

KEK 電子陽電子入射器における Archiver Appliance 運用に関するトラブル及び対処

PROBLEMS AND SOME APPROACHES DURING OPERATION OF ARCHIVER APPLIANCE IN KEK ELECTRON POSITRON INJECTOR LINAC

佐武いつか^{#, A)}, 佐藤政則^{A, B)}, 佐々木信哉^{A)}, 廣瀬雅哉^{C)}, 工藤拓弥^{D)}, 草野史郎^{D)}, 王迪^{B)}
Itsuka Satake^{#, A)}, Masanori Satoh^{A, B)}, Shinya Sasaki^{A)}, Masaya Hirose^{C)}, Takuya Kudou^{D)}, Shiro Kusano^{D)}, Di Wang^{B)}

^{A)} High Energy Accelerator Organization (KEK), Accelerator Laboratory

^{B)} The Graduate University for Advanced Studies (SOKENDAI), Department of Accelerator Science

^{C)} Kanto Information Service Co., Ltd (KIS)

^{D)} Mitsubishi Electric System & Service Co., Ltd (MSC)

Abstract

Electron and positron beams are delivered to five different rings of SuperKEKB electron/positron/positron damping rings, PF, PF-AR in the KEK electron/positron injector linac. In 2019, we realized top-up injection of 4 rings at the same time, and we are still continuing various technical research and development to further improve beam performance. In addition, since the status of many devices is monitored and controlled, the monitoring target and data size are steadily increasing. KEK injector linac has been operating the Archiver Appliance since November 2019 as data collection software, along with the CSS archiver that has been in operation. The number of archived items has reached about 120,000, and these data are recorded by the archive system. Archiver Appliance consumes less disk and can read data very quickly. However, there is a problem that it will stop every few months in 2020. Using Kibana, a data visualization tool, for the research, we started to monitor the system information on the Archiver Appliance operation server, including the Broadcast information on the KEK injector control network. In this paper we report in detail the research of cause and some attempted countermeasures.

1. はじめに

KEK 電子陽電子入射器では、電子及び陽電子ビームを SuperKEKB 電子/陽電子、陽電子ダンピングリング、PF、PF-AR の異なる 5 つのリングに供給している。2019 年には 4 リング同時トップアップ入射を実現し、現在もさ

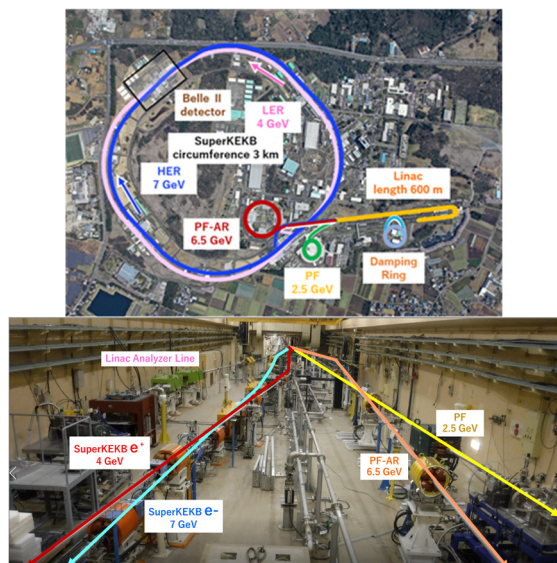


Figure 1: Photograph of KEK injector linac and related accelerators.

らなるビーム性能向上のため、様々な技術研究開発を続けている(Fig. 1) [1]。KEK 入射器ではデータ収集ソフトウェアとして、従来からの CSS archiver [2-4]とともに、2019 年秋から Archiver Appliance(AA) [5]を運用している[6]。現在のアーカイブ対象は約 12 万点に達し、これらのデータをアーカイバーで記録している。AA は、CSS archiver と比べてディスク消費量も少なく、速いデータ読み出しが実現できている。しかしながら、2020 年頃より数ヶ月ごとに AA が停止するという問題が発生している。調査のためにデータの可視化ツールである Kibana[7]を用いて、KEK 入射器制御ネットワークにおける Broadcast の調査も含め、AA 運用サーバーにおけるシステム情報の監視を始めた。現在の運用状況に加えて、原因調査及びいくつか試みた対処について詳細を報告する。

2. Archiver Appliance システム概要

AA は、SLAC、BNL、MSU のコラボレーションによって開発されたアーカイブシステムである。データファイルはバイナリファイルとして記録され、短期(STS)、中期(MTS)、長期(LTS)データ保存用に分かれている。これらの間でデータを移動する組み込みプロセスが動いている。それぞれのデータは Ramdisk, SSD, SAN などの異なる記憶媒体を使い分けることが可能である。短期データ保存用に SSD などの高速ストレージを使用することで、ストレージの高速化が実現でき、速い読み出しを可能にしている。Table 1 にはソフトウェア環境を示す。Figure 2 に、

