

Переход в облако

Практическое руководство
по организации облачных вычислений
для ученых и IT-специалистов

ХУАН АНЬЕЛЬ, ДИЕГО МОНТЕС,
ХАВЬЕР РОДЕЙРО ИГЛЕСИА



Хуан Анбель

**Переход в облако. Практическое
руководство по организации
облачных вычислений для
ученых и IT-специалистов**

«Альпина Диджитал»

2020

Аньель Х.

Переход в облако. Практическое руководство по организации облачных вычислений для ученых и IT-специалистов /
Х. Аньель — «Альпина Диджитал», 2020

Многим ученым в ходе работы приходится делать высокопроизводительные вычисления и хранить большие объемы данных. Для этого нужно дорогостоящее оборудование, а порой и целый IT-отдел, занимающийся его поддержкой и настройкой. Эффективным решением могут стать облачные сервисы, которые быстро предоставляют доступ к необходимым ресурсам, позволяют легко обмениваться данными и хранить их вне локальных компьютеров и значительно снижают затраты на техническое оснащение лабораторий и исследовательских центров. Облачные технологии уже сегодня широко используются в математике, физике, археологии, исследованиях атмосферы и климата, медицине, космических исследованиях, энергетике и других областях. Они расширяют возможности научных исследований за счет собственных мощностей, способствуют развитию науки в неблагополучных странах и областях, помогают развивать международное научное сотрудничество, предоставляют высокий уровень кибербезопасности, позволяют оптимизировать расходы и воздействие на окружающую среду. Облачные сервисы также внедряют в небольшие стартапы и крупные бизнесы с большим количеством сотрудников, клиентов и поставщиков, так как они обеспечивают постоянную синхронизацию и мобильность. Над книгой «Переход в облако» работали доктора наук и профессора Хуан Аньель, Диего Монтес и Хавьер Иглесиас, которые много лет используют облачные решения для своих проектов. В этом издании они собрали всю информацию, необходимую для перехода к облачной инфраструктуре, включая анализ и сравнение различных сервисов, описание доступных инструментов и программ, этические и юридические аспекты облачных технологий и другие подробности.

© Анъель Х., 2020

© Альпина Диджитал, 2020

Содержание

Вступительное слово	7
Предисловие	8
Введение	10
Благодарность	11
Список сокращений	12
Глава 1	14
Конец ознакомительного фрагмента.	16

**Хуан Анъель, Диего Монте,
Хавьер Родейро Иглесиа**
**Переход в облако. Практическое
руководство по организации облачных
вычислений для ученых и IT-специалистов**

Издано при содействии SberCloud

Руководитель проекта *И. Позина*

Научный редактор *В. Яценков*

Дизайнер *А. Маркович*

Корректор *Е. Жукова*

Компьютерная верстка *Б. Руссо*

First published in English under the title Cloud and Serverless Computing for Scientists; A Primer by Juan A. Añel, Diego P Montes and Javier Rodeiro Iglesias, edition: 1

Copyright © Springer Nature Switzerland AG, 2020 *

This edition has been translated and published under licence from Springer Nature Switzerland AG.

Springer Nature Switzerland AG takes no responsibility and shall not be made liable for the accuracy of the translation.

© Перевод. Смилевска Л., 2022

© Оформление, издание на русском языке. ООО «Альпина ПРО», 2022

Все права защищены. Данная электронная книга предназначена исключительно для частного использования в личных (некоммерческих) целях. Электронная книга, ее части, фрагменты и элементы, включая текст, изображения и иное, не подлежат копированию и любому другому использованию без разрешения правообладателя. В частности, запрещено такое использование, в результате которого электронная книга, ее часть, фрагмент или элемент станут доступными ограниченному или неопределенному кругу лиц, в том числе посредством сети интернет, независимо от того, будет предоставляться доступ за плату или безвозмездно.

Копирование, воспроизведение и иное использование электронной книги, ее частей, фрагментов и элементов, выходящее за пределы частного использования в личных (некоммерческих) целях, без согласия правообладателя является незаконным и влечет уголовную, административную и гражданскую ответственность.

* * *

Вступительное слово

Перед прочтением этой книги задайте себе вопрос: знаете ли вы, что такое облака? Если в первую очередь вам пришло на ум облачное хранилище файлов, то вы правы. Но это лишь малая толика возможностей облачных технологий не только для бизнеса, но и для частных пользователей. Предприятия с помощью «облаков» могут как создавать удаленные рабочие места и виртуальные сети, так и помогать тестировать и разрабатывать приложения в облачных средах, использовать технологии машинного обучения. Облачный рынок даже во время кризиса и пандемии Covid-19 не потерял высоких темпов развития.

