RIBF データアーカイブシステムの高性能化

IMPROVEMENT OF DATA ARCHIVE SYSTEM AT RIBF

内山暁仁#, A), 込山美咲 A)
Akito Uchiyama #, A), Misaki Komiyama A)
A) RIKEN Nishina Center

Abstract

Since 2009, PostgreSQL-based data archive system, which has developed by RIKEN Nishina Center, has been utilized for RIBF. The archived data is used for troubleshooting and checking operation status during accelerator operation. Although PCI Express-based SSD was taken as the main storage of the archive system because of the reading speed, there was a problem of redundancy. If the disk failure occurred in the PCI Express-based SSD, it takes time to recover the system. In updating the server hardware, we improve the data reading speed while securing redundancy by separating the writing and reading servers of the archive system by using the replication function of PostgreSQL. In addition, another MySQL-based data archive system for non-EPICS-based system also stores the data into the PCI Express-based SSD by using the replication function. In this report, we report its system design and the performance in detail.

1. はじめに

理化学研究所仁科センターRI ビームファクトリー（RIBF）はサイクロトロンを用いた重イオンビーム実験施設である。超伝導リングサイクロトロン SRC を主加速器とし、入射器に AVF、RILAC、RILAC2、また中間段加速に RRC, fRC, IRC を用いた多段加速器システムで構成されている[1]。真空機器や電磁石電源といった加速器を構成するコンポーネントの履歴データを蓄積するシステムをデータアーカイブシステムと呼ぶ。これらは運転のトラブル時における原因追究や各運転パラメータ間の相関の検索、また性能向上を目的としたデータ解析等で利用されている。RIBF 制御系は主に EPICS (Experimental Physics and Industrial Control System) を用いた分散制御システムで構築されている。この EPICS で制御されている機器用のデータアーカイブシステムが RIBFCAS であり、2009年より運用を開始した。一方、RIBF ではシステムの導入コストを抑えるため、商用システムやクライアントとネットワーク機器から成る二層システムも利用されている。主にこれら非 EPICS システムのデータを格納するためのアーカイブシステムとして MyDAQ2 が用いられている。

RIBFCAS はデータ呼び出し速度を重視したため PCI Express ベースの SSD を採用しているが、これは RAID 構成になっていない事から冗長性が低いという欠点があった。一方 MyDAQ2 に関しては読み出し速度の向上が求められていた。なぜなら、何らかのトラブル時において、アーカイブデータの表示速度は履歴調査や解析の作業効率に直結し、復旧時間に影響があるからである。またデータを呼び出し、視覚化するクライアントソフトウェアについて MyDAQ2 クライアントからも RIBFCAS のデータを簡便に比較可能にさせるシステムも求められていた。上記問題点を解決するために、システム改善を行った。

2. RIBFCAS

2.1 システム概要

RIBFCAS は2017年7月現在運用中の IOC 62台から得られる約3400点のデータを、機器の種類に応じて～20秒間隔で取得し、PostgreSQL ベースのデータベースへ格納する[2]。ソフトウェアは Java で実装され、EPICS からのデータ取得には J-PARC 制御グループが開発した Java Channel Access Light Library (JCAL)[3] を使用している。データを表示するクライアントアプリケーションは Adobe AIR Runtime 上で実行される。

2.2 従来のシステム構成

データの呼び出し速度を最優先に考えて設計したため主ストレージには PCI Express ベースの SSD である OCZ RevoDrive3 X2[4]を採用していた。一般的にアーカイブシステムのストレージでは RAID 等で冗長性を高めるが、PCI Express ベースの SSD において RAID1/5/10 を実現するには高コストで、かつ書込寿命の考慮をしなくてはいけないことから RAID 構成にはしなかった。よってデータベースに冗長性はなく、仮に主ストレージである SSD が故障した場合、復旧に時間を要するシステムとなっていた。サーバ仕様を Table 1 に示す。

