

Применение методов визуальной аналитики для кластеризации задач анализа и обработки данных эксперимента ATLAS на Большом Адронном Коллайдере

Т.П. Галкин^{1,A}, М.А. Григорьева^{2,B,D}, А.А. Климентов^{3,B,C}, Т.А. Корчуганова^{4,D},
И.Е. Мильман^{5,A}, В.В. Пилюгин^{6,A}, М.А. Титов^{7,B}

^A Национальный исследовательский ядерный университет “МИФИ”, Россия

^B Национальный исследовательский центр “Курчатовский институт”, Россия

^C Брукхейвенская Национальная Лаборатория, США

^D Национальный исследовательский Томский политехнический университет, Россия

¹ ORCID: 0000-0003-2859-6275, TPGalkin@mephi.ru

² ORCID: 0000-0002-8851-2187, Maria.Grigorieva@cern.ch

³ ORCID: 0000-0003-2748-4829, Alexei.Klimentov@cern.ch

⁴ ORCID: 0000-0001-5792-8182, Tatiana.Korchuganova@cern.ch

⁵ ORCID: 0000-0001-9705-9401, Igal.Milman@gmail.com

⁶ ORCID: 0000-0001-8648-1690, VVPilyugin@mephi.ru

⁶ ORCID: 0000-0003-2357-7382, Mikhail.Titov@cern.ch

Аннотация

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 уже 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 - анализа данных с использованием способствующего интерактивного визуального интерфейса. В настоящее время визуально-аналитические методы анализа данных активно применяются в различных областях – в научных исследованиях, проектно-конструкторских работах, управлении, финансовом мониторинге, информационной безопасности. Благодаря средствам анализа, которые поддерживают визуализацию данных, полученные выводы может анализировать любой человек, хорошо осведомленный об объекте исследования, но не обязательно владеющий детальной информацией о внутреннем устройстве модели данных. Кроме того, визуальная аналитика существенно упрощает работу с полученными результатами анализа: данные выводятся в виде графических объектов, с которыми можно взаимодействовать привычными для человека способами - мышью или сенсорными экранами. При этом пространственно-образное мышление человека активно используется для выявления новых тенденций и закономерностей, которые открывают собранные данные, оставляя в стороне необходимость “борьбы” с инструментальными аналитическими средствами.

Возможность использования методов визуальной аналитики в данной работе продемонстрирована на примере данных в эксперименте ATLAS на LHC для кластерного анализа вычислительных задач/заданий системы управления загрузкой PanDA. 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.

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

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 может быть существенно упрощено с использованием 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. Такие приложения как HBOOK[1], PAW[2], ROOT[3], Garglia[4], Geant4[5] имеют средства визуализации, которые могут быть использованы для визуального анализа данных.

В представленной работе исследуются метаданные эксперимента ATLAS [6]. Информация, накопленная за многие годы работы системы обработки данных эксперимента ATLAS (ProdSys2/PanDA [7,8]), содержит данные о ходе выпол-

нения более чем 10 миллионов заданий и порядка 3 миллиардов 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 и их возможных причин.

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 3x-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) [9], а также ресурсы облачных вычислений и суперкомпьютеры.

Система обработки данных эксперимента ATLAS второго поколения (Production System - ProdSys2), предназначенная для выполнения комплексных вычислительных приложений, характеризуется концепцией вычислительных заданий и задач.

- **Задание (task)** содержит описание обработки данных одного датасета (группы файлов, содержащих статистически эквивалентные физические события) в эксперименте ATLAS, которое использует заданные пользователем/координатором параметры обработки и параметры, описывающие условия обработки. Задание также содержит программу преобразования/трансформации входных файлов для получения файлов приведенных данных и версии программного обеспечения. Задание состоит из набора задач. Результаты выполнения заданий, т.е. файлы, образуют входные датасеты.

- ProdSys2/PanDA управляет потоками заданий, включая цепочки заданий, которые выполняются в некоторой последовательности на одних и тех же датасетах. Система управляет заданиями относящимися к моделированию событий (генерация событий, моделирование детектора и др.) и к их реконструкции и анализу. Задания разбиваются на задачи,

которые затем выполняются в распределенной компьютерной среде. Результатом выполнения задания могут быть, например, файлы, формирующие выходной датасет.

- Процессы создания цепочек заданий, их фрагментации на задачи, планирование выполнения задач, и их запуск на вычислительных ресурсах, выполняются полностью автоматизировано в системе ProdSys2/PanDA.

- **Задача (job)** представляется как единица обработки (т.н., "рабочая нагрузка" для вычислительного узла). Задача включает в себя входные и выходные файлы, исполняемый код, и, после выполнения, параметры вычислительного узла на котором задача выполнялась, включая некоторые метрики производительности.

В текущем 2018 году вычислительную нагрузку распределенной системы анализа и обработки данных ProdSys2/PanDA можно охарактеризовать следующими метриками: система управляет в среднем 350 тыс. заданиями в день, выполняющимися на более чем 200 вычислительных центрах (более 300 тыс. вычислительных узлов) тысячами пользователей [7]. Рисунок 1 показывает, что ежемесячно система обрабатывает в среднем 129 Петабайт данных, а на рисунке 2 отражено количество выполненных задач - около 23.8 миллионов ежемесячно. Стоит также отметить, что в периоды, предшествующие важным событиям (например, международным конференциям по физике элементарных частиц и ядерной физике), количество выполненных задач физического анализа может достигать 2 миллионов в день.

