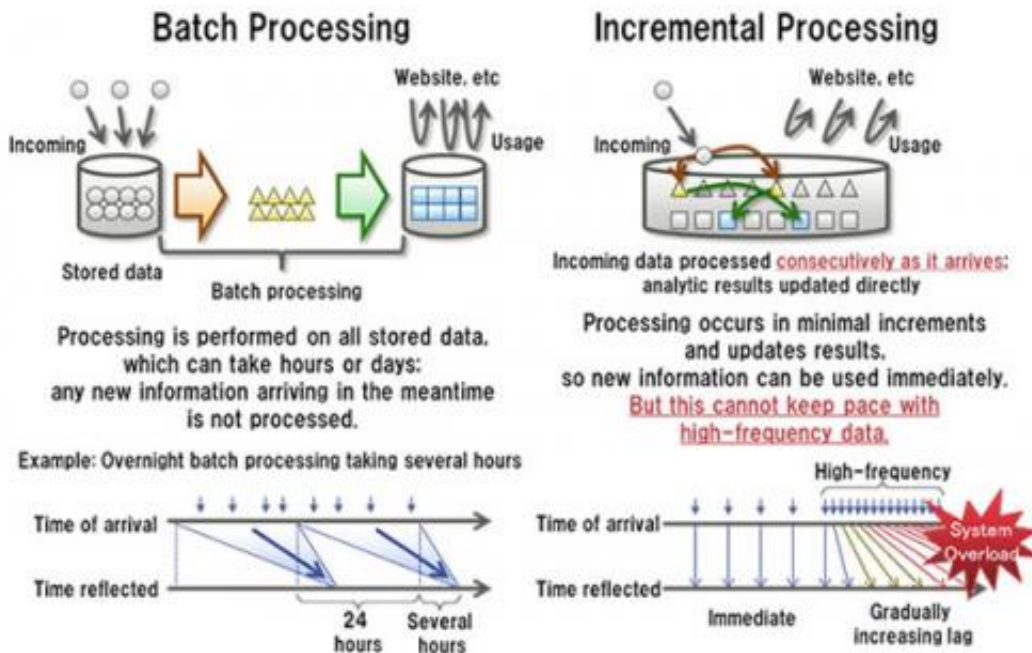


Figure 2: Problems with batch and incremental processing

In recent years, the amount of large-volume, diverse data, particularly chronological data such as sensor data and human [location data](#), continues to grow at an explosive pace. There is a strong demand to take this type of "big data" and efficiently extract valuable information that can be put to immediate use in delivering services, such as various navigation services.

A number of data-processing techniques have emerged for handling big data (Figure 1). One of these, parallel batch processing, as in Hadoop, has become a focus of attention. In parallel batch processing, the dataset is divided and quickly processed by multiple servers.

Another technology that has also received interest is complex event processing (CEP), which handles a stream of incoming data in real time. This has the benefit of being extremely fast because it processes data in

memory.

The goal of extracting valuable information more quickly, from larger datasets, requires a data-processing technology that is disk-based and can quickly produce analytic results. While there are both batch and incremental disk-based processing techniques, obtaining analytic results from either one quickly (responsiveness) remains a problem.

Because batch techniques perform a batch process on a snapshot of the data, there will always be a fixed lag-time before new information can be reflected in the analytic results.