Table 1: Specification of Previous PostgreSQL Server System for RIBFCAS

OS	Windows Server 2008R2
CPU	Intel Xeon E5506 2.13 GHz ×2
Storage	RAID1: 500 GB (Operating System) RAID5: 3.6 TB (Old DB Data) RevoDrive3 X2 : 960 GB (Current DB Data)
Memory	16 GB
DB	PostgreSQL 9.1

a-uchi@riken.jp

2.3 システムデザイン

更新されたシステムでは、呼び出し性能は落とさずに、かつ冗長性を高めるために書込用サーバ(マスタ)と読込用サーバ(スレーブ)2台で運用する事にし、レプリケーション機能を用いる事で両サーバ間のデータを同期させた。マスタとスレーブの仕様を Table 2 と Table 3 に示す。スレーブの主ストレージには PCI Express ベースの SSD である MEMBLAZE PBlaze4[5](Figure 1 参照)を採用した。

RIBFCAS では年度別にデータベースを分け運用しているが PostgreSQL のレプリケーション機能はデータベース単位では行えない。一方で全てのデータをマスタ・スレーブ間で同期させ、スレーブにインストールされた PBlaze4 からデータ提供するには高コストになってしまう。したがって利用頻度が高いデータは直近の1年間である事から、マスタ用サーバで PostgreSQL のプロセスを2つ立ち上げ、最新の1年分のみをスレーブと同期し、他の年度のデータはマスタの RAID10 構成された SAS (Serial Attached SCSI)ディスクに展開、サービスを提供する事とした。更新されたシステム図を Figure 2 に示す。

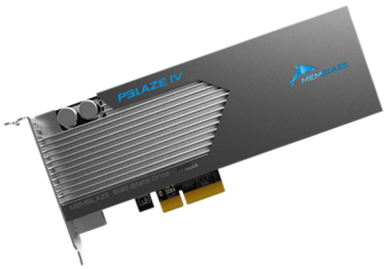

Figure 1: Photograph of PBlaze4 that is installed as main storage on RIBFCAS slave server.

Table 2: Server Specification for Master PostgreSQL Service of RIBFCAS

OS	CentOS 6.8 x86_64
CPU	Intel Xeon E5-2603 v4 1.7 GHz (6 core)
Storage	RAID10: 12 TB (OS, All PostgreSQL Data)
Memory	16 GB
DB	PostgreSQL 9.6.1

Table 3: Server Specification for Slave PostgreSQL Service of RIBFCAS

OS	CentOS 6.8 x86_64
CPU	Intel Xeon E3-1230 v5 3.4 GHz (4 core)
Storage	RAID1: 450 GB (Operating System) PBlaze4 : 1.2 TB (Current DB Data)
Memory	32 GB
DB	PostgreSQL 9.6.1

3. MyDAQ2

3.1 システム概要

MyDAQ2 は SPring-8 で開発されたデータ収集、表示システムである[6]。MADOCA 互換メッセージを TCP ソケット通信でデータ収集サーバに送り、データを保存する。送られたデータは MySQL ベースのデータベースで管理され、データの閲覧はウェブブラウザを利用して行われる。もともとシンプルな構成で個人 PC においても運用できるように設計されていたが、MyDAQ2 の有用性か

