BEAM DATA LOGGING SYSTEM BASE ON NoSQL DATABASE AT SSRF*

Y.B. Yan, Y.B. Leng#, L.W. Lai, Z.C. Chen, H.L. Geng Shanghai Synchrotron Radiation Facility (SSRF) Shanghai Institute of Applied Physics, Chinese Academy of Sciences
Shanghai 201204, P.R. China

Abstract

To improve the accelerator reliability and stability, a beam data logging system was built at SSRF, which was base on NoSQL database Couchbase. The Couchbase is an open source software, and can be used both as document database and pure key-value database. The logging system stores beam parameters under predefined conditions. It is mainly used for the fault diagnosis, beam parameters tracking or automatic report generation. The details of the data logging system will be reported in this paper.

OVERVIEW

Shanghai Synchrotron Radiation Facility (SSRF) is a low emittance third generation light source located at Shanghai, China. It includes a 150MeV LINAC, 150MeV to 3.5GeV Booster, LINAC to Booster transfer line, Booster to storage ring transfer line and 3.5GeV storage ring. To improve the accelerator reliability and stability, a beam data logging system was built, mainly based on the beam instrumentation system.

SSRF beam instrumentation system consists of more than 200 devices, which covered the beam position measurement, beam charge & current measurement, beam size & length measurement, fill pattern measurement and so on [1]. All these parameters are very important during the accelerator commissioning, operation and machine studies. More than 20k scalar process variables and hundreds of 2k-points waveform records are published online every second. With proper storage and analyze tool-kits, these data could be invaluable. Otherwise the potential of various new electronics will be wasted.

On the other hand, various hardware and software failures have been recorded during the past few years, such as global orbit disturbance, random glitch or offset jump of individual position readings [2]. All these failures affected the reliability and stability of the entire machine. There were no effective tools to analyze the reason due to lack of adequate raw data. The regular sampling rate of achieved data is about 1Hz. History of broadband data such as turn-by-turn (several hundreds kHz) orbit data or bunch-by-bunch data are required in this case. Due to the huge size, the data are not likely to be stored periodically. A logging system, which stores the raw data under some predetermined conditions, is urgently needed.

* Work supported by the Knowledge Innovation Program of Chinese Academy of Sciences and National Natural Science Foundation of China (No. 11105211, 11305253)
# lengyongbin@sinap.ac.cn

SYSTEM ARCHITECTURE

The data logging system is based on the Couchbase [3], which is an open source, distributed NoSQL database. It provides key-value or document access with low latency and high sustained throughput. The system architecture is shown in Figure 1.

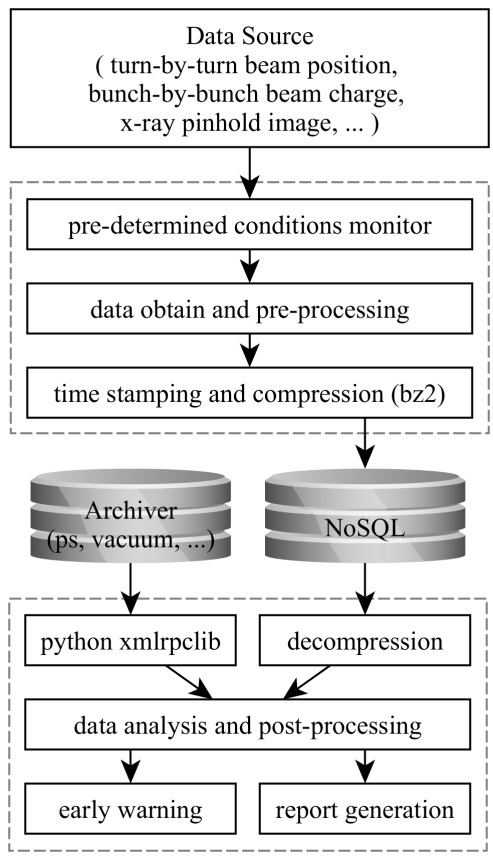

Figure 1: The architecture of data logging system.

In this system IBM System x3550 M4 server and IBM Storwize V3700 storage system are adopted, which is a cost-effective option to achieve high performance. All the software run on the Linux operating system, and written using Python, a widely used general-purpose, high-level programming language.

