The Hong Kong Observatory's Operational Data Management Systems

Y.C. Cheng – Hong Kong Observatory, Hong Kong, China

Abstract

Observational data collected by the Hong Kong Observatory (HKO) are stored in databases for use by a wide range of applications including NWP models, nowcasting systems, graphical display systems, and meteorological information dissemination systems. Owing to different usage patterns and characteristics of individual applications, database management systems are designed and set up in specific ways to meet the needs for performance, availability and scalability. Oracle's Real Application Clusters (RAC) and Data Guard technologies, which are employed separately in two typical applications in at HKO, will be discussed.

Background

Observational data is the lifeblood of a meteorological centre. To support operations, the HKO maintains several database management systems (DBMS) to manage observational data collected over the Global Telecommunication System (GTS), Automatic Weather Station (AWS) and rain-gauge networks, and systems developed under regional data exchange programmes. The collected data are used by limited area numerical models, nowcasting systems and graphical display system that form the building blocks of the Decision Support System (DSS) of the HKO.

Another DBMS used at HKO is in association with the dissemination of weather information. To serve the public and end-users, the observational data, forecasts and warnings, or in their derived forms, are prepared and dispatched using a Meteorological Information Dissemination System (MINDS) via various channels including leased line circuits, fax, SMS, Interactive Voice Response telephone system, and the web. A DBMS is used to manage different kinds of data in the MINDS including bulletins prepared by forecasters, products disseminated to end-users for different channels, client information and customer data, as well as configuration data for the workflow module.

HKO's GTS DBMS

The system was first developed in 2003 and was deployed for operational use in 2004. It was built on a high-availability hardware platform employing IBM's HACMP technology (Figure 1) with a backup system installed on another server. The system ingests all data into its database. Data are received from external parties including RTH Tokyo, RTH Beijing, NMC Macau, meteorological centres of nearby cities like Guangzhou, Shenzhen, etc., and also HKO's internal systems. The data, both alphanumeric and binary, are stored in the database using large object (LOB) data type. The routing rules for individual bulletins are stored in a table in the system. When a bulletin arrives, it will be routed to the desired destination according to the rules defined in the routing table. This approach could minimize the amount of database query as data are automatically 'pushed' to user applications instead of requiring the user applications to pull data from the database regularly.

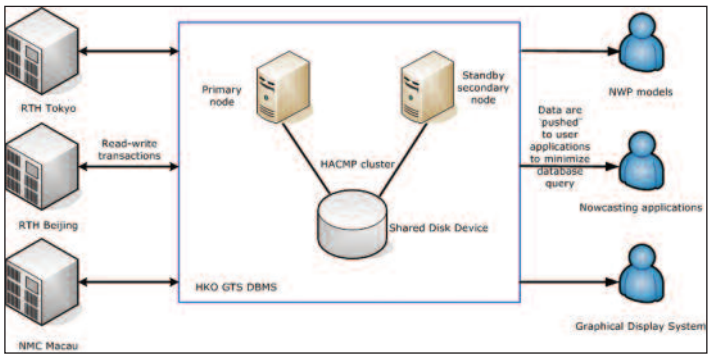

Fig. 1 HKO GTS Data Management System

As the data volume continues to grow, the response time of the system becomes slower. Moreover, the needs of online database query continue to increase as new components of the DSS are introduced. Both changes call for an upgrade to meet future's needs. The challenge of the upgrade is to improve system performance in a cost-effective way.

An upgrade system is planned for launch in early 2010. Instead of storing bulletins in the database, the upgrade system will store the link of a bulletin in the database in attempt to improve the read/write performance for small files. The upgrade system will also employ Oracle's Data Guard technology instead of HACMP to maintain high system availability (Figure 2). In Data Guard, no shared disk device is required and hence the implementation cost could be reduced. Moreover, different nodes of Data Guard are only required to connect together through LAN and hence there is higher flexibility in deployment. In 11g version, the standby node of Data Guard could be turned

active. For some applications particularly the internal decision-support systems that need to read data from database only, write transactions on the database are not required. These read-only transactions are off-loaded to the standby databases, thereby largely increased the capacity of the system in handling online query without having significant performance impact on the primary database.