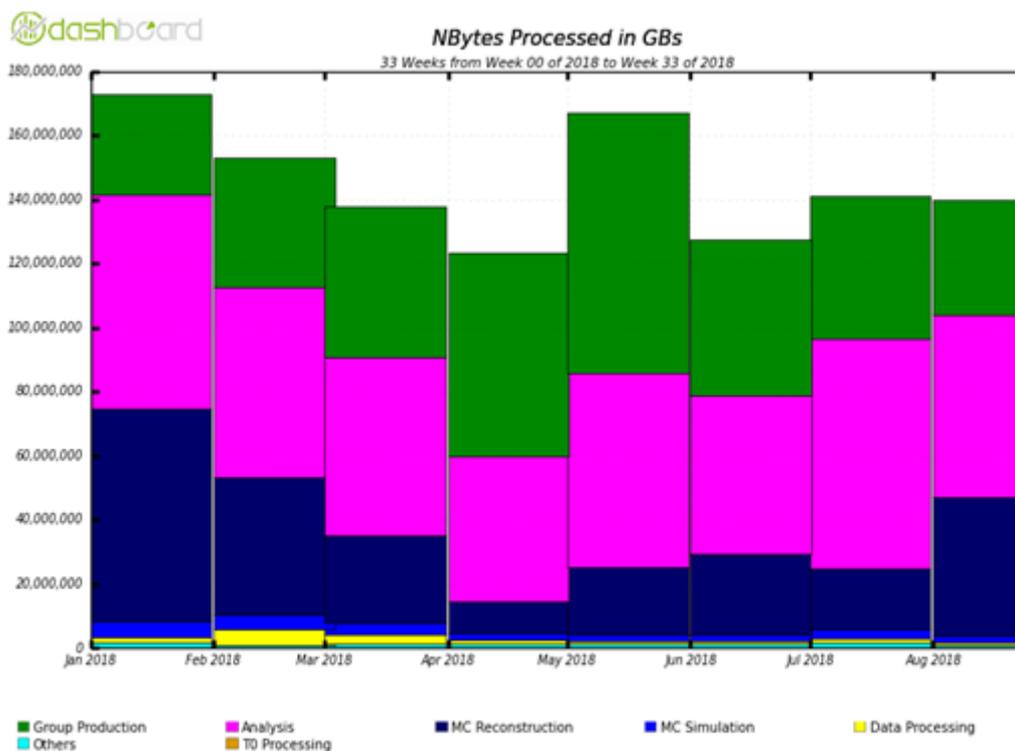


Рис. 1. Объем обработанных данных системой ProdSys2/PanDA в эксперименте ATLAS (в период с января по август 2018 года)

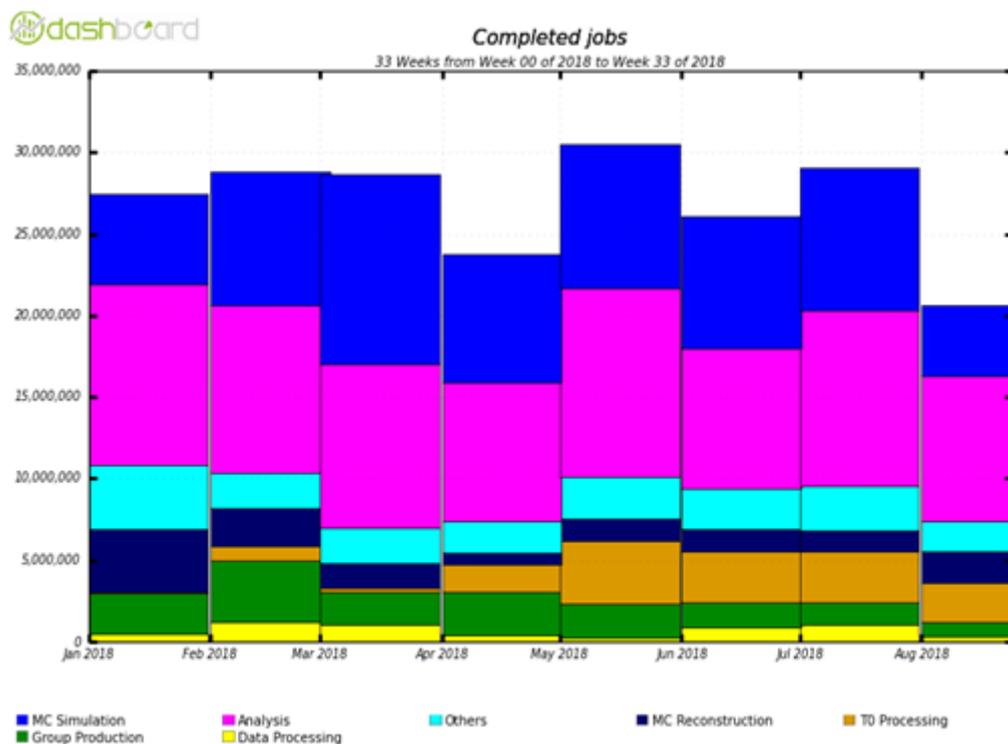


Рис. 2. Количество выполненных задач в эксперименте ATLAS (в период с января по август 2018 года)