Data Source

For the particle accelerator, the data can be divided into two categories, hardware device related and beam related parameters. The hardware device related (such as vacuity,

ISBN 978-3-95450-146-5

magnet current, undulator gap or shift, etc.) and scalar, primary beam parameters (such as beam current, close orbit, etc) have been achieved in the regular archiver system. The logging system mainly stores the waveform records, including the raw data (such as turn-by-turn orbit data, bunch-by-bunch charge data, synchrotron light ccd image, etc) and processed data (such as beam spectrum data, beta function measured, etc).

Data Storage

All the data are obtained under some pre-determined conditions, such as the global orbit disturbance, beam injecton, etc. The dedicated routines decide whether the conditions have been satisfied.

Before storing, some of data are pre-processed using some algorithms, then all data under the pre-determined condition are packaged and stored in the database as a entry. In order to reduce the size, the bzip2 is adopted, which is a free and open-source file compression program and uses the Burrows-Wheeler algorithm.

Data Processing

A lot of data pre-processing algorithms, such as the correlation analysis, the cluster analysis and the principal component analysis, can be used to extract the useful information from raw data. Beside the standard library, Python supports a large number of 3rd party libraries. It makes these algorithms can be implemented easily.

The signals from various probes are different aspects of a single measurement procedure. The correlation analysis of the overall signals fits the characteristic function such as the beta function and the dispersion function to the data, which would effectively increase the usage of the original information and promote the accuracy, reliability and feasibility of the output results [4].

The principal component analysis (PCA) finds a small number of uncorrelated principal components that can account for the maximum amount of observed variances and covariances in the data. Each principal component is a linear combination of the observed signals and retains the maximum variance along its direction. It can be achieved by an singular value decomposition (SVD) of the data matrix, such as all the turn-by-turn orbit data of the storage ring. The spatial and temporal vectors can be used to identify the betatron motion, energy motion and others, such as electronics noise [5, 6].

The post-processing is mainly used for fault diagnosis, beam parameters tracking, which will be discussed in the later part.

Early Warning

As a user facility, the reliability and stability are very important. Before the failure occurs, if the qualitative or quantitative forecasting based on the achieved data can be made, it will effectively extend the mean time between failures of the accelerator. Especially for some slow drift, the early warning will be much helpful for the operators and physicists to optimize machine parameters.

Report Generation

For the regular data analysis, some general algorithms and scripts have been used, just called with different arguments. Actually more functions can be also added into the scripts, such as data retrieving from the database (Archiver or Couchbase), data processing and graphing, files generation, etc. According to the requirement, the files will be automatically generated daily, or at some specific time. As the instance, a multi-page document about the beam position distortion during the undulator gap adjustment is given in Figure 2.

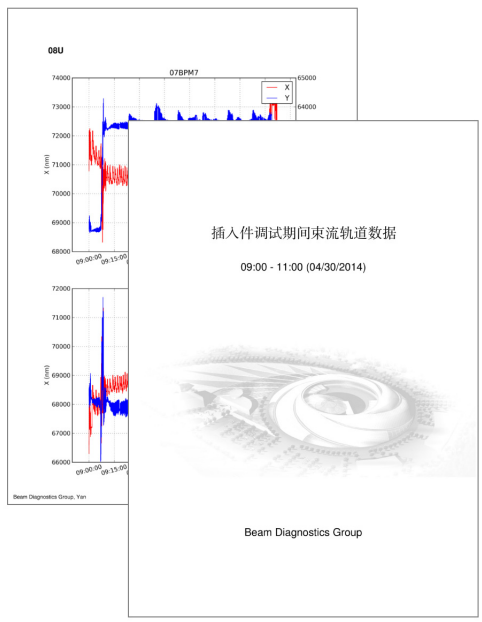

Figure 2: The auto-generated documentation.

The matplotlib and reportlab library are adopted in Python. The matplotlib is a plotting library and can save plot to image file instead of displaying it. The reportlab library allows rapid creation of portable document format (PDF) documents. Its open-source version is available under the BSD license [7]. The final reports include text (such as title, comments, calculation results, etc) and the previous generated images.

FAULT DIAGNOSIS

The particle accelerators are complicated system, with a large number of various components. The fault detection and diagnosis are a long and difficult task, and the data logging system will be helpful.

There is a typical example, which happened after a summer shutdown. The beam position monitor (BPM) cables and part of the electronics were upgraded. As mentioned above, the beta function of the storage ring can be stored using some data pre-processing algorithms. But the data is abnormal at one position, shown in Figure 3. Finally the cause was found out, a cable connection error. The neighbouring cables (Channel C and D) were cross-connected by mistake.