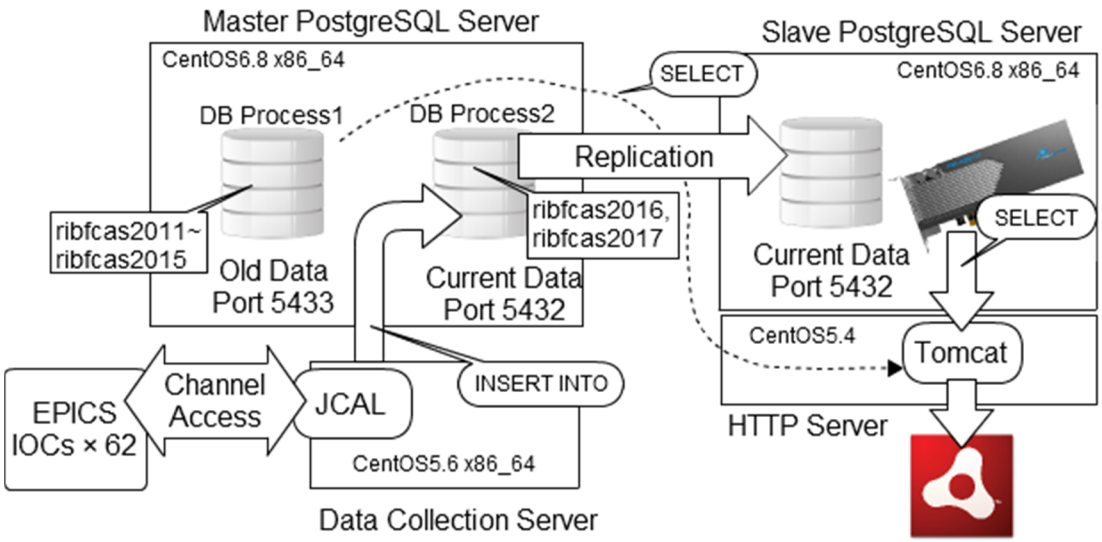

Figure 2: System diagram of upgraded RIBFCAS. The current data in the master server is synchronized to the slave server by the replication function. Adobe AIR-based client application for RIBFCAS obtains the data via Tomcat. Current data is provided from slave server, and old data is provided from port 5433 by master server.

ら RIBF では主に非 EPICS のデータを収集する目的で独自に機能拡張を行い、大規模に運用を行っている[7]。

3.2 システムデザイン

更新された MyDAQ2 システムは MySQL 5.1 を採用、RIBFCAS 同様、書込用サーバ(マスタ)と読込用サーバ(スレーブ)に分け、データベースのレプリケーションを行っている。スレーブに PBlaze4 をインストールし、レプリケーションを行う事で冗長性の確保だけでなく、データの検索、呼び出し速度を向上させた。システム構成を Figure 3 に示す。

本システムでは以下の理由で仮想環境[8]にもスレーブサーバを構築している。ゼロから MySQL のレプリケーション環境を構築する時、データベースの書き込みを停止させた後、スレーブにデータ全てをコピーしてから開始させる。しかしマスタの停止中は新しいデータを追記できない事になるので、運用中は行う事が難しい。よって仮想環境にもスレーブがあれば、マスタを停止させなくとも仮想環境のスレーブを停止させ、データをコピーすれば済むため、データの不整合やディスク障害といったトラブル時でも対応が可能となる。

また、データベースだけでなくデータを閲覧する Apache HTTP サーバも PBlaze4 上に展開する事でチャート表示速度の向上も実現している。

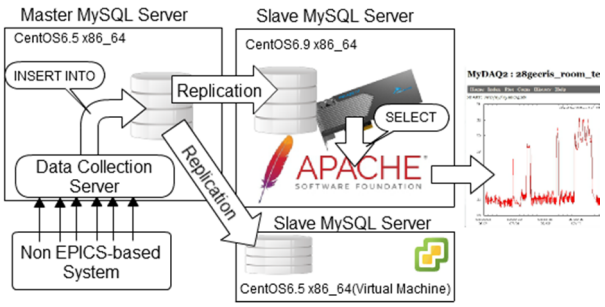

Figure 3: System diagram of upgraded MyDAQ2. The whole data in the master server is synchronized to the slave server by the replication function.

