

Правда, искажающая истину. Как следует анализировать Top500?

С.М. Абрамов

Институт программных систем имени А.К. Айламазяна Российской академии наук

После каждого выпуска рейтинга Top500 [1] выполняются подсчеты и публикуются суждения, вида: «Подавляющее большинство суперкомпьютеров списка Top500 используются в индустрии». Или другие подобные подсчеты и суждения о долях в списке Top500: (i) разных типов процессоров; (ii) различных типов интерконнекта; (iii) производителей суперкомпьютеров; (iv) стран и т. п. Часто на базе подобных подсчетов и суждений принимаются серьезные решения, в том числе и на правительственном уровне. В данной работе показано: все, что сказано в подобных суждениях — это правда, однако эта правда серьезно искажает истину: искажает истинное положение в суперкомпьютерной отрасли. В работе дается анализ причины серьезного отличия «правды» от «истины», приводятся методика корректного анализа данных Top500 и результаты такого анализа.

1. Введение

Начиная с июня 1993 года, два раза в год публикуется список пятисот самых мощных суперкомпьютеров мира — мировой рейтинг Top500 [1]. Всего за истекшие 20 лет опубликовано сорок выпусков Top500. Каждая публикация рейтинга является серьезным новостным событием, а также поводом для анализа состояния и тенденций суперкомпьютерной отрасли¹.

После выхода новой редакции рейтинга (или одновременно с этим) многие выполняют различные подсчеты и публикуют суждения, основанные на результатах таких подсчетов. Довольно часто подсчеты посвящены вычислению различных долей в списке Top500. Например, вычисляют, какие доли приходятся на различные области применения суперкомпьютеров в рейтинге Top500. Или, какие доли приходятся на суперкомпьютеры, использующие те или иные микропроцессоры. Анализируют и другие распределения долей: доли различных архитектур, доли производителей суперкомпьютеров, доли стран и т. п. ...

Среди прочих, таким анализом занимаются и сами издатели рейтинга — на портале Top500 публикуют одновременно и сам список, и плакат, посвященный выходу в свет новой редакции рейтинга. Обратим внимание на плакат, выпущенный в ноябре 2012 года², рассмотрим диаграмму «Installation Type» с этого плаката — Рис. 1. Мы видим диаграмму из 40 столбцов — каждый столбец соответствует одному выпуску рейтинга, два столбца (июнь и ноябрь) приходятся на один год. Столбец состоит из частей разных цветов; размер частей определяется долями различных сегментов применения суперкомпьютеров из соответствующего рейтинга Top500. Различают шесть значений для задания сегментов применения: «Vendor», «Research», «Industry», «Government», «Classified» и «Academic».

Действительно, легко взять полную Excel-таблицу³ списка Top500 за ноябрь 2012 года и посчитать, сколько суперкомпьютеров в нем в колонке «Segment» имеют то или иное значение области применения. Результат представлен ниже (Таблица 1). Доли, посчитанные в третьей колонке, естественно, в точности соответствуют длинам цветных частей в правого столбца диаграммы «Installation Type» (Рис. 1). Тем самым, будет справедливым сделать следующее суждение:

¹ Здесь и далее используется широкое толкование суперкомпьютерной отрасли, что включает исследование, разработку, изготовление, эксплуатацию суперкомпьютерных технологий и охватывает аппаратные решения, программное обеспечение — системное, инструментальное, прикладное, — и суперкомпьютерные сервисы.

² http://s.top500.org/static/lists/2012/11/TOP500_201211_Poster.pdf

³ http://s.top500.org/static/lists/2012/11/TOP500_201211.xls

§1 В ноябре 2012 года большая часть (49.40%) суперкомпьютеров использовалось в индустрии (Segment=Industry). При этом индустриальное использование превосходило применение суперкомпьютеров для науки — 44.2% = 24.60% + 19.60% (Segment=Research и Segment= Academic).

По результатам подобного подсчета для Top500 за ноябрь 2009 года (обратите внимание на Рис. 1 на столбец, соответствующий ноябрю 2009 года) можно сказать еще сильнее:

§2 В ноябре 2009 года **подавляющая** часть (62.40%) суперкомпьютеров использовалась в индустрии (Segment=Industry). При этом индустриальное использование **значительно (почти вдвое)** превосходило применение суперкомпьютеров для науки — 34.0% = 18.20% + 15.80% (Segment=Research и Segment= Academic).

Таблица 1. Распределение суперкомпьютеров по различным сегментам применения — по сведениям Top500 за ноябрь 2012 года

Значения в колонке «Segment»	Количество систем	Доля
Vendor	12	2.40%
Research	123	24.60%
Industry	247	49.40%
Government	16	3.20%
Classified	4	0.80%
Academic	98	19.60%
ВСЕГО	500	100.00%

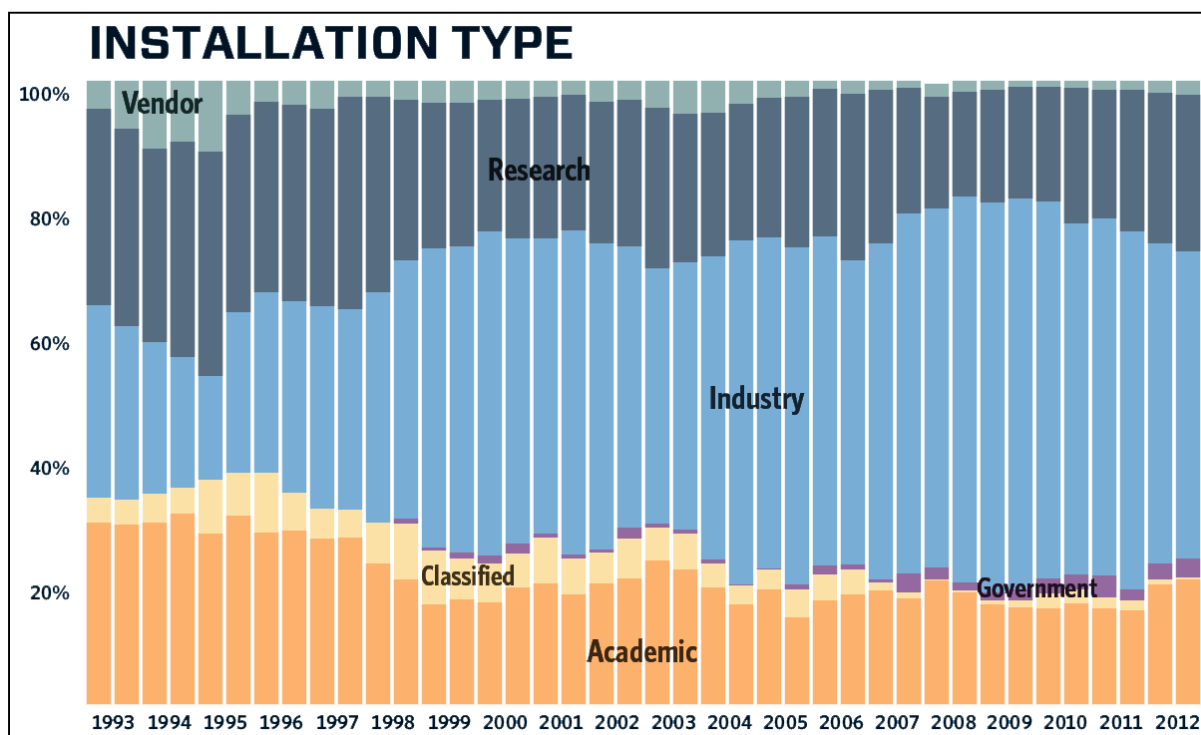


Рис. 1. Диаграмма «Installation Type» с плаката рейтинга Top500, выпущенного в ноябре 2012 года

Подобные вычисления и суждения (§1 и §2) легко могут быть построены — для этого не нужно быть большим специалистом, достаточно начальных навыков владения Excel-ем. Более того, график «Installation Type» просто входит в официальный плакат рейтинга Top500 и очень наглядно иллюстрирует распределение суперкомпьютеров по сегментам применения и то, как с течением времени меняется это распределение.

И подобные суждения (§1 и §2) и график «Installation Type» широко обсуждаются в различных публикациях, которые читают и специалисты, и обыватели, и лица, принимающие решения. Как результат, суждения, подобные §1 и §2, находим в правительственной переписке самого высокого уровня, посвященной суперкомпьютерам. Естественно, в контексте §1 и §2 следующие управленческие решения *на первый взгляд кажутся вполне разумными*:

- §3 *Государственная поддержка должна стимулировать создание суперкомпьютеров в большей степени (почти в два раза) не в научных российских центрах, а в промышленных.*
- §4 *В деле развития российской суперкомпьютерной отрасли представляется правильным перераспределить ресурсы, роли и ответственность с переносом центра тяжести к министерствам и ведомствам, связанным с индустрией, а не с наукой.*
- §5 *При создании суперкомпьютеров следует стремиться к таким долям государственного финансирования и привлекаемых из индустрии внебюджетных средств (ВБС): около 35% от государства, около 65% ВБС от индустрии — см. §2.*

Ключевым обстоятельством в данной статье является следующее: график «Installation Type» (Рис. 1), суждения §1 и §2 — все они являются правдой, но **эта правда существенным образом искажает истинное положение дел в суперкомпьютерной отрасли**. И как результат — сплошь и рядом приводит к ошибочным управленческим решениям.

