Кластерный анализ вычислительных задач системы управления загрузкой эксперимента ATLAS на LHC с использованием визуальной аналитики

Т.П.Галкин[1], М.А.Григорьева[2,3], А.А.Климентов[2], Т.А.Корчуганова[3], И.Е.Мильман, В.В.Пилюгин[1], М.А.Титов[2]
z@wqc.me | magsend@gmail.com | alexei.klimentov@cern.ch | tatiana.korchuganova@cern.ch |
igal.milman@gmail.com | VVPilyugin@mephi.ru | mikhail.titov@cern.ch
[1]Национальный исследовательский ядерный университет «МИФИ», Москва, Россия;
[2]Национальный исследовательский центр «Курчатовский институт», Москва, Россия;
[3]Национальный исследовательский Томский политехнический университет, Томск, Россия.

При проведении экспериментов на научных установках, таких как LHC, RHIC, KEK, для решения задач в области физики высоких энергий (ФВЭ) и ядерной физики (ЯФ) получены сотни петабайт экспериментальных данных. По мере модернизации ускорителей (повышение энергии и светимости), объемы данных стремительно растут и достигли эксабайтной отметки, что также способствует увеличению количества выполняемых задач анализа и обработки данных, непрерывно конкурирующих между собой за вычислительные ресурсы. Последнее обуславливает повышение производительности вычислительной инфраструктуры привлечением высокопроизводительных вычислительных ресурсов, образуя гетерогенную распределённую вычислительную среду (сотни распределённых вычислительных центров). При распределённой модели обработки и анализа данных, оптимизация работы систем управления данными и загрузкой становится фундаментальной задачей, и отсутствие её своевременного решения приводит к экономическим, функциональным и временным потерям. Данная работа описывает первый этап исследований, направленных на решение задачи повышения стабильности и эффективности функционирования распределённых систем обработки данных экспериментов класса мега-сайенс с использованием методов визуальной аналитики. На примере обработки данных эксабайтного диапазона в эксперименте ATLAS на LHC будет продемонстрирована возможность использования визуальных методов для кластерного анализа вычислительных задач/заданий системы управления загрузкой PanDA. При этом будут исследованы и графически интерпретированы взаимозависимости и корреляции между различными параметрами упомянутых задач/заданий в N-мерном пространстве с использованием трёхмерных проекций. Визуальный анализ позволит выявлять схожие (подобные) задачи, а также аномальные задачи, при этом определять, чем обусловлена аномальность. Дальнейшее развитие работ в данном направлении будет основано на увеличении количества анализируемых вычислительных задач и разработке соответствующего программного инструментария.

Ключевые слова: *визуальная аналитика, физика высоких энергий, ядерная физика, эксперимент ATLAS, кластерный анализ.*

Visual Cluster Analysis for Computing Tasks at Workflow Management System of the ATLAS Experiment at the LHC

T.P.Galkin[1], M.A.Grigorieva[2,3], A.A.Klimentov[2], T.A.Korchuganova[3], I.E.Milman, V.V.Pilyugin[1], M.A.Titov[2]
on behalf of the ATLAS Collaboration
z@wqc.me | magsend@gmail.com | alexei.klimentov@cern.ch | tatiana.korchuganova@cern.ch |
igal.milman@gmail.com | VVPilyugin@mephi.ru | mikhail.titov@cern.ch
[1]National Research Nuclear University “MEPHI”, Moscow, Russian Federation;
[2]National Research Center “Kurchatov Institute”, Moscow, Russian Federation;
[3]National Research Tomsk Polytechnic University, Tomsk, Russian Federation.