4. クライアントシステムアップグレード

MyDAQ2 データの閲覧表示用クライアントソフトウェアはウェブアプリケーションで実現されている。オリジナルではデータのチャート表示に gnuplot[9]を利用し、ブラウザには画像ファイルとして表示させる。RIBF での運用において、複数データの表示、比較を簡便に行うために独自に JavaScript チャートを実装し、機能拡張が行われている。また HTTP 経由での CSV ファイルでのデータ取得のためにいくつかのウェブアプリケーションも新たに開発した。

一方で MyDAQ2 クライアントソフトウェアにおいて、RIBFCAS と MyDAQ2 双方のデータを同じ時系列でチャートとして表示させる事はオペレーション時に有用である事から、MyDAQ2 クライアントソフトウェアからも RIBFCAS のデータベースにアクセスしてデータを取得できる様に改善を行った。RIBFCAS データは、MyDAQ2 JavaScript チャート用に JSON フォーマットに変換され利用されている。RIBFCAS と MyDAQ2、双方のデータを検索、表示させた時の JavaScript チャートの例を Figure 4 に示す。

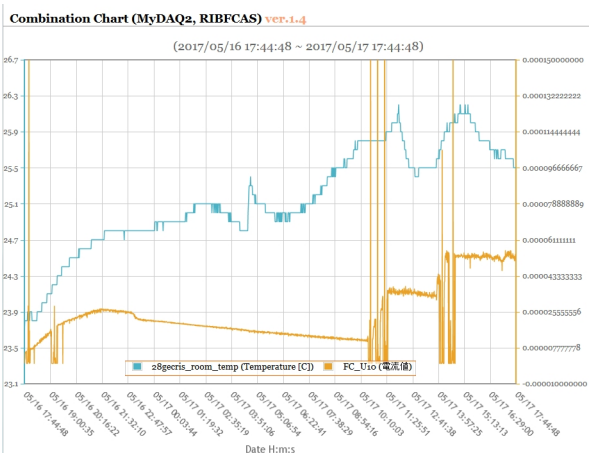

Figure 4: Screenshot of JavaScript chart showing data of both MyDAQ2 and RIBFCAS. In this example, the blue line is the data of MyDAQ2, the orange line is the data of RIBFCAS.

5. パフォーマンス測定

RIBFCAS に関して、２０秒に１回の周期でデータ取得を行っている機器について、データベースへ SELECT 文を発行し検索に要した時間を測定する事でパフォーマンスの比較を行った。この時、検索結果にキャッシュされた値が反映されないようにしている。測定結果を Figure 5 に示す。１０日間分のデータを検索した時、マスタ（SAS・RAID10, Xeon E5-2603）は約２５秒、スレーブ（PBlaze4, Xeon E3-1230）は約５秒、旧システム（RevoDrive3 X2, Xeon E5506）は約１２秒要した。

同様に MyDAQ2 についても、５秒に１回の周期でデータ取得を行っている機器についてデータベースへ SELECT 文を発行し検索に要した時間を測定する事でパフォーマンスの比較を行った。測定結果を Figure 6 に示す。３０日間分のデータを検索した時、マスタ（SAS・RAID10, Xeon E5-2420, 16 GB memory）は２．５４秒、スレーブ（PBlaze4, Xeon E3-1225, 16 GB memory）は１秒要した。また、同様のデータにおいて gnuplot チャートを１ヵ月間分表示させた時の時間を測定（Firefox 54.0.1、Firebug 2.0.19）、比較した。結果としてマスタは表示に約１１．６秒、スレーブは表示に約５．８秒要した。

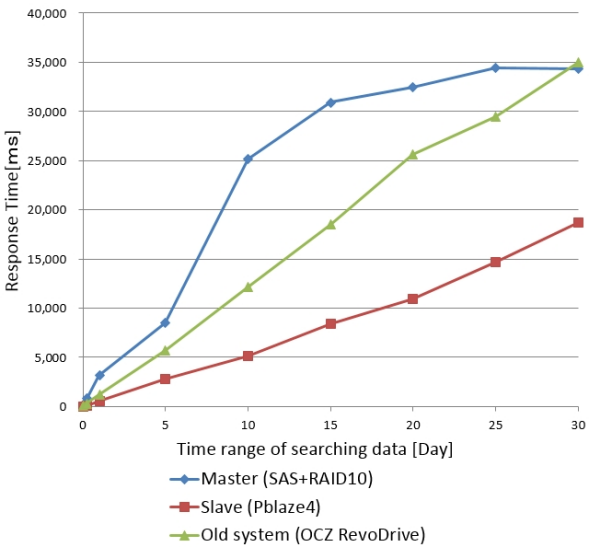

Figure 5: Comparison of response time of PostgreSQL-based database on the master server, the slave server, and the previous system.