Если посчитать истинные доли, приходящиеся на различные сегменты применения суперкомпьютеров¹ (Таблица 2), то мы увидим, что различие² между «правдой» (колонка А) и «истиной» (колонка В) оказывается весьма значительной — в разы. И так же от §1 разительно отличается истинное суждение:

- §6 *В ноябре 2012 года подавляющая доля — 77.67% = 59.23% + 18.44% — была у использования суперкомпьютеров в науке (Segment=Research и Segment=Academic), что многократно (в 4,4 раза) большие доли (17.56%) использования суперкомпьютеров в индустрии (Segment=Industry).*

Таблица 2. Истинное распределение долей различных сегментов применения суперкомпьютеров — по сведениям Top500 за ноябрь 2012 года

Сегмент применения	(А) «Правда»: доля систем (Таблица 1)	(В) «Истина»: истинная доля	Степень искажения истины
Research	24.60%	59.23%	↓2.41
Academic	19.60%	18.44%	↑1.06
Vendor	2.40%	2.22%	↑1.08
Industry	49.40%	17.56%	↑2.81
Government	3.20%	2.00%	↑1.60
Classified	0.80%	0.55%	↓2.41
ВСЕГО	100.00%	100.00%	

The chart displays two horizontal stacked bars representing the distribution of supercomputer usage segments. The top bar, labeled «Истина», shows a dominant Research segment (59.23%, pink), followed by Academic (18.44%, red), Vendor (2.22%, green), Industry (17.56%, light green), Government (2.00%, blue), and Classified (0.55%, yellow). The bottom bar, labeled «Правда», shows a more balanced distribution: Research (24.60%, pink), Academic (19.60%, red), Vendor (2.40%, green), Industry (49.40%, light green), Government (3.20%, blue), and Classified (0.80%, yellow). A legend on the right identifies the colors for each segment.

¹ Как следует вычислять истинные доли — будет обсуждено в последующих разделах.

² Степень искажения истины — самая правая колонка, вычисляется как $\max(A,B)/\min(A,B)$, — что указывает во сколько раз «правда» (А) приуменьшает (знак «↓» перед числом) или преувеличивает (знак «↑» перед числом) «истину» (В).

Серьезное отличие (в разы) «правды» от «истины» влечет опасность и недопустимость использования графика «Installation Type» (Рис. 1), суждений §1 и §2 для обоснования любых управленческих решений: на их основе легко сделать ложные выводы и, как результат — вредные управленческие решения, например §3–§5.

Распределение долей вычисляют не только для сегментов применения суперкомпьютеров. В общем случае, если суперкомпьютеры некоторой редакции Top500 каким-то образом разбиты на категории, то можно двумя способами посчитать доли этих категорий: (А) по общепринятой процедуре — доли числа суперкомпьютеров (среди всех 500 систем) соответствующей каждой категории; (В) истинные доли категорий — методика подсчета обсуждается в разделе 3.

На основе данных Top500 за ноябрь 2012 года были построены таблицы таких долей для следующих категорий:

- Таблица 3: используемая технология интерконнекта: Infiniband, Ethernet, Myrinet и Custom¹.
- Таблица 4: компания-производитель: IBM, Hewlett-Packard, Cray Inc. и Others².

Таблица 3. Распределение долей между разными технологиями интерконнектов, используемых в суперкомпьютерах — по сведениям Top500 за ноябрь 2012 года

Интерконнект	Число систем	(А) «Правда»: доля систем в Top500	(В) «Истина»: истинная доля	Степень искажения истины
Infiniband	224	44.80%	32.51%	↑1.38
Ethernet	189	37.80%	12.60%	↑3.00
Myrinet	3	0.60%	0.21%	↑2.79
Custom	84	16.80%	54.68%	↓3.25
Всего	500	100.00%	100.00%	

Во всех рассмотренных случаях очень часто «правда» сильно (в разы) отличается от «истины»: Таблица 2–Таблица 4, см. правую колонку. Вот несколько примеров:

- Таблица 2: в колонке «А» доля сегмента «Industry» преувеличена в 2.8 раза; доля сегмента «Research» преуменьшена в 2.4 раза.
- Таблица 3: в колонке «А» доля технологии «Infiniband» преувеличена в 1.38 раза, доля технологии «Ethernet» преувеличена в 3 раза, доля коммерчески недоступных решений — «Custom», — преуменьшена в 3.25 раз.
- Таблица 4: в колонке «А» доля компании Hewlett-Packard преувеличена в 2.61 раза, доля компании Cray Inc. преуменьшена в 2.8 раза.

Какой должна быть корректная методика вычисления истинных долей? По какой причине «правда» так сильно отличается от «истины»? — Все это будет обсуждено ниже:

- В разделе 2 обсуждаются основные определения и понятия.
- В разделе 3 определяется методика вычисления истинных долей.
- В разделе 4 обсуждается причина сильного отличия «правды» от «истины».

¹ Интерконнект, коммерчески недоступный на рынке. Недоступный, по крайней мере, как отдельное изделие. Интерконнект, который невозможно купить. Если надо использовать такой, то придется (аналог) разработать самостоятельно.

² Все остальные компании-поставщики.

- В разделах 5–8 исследуются истинные доли для таких понятий как «сегменты применения суперкомпьютеров», «технологии процессоров, используемых в суперкомпьютерах», «компании-производители», «технологии интерконнекта».
- В разделе 9 формулируется заключение по результатам работы.

Таблица 4. Распределение долей между разными компаниями-производителями суперкомпьютеров — по сведениям Top500 за ноябрь 2012 года

Компания-производитель	Число систем	(А) «Правда»: доля систем в Top500	(В) «Истина»: истинная доля	Степень искажения истины
IBM	193	38.60%	40.84%	↓1.06
Hewlett-Packard	146	29.20%	11.18%	↑2.61
Cray Inc.	31	6.20%	17.39%	↓2.80
Others	130	26.00%	30.59%	↓1.18
Всего	500	100.00%	100.00%	

2. Высокпроизводительные вычисления, суперкомпьютеры

Для того чтобы исключить неверное толкование, приведем используемые нами определения некоторых терминов.

2.1 Производительность

Среди важнейших технических характеристик компьютеров традиционно выделяют **производительность** — количество операций с плавающей точкой, выполняемых вычислителем за секунду¹. Различают

- **пиковую производительность** — максимальное число операций в секунду, которое может выполнить установка в идеальном случае — в принципе;
- **реальную производительность на некоторой задаче** — реальное количество операций, выполненных при решении задачи, деленное на реальное время решения задачи.

Пиковую производительность оценивают теоретически, исходя из состава оборудования компьютера. Реальную производительность измеряют опытным путем, решая на системе некоторую задачу. На разных задачах реальная производительность одного и того же компьютера могут быть разными.

Для сравнения производительности различных суперкомпьютеров между собой чаще всего используют реальную производительность на задаче Linpack². В последнее время набирают популярность и другие тесты реальной производительности суперкомпьютеров, например, основанные на задачах с интенсивной обработкой данных³.

¹ 1 Gflops — гигафлопс, 10^9 операций в секунду, 1 Tflops — терафлопс, 10^{12} операций в секунду, 1 Pflops — петафлопс, 10^{15} операций в секунду, 1 Eflops — эксафлопс, 10^{18} операций в секунду.

² Решение системы линейных уравнений с большим числом неизвестных. Используется в мировом рейтинге суперкомпьютеров Top500 [1].

³ Например, задача поиска в большом графе в ширину — используется как тест в другом мировом рейтинге суперкомпьютеров — Graph500 [2].

2.2 Суперкомпьютеры

Отметим интересный факт: если в известной сетевой энциклопедии Wikipedia [3] попытаться просмотреть термин «High-performance computing — высокопроизводительные вычисления», то, как результат, будет перенаправление на страницу «Supercomputer — Суперкомпьютер». Это верно и для англоязычной и для русскоязычной версии Wikipedia. Тем самым, подчеркивается синонимичность понятий «высокопроизводительный компьютер» и «суперкомпьютер». Дадим формальное определение.

К вычислительным системам высокой производительности — к *суперкомпьютерам*, — отнесем вычислительные машины, значительно превосходящие по своей реальной производительности большинство существующих компьютеров.

То есть, в каждый момент времени, если среди всех существующих компьютеров отобрать самые мощные — например, *пятьсот самых производительных*, — то они и определяют термин «суперкомпьютер» на данный момент времени.

Значит, начиная с июня 1993 года, можно установить тесную связь между понятием «суперкомпьютер» и рейтингом Top500. В принципе можно сказать, что вычислительная система является суперкомпьютером, если она была включена¹ в некоторый выпуск рейтинга Top500 — и только в этом случае.

Тем самым, каждую редакцию Top500 можно рассматривать как исчерпывающее описание текущего состояния суперкомпьютерных технологий. А всю совокупность выпусков рейтинга можно рассматривать как исчерпывающую хронологию суперкомпьютерной отрасли за последние 20 лет.

2.3 Top500 — источник знаний о суперкомпьютерной отрасли

Редакции рейтинга Top500 публикуются дважды в год (в июне и ноябре) начиная с июня 1993 года. Рейтинг основан на реальной производительности суперкомпьютеров на задаче Linpack. На сегодня в открытом доступе [1] имеются данные 40 выпусков рейтинга — с июня 1993 года по ноябрь 2012 года. Можно выгрузить каждый выпуск рейтинга в виде Excel-таблицы. В этом случае предоставляется самая полная информация. Если свести все 40 Excel-таблиц вместе, то получим таблицу с $40 \times 500 = 20\,000$ строками и с 40 колонками² (полями). Профессиональный анализ выпусков списка Top500 позволяет строить весьма достоверные суждения о состоянии и перспективах суперкомпьютерных технологий в мире и в России.

Обратим внимание, что при проведении анализа иногда приходится совместно обрабатывать несколько полей одной записи. Так, совместная обработка полей «Segment» и «Application Area» позволяет установить область использования суперкомпьютера более точно, чем это указано в поле «Segment». Чтобы точнее понять устройство интерконнекта, имеет смысл обрабатывать тоже два поля: «Interconnect» и «Interconnect Family». Для точного определения используемого процессора надо рассмотреть шесть полей: «Processor», «Processor Family», «Processor Generation», «Processor Technology», «Proc. Frequency», «Cores per Socket».

Понятно, что вручную выполнить тонкий анализ такого количества данных (20 000 записей с 40 полями) невозможно. Поэтому автор в 2009 году в инициативном порядке создал и до сих пор развивает программу Top500 Analyzer [4] для анализа рейтинга Top500. Все иллюстрации (за исключением Рис. 1 и Рис. 7) и все данные для расчетов в данной работе подготовлены при помощи этой программы.