Масштабы описанной выше системы, ее внутренняя сложность, гетерогенность и распределенность, а также объемы обрабатываемых данных - все это требует нетривиальных аналитических

средств для анализа и прогнозирования состояния функционирования системы. Первый шаг, который необходимо сделать в этом направлении - разработать инструментарий, предоставляющий

удобный и понятный способ кластеризации данных. Это позволит не просто применять различные алгоритмы кластеризации, которые зачастую являются “черными ящиками”, но и обеспечит исследователя возможностью активного взаимодействия с параметрами кластеризации и визуальным представлением их результатов. Это позволит усилить человеческий контроль над процессом анализа сложных многомерных данных и сделает этот анализ более осмысленным.

3. Исследуемые метрики вычислительных задач системы управления загрузкой

Первым этапом работ является поиск релевантных данных и их подготовка для дальнейшего анализа. В нашем случае были выбраны метрики производительности задач в системе ProdSys2/PanDA, принадлежащих определенным заданиям. ProdSys2/PanDA сохраняет состояние и все текущие параметры задач в базе данных Oracle. Сами задачи, в свою очередь, отправляют информацию обратно в базу Oracle, откуда затем данные могут быть запрошены и подготовлены для анализа.

В последствии осуществляется построение трехмерных визуальных проекций на основе значений параметров задач и их анализ.

Это позволяет исследовать выполнение процесса обработки данных с одинаковым исполняемым кодом и размером входных наборов данных. Для описания вычислительных задач были выбраны следующие пробные параметры, характеризующие потребление ресурсов:

- Идентификатор задания в системе ProdSys2 / PanDA: **ID** (integer)
- Время выполнения задачи: **duration** (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)
 - $cpu_eff = cpu_time / (wall_time * num_of_CPU)$
- 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

Для проведения кластерного анализа средствами визуальной аналитики, был использован метод последовательных проекций [10] с отображением расстояний в виде характеристики геометрической фигуры. Данный метод позволяет исследовать данные непосредственно в исходном многомерном пространстве, не прибегая к методам снижения размерностей. Подробнее этот метод был описан в журнале “Научная визуализация” [11], а также был прочитан доклад на международной конференции ГрафиКон’2014 и опубликована статья по материалам конференции [12]. Данный метод был опробован на данных о кредитных организациях, где позволил выявить ряд аномальных кредитных организаций [13].

Параметры вычислительных задач (jobs) системы ProdSys2 / PanDA были представлены в виде многомерных табличных данных. Далее каждая строка этой таблицы была поставлена в соответствие точке в многомерном пространстве E_n , координатами которой являются нормированные значения параметров: $p_i = (p_i^1, p_i^2, \dots, p_i^n) \in E_n$. В данной ра-

боте в качестве метрики различия задач p и q выбрано Евклидово расстояние $d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$ 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.

Далее осуществляется проецирование исходного множества точек на одно из трёхмерных пространств. Многомерная точка p_i проецируется в трёхмерную точку, которая затем отображается в виде сферы S_i (при этом выполняется цен-

тральная проекция, т.е. все координаты, кроме 3-х используемых, приравниваются к нулю). Графическая проекция полученной пространственной сцены представлена на рисунке 3. При этом у пользователя присутствуют инструменты для управления графическим изображением и пространственной сценой (выполнение аффинных преобразований, получение информации о точках, соответствующих сферам).