Вокруг темы «Облачные технологии» до сих пор ходит достаточно много мифов и домыслов. К сожалению, нельзя сказать, что книжный рынок изобилует современными изданиями, где простыми словами объясняется самая суть облаков.

Компания SberCloud – один из российских и достаточно известных облачных провайдеров – создана в январе 2019 года. Мы неоднократно получали награды, в том числе, международные, за свои продукты и решения. Мы предоставляем облачную IT-инфраструктуру в аренду, предлагаем 50+ платформенных сервисов, а также являемся разработчиком самого мощного в России суперкомпьютера Christofari, доступного из облака, и платформы ML-разработки – ML Space. Наши услуги популярны как среди бизнеса и государственных организаций, так и среди физических лиц.

Принимая участие в новом для нас проекте, мы использовали нашу накопленную экспертизу в области облачных технологий, чтобы в легкой, интересной, а главное правильной форме донести для читателя все возможности, преимущества и факты об облаках. Эта книга может стать подспорьем для студента, захватывающим чтением для любителя новых технологий, а эксперт, надеемся, также найдет в ней что-то полезное для себя. Приятного чтения!



Предисловие

Как исследователь в различных областях, получивших одну лишь пользу от развития распределенных вычислений, могу сказать: очевидно, что мы еще долгое время будем продолжать движение по пути открытий и разработки новых инструментов и сервисов. С появлением ПК мы отошли от такой исторической парадигмы, как мейнфрейм, затем последовательно переходили на все более мощные рабочие станции, универсальные высокопроизводительные вычисления, распределенные вычисления, и теперь – в «облако». Справедливости ради, одним из недостатков таких изменений является то, что сторонники этих новых архитектур зачастую подают последнюю парадигму как наилучшую практически во всех отношениях, а значит, она должна заменить все предыдущие архитектуры, переводя пользователей на новейшие, величайшие и наилучшие системы. Это явно не тот случай, когда шумиха не имеет значения, и потому предназначением данной книги можно считать просвещение читателя: в каждой работе должен использоваться свой, подходящий именно ей инструмент.

Читая эту книгу, я был поражен разнообразием целевой аудитории в предметной области; это к лучшему, потому что сообщества практикующих специалистов, которым так важно использовать облачные технологии, найдут для себя подходящие примеры. Однако я также очень надеюсь, что книгу прочтает еще одна группа – руководители проектов или старшие преподаватели: книга должна стать для них обязательной к прочтению. Наряду с выполнением традиционных функций они сегодня несут и юридическое обязательство – по соблюдению требований об открытости данных и воспроизводимости; для этих целей как раз подходит перенос сервисов в облако и хранение в нем данных. Кроме того, следует понимать преимущества и, что не менее важно, проблемы, которые может принести облачная стратегия.

Работая с Хуаном и Диего еще до написания этой книги, мы уделили немало времени проработке различных моделей применения IaaS в сфере наук об окружающей среде, и в частности – о климате. Следовательно, многие замечания, сделанные в книге, были выработаны авторами на практике, а не в ходе теоретических изысканий. Помимо этого, с учетом моего опыта разработки общеевропейских федеративных облачных сервисов для исследований, я отчетливо вижу, что уроки об особенностях использования облака, извлеченные в том числе из неудач, можно было усвоить быстрее и легче, если бы тогда существовал подобный текст, которым мы могли бы поделиться с пользователями. Я считаю также, что эта книга будет крайне полезна не только тем исследователям, кто со временем станет пользователем крупных публичных облачных служб, но и тем, кто выберет облачный проект European Open Science Cloud, который становится все более популярным инструментом среди европейских исследователей.

Наконец, я хотел бы подчеркнуть уже высказанную авторами мысль: в динамичном и изменчивом мире облачных вычислений все устаревает крайне быстро. Примеров тому предостаточно; однако вот что важно: наличие различных моделей развертывания облачных вычислений означает, что, описывая свой метод использования облака, двое исследователей отнюдь не обязательно говорят об одном и том же. Вот почему необходимо убедиться в том, все ли понимают базовую модель облачных вычислений, как это сделано в начале данной книги. В заключение добавлю: как только читатели поймут, как можно использовать в их исследованиях облачные вычисления, им не мешает сделать паузу в чтении и поразмыслить над тем, как они хотели бы двигаться вперед с облачными решениями, проведя тесты и убедившись, что они действуют в соответствии с обоснованными научными и техническими принципами развертывания сервисов в облаке. Да, облако не палочка-выручалочка, но в долгосрочной перспективе при грамотном развертывании может ощутимо упростить все процессы.