¹ Либо ее технические показатели позволяли ее включить в рейтинг, но это не было сделано по некоторым причинам.

² Приведем имена этих полей: Accelerator, Accelerator Cores, Application Area, Architecture, Computer, Continent, Cores, Cores per Socket, Country, Efficiency(%), First Appearance, First Rank, Interconnect, Interconnect Family, Manufacturer, Measured Size, Mflops/Watt, Name, Nhalf, Nmax, Operating System, OS Family, Power, Previous Rank, Proc. Frequency, Processor, Processor Cores, Processor Family, Processor Generation, Processor Technology, Rank, Region, RMax, Rpeak, Segment, Site, System Family, System Model, Year.

2.4 О частичной неполноте и о частичной недостоверности Top500

Время от времени случаются публикации [5], указывающие на *частичную недостоверность данных* в рейтинге Top500: в рейтинг попадают установки до момента, когда они реально создаются, или установки, которые прекратили свое существование. Бывает.

Кроме того, всегда и во всех странах существуют суперкомпьютеры, которые не включают в рейтинг Top500 из соображений государственной безопасности или по каким-то другим причинам. Значит можно говорить о *частичной неполноте данных* в рейтинге Top500.

Однако можно предполагать, что эти обстоятельства:

- не существенны;
- более-менее равномерно влияют на различные категории суперкомпьютеров.

Тем самым, выводимые из данных Top500 относительные оценки оказываются весьма достоверными.

Это подобно тому, что вполне достоверно можно сравнивать между собою айсберги, основываясь на неполной информации — анализируя только их надводные (видимые) части.

3. Методика вычисления истинных долей

Почему правильное суждение (абсолютная правда) «В ноябре 2012 года большая часть (247 из 500) суперкомпьютеров использовалась в индустрии (Segment=Industry)» не может использоваться для вычисления «истинной доли» индустриального применения суперкомпьютеров в лоб — по формуле $247/500 = 49.40\%$?

Совсем небольшое размышление приводит к правильному ответу: суперкомпьютеры нельзя мерить штуками.

Пять одних суперкомпьютеров могут сильно отличаться от пяти других, в любом смысле: в стоимостном (при оценке долей рынка), в смысле технической сложности (при оценке доли в общем количестве процессоров/ядер или доли в общем числе портов интерконнекта) и т. п.

Вычисляя «истинные доли», следует оперировать не количеством суперкомпьютеров «в штуках», а такими числовыми характеристиками суперкомпьютеров, которые наиболее верно отражают **наиважнейшую характеристику суперкомпьютеров**, как изделий. Точно так же, например, когда сравнивают торговые флоты разных стран, измеряют размеры флотов не «в штуках», а в суммарном тоннаже.

Самая важная числовая характеристика суперкомпьютеров очевидна (даже просто в силу самого определения понятия «суперкомпьютер», см. раздел 2.2) — это реальная производительность. Конечно, лучше было бы при этом оперировать реальной производительностью на некоторых целевых (интересующих того или иного заказчика), задачах. Но если таких сведений нет, то будем довольствоваться Linpack-производительностью, сведения о которой имеются в записях Top500 — поле RMax.

3.1 Linpack-производительность, как истинная мера при измерении долей

Реальная производительность — в частности, Linpack-производительность,— главная, определяющая характеристика суперкомпьютеров. По ней разграничиваются суперкомпьютеры от «просто компьютеров». Кроме того, по сравнению со «штуками», Linpack-производительность гораздо точнее¹ коррелирует с такими характеристиками, как:

- Научно-технический уровень системы.
- Стоимость системы — что важно для правильной оценки распределения долей рынка.
- Объемы различных подсистем и смежные технические параметры. Например, размер подсистемы интерконнекта (количество портов), количество процессоров или ядер и т. п.

Таким образом, мы приходим к методике расчета истинных долей через *вычисление доли суммарной Linpack-производительности*.

¹ Разницу можно оценить в два порядка — до ~250 раз,— как увидим далее.

3.2 Формальное описание метода вычисления истинных долей

Пусть $n \in [1..40]$ — номер редакции Top500, $i \in [1..500]$ позиция, занятая некоторым суперкомпьютером в рейтинге, обозначим $RMax(n, i)$ — Linpack-производительность данной системы в n -ой редакции Top500.

Рассмотрим некоторую категорию суперкомпьютеров, например, все суперкомпьютеры индустриального использования: $Segment=Industry$. Пусть $C = \{... i ... \} \subseteq [1..500]$ множество всех позиций, которые суперкомпьютеры из данной категории занимают в n -ой редакции Top500.

В n -ой редакции Top500 *истинную долю суперкомпьютеров заданной категории* определим как долю суммарной Linpack-производительности суперкомпьютеров данной категории в суммарной Linpack-производительности всего списка:

$$\frac{\sum_{i \in C} RMax(n, i)}{\sum_{i \in [1..500]} RMax(n, i)}$$

Рассмотрим некоторый *подсписок* в n -ой редакции Top500, заданный множеством позиций $J = \{... i ... \} \subseteq [1..500]$. Например, «первую сотню» Top1–100: $J = [1..100]$.

В n -ой редакции Top500 *истинную долю суперкомпьютеров заданной категории в указанном подписке* определим как долю суммарной Linpack-производительности суперкомпьютеров данной категории из подписка в суммарной Linpack-производительности всего подписка:

$$\frac{\sum_{i \in (C \cap J)} RMax(n, i)}{\sum_{i \in J} RMax(n, i)}$$

4. Причина сильного отличия «правды» от «истины»

Используя обозначения раздела 3.2, посчитаем «правду» — долю категории C по традиционной методике, «в штуках»:

$$p_1 = \frac{\sum_{i \in C} 1}{\sum_{i \in [1..500]} 1} = \sum_{i \in C} 1/500 = \sum_{i \in C} 0.2\%$$

Таким образом, при такой методике в общую копилку доли p_1 категории C каждый суперкомпьютер вносит один и тот же вклад — 0.2%, — вне зависимости от того, крупный это суперкомпьютер или небольшой, дорогой или бюджетный и т. п.

Введем обозначение для доли Linpack-производительности одного суперкомпьютера $RMax(n, i)$ в суммарной Linpack-производительности всего списка:

$$pRMax(n, i) = \frac{RMax(n, i)}{\sum_{i \in [1..500]} RMax(n, i)}$$

Тогда истинную долю категории C можно записать таким образом:

$$p_2 = \frac{\sum_{i \in C} RMax(n, i)}{\sum_{i \in [1..500]} RMax(n, i)} = \sum_{i \in C} pRMax(n, i)$$

Сравним между собою «правду» $p_1 = \sum_{i \in C} 0.2\%$ и «истину» $p_2 = \sum_{i \in C} pRMax(n, i)$. Ясно, что если бы все суперкомпьютеры не очень сильно отличались бы между собою по Linpack-производительности, то все $pRMax(n, i)$ были бы близки к 0.2%, а «правда» p_1 не сильно бы отличалась от «истины» p_2 .

Однако суперкомпьютеры в одном и том же рейтинге Top500 имеют огромный разброс в Linpack-производительности $RMax(n, i)$ и, как следствие, огромный разброс $pRMax(n, i)$ — от 10.849% до 0.047% для 40-ой редакции рейтинга Top500 — разница в 230 раз!

Такое гигантское расслоение суперкомпьютеров по параметру Linpack-производительности определяет огромное отличие «правды» от «истины». Это расслоение делает осмысленным введение отдельных уровней (слоев, классов) суперкомпьютеров.

4.1 Различные уровни суперкомпьютерных систем

В работе [6] были введены 4 уровня суперкомпьютеров — Top1–20, Top21–100, Top101–250, Top251–500,— цитата:

- *суперЭВМ в крупнейших национальных центрах* — единичные установки в стране, соответствующие местам 1–20 в мировом рейтинге Top500;
- *суперЭВМ в крупнейших региональных и отраслевых центрах* — два–четыре десятка установок в стране, соответствующих местам 21–100 в мировом рейтинге Top500;
- *суперЭВМ в крупных региональных и корпоративных центрах* — от четырех десятков до сотни установок в стране, соответствующих местам 101–250 в мировом рейтинге Top500;
- *суперЭВМ предприятий и научных учреждений* — одна–три сотни установок в стране, соответствующих местам 251–500 в мировом рейтинге Top500.

В работе [7] предлагается и обосновывается выделение из первого уровня отдельной группы *сверхвысокопроизводительных систем* — Top1–10.

Везде далее обсуждаются эти пять уровней суперкомпьютеров: Top1–10, Top11–20, Top21–100, Top101–250 и Top251–500.

4.2 Резкое расслоение в суперкомпьютерной отрасли по Linpack-производительности

Для оценки глубины расслоения суперкомпьютерной отрасли рассмотрим разницу в Linpack-производительности у суперкомпьютеров разных уровней (по данным редакции Top500 за ноябрь 2012 года, Таблица 5). Разница по Linpack-производительности самой мощной и самой слабой системы в классе Top1–20 (20 систем) составляет 16.7 раза¹, в классе Top21–100 (80 систем) — 4.3 раза, в классе Top101–250 (150 систем) — всего 2.0 раза, в классе Top251–500 (250 систем) — 1.4 раза.

Таким образом, системы из Top1–20 (и особенно — системы Top1–10) радикально отличаются от других, а в классах Top21–100, Top101–250 и Top251–500 расположены системы, не принципиально отличающиеся друг от друга по производительности.