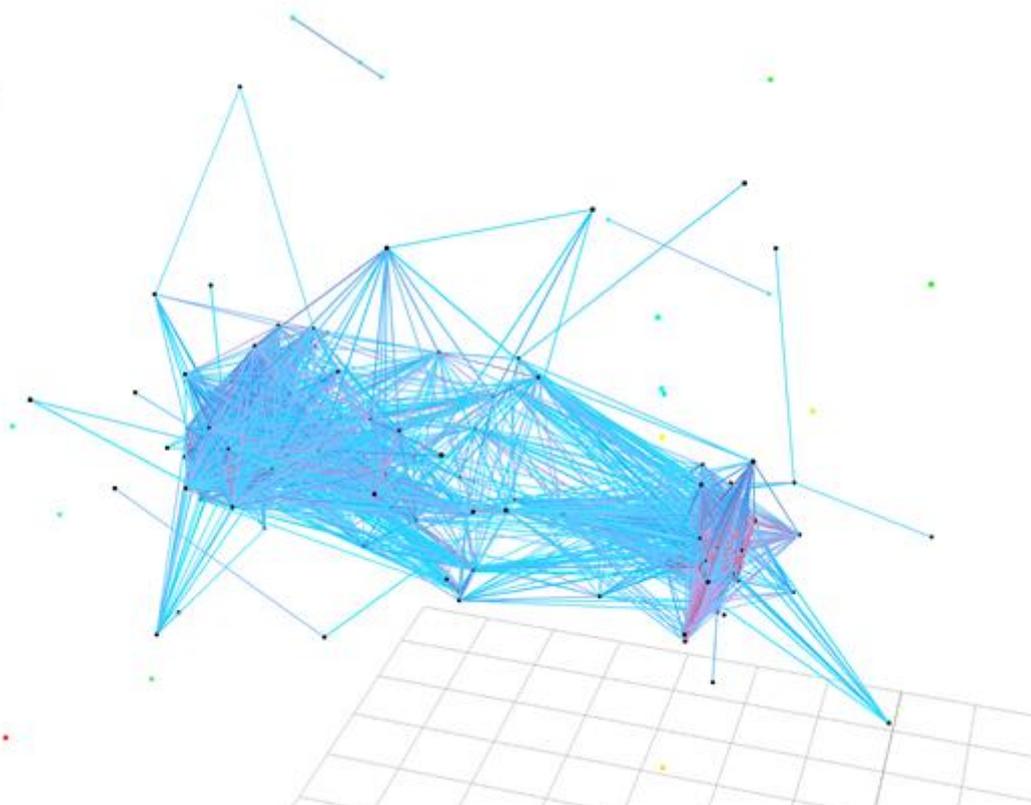


Рис. 3. Проекция пространственной сцены

Для визуального представления связности объектов вводится некоторое пороговое расстояние d (максимальное внутрикластерное расстояние), задаваемое в интерактивном режиме (как показано на рисунке 4). Если расстояние между точками n -мерного пространства меньше d , то эти точки соединяются отрезком, который дополняется до цилиндра, цвет которого меняется от красного (малое расстояние) до синего (расстояние, близкое к d). Группы соединенных точек формируют кластеры. Отдельные (удаленные) точки, расположенные далеко от всех других точек в многомерном пространстве, представляют собой аномальные объекты данных, которые будут объектом более детальных исследований.

Таким образом, визуальное фиксируемое наличие цилиндров между сферами позволяет аналитику фиксировать близкие точки в пространстве, образующие кластер, и соответствующие им близкие исходные объекты, также образующие свой кластер. Задание различных цветов цилиндрам позволяет делать суждения о расстоянии в исходном n -мерном пространстве при визуальном анализе пространственной сцены 3-х мерного пространства.

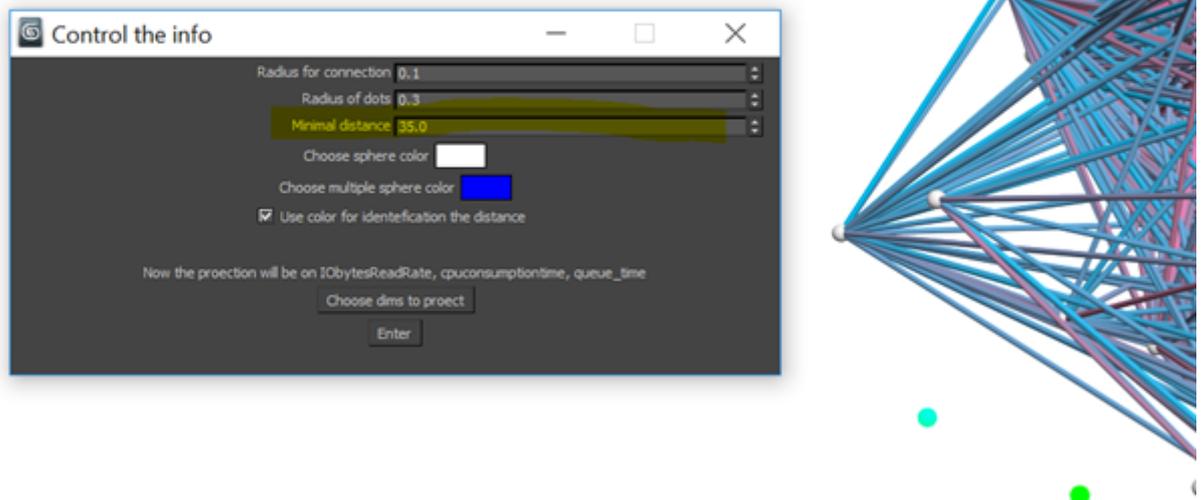


Рис. 4. Окно управления геометрическими и оптическими характеристиками объектов. (Жёлтым цветом выделено задаваемое пороговое расстояние d)

5. Программная реализация метода визуальной аналитики для кластерного анализа

Следует отметить, что внимательное изучение литературы, посвященной описанию конкретных приложений анализа многомерных данных с применением визуальных методов, позволяет утверждать, что в реальности интерактивным системам работы с многомерными данными зачастую придается меньшее значение по сравнению с системами отображения результатов применения методов Data Analysis. В качестве примера можно привести такие системы, как система ситуационного оповещения AdAware [14], система визуального анализа в задачах самолетостроения [15], система визуального анализа текстовой информации VxInsight, программный комплекс SAS Visual Analytics [16], предназначенный для обработки и анализа больших объемов финансовой и экономической информации. Все указанные системы носят промышленный характер, являются коммерческими,

предоставляют пользователю огромное количество интерфейсов и возможностей визуального представления данных. Как показывает практика [17,18], для такого визуального представления многомерных данных широко используются классические методы параллельных координат, кривых Эндруса, лиц Чернова [19] и других подобных мнемонических графических изображений. Эти изображения имеют набор параметров, соответствующих координатам рассматриваемых многомерных точек.