[#] Itsuka.satake@kek.jp

Table 1: Software Environment of Archiver Appliance

Archiver Appliance	Nov 2019 Release
OS	CentOS 7.6.1810
Tomcat	apache-tomcat-9.0.36
Java	12.0.2 (OpenJDK 12.0.2)
Database	MySQL 5.7.25 (mysql-connector-java 8.0.15)

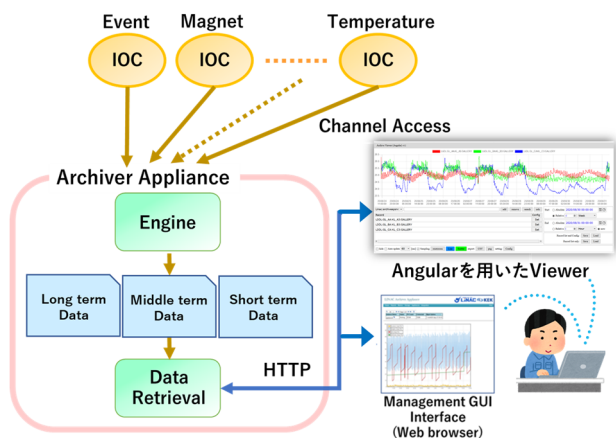


Figure 2: Schematic view of Archiver Appliance in the KEK injector linac.

KEK 入射器における AA のシステム概要を示す。クライアント及び Retrieval 間の通信には HTTP を用いており、JSON や CSV などのフォーマットでのデータ取得が可能である。また、Web ブラウザを用いたシステム管理・監視機能をもつインターフェイスや Business Process Logics (BPL)と呼ばれるAPIを利用した処理スクリプト(Pythonを用いた)が用意されている。PV の登録情報は、データベースであるMySQLで管理しており、データ表示には、Angularを用いたViewerを使用している。現在のアーカイブシステムにおける PV 登録数は約 12 万点あり、AA での一年分のデータサイズは約 3.7 TB である。

3. AA の停止問題

2020 年から 1,2 ヶ月ごとに AA が停止してしまうというトラブルが生じていた。

2020.10.20 (火)	AA 停止(サーバー停止なし)
2020.11.02 (月)	AA 停止(サーバー停止)
2020.12.14 (月)	AA 停止(サーバー停止)
2020.12.22 (火)	サーバー修理 メモリ交換実施
2021. 2.28 (日)	AA 停止、再起動して復旧
2021. 3.10 (水)	対処①
2021. 4.21 (水)	対処②
2021. 5.26 (水)	対処③

おもにネットワーク機器情報の監視・可視化ツ

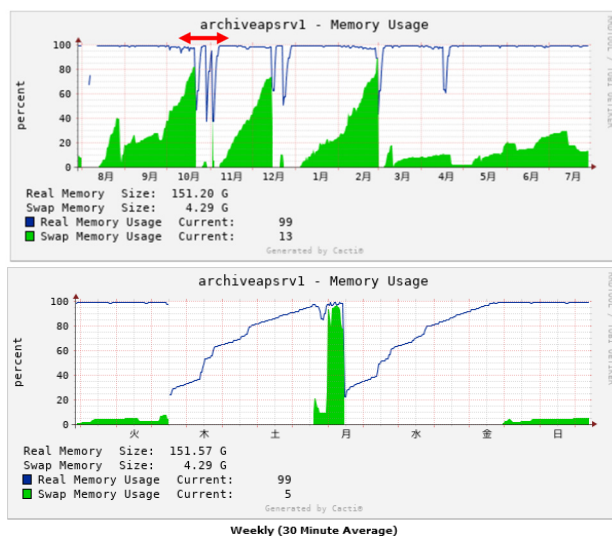


Figure 3: Screenshot of Cacti for memory usage on Archiver Appliance server in the recent past months (top) and around 2020.11.02 (top (red arrow term), bottom).

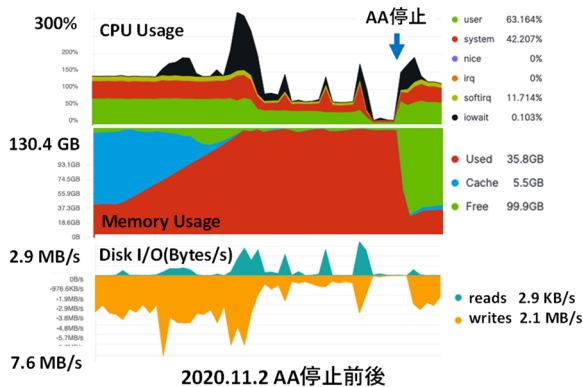


Figure 4: System Information of Kibana for Archiver Appliance server before server repair.