Таблица 5. Разница в Linpack-производительности между суперкомпьютерами разных уровней — по данным редакции Top500 за ноябрь 2012 года

Места в Top500	Linpack-производительность, max–min (Tflops)	Разница Linpack-производительности, max/min (разы)	Разница Linpack-производительности от Top1 (разы)
Top1–10	17 590–1 515	11.6	1–12
Top11–20	1 359–1 050	1.3	13–17
Top21–100	1 043–244	4.3	17–72
Top101–250	240–111	2.2	73–159
Top251–500	111–76	1.4	159–230

Наглядно оценить резкое расслоение суперкомпьютерной отрасли сегодня позволяют график функции $f(i) = \frac{RMax(40,i)}{RMax(40,1)}$, где $i \in [1..500]$ (Рис. 2), а так же график функции $g(n) = \frac{\sum_{i \in [1..n]} RMax(40,i)}{\sum_{i \in [1..500]} RMax(40,i)}$, где $n \in [1..500]$ (Рис. 3), иллюстрирующий, какую долю суммарной Linpack-

¹ Top1–10 — 11.6 раза, Top11–20 — 1.3 раза.

производительности всего списка Top500 обеспечивают первые n систем из списка. Видно, что глубокое расслоение суперкомпьютерной отрасли обеспечивает почти точное выполнение принципа Вильфредо Парето¹.

Подчеркнем, что анализируя приведенные данные (Таблица 5, Рис. 2, Рис. 3), уместно помнить, что отличия (сильные или слабые) суперкомпьютеров по Linpack-производительности влекут подобные же (сильные или слабые) отличия по цене, технической сложности, объему оборудования в различных подсистемах суперкомпьютеров.

Например, скорее всего (Рис. 3) суммарная стоимость первых 50 суперкомпьютеров в Top500 примерно равна суммарной стоимости остальных 450 суперкомпьютеров.

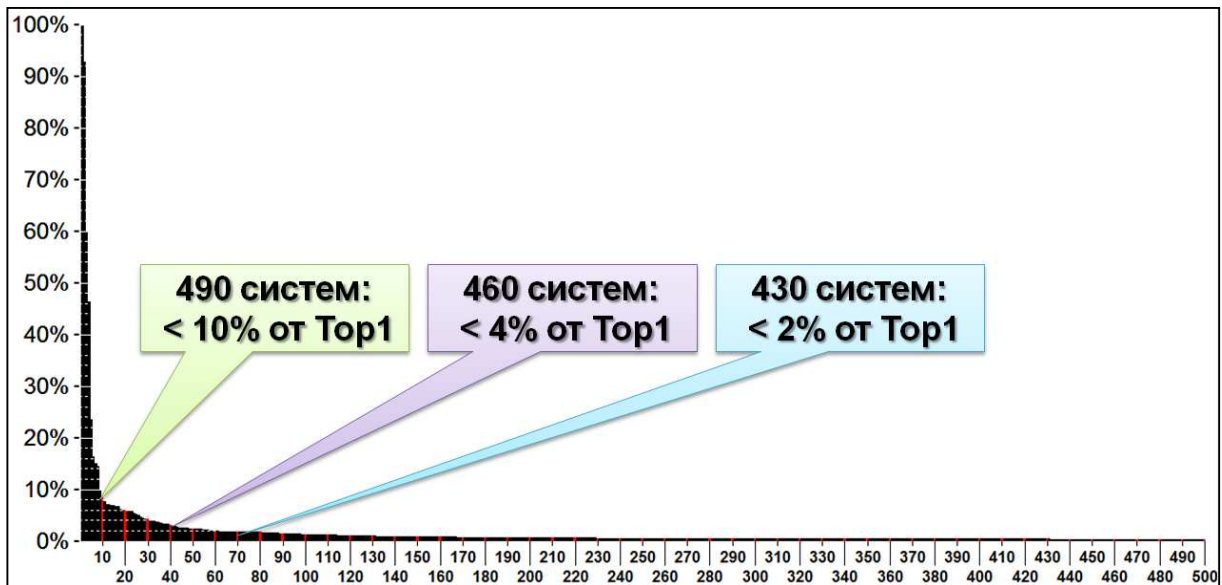


Рис. 2. Относительная Linpack-производительность в Top500 i -той системы, $i \in [1..500]$ — за 100% принята Linpack-производительность Top1. По данным Top500 за ноябрь 2012 года

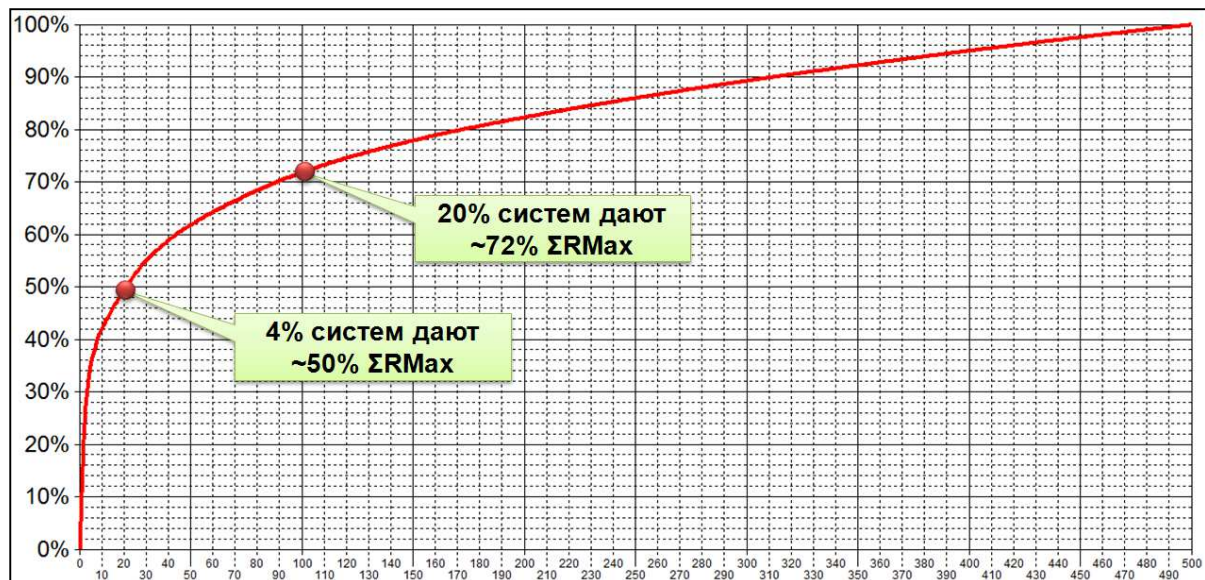


Рис. 3. Какую долю суммарной Linpack-производительности всего списка Top500 обеспечивают первые $n \in [1..500]$ систем из списка. По данным Top500 за ноябрь 2012 года

¹ Принцип Вильфредо Парето часто формулируют так: «20% усилий дают 80% результата».

5. Восстановление истины: сегменты применения суперкомпьютеров

Выведа и обосновав (раздел 3) методику вычисления истинных долей различных категорий, обсудив (раздел 4) причины серьезного различия истинных долей от долей, рассчитанных «в штуках», далее, в этом и последующих разделах мы проведем исследование долей по различным категориям. Исследования будут выполняться при помощи программы Top500 Analyzer [4].

5.1 Анализ сегментов применения суперкомпьютеров

Начнем с анализа сегментов применения суперкомпьютеров. Все суперкомпьютеры разбиваются по категориям, в зависимости от указанных значений поля «Segment» в Top500. В данном поле всегда указывают одно из шести значений — «Research», «Academic», «Vendor», «Industry», «Government», «Classified». Соответственно получаем шесть категорий суперкомпьютеров. Диаграмма (Рис. 1) долей «в штуках» этих категорий входит в официальный плакат рейтинга Top500, выпущенного в ноябре 2012 года. Ниже (Рис. 4) показаны для сравнения диаграммы, построенные программой Top500 Analyzer. Левая часть рисунка — доли «в штуках», — в точности совпадает с диаграммой с официального плаката. Правая диаграмма показывает истинные доли сегментов применения суперкомпьютеров.

Сравнивая левую и правую части рисунка, мы видим, что в левой части доля категории «Research» существенно занижалась в каждом выпуске рейтинга, а доля категории «Industry» — существенно преувеличивалась.

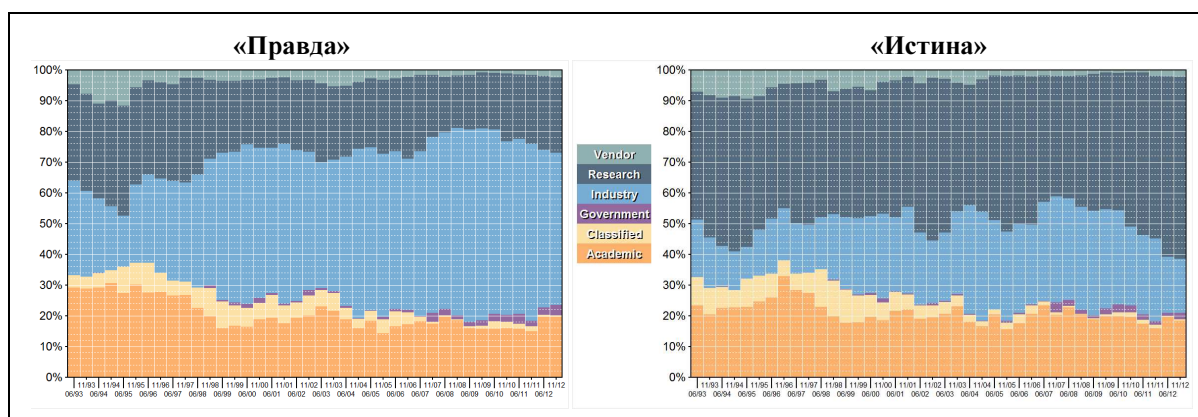


Рис. 4. Доли сегментов применения суперкомпьютеров по данным всех 40 редакций рейтинга Top500 (от июня 1993 года по ноябрь 2012 года). Слева: доли «в штуках» (доли от общего числа суперкомпьютеров), справа истинные доли (доли в Linpack-производительности)

5.2 Анализ областей использования суперкомпьютеров

Понятие «сегмент применения суперкомпьютеров» определяется напрямую значением одного поля «Segment» в рейтинге Top500. В программе Top500 Analyzer кроме этого поддерживается понятие «область применения суперкомпьютера», которое определяется за счет анализа двух полей: «Segment» и «Application Area». В результате все суперкомпьютеры программой Top500 Analyzer относятся в одну из 4 категорий: «RnD»¹ — использование для фундаментальных исследований и НИОКР; «Industry» — использование в промышленности и в других областях реальной экономики (например, в индустрии развлечений и т. п.); «Gov.Mil» — использование для государственных и военных нужд; «Unknown» — недостаточно информации для отнесения в категории «RnD», «Industry» или «Gov.Mil».