В то же время, большинство подобных систем по сути настроены на внутреннюю обработку многомерных данных и представление их пользователю в удобном для него виде, не предоставляя возможности пользователю непосредственно работать с многомерными данными с использованием визуальных отображений этих данных.

Рассмотренный выше метод кластерного анализа многомерных данных был реализован как приложение IVAMD (Interactive Visual Analysis of Multidimensional Data) [20]. Это приложение позволяет пользователю непосредственно работать с исходными многомерными

данными. При этом не выполняется изначальной численной обработки исходных многомерных данных, а производится целенаправленное непосредственное манипулирование аналитиком исходными данными и визуальный анализ получаемых результатов.

Приложение IVAMD реализовано на скриптовом языке MAXScript^[1] 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 ис-

пользованием 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, y_i) и затем просматриваются все эти проекции при различных i . Результаты кластеризации можно увидеть во время использования программы, а также они могут быть экспортированы в Excel-таблицу, размеченную различными цветами, соответствующими выделенным кластерам и аномальным точкам. Пример такой таблицы представлен на рисунке 5.

ID	avgpss	avgrss	avgvmem	cpu_eff	cpuconsumptiontime
378	3,02155	0,540255	0,0968087	0,340742	5,01355
2219	0,710337	0,134811	0	0,370008	6,94845
2460	1,59649	0,263486	0,35708	0,129835	5,43723
2463	1,44083	0,17454	0,268137	0,391832	6,4773
2494	1,40164	0,146246	0,230384	0,25998	7,76454
2510	2,00695	0,575056	0,115821	0,286225	7,20546
7700	47,9374	45,687	49,1948	51,373	23,8059
9437	44,9547	45,6759	49,3565	51,3017	23,6733
9438	46,4195	45,9381	49,5506	51,662	26,3812
10219	47,1236	45,9035	49,7007	51,6055	26,4637
10220	45,6879	45,6556	49,1678	51,2164	26,7734
10371	49,3227	46,2934	49,647	51,2322	29,2729
10976	97,3758	99,5828	99,5562	80,0312	72,2981
11007	99,2486	99,502	99,3153	100	74,6276
11247	1,22865	0,305296	0,335525	0,438041	39,7412
11272	1,49591	0,337036	0,426979	0,116061	46,876

Рис.5. Результирующая таблица с цветовым выделением кластеров и аномальных объектов

В силу особенностей метода, необходимое для решения задачи количество трехмерных объектов составляет $m + m \times (m - 1) / 2$, где m – количество сфер (равно количеству строк). Для отображения 100 объектов необходимо отобразить около 5050 примитивов 3DSMax (сфер и цилиндров). Описываемое программное средство является прототипом реализации метода и имеет ограничение на количество обрабатываемых объектов. Дальнейшая разработка и усо-

вершенствование прототипа, и оптимизация в рамках высокопроизводительной программно-аппаратной инфраструктуры позволит устранить текущие ограничения.

Основным преимуществом данного программного прототипа является интерактивное взаимодействие пользователя как с данными, так и с пространственной сценой. Итерационно меняя пороговое расстояние, исследователь может визуально исследовать измене-

ние кластерной структуры данных, отслеживать появление или исчезновение аномальных объектов данных. Кроме того, все сферы пространственной сцены кликабельны, что позволяет в любой момент времени узнать их многомерные координаты. Так исследователь может оценить какие координаты вносят наибольший вклад в формирование кластеров.

6. Применение метода визуальной аналитики

В качестве тестового испытания был выполнен визуальный анализ вычисли-

тельных 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 семимерное пространство значений было кластеризовано. Трёхмерная модель выполнена на проекции параметров: avgPSS, duration и outputFileBytes (представлено на рисунке 6). В результате выделено 6 кластеров (мощностью 31, 17, 8, 5, 3 и 2), а также 2 одиночные точки (кластеры мощностью 1).