*Профессор Дэвид Уоллом
доцент и заместитель директора по инновациям Оксфордского центра
электронных исследований,
Оксфорд, Великобритания*

Введение

Эту книгу мы написали для ученых, инженеров и для всех, кто хочет ближе познакомиться с облачными вычислениями, чтобы узнать о них больше или оценить облако как альтернативное или дополнительное решение для собственных вычислительных потребностей. Кроме того, книга может быть полезна IT-специалистам, например архитекторам программных решений, желающим быть в курсе текущих и будущих потребностей научного сообщества, а также понимать, как можно удовлетворить их с помощью облачных технологий – и как следствие предлагать своим клиентам более подходящие решения. Таким образом, этот текст может стать своеобразным мостом между видением пользователя и поставщика.

В этой книге мы попытались дать общее представление о современном состоянии облачных вычислений. Текст может послужить вводным курсом для выпускников, студентов магистратуры и аспирантов, обучающихся любым дисциплинам: он даст представление о потенциальных возможностях использования облачных технологий и о том, как они могут помочь в разработке проектов и будущей работе.

Хуан Анъель

Диего Монтес

Хавьер Родейро Иглесиа

Оренсе, Испания

Ноябрь 2019 г.

Благодарность

Эту книгу мы хотели бы посвятить нашим семьям и друзьям за их ежедневную поддержку и понимание, позволяющие нам отдавать себя исследовательской работе, что слишком часто происходит за счет времени, которое мы могли провести с ними.

Кроме того, мы хотим поблагодарить Springer за интерес к этой книге и возможность опубликовать ее, и особенно Закари Романо, нашего ответственного редактора, а также редакторскую службу АЖЕ. Особая благодарность Пауло Родригесу из Dropbox за полезные дискуссии и содействие в работе, за внесенный в эту книгу вклад. Кроме того, мы хотели бы поблагодарить Google Enterprise, Microsoft Research и Amazon Web Services за предоставленные вычислительные ресурсы и техническую поддержку для проведения экспериментов с использованием их облачных платформ. Том Грей, Барак Регев и Шайлеш Рао из Google, Кендзи Такеда из Microsoft Research – мы выражаем вам глубокую признательность. Также мы благодарим коллег из ClimatePrediction.net в Оксфордском университете, особенно Дэвида Уоллома, всех сотрудников Галисийского центра суперкомпьютерных вычислений (CESGA), в частности Карлоса Фернандеса, а также Томаса Фернандеса Пену из Университета Сантьяго-де-Компостела.

Мы благодарим преподавателя кафедры физиологии Университета Сантьяго-де Компостела Сулаю Тóвар и профессора Хосе Агустина Гарсию из Университета Эстремадуры за отзывы о черновых версиях этой книги. Кроме того, мы выражаем признательность Александре Иглесиас за ее вклад в дизайн обложки.

Часть исследований, проведенных в течение многих лет и позволивших нам накопить знания для создания этой книги, финансировались различными агентствами и учреждениями, в т. ч. Европейским фондом регионального развития (ERDF). Кроме того, в последние годы Хуану Анхелю была оказана поддержка в виде гранта Рамона-и-Кахаля от правительства Испании (RYC-2013–14560).

Список сокращений

AI	Artificial Intelligence	Искусственный интеллект
API	Application programming interface	Программный интерфейс приложения
AWS	Amazon Web Services	Облачный сервис; коммерческое публичное облако, поддерживаемое Amazon
BOINC	Berkeley Open Infrastructure for Network Computing	Открытая программная платформа университета Беркли; некоммерческое межплатформенное ПО для организации распределенных вычислений
CDC	USA Centers for Disease Control and Prevention	Центры по контролю и профилактике заболеваний США
CERN	Conseil Européen pour la Recherche Nucléaire	Европейская организация ядерных исследований (ЦЕРН)
CLI	Command Line Interface	Интерфейс командной строки
CNCF	Cloud Native Computing Foundation	Фонд, созданный при поддержке Linux Foundation для разработки и продвижения облачных технологий
DoD	US Department of Defense	Министерство обороны США
EC2	Elastic Cloud Compute	Один из сервисов AWS, позволяющий пользователю арендовать виртуальный сервер с настраиваемым объемом ресурсов
ERP	Enterprise Resource Planning	Планирование ресурсов предприятия; ПО для управления бизнес-процессами
ESA	European Space Agency	Европейское космическое агентство (ЕКА)
EU	European Union	Европейский союз
FaaS	Function as a Service	«Функция как услуга»
FDA	U.S. Food and Drug Administration	Управление по санитарному надзору за качеством пищевых продуктов и медикаментов США
FMI	Finnish Meteorological Institute	Финский метеорологический институт
GCP	Google Cloud Platform	Набор облачных служб, предоставляемый Google
GCS	Google Cloud Storage	Служба хостинга файлов в инфраструктуре GCP
GPL	GNU Public License	Генеральная общедоступная лицензия GNU, проекта по разработке свободного ПО