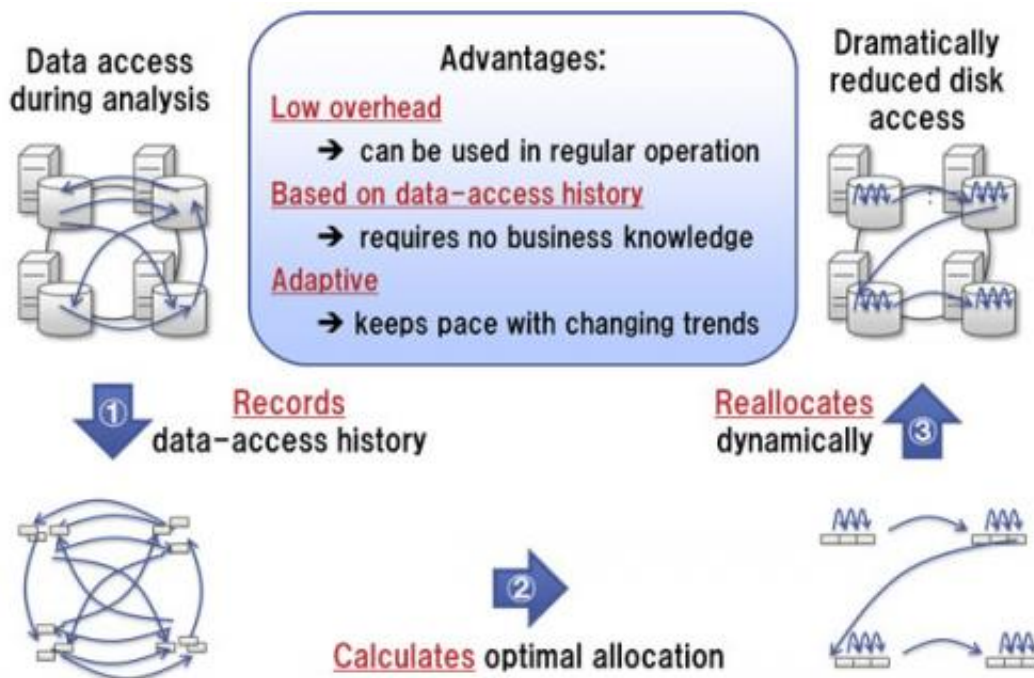


Figure 3: Adaptive locality-aware data reallocation

Conversely, with incremental processing, new data is processed

consecutively as it arrives, but updating the analytic results directly requires the disk to be accessed numerous times. This creates a bottleneck for analytic processing overall, which ultimately cannot keep up with the pace of incoming data (Figure 2). Quickly reflecting new data in analytic results, therefore, required addressing the problem of reducing the number of disk accesses.

Fujitsu has developed a technology it calls "adaptive locality-aware data reallocation," which dramatically reduces the number of accesses, along with distributed parallel middleware for incremental processing.

With adaptive data localization, data is optimally allocated by the following three steps (Figure 3):

- Record data-access history: Records sets of continuously accessed data.
- Calculate optimal allocation: Based on step 1, group sets of data that tend to be accessed continuously.
- Reallocate data dynamically: Based on step 2, specify a location on disk for data belonging to a group and allocate it there.

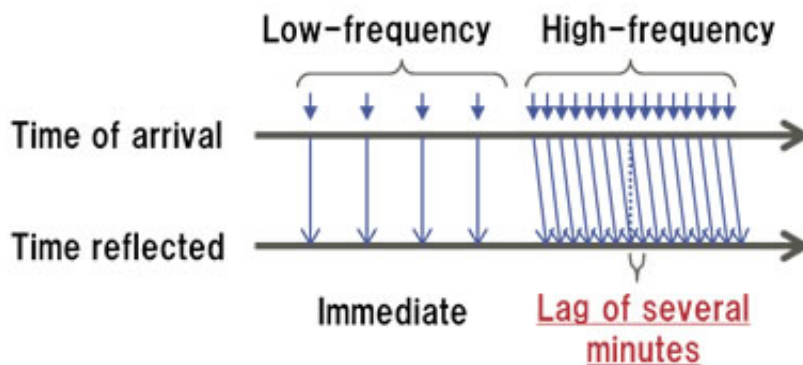


Figure 4: Results using this technology

This makes it possible to acquire desired data through a fewer number of continuous accesses, not numerous random accesses, which vastly increases overall throughput in a distributed-processing system. Also, by monitoring and automatically recognizing patterns of data access, this technology can gradually accommodate the hard-to-anticipate data characteristics of social-infrastructure systems.

This technology can perform analytic processing on big data using incremental processing while accepting data as quickly as it arrives, allowing for rapid analytic processing of current data.

This technology was used in the analytical processing portion of an electronic commerce recommendation system, where it was shown to operate with about one-tenth the number of disk accesses of previous technologies. Consequently, whereas batch processing had conventionally been used for analytical processing of large data volumes, incremental processing is now suitable. This greatly reduces the time required for new data to be reflected in analytical results. When applied to analytic processes that had been run as overnight batches because of the hours-long processing time required with batch processing, this technology can be used to utilize analytical results in a matter of minutes.

[Fujitsu](#) Laboratories plans to move forward to make further performance enhancements to the technology and conduct verification testing with the aim of applying it to commercial products and services in fiscal 2013.

Source: Fujitsu

Citation: Fujitsu technology puts big data to use in minutes (2012, April 5) retrieved 26 April

2024 from <https://phys.org/news/2012-04-fujitsu-technology-big-minutes.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.