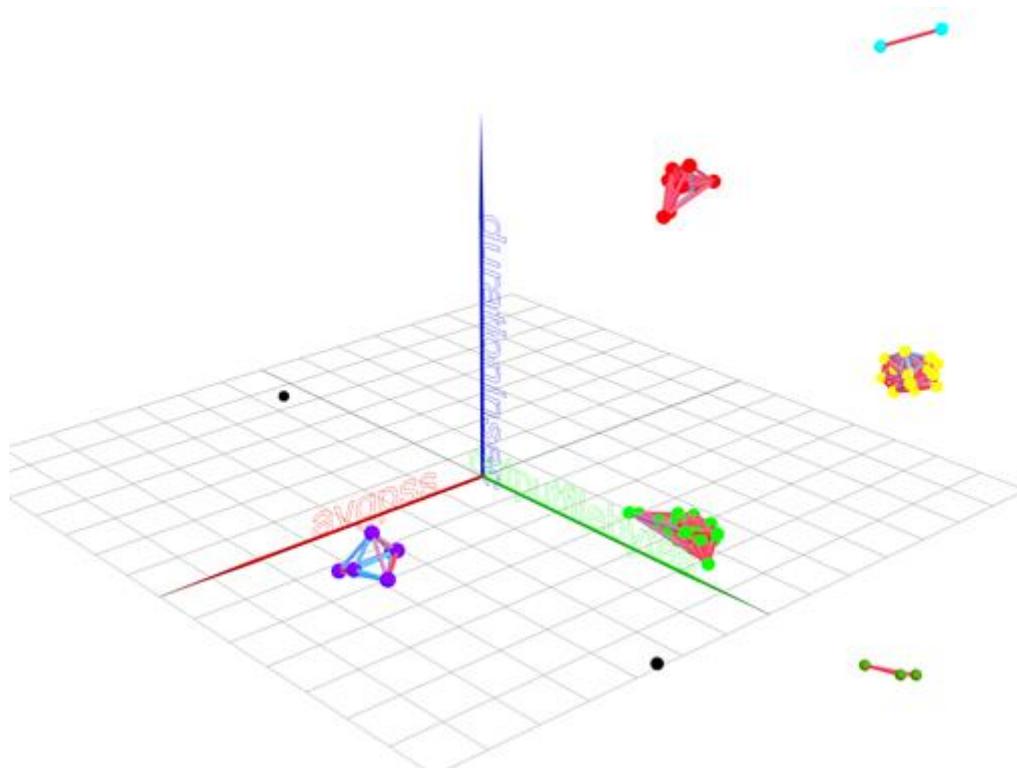


Рис. 6. Разбиение исходных точек на кластеры в трёхмерном пространстве (красная ось - avgPSS, зелёная ось - outputFileBytes, синяя ось - duration in sec)

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).

На рисунке 7 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.

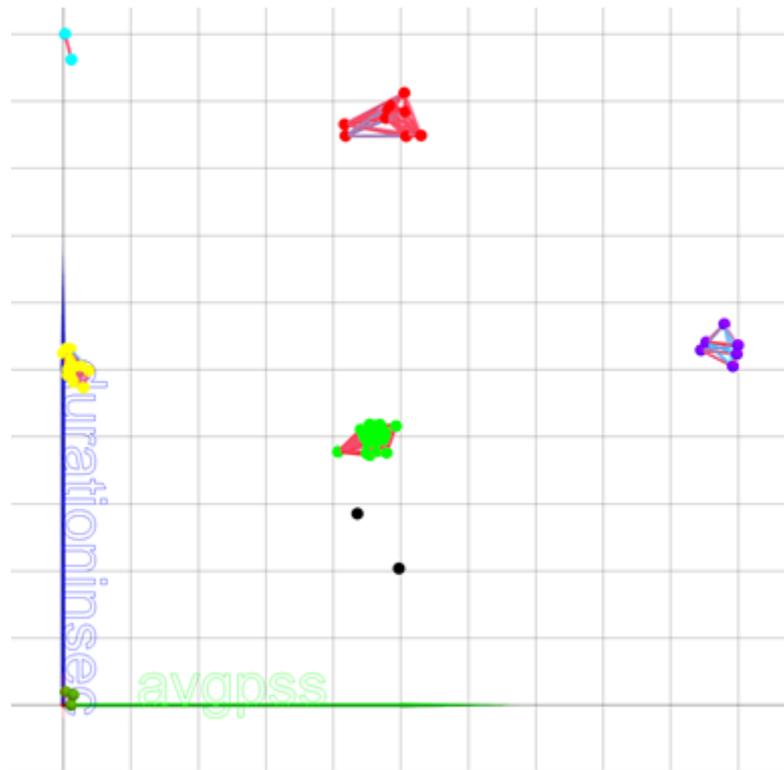


Рис. 7. Графическая проекция на плоскость (avgPSS, duration)

Рассмотрим кластеры в другом подпространстве. На рисунке 8 видно, что различие в потреблении процессорного времени (CPU Consumption) прямо пропорционально влияет на duration, как и ожидалось.

На рисунке 9 видно, что степень влияния объёма входных файлов (inputFileBytes) на длительность выполнения задач обработки и анализа данных не является решающей при распределённой обработке. Это также применимо и к выходным данным/файлам outputFileBytes (рисунок 10). Для количественной оценки влияния объёма данных на длительность выполнения задач требуются дополнительные исследования на большей статистике, которые будут проведены в дальнейшем.