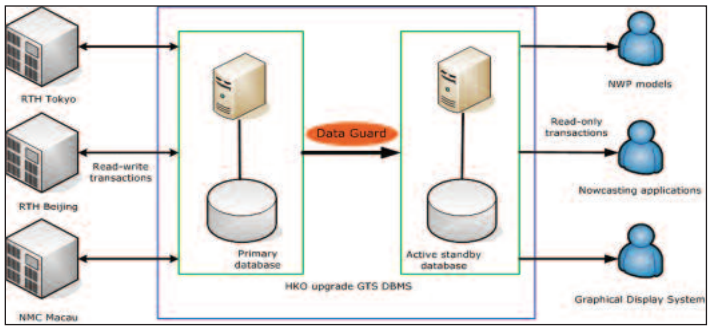

Fig. 2 HKO upgrade GTS Data Management System

HKO's MINDS DBMS

The main goal of the MINDS is to provide a platform with high level of availability and fault resilience for compilation and dissemination of weather report, forecast and warning bulletins to support the public weather service. In contrary to the large amount of global observational data processed by the HKO's GTS DBMS, data processed by the MINDS are much slimmer in terms of data volume. However, transactions need to be handled by the MINDS remain high. As this system serves the public and end-users by delivering products to them directly, product processing status in all dissemination channels has to be updated in a real-time manner on the database to support customer service. Moreover, system loading could show many-fold increase under inclement weather conditions compared with normal hours. Thus, the availability and scalability of the database are the most challenging requirements in the system design of MINDS.

The Oracle's Real Application Cluster (RAC) technology was employed in setting up the MINDS database. The RAC cluster, currently composed of two nodes, offers high level of availability and fault resilience to the middle-tier application server clients (Figure 3). To meet the demands of usage growth, additional nodes could join the cluster to expand the processing capacity of the system. To further improve system availability with complete redundancy, the RAC database was replicated to standby databases at other sites using Data Guard technology.

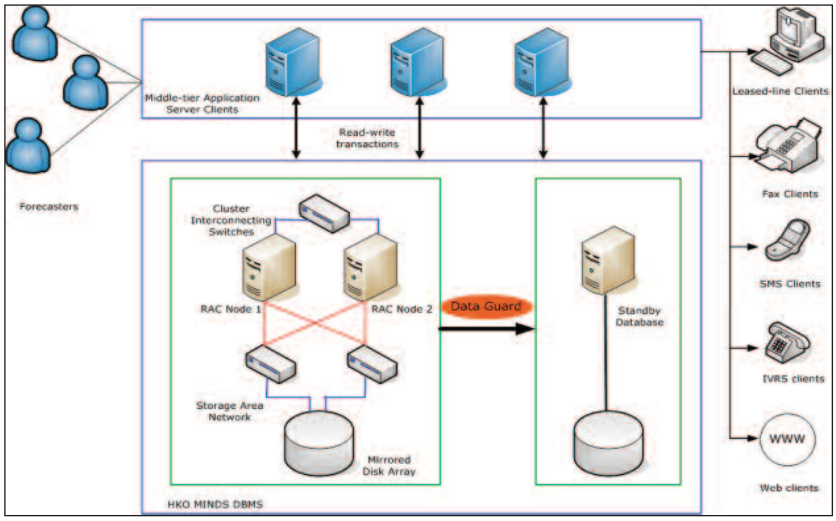

Fig. 3 HKO MINDS DBMS

Conclusions

The MINDS is a mission-critical system in the HKO operating round-the-clock to support quality public weather services. To minimize down-time and for future expansion, the database was set up in RAC cluster protected with standby database through Data Guard to provide high level of availability and scalability.

In circumstances where read-only applications could be identified and are isolated from read-write transactions, scalability of the system could be largely increased by setting up active standby database for read-only access. Such an approach could be adopted in setting up the DBMS of GTS, AWS and rain-gauge data in the Observatory to improve system performance and availability in a cost-effective way.