Hundreds of petabytes of experimental data in high energy and nuclear physics (HENP) have already been obtained by unique scientific facilities, such as LHC, RHIC, KEK. As the accelerators are being modernized (energy and luminosity are increasing), data volumes are rapidly growing and have reached the exabyte scale, that also leads to an increasing the number of analysis and data processing tasks, that are competing continuously for computational resources. The increase in processing tasks causes an increase in the performance of the computing infrastructure by the involvement of high-performance computing resources and forming a heterogeneous distributed computing environment (hundreds of distributed computing centers). With a distributed model of data processing and analysis, the optimization of data management and workload systems becomes a critical task, and the absence of an adequate solution leads to economic, functional and time losses. This work describes the first stage of the study aiming at solving the task of increasing the stability and efficiency of workflow management systems for mega-science experiments by using visual analytics methods. Using the case of the ATLAS experiment at LHC the visual methods for cluster analysis of the workload management system computing tasks/jobs will be applied. The interdependencies and correlations between various tasks/jobs parameters will be investigated and graphically interpreted in N-dimensional space using 3D projections. Visual analysis allows to identify similar jobs, as well as anomalous jobs, and to determine what causes such anomaly. A further evolvement of the work in this direction will be focused on the increasing the amount of analysed computing jobs and the development of the appropriate infrastructure.

Keywords: *visual analytics, high-energy physics, nuclear physics, ATLAS experiment, cluster analysis.*

1. Введение

Научные установки, используемые для решения задач в области физики высоких энергий (ФВЭ) и ядерной физики (ЯФ), генерируют огромные объёмы данных. Эти научные области одними из первых столкнулись с необходимостью анализа и обработки эксабайтных объёмов данных и сопутствующих метаданных.

Стремительное увеличение и усложнение распределённой вычислительной инфраструктуры современных научных экспериментов в области ФВЭ и ЯФ и экспоненциальный рост объёмов обрабатываемых данных обусловили появление новых задач, решение которых

Работа опубликована при финансовой поддержке РФФИ, грант 18-07-20045\18

затруднено или невозможно без визуальной аналитики. Исторически, научные эксперименты в области ФВЭ и ЯФ использовали различные пакеты визуализации для представления данных и решения различных классов задач: моделирование работы детектора, анализ событий, представление результатов исследований для обмена информацией в научном сообществе (HBOOK[6], PAW[5], ROOT[3], MINUIT[7], Ganglia[9], GEANT[2]).

В представленной работе исследуются данные эксперимента ATLAS [1] на LHC - крупнейшего эксперимента в области ФВЭ и ЯФ. Информация, накопленная за многие годы работы системы обработки данных эксперимента ATLAS (ProdSys2 / PanDA [4,8]), содержит данные о ходе выполнения более чем 10 миллионов заданий и порядка 3000 миллионов задач (см. подробнее о заданиях и задачах в разделах 2 и 3). Существующие программно-аппаратные средства позволяют осуществлять контроль, мониторинг и оценку многих параметров и метрик в реальном времени. Однако текущая инфраструктура мониторинга не имеет инструментов оценки корреляций между многочисленными свойствами объектов, в том числе для анализа временных задержек при выполнении вычислительных задач в распределённой компьютерной среде.

Для решения этих проблем будут применены методы визуальной аналитики для поиска новых (неявных) знаний об объектах и обеспечения эффективного взаимодействия с данными, соответствующего человеческим когнитивным системам при обработке сложной информации.

В данной статье описывается первый этап работ по применению визуальной аналитики к исследованию функционирования систем управления загрузкой эксперимента ATLAS: кластерный анализ вычислительных задач с использованием визуально-аналитических методов. Этот анализ позволит пользователю визуально интерпретировать наиболее близкие по параметрам задачи с использованием 3х-мерных проекций, отслеживая при этом корреляции различных комбинаций параметров.