¹ От английского: Research and Development — исследования и разработки (НИР и ОКР).

При помощи программы Top500 Analyzer построены диаграммы (Рис. 5) данных категорий. Левая часть рисунка — доли областей использования суперкомпьютеров «в штуках», правая диаграмма — истинные доли областей использования суперкомпьютеров.

Сравнивая левую и правую части рисунка, легко заметить, что в левой части доля категории «RnD» существенно занижалась в каждый момент времени, а доля категории «Industry» — существенно преувеличивалась. Более того, правая часть рисунка явно выявляет тенденцию последних лет на сокращение истинной доли индустриального использования суперкомпьютеров.

При помощи программы Top500 Analyzer для редакции Top500 за ноябрь 2012 года построим распределение (Рис. 6) областей использования суперкомпьютеров по пяти уровням суперкомпьютеров: Top1–10, Top11–20, Top21–100, Top101–250 и Top251–500. Поясним структуру этой диаграммы. Левые пять столбцов иллюстрируют истинные доли (доли суммарной Linpack-производительности — RMax) областей использования для пяти отдельных уровней суперкомпьютеров: от Top1–10 до Top251–500. Предпоследний столбец на Рис. 6 иллюстрируют истинные доли областей использования для всего списка Top500 (ноябрь 2012 года) — что, по сути, совпадает с правым столбцом в правой части Рис. 5. Последний столбец на Рис. 6 иллюстрируют доли «в штуках» областей использования для всего списка Top500 (ноябрь 2012 года) — что, по сути, совпадает с правым столбцом в левой части Рис. 5.

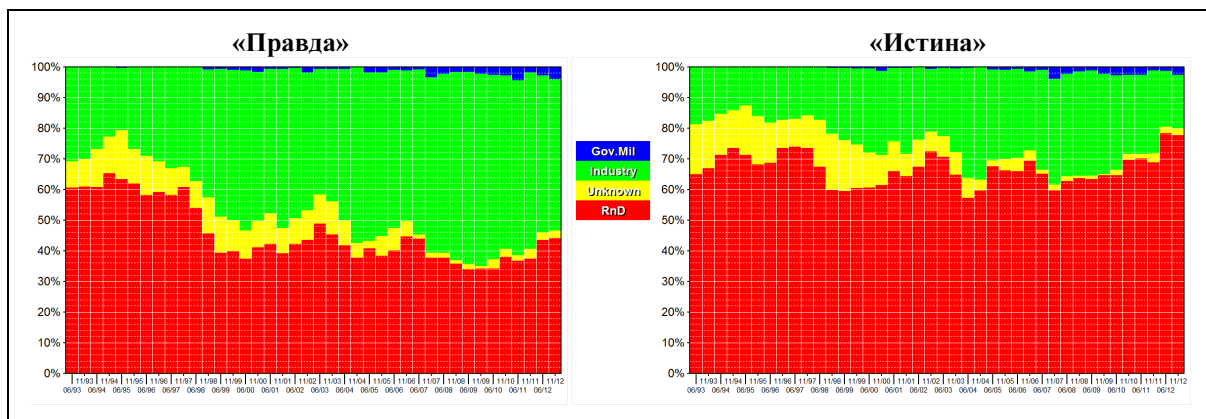


Рис. 5. Доли областей использования суперкомпьютеров по данным всех 40 редакций рейтинга Top500 (от июня 1993 года по ноябрь 2012 года). Слева: доли «в штуках» (доли от общего числа суперкомпьютеров), справа истинные доли (доли в Linpack-производительности)

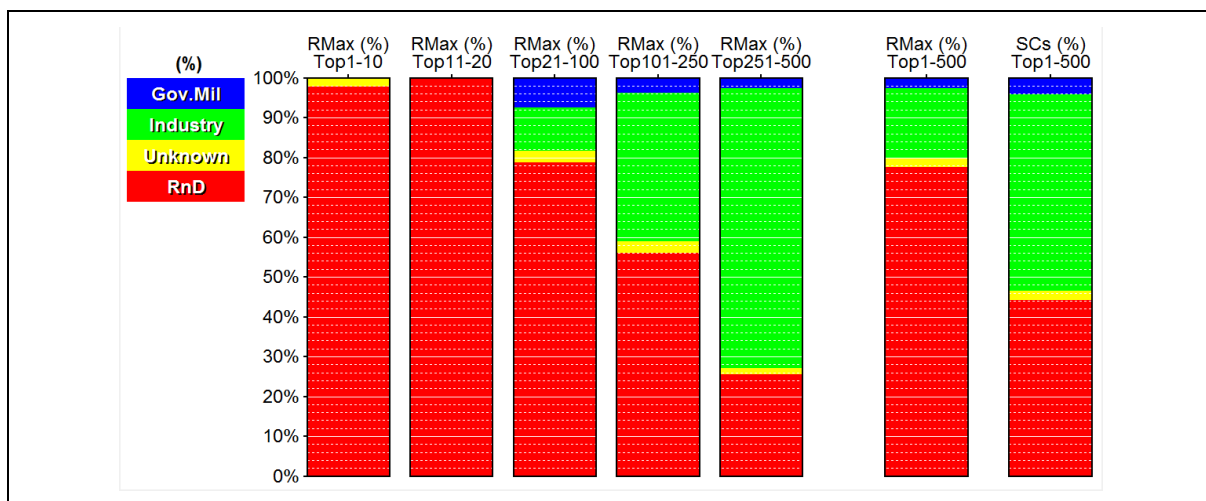


Рис. 6. Распределение областей использования суперкомпьютеров по уровням Top1–10, Top11–20, Top21–100, Top101–250, Top251–500. По данным Top500 за ноябрь 2012 года

Распределение по уровням (см. левые пять столбцов на Рис. 6) позволяет понять резкое отличие «истины» и «правды» (см. два правых столбца на Рис. 6) за счет явного изображения «ареалов обитания» каждой категории на различных уровнях суперкомпьютеров. Так, видно, что в индустрии совсем не применяются суперкомпьютеры первого и второго уровня, использование систем третьего и четвертого уровня незначительно, и только самые слабые и самые многочисленные системы — пятый уровень, 250 систем с производительностью в 150–230 раз хуже, чем у Top1,— характерны для категории «Industry».

6. Восстановление истины: типы микропроцессоров, используемых в суперкомпьютерах

На плакатах рейтинга Top500 кроме диаграммы «Installation Type» (доли сегментов применения) традиционно размещают диаграмму «Chip Technology» — диаграмму долей различных технологий процессоров, используемых в суперкомпьютерах. Данная диаграмма на плакате за ноябрь 2012 года (Рис. 7) состоит из 40 столбцов — каждый столбец соответствует одному выпуску рейтинга, два столбца (июнь и ноябрь) приходятся на один год. Каждый столбец состоит из частей разных цветов; размер частей определяется долями различных технологий процессоров из соответствующего рейтинга Top500. Различают восемь значений (категорий) для обозначения технологий процессоров: «Alpha», «IBM», «HP», «Intel», «MIPS», «SPARC», «AMD» и «Proprietary».

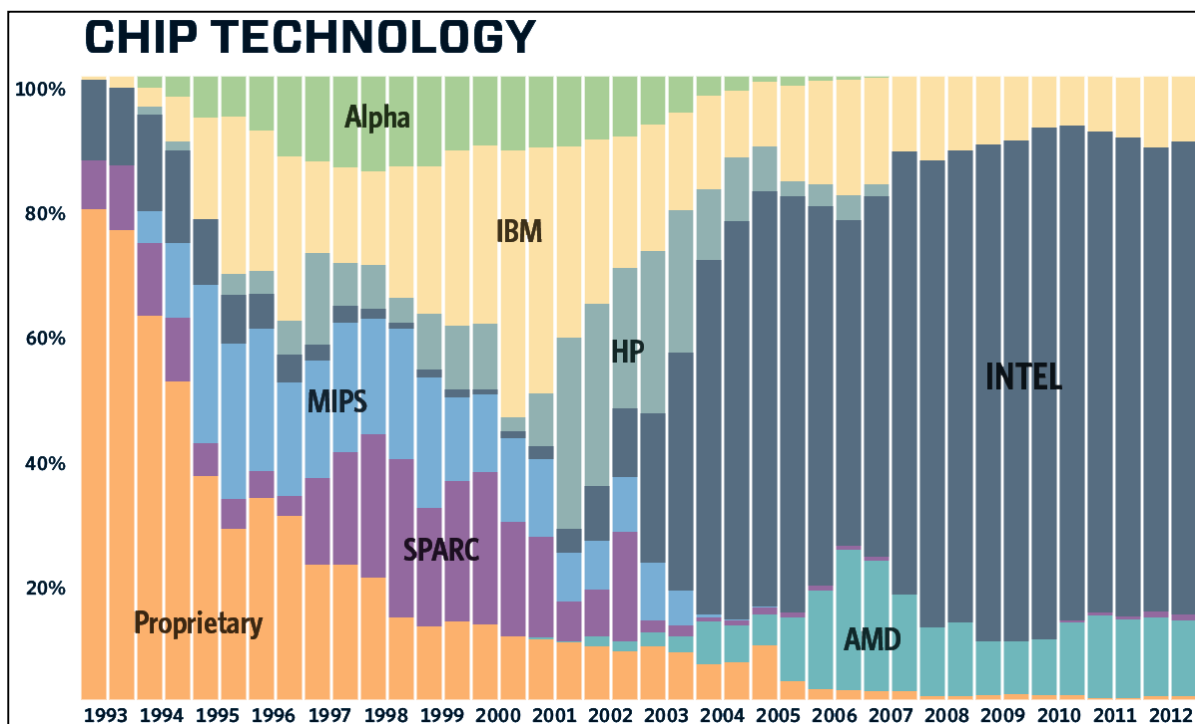


Рис. 7. Диаграмма «Chip Technology» с плаката рейтинга Top500, выпущенного в ноябре 2012 года

Рассматривая эту диаграмму, легко сделать весьма ошибочные суждения, например:

§7 К ноябрю 2012 года в суперкомпьютерах Top500 на процессоры Intel приходится подавляющая доля (76%). Отрыв от ближайших преследователей весьма значительный: почти в 7 раз от IBM (11%) и почти в 6 раз от AMD (13%).

