Temat (EN): System for the data-agnostic quality control for the CERN ALICE experiment

Temat (PL): System do kontroli jakości dowolnego rodzaju danych w eksperymencie ALICE CERN

The next period of operation of the LHC accelerator will be characterized by much larger amounts of generated data in relation to the previous periods of operation. Therefore, ALICE Collaboration decided to upgrade detector [1] and implement a new computing system O2 [2] for the online-offline data processing. One of the parts of the O2 system is the Quality Control (QC), which is being implemented [3].

The QC system plays an important role in data processing. It is used to monitor data quality at each reconstruction step and can rise alarms in case of any anomaly detection. Algorithms to check data quality are usually written for a given detector which requires a deep knowledge of detector response to traversing particles. . However, during early detector development or commissioning one could benefit from a general data-agnostic QC to derive features based on the binary data blobs. In addition, it can be used to monitor data acquisition and processing software, thus helping to investigate potential issues in the system during data taking.

The scope of work includes the development of a system for the data-agnostic QC in the O2 framework. The work consists of the following elements:

1. Getting general knowledge on the O2 programming environment and the Calibration Constant Data Base (CCDB) along with the analysis of the structure and type of data stored in the CCDB and the methods for obtaining data from CCDB.
2. Analysis of the existing environments for scientific visualization, including the Grafana environment,
3. Proposing and developing several methods for analysis and scientific visualization using artificial intelligence.
4. Integration of the developed methods with the selected environment for scientific visualization. The use of plug-in modules and container technology is recommended.

When developing a system, it is necessary to take into account the specific requirements of the user resulting from the scope of data analysis, his practice related to the scientific discipline and the volume of data intended for analysis and visualization. Due to the diversity and volume of data, artificial intelligence methods are to be considered. It is necessary to propose a set of data analysis and visualization methods, with the possibility of easy integration and extension. Their diversity should ensure the use of different analysis and visualization criteria aimed at discovering previously unknown artefacts and phenomena occurring during particle collisions. The research goal is to select the method of data analysis and visualization to the problem represented by the data. The scope of the work, therefore, corresponds to the 4th paradigm of scientific research.

[1] <https://cds.cern.ch/record/1475243>

[2] <https://aliceo2group.github.io>

[3] <https://github.com/AliceO2Group/QualityControl>

Temat (PL): System do kontroli jakości dowolnego rodzaju danych w eksperymencie ALICE CERN

Temat (EN): System for the data-agnostic quality control for the CERN ALICE experiment

Kolejny okres pracy akceleratora LHC będzie charakteryzował się dużo większymi ilościami generowanych danych w stosunku do poprzednich okresów pracy. Dlatego w ramach ALICE Collaboration zdecydowano się zaktualizować detektor [1] i wdrożyć nowy system obliczeniowy O2 [2] do przetwarzania danych online-offline. Jedną z części systemu O2 jest kontrola jakości (QC), która jest aktualnie wdrażana [3].

System kontroli jakości odgrywa ważną rolę w przetwarzaniu danych. Służy do monitorowania jakości danych na każdym etapie rekonstrukcji i może wywołać alarm w przypadku wykrycia jakiejkolwiek anomalii. Algorytmy sprawdzania jakości danych są zwykle tworzone dla danego detektora, co wymaga głębokiej wiedzy na temat odpowiedzi detektora na przechodzące cząstki. Jednak podczas wczesnych etapów budowy czy uruchamiania detektora można odnieść korzyści z ogólnej kontroli jakości niezależnej od rodzaju danych na podstawie binarnych obiektów typu blob. Ponadto taki system może być używany do monitorowania oprogramowania do odczytu i przetwarzania danych, pomagając w ten sposób badać potencjalne problemy w systemie podczas zbierania danych.  
  
Zakres prac obejmuje opracowanie systemu do kontroli jakości danych agnostycznych w ramach O2. W skład pracy wchodzą następujące elementy:

1. Uzyskanie ogólnej wiedzy o środowisku programistycznym O2 i o bazie danych kalibracyjnych, Calibration Constant Data Base (CCDB,) wraz z analizą struktury i rodzaju danych przechowywanych w CCDB oraz metodach pozyskiwania danych z CCDB.
2. Analiza istniejących środowisk do wizualizacji naukowej, w tym środowiska Grafana,
3. Zaproponowanie i opracowanie kilku metod analizy danych i wizualizacji naukowej z wykorzystaniem sztucznej inteligencji.
4. Integracja opracowanych metod z wybranym środowiskiem do wizualizacji naukowej. Zalecane jest użycie modułów dodatkowych i technologii kontenerów.

Tworząc system należy wziąć pod uwagę specyficzne wymagania użytkownika wynikające z zakresu analizy danych, jego praktyki związanej z dyscypliną naukową oraz objętości danych przeznaczonych do analizy i wizualizacji. Ze względu na różnorodność i ilość danych należy rozważyć metody sztucznej inteligencji. Konieczne jest zaproponowanie zestawu metod analizy i wizualizacji danych, z możliwością łatwej integracji i rozbudowy. Ich różnorodność powinna zapewnić stosowanie różnych kryteriów analizy i wizualizacji, mających na celu odkrywanie nieznanych wcześniej artefaktów i zjawisk zachodzących podczas zderzeń cząstek. Celem badań jest dobór metody analizy i wizualizacji danych do problemu reprezentowanego przez dane. Zakres pracy koresponduje zatem z czwartym paradygmatem badań naukowych.

[1] <https://cds.cern.ch/record/1475243>

[2] <https://aliceo2group.github.io>

[3] <https://github.com/AliceO2Group/QualityControl>