2. Система управления загрузкой эксперимента ATLAS

Для обработки больших объемов данных, эксперименты на LHC используют вычислительную Грид инфраструктуру Worldwide LHC Computing Grid (WLCG)[1], а также ресурсы облачных вычислений и суперкомпьютеры. Система обработки данных эксперимента ATLAS второго поколения (Production System - ProdSys2) предназначена для выполнения комплексных вычислительных приложений, и характеризуется следующими показателями: в среднем выполняется 350К задач в день в более 200 вычислительных центрах (более 250К узлов) тысячами пользователей [4]. Основные составляющие ProdSys2:

- Интерфейс запросов (Request Interface). Отвечает за определение параметров запуска вычислительных заданий и/или группы заданий в форме соответствующего запроса.
- DEfT (Database Engine for Tasks). Данный компонент формирует отдельные задания (tasks), цепочки заданий (пример этапов выполнения: генерация, симуляция, реконструкция и т.д.) или группы заданий на основе установленных параметров в запросе.
- JEDI (Job Execution and Definition Interface). Компонент отвечает за управление полезной нагрузкой на уровне заданий; в основе данного решения лежит динамическое создания задач (jobs) на основе сформированных заданий и управление их выполнением (позволяет оптимизировать загруженность ресурсов), включает в себя механизм принятия решений. Данный компонент также является частью системы PanDA.
- PanDA - система управления загрузкой гетерогенной вычислительной инфраструктуры, отвечает за выполнение задач (внутренний планировщик определяет порядок запуска задач и распределяет по ресурсам). Данная система разработана на основе концепции пилотных задач (pilots / pilot jobs), обеспечивающих контроль состояния ресурсов и назначение рабочей нагрузки (т.н. “поздняя привязка” полезной нагрузки).

3. Исследуемые метрики вычислительных задач системы управления загрузкой

Для запуска цикла обработки данных, который представляет собой конвейер заданий по трансформации и анализу данных, менеджер группы формирует запрос к системе ProdSys2. Запрос преобразуется в набор заданий (tasks) и задач (jobs), которые распределяются системой управления загрузкой PanDA на ресурсах Грид инфраструктуры. Каждое задание разбивается на вычислительные задачи (от нескольких единиц до нескольких тысяч). На первом этапе исследований анализируются визуальные трёхмерные проекции задач, принадлежащих определенным заданиям. Для описания вычислительных задач были выбраны следующие пробные параметры, характеризующие потребление ресурсов:

- Идентификатор задания в системе ProdSys2 / PanDA: ID (integer)
- Время выполнения задачи: duration = endtime - starttime (integer)
- Объём входных данных для задачи: inputFileBytes (integer)
- Объём выходных данных задачи: outputFileBytes (integer)
- Эффективность процессора (отношение общего процессорного времени к произведению времени выполнения запроса на количество ядер): CPU eff per core (integer)
- Потребление процессорного времени: CPU consumption (integer)
- Средний объём общей памяти, который хранится в ОЗУ для процесса: avgRSS (integer)
- Средний размер потребляемой оперативной памяти, в котором учитывается, что страницы памяти могут быть разделены между несколькими процессами: avgPSS (integer)
- Средний размер выделенной виртуальной памяти: avgVMEM (integer)

4. Метод визуальной аналитики для кластерного анализа

Параметры вычислительных задач (jobs) системы ProdSys2 / PanDA можно представить в виде многомерных табличных данных. Тогда каждую строку этой таблицы можно поставить в соответствие с точками в многомерном пространстве E_n, координатами которых являются нормированные значения параметров: $p_i = (p_i^1, p_i^2, \dots, p_i^n) \in E_n$. В данной работе в качестве метрики различия задач выбрано Евклидово расстояние $d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$

[1] The Worldwide LHC Computing Grid, http://wlcg.web.cern.ch

The publication of the conference proceedings was supported by the RFBR, grant No. 18-07-20045\18

между точками $p(p_1, \dots p_n)$ и $q(q_1, \dots q_n)$. Для анализа расстояния между точками n-мерного пространства предлагается использовать визуальное отображение этих точек. Нормализация всех исследуемых значений параметров выполняется по следующей формуле: $x_n = (x - x_{min})/(x_{max} - x_{min})$, где x_{min}, x_{max} - минимальное и максимальное значения параметра x соответственно.