Для выявления истинного положения с помощью программы Top500 Analyzer построим диаграммы долей различных технологий процессоров, используемых в суперкомпьютерах (Рис. 8). Для вычисления технологии процессора для каждой записи в Top500 анализируется 4 поля: «Processor», «Processor Family», «Processor Generation», «Processor Technology». Левая часть рисунка — доли «в штуках», — в точности совпадает с диаграммой «Installation Type» с

официального плаката. Правая диаграмма показывает истинные доли различных технологий процессоров, используемых в суперкомпьютерах.

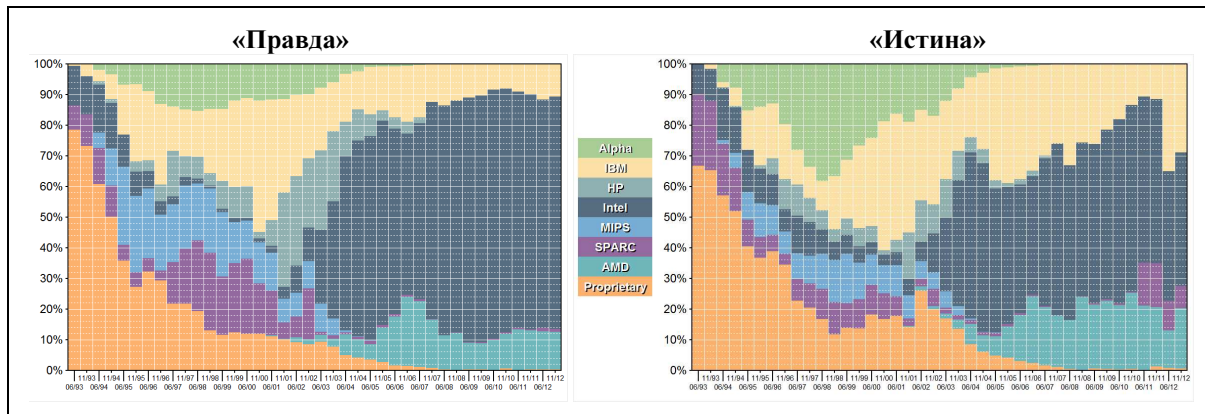


Рис. 8. Доли различных технологий процессоров, используемых в суперкомпьютерах по данным всех 40 редакций рейтинга Top500 (от июня 1993 года по ноябрь 2012 года). Слева доли «в штуках» (доли от общего числа суперкомпьютеров), справа истинные доли (доли в Linpack-производительности)

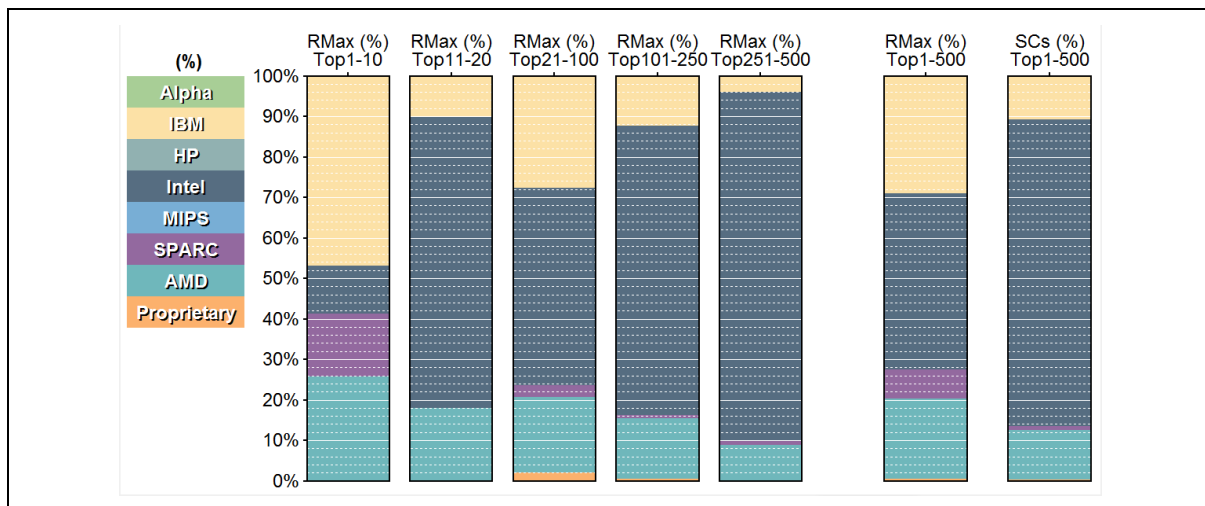


Рис. 9. Распределение технологий процессоров по уровням Top1–10, Top11–20, Top21–100, Top101–250, Top251–500. По данным Top500 за ноябрь 2012 года

Налицо явная и весомая разница между «правдой» и «истиной». Разберемся с этой разницей на примере редакции Top500 за ноябрь 2012 года. При помощи программы Top500 Analyzer построим распределение (Рис. 9) технологий процессоров по пяти уровням суперкомпьютеров: Top1–10, Top11–20, Top21–100, Top101–250 и Top251–500.

Видно, что для категории «Intel» ареал распространения в левых пяти колонках напоминает треугольник, с вершиной слева и с основанием — справа. То есть, процессоры Intel тем лучше представлены в суперкомпьютерах, чем к более слабому уровню они относятся — где суперкомпьютеров по количеству много, а по производительности они слабые. Для категорий «IBM», «AMD» и «SPARC» ареалы распространения в левых пяти колонках смещены к старшим уровням суперкомпьютеров — где суперкомпьютеров по количеству мало, а по производительности они мощные. В результате на официальном плакате истинные доли категорий «IBM», «AMD» и «SPARC» сильно приуменьшены, а доля категории «Intel» серьезно преувеличена. Истинное суждение (поправляющее заблуждение §7) будет таким:

§8 К ноябрю 2012 года в суперкомпьютерах Top500 на процессоры Intel приходится значительная доля (44%). Однако, отрыв от ближайших преследователей не такой уж и большой: в 1.5 раз от IBM (29%), в 2.2 раза от AMD (20%). Заметная доля (7%) приходится на процессоры SPARC.

7. Восстановление истины: компании-производители суперкомпьютеров

Для суперкомпьютеров, вошедших в Top500, проанализируем показатель «компания-производитель»; соответствующее поле в записях Top500 — «Manufacturer». Безусловные лидерские позиции здесь принадлежат трем компаниям, поэтому в программе «Top500 Analyzer» введем четыре категории¹: «Cray» — суперкомпьютер изготовлен компанией Cray Inc.; «IBM» — суперкомпьютер изготовлен компанией IBM; «HP» — суперкомпьютер изготовлен компанией Hewlett-Packard, «Other» — суперкомпьютер изготовлен любой иной компанией.

С помощью программы Top500 Analyzer построим диаграммы долей компаний-производителей (Рис. 10). Как обычно, левая часть рисунка — доли «в штуках», правая диаграмма показывает истинные доли компаний-производителей. Опять налицо явная и серьезная разница между «правдой» и «истиной». Среди прочего видно, что в последнее пятилетие истинная доля категории «HP» существенно (в разы) преувеличивается, а доля категории «Cray» существенно преуменьшается. Основываясь на вычислении долей «в штуках» за ноябрь 2008 года можно сделать следующее утверждение, которое, несомненно, является правдой:

§9 По данным редакции Top500 за ноябрь 2008 года компания Hewlett-Packard поставила больше всех других компаний суперкомпьютеров (42%), вошедших в данную редакцию Top500. Ближайшие конкуренты: IBM (37%, отставание в 1.13 раза) и Cray (5%, отставание в 8.4 раза). Все остальные компании-производители, вместе взятые, серьезно уступают лидеру (16%, отставание в 2.6 раза).

Истинное положение дел в ноябре 2008 года серьезно отличается от утверждения §9:

§10 По данным редакции Top500 за ноябрь 2008 года суперкомпьютеры компании IBM обеспечили 38% всей суммарной Linpack-производительности списка Top500. Это серьезно превышает доли ближайших конкурентов. Так, доля суперкомпьютеров компании Hewlett-Packard — 25% (отставание в 1.5 раза), компании Cray — 15%, всех остальных компаний-производителей, вместе взятых — 22%.

Сравнивая утверждения §9 и §10, отметим, что в утверждении §9 истинная доля категории «HP» была серьезно (в 1.76 раз) преувеличена, истинная доля категории «Cray» была серьезно (в 3 раза) преуменьшена, истинная доля категории «Others» была преуменьшена в 1.4 раза. Кроме того, был просто неверно указан лидер отрасли.

Уместно напомнить, что отличия суперкомпьютеров по Linpack-производительности влекут подобные же отличия по цене, технической сложности, объему оборудования в различных подсистемах суперкомпьютеров. Тем самым, утверждение §10 дает лучшее представление о распределении между компаниями долей (в денежном исчислении) рынка суперкомпьютеров. Именно такая информация важна для потенциальных инвесторов.