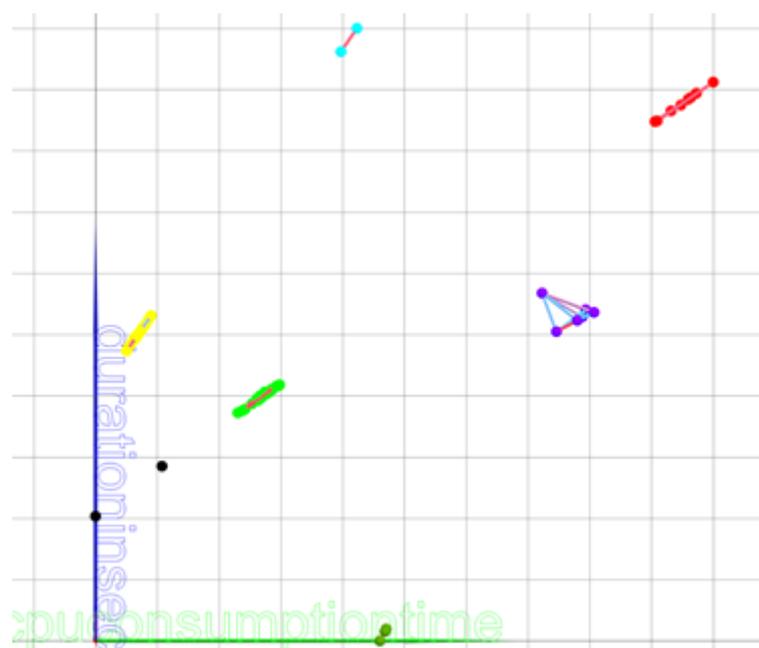


Рис. 8. Графическая проекция на плоскость (cpuConsumptionTime, duration)

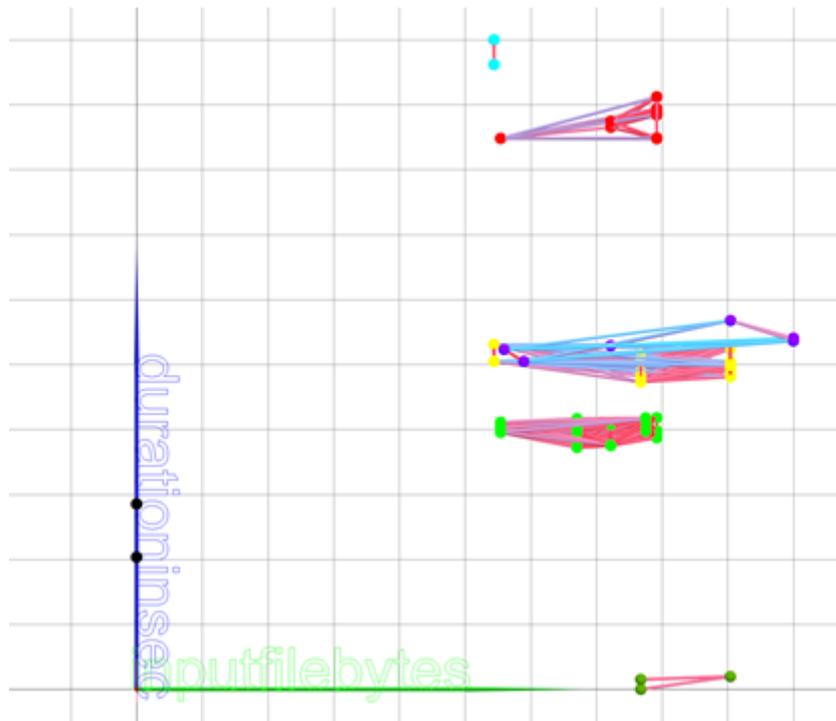


Рис. 9. Графическая проекция на плоскость (inputFileBytes, duration)

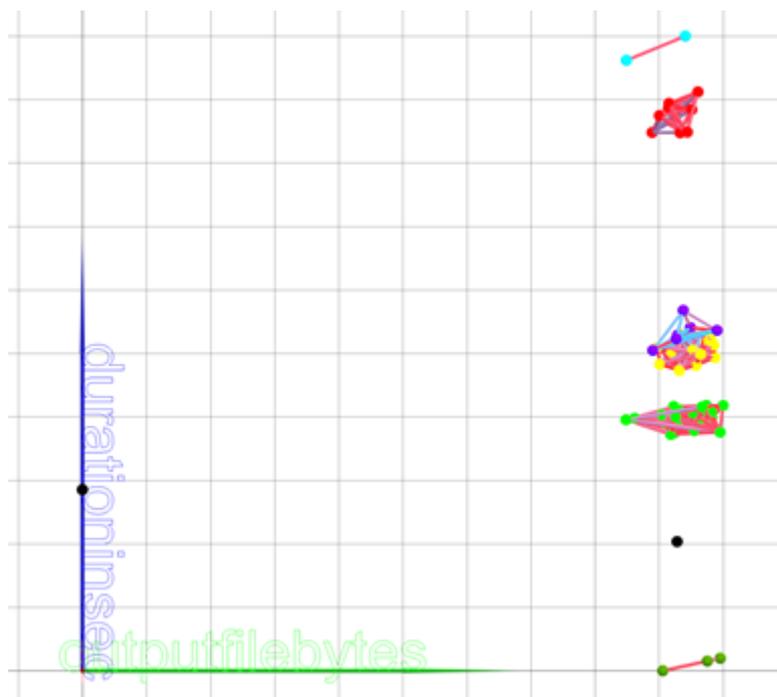


Рис. 10. Графическая проекция на плоскость (outputFileBytes, duration)

7. Заключение

Данная работа является первой попыткой применения методов визуальной аналитики к анализу метаданных распределенной системы обработки эксперимента ATLAS. Ввиду огромного количества данных, их сложности и многомерности, существующие методы анализа данных, такие как машинное

обучение и статистика, нуждаются в средствах визуализации, позволяющих увеличить человеческий контроль за процессом анализа. Данная статья демонстрирует метод визуальной аналитики как кластеризацию и категоризацию задач системы управления загрузкой эксперимента ATLAS. Этот метод основан на геометрическом представлении многомерных данных в 3-мерном

пространстве в виде сфер с вычисленными многомерными расстояниями между ними. Наиболее близкие расположенные сферы соединяются отрезками, формируя кластеры. Аналитик взаимодействует с пространственной сценой, изменяет пороговое расстояние и наблюдает изменение структуры кластеров.