Далее осуществляется проецирование исходного множества точек на одно из трёхмерных пространств. Многомерная точка p_i проецируется в сферу S_i (при этом все координаты, кроме 3-х используемых, приравниваются к нулю). Для визуального представления связных объектов вводится некоторое пороговое расстояние d, задаваемое в интерактивном режиме. Если расстояние между точками n-мерного пространства меньше d, то эти точки соединяются цилиндром, цвет которого меняется от красного (малое расстояние) до синего (расстояние, близкое к d).

На следующем шаге выполняется графическое проецирование сфер и цилиндров на картинную плоскость с последующим их визуальным анализом. Результирующая совокупность сфер и цилиндров образует пространственную сцену с заданной геометрией и оптическими (цветовыми) характеристиками.

5. Программная реализация метода визуальной аналитики для кластерного анализа

Для визуализации и решения задачи анализа используется программа IVAMD [10]. Данная программа реализована на скриптовом языке MAXScript с дополнительным модулем, написанным на языке программирования C#. Основной функционал программного средства включает в себя: отображение пространственной сцены с использованием пользовательских параметров визуализации (порогового расстояния d, радиусов сфер и цилиндров, трёхмерного пространства для проекции), проведение аффинных преобразований трёхмерного пространства, расчёт расстояния в исходном n-мерном пространстве, разбиение на кластеры (обозначаются с использованием различных цветов), проведение микроанализа пространств. При микроанализе, а именно, анализе удаленных точек, важным является то, какие именно координаты вносят больший вклад в расстояние — происходит ли это за счёт всех координат или за счёт большого отличия только нескольких координат. Для определения этого строятся графические проекции исходного множества на плоскости (x_i, x_i) и затем просматриваются все эти проекции при различных i. Результаты кластерного анализа можно получить в виде таблицы, где строки отмечены цветом в зависимости от заданного цвета кластера.

В силу особенностей метода, необходимое для решения задачи (n строк, m столбцов) количество объектов для визуализации: n сфер (равно количеству строк), количество цилиндров равно $n \times (n-1)/2$. Итого $n \times (n+1)/2$, т.е., для отображения 100 строк необходимо отобразить около 5000 объектов. Описываемое программное средство является прототипом реализации метода и имеет ограничение на количество обрабатываемых объектов. Дальнейшая разработка и усовершенствование прототипа, и оптимизация в рамках высокопроизводительной программно-аппаратной инфраструктуры позволит устранить текущие ограничения.

6. Применение метода визуальной аналитики

В качестве тестового испытания был выполнен визуальный анализ вычислительных заданий, состоящих из множества задач. Например, задание №12196428 из данных за 2017-10-02 состоит из 74 задач. Первым шагом был проведен макроанализ всего подпространства, т.е. всё исходное 9-и мерное пространство значений было кластеризовано. Трёхмерная модель выполнена на проекции трех параметров: avgPSS, duration, outputFileBytes (представлено на рисунке 1). В результате выделено 6 кластеров (мощностью 31, 17, 8, 5, 3 и 2), а также 2 одиночные точки (кластеры мощностью 1).

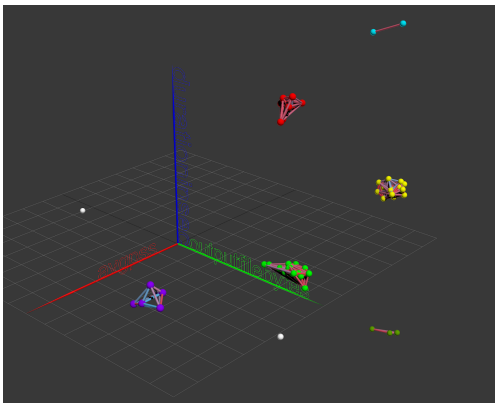

Рис. 1. Разбиение исходных точек на кластеры в трёхмерном пространстве (красная ось - avgPSS, зелёная ось - outputFileBytes, синяя ось - duration in sec).