ルである Cacti と Kibana を用いて調査した。Cacti より、停止するまでにメモリ使用率が急増し、停止に至っているように見える(Fig. 3)。Kibana でシステム情報を確認すると、AA が停止する前には、メモリ使用率が 90%以上となっていた(Fig. 4)。AA のプロセスログにおいても、"Out of Memory Error" というログが残っている。また、Disk I/O も増加しており、停止前には CPU の IO wait も高くなっている。/proc/meminfo 情報を確認すると、Slab キャッシュが約 136 GB となっており、システムメモリ(148 GB)のほとんどが使用されていた。Slab キャッシュは、カーネル内のメモリ領域であり、カーネル空間内の様々なメモリ資源を効率的に利用するため、資源ごとにキャッシュをする仕組みである。そのため、Slab キャッシュを解放するコマンドを実行したが、

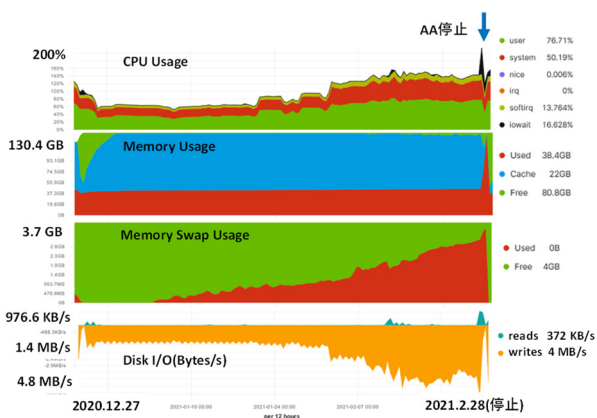


Figure 5: System Information of Kibana for Archiver Appliance server after server repair.

解放されなかった。Slab キャッシュには SReclaimable (回収可能)と SUNreclaim (回収不可)なものがあり、このときのキャッシュは SUNreclaim がほとんどを占めていたためと思われる。その中でも xfs_inode が多くを占めていた。これはファイルシステムに関連している。

以下についても調査したが、原因の特定には至らなかった。

- lsof コマンドにより、大量のファイルを開いているプロセスを特定する。
- iotop コマンドにより、Disk I/O 負荷の高いプロセスを特定する。

2020年12月22日には、サーバーのメモリ障害が判明し、メモリ交換を実施した。しかし、メモリ交換後も Swap メモリの使用量が増加し続けており、Swap メモリサイズ (4GB) の約 79.4%到達以降に動作を停止した(Fig. 5)。起動してから約 9 週間で停止に至った。停止後は、システム再起動により復旧している。サーバー修理後も AA 停止は続いたため、大きく分類して 3 つの対処を試みることにした。

4. 対処とその結果

Disk I/O 増加や xfs_inode が関連している可能性が高いことから、NAS 経由でデータのやりとりをおこなっていることが原因の 1 つではないかと思われる。そのため、1 つ目の対処として、インストールディレクトリを NAS 上からローカルディスク上に移動して実行するよう変更した。その結果、Swap メモリ使用量の増加率は下がったが、増加傾向のままであった(Fig. 6)。

次に、AA のデータである sts, mts, lts は、全て HDD を指定していた。そこで 2 つ目の対処として、1 日分のデータである sts データのみ、HDD から RAM ディスク(tmpfs)を指定するよう変更した。RAM ディスクは、メモリ上に作成されるファイルシステムであるため、読み書きの速度が高速になる。RAM ディスクに格納することで速度に違いがあるか、dd コマンドで簡易的に確認したところ、書き込み速度は約 3 倍、読み込み速度は約 33 倍速くなること

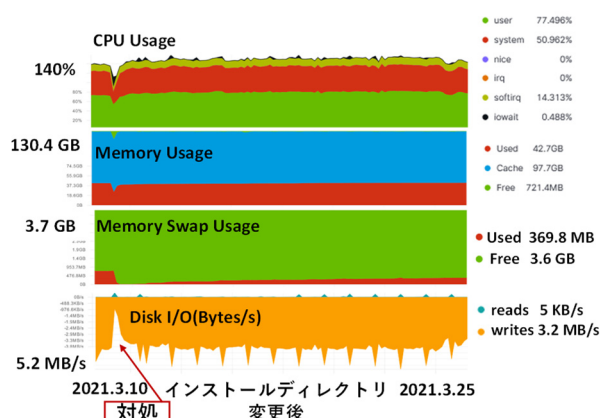


Figure 6: System Information of Kibana for Archiver Appliance server after changing the install directory on local disk.