В ходе дальнейших работ планируется построение модели многоуровневой интерактивной визуальной кластеризации. Эта модель подразумевает концепцию суперточек (Superpoints): поскольку объем исходных данных может быть слишком велик для визуализации всех точек, их количество может быть сокращено путем использования суперточек, которые представляют собой наборы схожих точек [21]. Суперточки вычисляются с использованием алгоритмов кластеризации и формируют первый уровень кластеризации. Затем осуществляется анализ суперточек отдельно, как наборов входящих в них объектов данных. Таким образом выполняется следующий уровень кластеризации. В дальнейшем планируется совершенствование разработанного инструментарий для визуализации 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, а также для разработки рекомендаций по оптимизации времени выполнения задач анализа и обработки данных.

8. Благодарности

Работа выполнена при поддержке гранта РФФИ №18-71-10003 от 02.08.2018 г. Данная работа выполнена в рамках программы исследований и разработок в области распределенных вычислений в коллаборации ATLAS, и мы благода-

рим коллаборацию за поддержку и сотрудничество.

9. Литература

1. R Brun and P Palazzi. Graphical Presentation for Data Analysis in Particle Physics Experiments: The HBOOK/HPLLOT Package // Proceedings Eurographics '80, pp.93--104, 1980.
2. R Brun, O Couet, C Vandoni and P Zanmarini. PAW, a general-purpose portable software tool for data analysis and presentation // Computer Physics Communications, vol.57, no.1, pp.432-437, 1989.
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. 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.
5. 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.
6. The ATLAS Collaboration, The ATLAS Experiment at the CERN Large Hadron Collider // Journal of Instrumentation, vol.3, S08003, 2008.
7. 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.
8. T Maeno et al. PanDA for ATLAS distributed computing in the next decade // Journal of Physics: Conference Series, vol.898, no.5, 052002, 2017.

9. The Worldwide LHC Computing Grid, <http://wlcg.web.cern.ch>
10. Alvaro R.De Pierro. From Parallel to Sequential Projection Methods and Vice Versa in Convex Feasibility: Results and Conjectures // Studies in Computational Mathematics Volume 8, 2001, Pages 187-201
11. О.П. Масленников, И.Е. Мильман, А.Э. Сафиуллин, А.Е. Бондарев, Ш.У. Низаметдинов, В.В. Пилюгин. Разработка системы интерактивного визуального анализа многомерных данных. Научная визуализация, 2014, №4, стр. 30-49
12. П. Масленников, И.Е. Мильман, А.Э. Сафиуллин, А.Е. Бондарев, Ш.У. Низаметдинов, В.В. Пилюгин. Интерактивный визуальный анализ многомерных данных. 24-я международная конференция по компьютерной графике и зрению ГрафиКон'2014, стр. 51-54
13. I.E.Milman, A.P.Pakhomov, V.V.Pilyugin, E.E.Pisarchik, A.A.Stepanov, Yu.M.Beketnova, A.S.Denisenko, Ya.A. Fomin. Data analysis of credit organizations by means of interactive visual analysis of multidimensional data. Scientific Visualization, 2015, n. 1, pp. 45-64
14. Y.Livnat, J.Agutter, S.Moon, S.Foresti, "Visual correlation for situational awareness", IEEE Symposium on Information Visualization (INFOVIS), pp.95--102, 2005.
15. D.Mavris, O.Pinon, D.Fullmer, "Systems design and modeling: A visual analytics approach", Proceedings of the 27th International Congress of the Aeronautical Sciences (ICAS), 2010.
16. SAS the power to know, [Online]. Available: http://www.sas.com/en_us/home.html [accessed on 10.02.2018].
17. K Xu, L Zhang, D Pérez, P H. Nguyen, A Ogilvie-Smith. Evaluating Interactive Visualization of Multidimensional Data Projection with Feature Transformation // Modal Technologies and Interaction, 2017, 1, 13; doi:10.3390/mti1030013
18. J. Poco, R. Etemadpour, F.V. Paulovich, T.V. Long, P. Rosenthal, M.C.F. Oliveira, L. Linsen and R. Minghim. A Framework for Exploring Multidimensional Data with 3D Projections // Eurographics / IEEE Symposium on Visualization 2011 (EuroVis 2011), Vol 30, Number 3
19. S Murray. Interactive Data Visualization for Web // O'Reilly Media, 2013, ISBN: 978-1-449-33973-9
20. 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.
21. B C Kwon, B Eysenbach, J Verma, K Ng, C deFilippi, W F Stewart, A Perer. Clustervision: Visual Supervision of Unsupervised Clustering // IEEE Transactions on Visualization and Computer Graphics, 2018, 10.1109/TVCG.2017.2745085