Далее был проведен микроанализ для определения вклада различных параметров в разбиение на кластеры, а также оценки влияния различных параметров объектов на продолжительность выполнения вычислительных задач (duration).

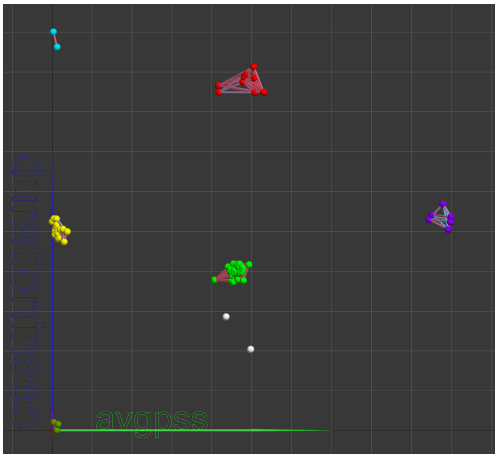

Рис. 2. Графическая проекция на плоскость (avgPSS, duration).

На рисунке 2 можно увидеть, что при одинаковом avgPSS, различие в duration у красного и зелёного кластера значительны. Аналогичная картина наблюдается для параметров avgRSS и avgVMEM, что позволяет сделать вывод о том, что перечисленные параметры вносят вклад в длительность выполнения задач, но только при средних значениях. Данная зависимость требует дополнительного исследования на большем количестве точек.

Рассмотрим кластеры в другом подпространстве. На рисунке 3 видно, что различие в потреблении процессорного времени (CPU Consumption) прямо пропорционально влияет на duration, как и ожидалось.

На рисунке 4 видно, что степень влияния объёма входных файлов (inputFileBytes) на длительность выполнения задач обработки и анализа данных не является решающей при распределенной обработке. Это также

Работа опубликована при финансовой поддержке РФФИ, грант 18-07-20045\18

применимо и к выходным данным/файлам (outputFileBytes). Для количественной оценки влияния объёма данных на длительность выполнения задач требуются дополнительные исследования на большей статистике, которые будут проведены в дальнейшем.

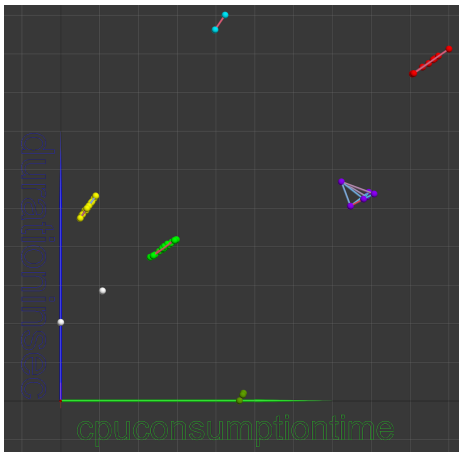

Рис. 3. Графическая проекция на плоскость (cpuConsumptionTime, duration).

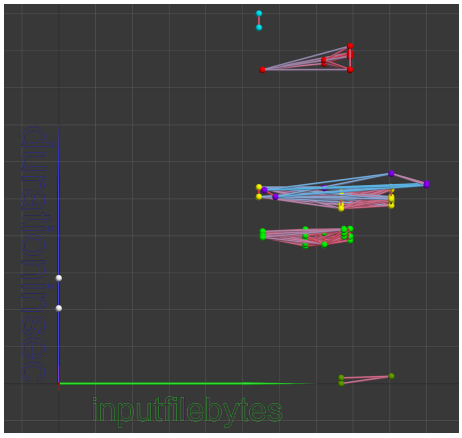

Рис. 4. Графическая проекция на плоскость (inputFileBytes, duration).