ISBN 978-3-95450-146-5

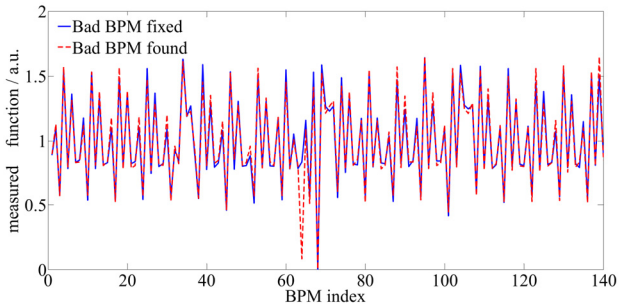

Figure 3: The beta function measured.

PARAMETERS TRACKING

The beam parameter tracking is very important for the accelerator operation or machine studies, such as the tune. Since December 2012, SSRF has operated in top-up mode, which improve the efficiency and quality of synchrotron light. The top-up injections are made continuously at the time interval of about ten minutes; each injection cycle takes about ten seconds. The tune can be archived during the injections.

The tune is extracted from the excitated turn-by-turn orbit data, which is the amplitude of resonance peak of betatron oscillation. The pre-determined condition is the gate signal of top-up injection and one entry is stored. Figure 4 shows the horizontal tune during one day. The rms can be less than 2e-4 in twelve hours. The jump may be derived from the adjustment of reference orbit.

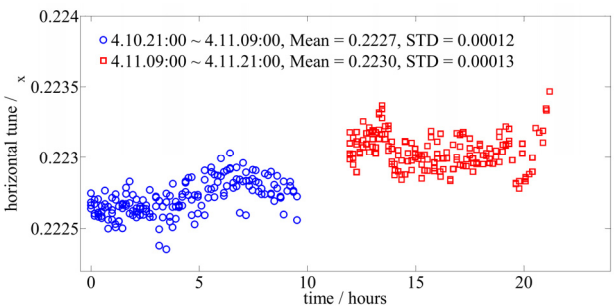

Figure 4: The horizontal tune during 24 hours.

The tune drift over one hundred days are shown in Figure 5. The blue points represent the horizontal tune, while the red represent vertical.

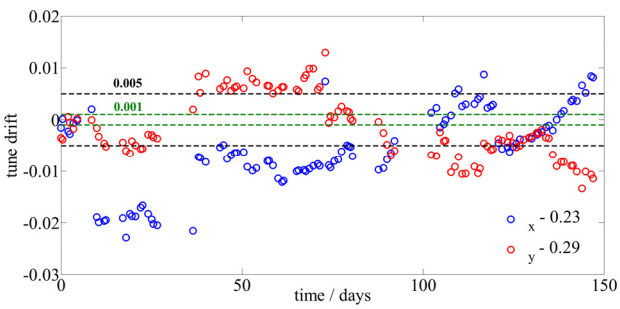

Figure 5: The tune drift during 150 days.

CONCLUSION

The beam data logging system has been implemented base on NoSQL database. More functionality will be added in the future. The preliminary applications show great potential for the fault diagnosis and parameters tracking. This will improve the efficiency of the operators and physicists.

With the increasing complexity of particle accelerator, the reliability and stability will become more and more crucial. Besides using high-reliability hardware, the rapid fault diagnosis, data mining and predictive analytics will also be effective ways to improve the efficiency of the accelerator.

ACKNOWLEDGMENT

We would like to thank everyone who contributed to this work through discussions and suggestions.

REFERENCES

[1] Y.B. Leng, K.R. Ye, et al.,"SSRF Beam Diagnostics System Commissioning", DIPAC'09, Basel, Switzerland (2009), http://www.jacow.org
[2] Y. B. Leng, Y. B. Yan, et al.,"Beam Diagnostics Global Data Warehouse Implementation and Application at SSRF", IPAC'11, San Sebastian, Spain (2011), http://www.jacow.org
[3] Couchbase website: http://www.couchbase.com
[4] Z.C. Chen, Y. B. Leng,"Study of Correlation Analysis of Global Data from a Storage Ring", Chinese Academy of Sciences, 2014
[5] J.H. Chen, Z.T Zhao,"Preliminary Application of Turn-by-Turn Data Analysis to the SSRF Storage Ring", Chinese Physics C, Vol. 33, No. 7, 2009
[6] N. Zhang, Y. Yang, et al.,"Application of Model Independent Analysis Based Method to Accelerator Bunch-by-Bunch Research", High Power Laser and Particle Beams, Vol. 26, No. 3, 2014
[7] ReportLab website: http://www.reportlab.com

ISBN 978-3-95450-146-5