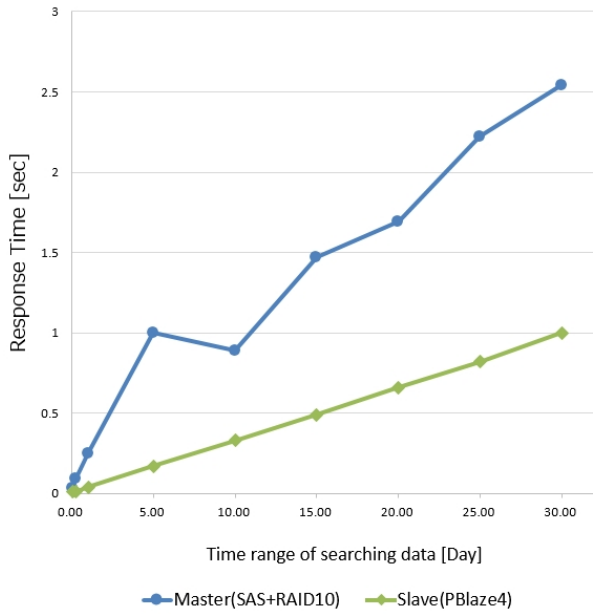

Figure 6: Comparison of response time of MySQL-based database on between the master server and the slave server.

6. まとめ

RIBF でアーカイブシステムとして利用されている RIBFCAS と MyDAQ2 においてシステムをアップグレードした。従来の RIBFCAS の問題点は冗長性の低さであったが、PostgreSQL のレプリケーション機能を用いる事で冗長性の確保に成功した。また PCI Express ベースの SSD である PBlaze4 を採用する事で、従来に比べデータベースからのレスポンスが２倍以上向上した。

MyDAQ2 のケースでは、PBlaze4 をインストールしたスレーブはマスタと比較し、２．５倍高いデータ呼び出し速度の性能を示している。また、ブラウザでのチャート表示も、１ヵ月間といった長期間のデータ表示させた場合、マスタに比べ２倍の表示速度の向上が見られた。

MyDAQ2 クライアントにおいて、RIBFCAS のデータを JSON フォーマットで取得可能にさせる機能拡張をした結果、MyDAQ2 と RIBFCAS、双方のデータを JavaScript チャートで表示、比較する事が簡便に利用可能になり、有用性が向上した。

参考文献

[1] O. Kamigaito *et al.*, Proceedings of IPAC2016, Busan, Korea, (2016), pp. 1281.
[2] M. Komiyama *et al.*, Proceedings of ICALEPCS2011, Grenoble, France, 2011, pp. 1423.
[3] H. Sako *et al.*, Proceedings of ICALEPCS2009, Kobe, Japan, (2009), pp. 842.
[4] https://www.ask-corp.jp/products/ocz/pci-express-ssd/ocz-revodrive3-x2.html
[5] http://www.memblaze.com/en/
[6] T. Hirono *et al.*, Proceedings of PCaPAC08, Ljubljana, Slovenia, (2008), pp. 55.
[7] A. Uchiyama, *et al.*, Proceedings of PCaPAC2016, Campinas, Brazil, (2016), No. WEPOPRPO12.
[8] A. Uchiyama *et al.*, Proceedings of 10th Annual Meeting of Particle Accelerator Society of Japan, Nagoya, Japan, (2013), pp. 1110.
[9] http://www.gnuplot.info/