Подобные (§9) мнимые признаки абсолютного лидерства в принципе дают компании аргументы для настойчивого продвижения своих решений, даже в тех сегментах, где, на самом деле, позиции компании весьма слабы. Например, это позволяет всерьез обращаться к лицам, принимающим решения, с предложением построить для России суперкомпьютер высшей производительности (Top1–5), аргументируя данное предложение своим лидерством в суперкомпьютерной отрасли. Для правильной оценки подобных предложений важно знать истинные позиции той или иной компании, причем на различных уровнях суперкомпьютерной отрасли.

На примере редакции Top500 за ноябрь 2012 года разберем распределение (Рис. 11) компаний-производителей по пяти уровням суперкомпьютеров: Top1–10, Top11–20, Top21–100, Top101–250 и Top251–500.

¹ Заинтересованный читатель-программист легко может в программе To500 Analyzer [4] изменить эти установки.

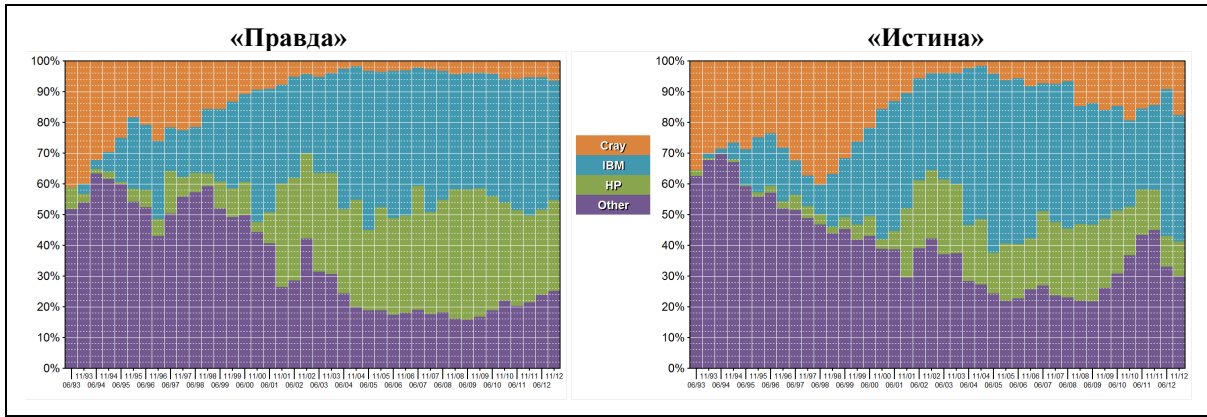


Рис. 10. Доли компаний-производителей суперкомпьютеров по данным всех 40 редакций рейтинга Top500 (от июня 1993 года по ноябрь 2012 года). Слева доли «в штуках» (доли от общего числа суперкомпьютеров), справа истинные доли (доли в Linpack-производительности)

Видно, что для категории «HP» ареал распространения в левых пяти колонках напоминает треугольник, с вершиной в третьем уровне и с основанием в пятом. Суперкомпьютеры Hewlett-Packard вообще отсутствуют в высших двух уровнях (Top1–10, Top11–20), слабо представлены на третьем уровне (Top21–100), заметно присутствуют на 4–5 уровнях (Top101–500) — там, где суперкомпьютеров по количеству много, а по производительности они слабые.

Для категорий «Others» и особенно «Cray» ареалы распространения смещены к старшим уровням суперкомпьютеров — где суперкомпьютеров по количеству мало, а по производительности они мощные. Для категории «IBM» ареал распространения занимает сравнимые доли на всех пяти уровнях.

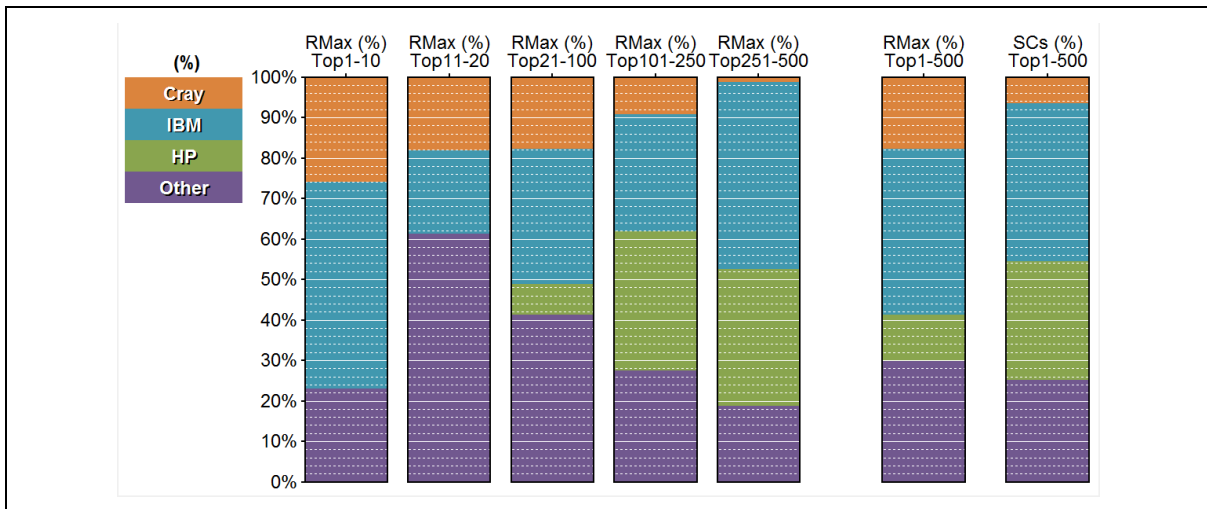


Рис. 11. Распределение долей компаний-производителей суперкомпьютеров по уровням Top1–10, Top11–20, Top21–100, Top101–250, Top251–500. По данным Top500 за ноябрь 2012 года

8. Восстановление истины: технологии интерконнекта

Для суперкомпьютеров, вошедших в Top500, проанализируем показатель «используемая технология интерконнекта». Для этого следует анализировать два поля: «Interconnect» и «Interconnect Family». В программе Top500 Analyzer введем шесть категорий для обозначения технологии интерконнекта.

Пять категорий из шести явно указывают используемую сетевую технологию: «Infiniband», «Ethernet», «Myrinet», «SCI» и «Quadrics». Все эти пять сетевых технологий являются коммерчески доступными: любой разработчик суперкомпьютеров может¹ приобрести отдельно

¹ В период производства соответствующих изделий.

соответствующие сетевые изделия — сетевые адаптеры, коммутаторы, кабели или даже микросхемы для адаптеров и коммутаторов,— и на этой базе разрабатывать свои собственные суперкомпьютеры.

Шестая категория — «Custom»,— объединяет сетевые технологии, которые коммерчески недоступны, как отдельные сетевые решения. Поясим: можно купить целиком суперкомпьютер IBM Blue Gene, но невозможно купить отдельно интерконнект, который используется в суперкомпьютерах IBM Blue Gene, и на базе такого интерконнекта разработать свой собственный суперкомпьютер. По факту, в категорию «Custom» попадают различные решения, которые, по сравнению с остальными, имеют более высокие технические показатели и расширенные функциональные возможности. При этом, данные технологии невозможно купить отдельно. Значит, если будет стоять задача разработки российского суперкомпьютера с подобным интерконнектом, то подобный интерконнект (аналог) придется разрабатывать самостоятельно, из-за невозможности покупки готового решения. При принятии решения о такой разработке, естественно, встает вопрос:

§11 *Надо ли тратить ресурсы на разработку российской технологии интерконнекта, подобной технологиям, представленным в категории «Custom»? Или коммерчески доступных технологий интерконнекта вполне достаточно для создания всех необходимых отечественных суперкомпьютеров?*

Давайте разберемся. С помощью программы Top500 Analyzer построим диаграммы долей технологий интерконнекта (Рис. 12). Как обычно, левая часть рисунка — доли «в штуках», правая диаграмма показывает истинные доли технологий интерконнекта.

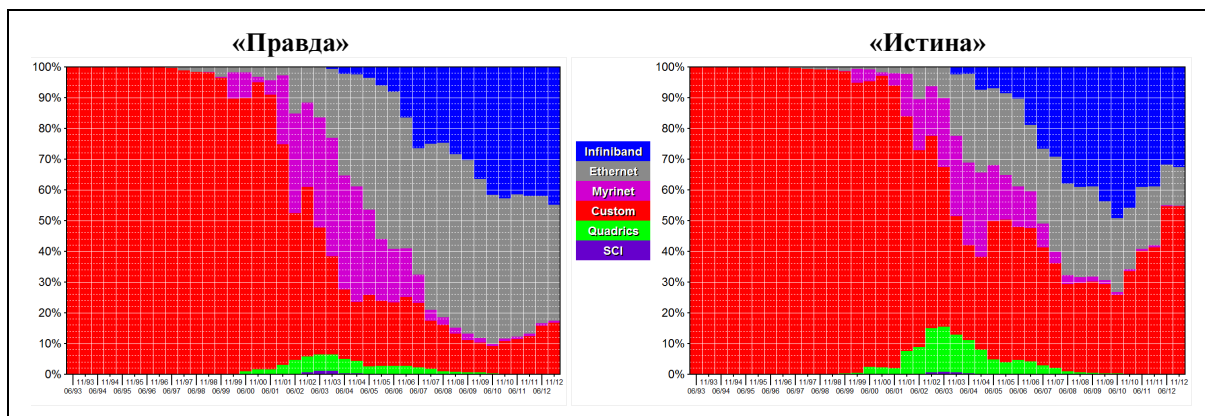


Рис. 12. Доли технологий интерконнекта в суперкомпьютерах по данным всех 40 редакций рейтинга Top500 (от июня 1993 года по ноябрь 2012 года). Слева доли «в штуках» (доли от общего числа суперкомпьютеров), справа истинные доли (доли в Linpack-производительности)

Опять имеем явную и серьезную разницу между «правдой» и «истиной». Среди прочего видно, что (на отрезке за последние годы) на левой диаграмме истинная доля категории «Ethernet» существенно (в разы) преувеличивается, а доля категории «Custom» существенно преуменьшается. Основываясь на вычислении долей «в штуках», можно сделать следующие утверждения, которые, несомненно, являются правдой:

§12 *По данным редакции Top500 за ноябрь 2012 года технологии Infiniband и Ethernet использовались в большинстве суперкомпьютеров, вошедших в данную редакцию списка (45%+38%=83%). Доли категорий «Custom» (16.7%) и Myrinet (0.3%) незначительны.*

§13 *Немногим ранее ситуация была еще радикальнее. По данным редакции Top500 за июнь 2010 года технологии Infiniband и Ethernet использовались в подавляющем большинстве суперкомпьютеров, вошедших в данную редакцию списка (41%+49%=90%). Доли категорий «Custom» (9%), Myrinet (0.5%) и Quadrics Myrinet (0.5%) — незначительны.*

На базе утверждений §12 и §13 легко принять глубоко ошибочное решение по вопросу §11:

§14 *Нецелесообразно тратить ресурсы на разработку российской технологии интерконнекта, подобной технологиям, представленным в категории «Custom». При*

разработке отечественных суперкомпьютеров вполне можно обойтись коммерчески доступными решениями Ethernet и Infiniband.