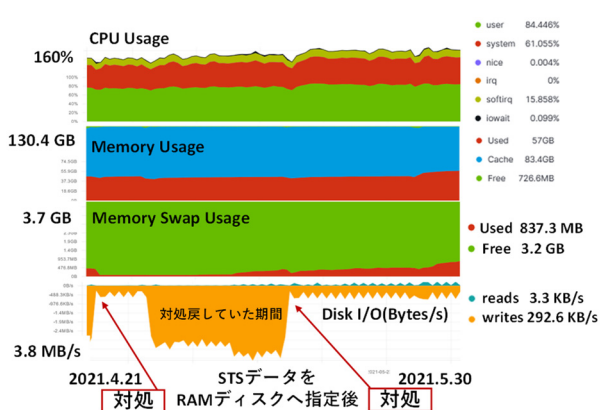


Figure 7: System Information of Kibana for Archiver Appliance server after changing the STS data from HDD to RAM disk.

がわかった。電源 OFF や再起動でファイルは消えるが、AA 再起動時に MTS へ移動されるため、AA の運用上は問題とならない。その結果、Swap メモリ使用量の増加率は下がったが、増加傾向のままであった(Fig. 7)。

3 つ目の対処として、クラスタ化することで各エンジンの負荷分散を試みた。メインのクラスタが 10 万以上の PV

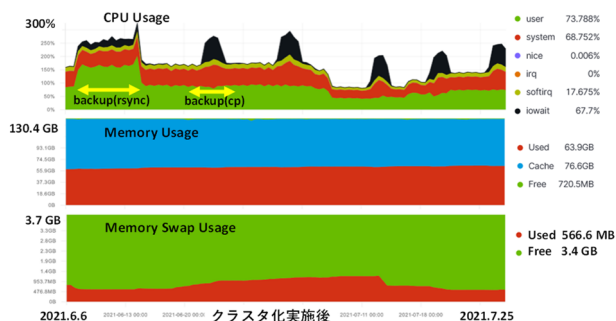


Figure 8: System Information of Kibana for Archiver Appliance server after clustering.

を持っていたため、これを分けることとした。その結果、CPU 負荷にはあまり変化は見られなかったが、Swap メモリ使用量の増加は止まったように見えている(Fig. 8)。エンジンの負荷が高いことも停止の要因だった可能性がある。このまま Swap メモリ使用量が一定レベルで維持されることを期待している。

5. まとめと展望

AA の Swap メモリ不足による AA が 1,2 ヶ月で停止するトラブルは、インストールディレクトリ変更、直近データの RAM ディスク指定及びエンジンの負荷分散をおこなうことで改善が見られた。夏期メンテナンスでサーバーを追加し(2 台体制へ)、さらにクラスタ化する予定である。

エンジン負荷分散の効果を秋の運転時で確認する。また、メモリ使用量増加の問題は長期的に監視する予定である。システム情報の可視化ツールである Kibana や Cacti が調査に役立った。

今後は監視システムをより強化して、安定したアーカイバースステムの運用と、調査及び問題解決の効率化を進めていく。

参考文献

- [1] K. Furukawa *et al.*, “Achievement of 200,000 hours of operation at KEK 7-GeV electron positron injector linac”, in Proceedings of the 17th Annual Meeting of Particle Accelerator Society of Japan, Sep. 2-Sep. 4, 2020, FRPP68.
- [2] Control System Studio;
<http://cs-studio.sourceforge.net/>
- [3] Control System Studio (CSS) at KEK;
<http://www-linac.kek.jp/cont/epics/css/>
- [4] T. Kudou *et al.*, “PRESENT STATUS OF CSS ARCHIVER AT KEK INJECTOR LINAC”, in Proceedings of the 12th Annual Meeting of Particle Accelerator Society of Japan, Tsuruga, Japan, Aug. 5-7, 2015, WEP113.
- [5] The EPICS Archiver Appliance;
https://slacmshankar.github.io/epicsarchiver_docs/
- [6] I. Satake *et al.*, “OPERATION STATUS OF ARCHIVER APPLIANCE IN KEK ELECTRON/POSITRON INJECTOR LINAC”, in Proceedings of the 17th Annual Meeting of Particle Accelerator Society of Japan, Sep. 2-Sep. 4, 2020, FRPP25.
- [7] Kibana;
<https://www.elastic.co/jp/kibana/>