Data Mining Approach for Commercial Data Classification and Migration in Hybrid Storage Systems

Authors : Mais Haj Qasem, Maen M. Al Assaf, Ali Rodan

Abstract : Parallel hybrid storage systems consist of a hierarchy of different storage devices that vary in terms of data reading speed performance. As we ascend in the hierarchy, data reading speed becomes faster. Thus, migrating the application' important data that will be accessed in the near future to the uppermost level will reduce the application I/O waiting time; hence, reducing its execution elapsed time. In this research, we implement trace-driven two-levels parallel hybrid storage system prototype that consists of HDDs and SSDs. The prototype uses data mining techniques to classify application' data in order to determine its near future data accesses in parallel with the its on-demand request. The important data (i.e. the data that the application will access in the near future) are continuously migrated to the uppermost level of the hierarchy. Our simulation results show that our data migration approach integrated with data mining techniques reduces the application execution elapsed time when using variety of traces in at least to 22%.

Keywords : hybrid storage system, data mining, recurrent neural network, support vector machine
Conference Title : ICWS 2016 : International Conference on Web Services
Conference Location : Madrid, Spain
Conference Dates : March 24-25, 2016