7. Заключение

Полученные результаты кластеризации задач с использованием разработанного метода и инструментария являются первым этапом работ по визуальной аналитике данных системы управления загрузкой эксперимента ATLAS. Стоит отметить, что данная работа является новаторской в рамках анализа данных системы ProdSys2 / PanDA. В ходе дальнейших работ планируется построение модели многоуровневой интерактивной визуальной кластеризации. При этом будут анализироваться целые классы задач и заданий анализа и обработки данных эксперимента, на которых будут опробованы различные методы кластеризации, включая методы машинного обучения. Разработанный инструментарий для визуализации будет усовершенствован в сторону обеспечения интерактивного переключения между кластерами различных уровней, обеспечив тем самым удобный для исследователя метод интерпретации результатов анализа разной степени детализации данных.

Предполагается, что полученные результаты могут быть использованы для визуального мониторинга системы управления загрузкой эксперимента ATLAS, для разработки рекомендаций по оптимизации времени выполнения задач анализа и обработки данных.

8. Благодарности

Работа выполнена при поддержке гранта РНФ №18-71-10003 от 02.08.2018г.

9. Литература

[1] The ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider // Journal of Instrumentation, vol. 3, S08003, 2008.

[2] S Agostinelli et al. Geant4 - a simulation toolkit // Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 506, no. 3, pp. 250--303, 2003.

[3] I Antcheva et al. ROOT - A C++ framework for petabyte data storage, statistical analysis and visualization // Computer Physics Communications, vol. 180, no. 12, pp. 2499--2512, 2009.

[4] F H Barreiro et al. The ATLAS Production System Evolution: New Data Processing and Analysis Paradigm for the LHC Run2 and High-Luminosity // Journal of Physics: Conference Series, vol. 898, no. 5, 2017.

[5] R Brun, O Couet, C Vandoni and P Zanarini. PAW, a general-purpose portable software tool for data analysis and presentation // Computer Physics Communications, vol. 57, no. 1, pp. 432--437, 1989.

[6] R Brun and P Palazzi. Graphical Presentation for Data Analysis in Particle Physics Experiments: The HBOOK/HPLOT Package // Proceedings Eurographics ’80, pp. 93--104, 1980.

[7] F James and M Roos. Minuit - a system for function minimization and analysis of the parameter errors and correlations // Computer Physics Communications, vol. 10, no. 6, pp. 343--367, 1975.

[8] A Klimentov et al. Migration of ATLAS PanDA to CERN // Journal of Physics: Conference Series, vol. 219, no. 6, 2010.

[9] M Massie, B Chun and D Culler. The ganglia distributed monitoring system: design, implementation, and experience // Parallel Computing, vol. 30, no. 7, pp. 817--840, 2004.

[10] D Popov, I Milman, V Pilyugin and A Pasko. A solution to a multidimensional dynamic data analysis problem by the visualization method // Scientific Visualization, vol. 8, no. 1, pp. 45--47, 2016.

Об авторах

Галкин Тимофей Петрович, инженер, Национальный исследовательский ядерный университет "МИФИ", email: z@wqc.me

Григорьева Мария Александровна, к.т.н., старший научный сотрудник, Национальный исследовательский центр “Курчатовский институт”, email: magsend@gmail.com

Климентов Алексей Анатольевич, к.ф-м.н., начальник лаборатории, Национальный исследовательский центр “Курчатовский институт”, email: alexei.klimentov@cern.ch

Корчуганова Татьяна Александровна, инженер, Национальный исследовательский Томский политехнический университет, email: tatiana.korchuganova@cern.ch

Мильман Игаль Евгеньевич, email: igal.milman@gmail.com

Пилюгин Виктор Васильевич, к.т.н., профессор, Национальный исследовательский ядерный университет "МИФИ", email: VVPilyugin@mephi.ru

Титов Михаил Анатольевич, PhD, научный сотрудник, Национальный исследовательский центр “Курчатовский институт”, email: mikhail.titov@cern.ch

The publication of the conference proceedings was supported by the RFBR, grant No. 18-07-20045\18