Истинное положение дел (и в ноябре 2012 года, и в июне 2010 года) серьезно — многократно, — отличается от утверждений §12 и §13. На это уже указывалось выше (Таблица 3). При этом не только многократно искажены доли технологий интерконнекта, но и неверно указана лидирующая категория: истинный и абсолютный лидер по данным Top500 за ноябрь 2012 года — категория «Custom» (55%). Еще раз упомянем: доля суперкомпьютера по Linpack-производительности коррелирует с технической сложностью, объемом оборудования в различных подсистемах суперкомпьютера (например, с числом портов интерконнекта).

Для правильной оценки роли той или иной технологии интерконнекта важно знать и распределение долей по уровням суперкомпьютерной отрасли. На примере редакции Top500 за ноябрь 2012 года разберемся с этим распределением (Рис. 13). Видно, что для категории «Ethernet» ареал распространения напоминает треугольник, с вершиной на третьем уровне и с основанием на пятом. Суперкомпьютеры в интерконнекте на базе Ethernet вообще отсутствуют в высших двух уровнях (Top1–10, Top11–20), слабо представлены на третьем уровне (Top21–100), заметно присутствуют на 4–5 уровнях (Top101–500) — где суперкомпьютеров по количеству много, а по производительности они слабые. Ареалы распространения категории «Infiniband» можно описать так: очень малое присутствие (менее 10%) на первом уровне, значительное присутствие (70%–50%–50%–40%) на втором–пятом уровнях. У категории «Custom» ареал распространения смещен к старшим уровням суперкомпьютеров — где суперкомпьютеров по количеству мало, а по производительности они мощные. Именно на данных решениях и строятся рекордные установки, обладание которыми стратегически важно для России.

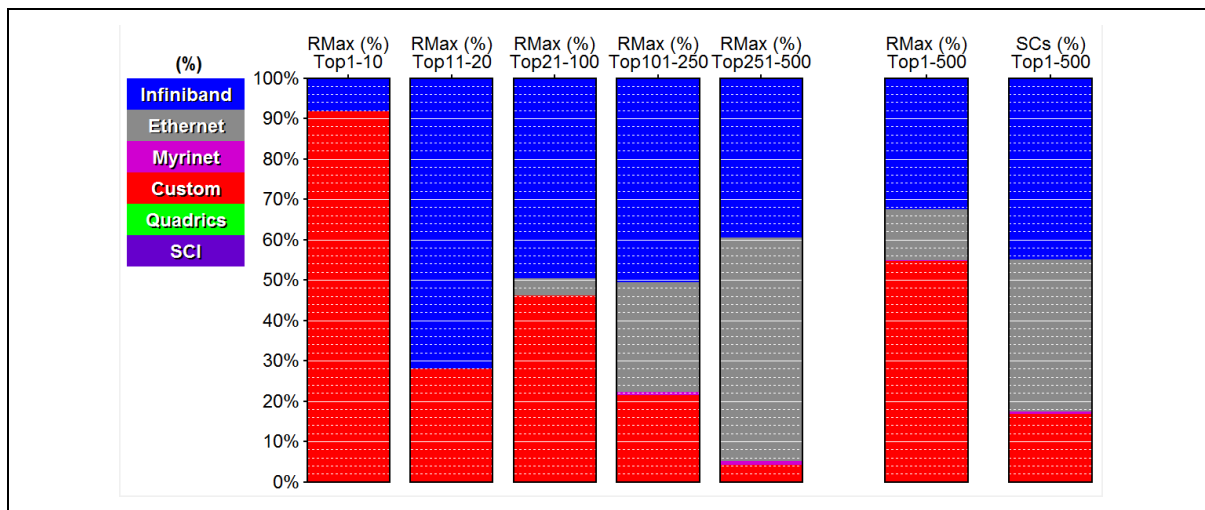


Рис. 13. Распределение долей технологий интерконнекта в суперкомпьютерах по уровням Top1–10, Top11–20, Top21–100, Top101–250, Top251–500. По данным Top500 за ноябрь 2012 года

Тем самым, обоснованное решение по вопросу §11 будет таким:

§15 *Технологи категории «Custom» обеспечивают подавляющую долю Linpack-производительности (55% по данным редакции Top500 от ноября 2012 года). А если говорить про самые мощные суперкомпьютеры — Top1–10 и Top10–20, — про суперкомпьютеры, которые России вряд ли будут проданы и которые предстоит построить самостоятельно, то эти системы практически всегда строятся на технологиях категории «Custom». При этом, технологии категории «Custom» не продаются как отдельные продукты. Таким образом, в России, безусловно, необходимо проводить разработку собственных технологий интерконнекта из категории «Custom».*

9. Заключение

При написании данной статьи целью было продемонстрировать читателю:

- насколько традиционный и широко распространенный способ анализа Top500 — подсчет долей в количестве суперкомпьютеров,— искажает истинное положение дел в суперкомпьютерной отрасли (разделы 1, 5–8);
- как важно разработать и грамотно применять правильные методики анализа данных Top500 (раздел 3);
- как легко на базе неверного поверхностного анализа совершаются ошибки в управленческих решениях с серьезными последствиями;
- насколько сильно сегодня расслоение (по реальной производительности) в суперкомпьютерном мире (раздел 4) — кажется, пока еще даже профессионалы (чисто психологически) не всегда в полной мере осознают всю глубину данного расслоения;
- насколько важно для профессионального анализа Top500 обладать правильно построенным инструментарием.

Хочется надеяться, что эти цели хотя бы частично были достигнуты. Конечно, формат статьи не позволяет продемонстрировать все возможности программы Top500 Analyzer. Так, распределений долей категорий (например, Рис. 13) по уровням (Top1–10, Top11–20, Top21–100, Top101–250, Top251–500) в программе можно просмотреть за каждую редакцию Top500. И можно это сделать в режиме анимации: один год (две редакции Top500) — за секунду. При таком просмотре можно разглядеть эпохи появления, расцвета и угасания той или иной категории, рассмотреть, как разные категории конкурируют за доли на том или ином уровне, как происходит их миграция с уровня на уровень.

Что касается дальнейшего развития работ, то есть планы по ряду улучшений в программе Top500 Analyzer. Будет хорошо, если найдутся коллеги, которые помогут в этом — советом или делом.

Конечно, было бы интересно применить методику анализа и программу Top500 Analyzer к рейтингу Graph500 [2] и к национальному рейтингу 50 самых мощных систем в СНГ [8]. Однако в последнем случае серьезными препятствиями являются:

- невозможность выгрузки редакций этого рейтинга в виде Excel-таблицы или в виде иного файла, с возможностью (приемлемого по сложности) разбора рейтинга по записям и по полям;
- предположительно малое число формализованных полей в записях рейтинга.

Завершая, хочется поблагодарить сотрудников ИПС имени А.К.Айламазяна РАН — Е.П. Лилитко и М.Г. Химшиашвили,— которые помогли автору при создании данной статьи.

Литература

1. TOP500 Supercomputer Sites — мировой рейтинг пятисот самых производительных (на тесте Linpack) вычислительных машин мира // Электронный ресурс в сети Интернет: <http://www.top500.org/>
2. Graph500 — мировой рейтинг самых производительных (на задаче поиска в большом графе в ширину) вычислительных машин мира // Электронный ресурс в сети Интернет: <http://www.graph500.org>
3. Википедия — свободная энциклопедия, которую может редактировать каждый // Электронный ресурс в сети Интернет, русскоязычный вариант — <http://ru.wikipedia.org> и англоязычный вариант — <http://en.wikipedia.org>.
4. Абрамов С.М. Top 500 Analyzer — программа для анализа данных рейтинга Top500 // Электронный ресурс в сети Интернет <http://skif.pereslavl.ru/psi-info/rcms-skif/top500analyzer>
5. Воейков Д. Рейтинг Top 500. Соревнование с гандикапом // PC Week/RE №27–28 (633–634) 22 июля–11 августа 2008, <http://www.pcweek.ru/themes/detail.php?ID=112308>

6. Абрамов С.М. Суперкомпьютерные технологии России: объективные потребности и реальные возможности // Журнал «CAD/cam/cae Observer» #2 (54), 2010, с. 1–11.
7. Абрамов С.М., Лилитко Е.П. Состояние и перспективы развития вычислительных систем сверхвысокой производительности // Труды Шестой Международной конференции «Параллельные вычисления и задачи управления». Москва, 24–26 октября 2012 г. (РАСО-2012), Институт проблем управления им. В.А. Трапезникова РАН. Том 1, М.: ИПУ РАН, 2012, с. 10–32. ISBN 978-5-91450-122-5 (т. 1).
8. Top50 суперкомпьютеров — рейтинг 50 вычислительных систем, установленных на территории СНГ и показавших наибольшую производительность на тесте Linpack // Электронный ресурс в сети Интернет: <http://top50.supercomputers.ru>