Глава 1

Почему вам нужна эта книга?

Активное использование компьютеров в научных целях постоянно стимулировало развитие технологий. В исследовательской работе компьютеры используются по-разному, но наиболее востребованными их функциями являются хранение данных и высокопроизводительные вычисления (НРС). При таком использовании наличие сетевого оборудования всегда являлось обязательным требованием; однако оно потеряло актуальность с развитием концепции облачных вычислений в последнее десятилетие и внесерверной обработки данных в последние годы. И действительно, отчет проекта [top500.org](https://www.top500.org/)¹ гласит: за период 2016–2017 гг. рынок НРС вырос примерно на 1,6 %, тогда как использование облачных НРС-сервисов возросло с 7 до 44 %. Скорее всего, объяснить эту разницу можно тем, что НРС как услуга (НРСaaS) используется для дополнения ресурсов локальных инфраструктур НРС, когда таких ресурсов недостаточно для удовлетворения нужд пользователя. Это, пожалуй, одна из самых привлекательных особенностей облачных вычислений, стимулирующая их использование в исследованиях: удовлетворение потребности в вычислительных ресурсах при реализации тех или иных проектов.

Облачные вычисления уже более десяти лет являются широко распространенной и внедряемой ИТ-отделами технологией. Приложения, решения и варианты практического применения облачных вычислений исследуются повсеместно. Однако далеко не всегда пользователи имеют полное представление об их потенциале, вариантах использования, устройстве и ситуации на рынке в целом. Так, они могли ознакомиться с основными характеристиками компьютера, ноутбука, рабочей станции и т. д. и получить общее представление о том, какими функциями они будут располагать при наличии такой технологии, а также о том, какую из технологий выбрать в соответствии с памятью или процессором своего устройства. Однако все это – пока не облачные вычисления. Многие еще просто-напросто не приспособились к этой новой парадигме, что привело к пробелу в знаниях между преобладающей сегодня технологией и тем, к чему привыкли пользователи. Тем не менее, мы все чаще используем облачные вычисления как для повседневной работы, так и для развлечений, например, открывая приложение на смартфоне. Также это касается и тех, кто работает в науке, занимается изысканиями, возглавляет лаборатории или исследовательские центры – вплоть до первых лиц. Все они могут получить весомые преимущества, если хотя бы частично задействуют облачные решения, как будет рассказано и показано на примерах далее. Они применимы практически к любой научной дисциплине. Что касается руководителей высокого уровня, то им, возможно, придется рассмотреть вопрос о переходе ИТ-инфраструктуры на облачные технологии; и, конечно, они захотят больше узнать о последствиях такого перехода. Отчет компании The Economist Intelligence Unit за 2016 г. содержал дискуссию о том, как недостаток знаний и недоверие становятся барьерами для внедрения облачных вычислений². И хотя сегодня эти показатели, скорее всего, улучшились, недавний опрос Евростата показал: всего 28 % европейских предприятий используют облачные технологии, а их более широкое внедрение ожидается только в будущем³.

¹ Feldman M. (2018) Cloud computing in HPC surges. <https://www.top500.org/news/cloudcomputing-in-hpc-surges/>. По состоянию на 29 ноября 2019 г.

² The Economist Intelligence Unit (2016) Trust in cloud technology and business performance: reaping benefits from the cloud, 26 pp.

³ Kaminska M., Smihily M. (2018) Cloud computing – statistics on the use by enterprises. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Cloud_computing_-_statistics_on_the_use_by_enterprises&oldid=416727. По состоянию на 9 апреля 2019 г.

Бессерверные вычисления – своего рода эволюция облачных: эта еще более свежая концепция призвана повысить уровень абстрагирования для пользователей, для которых вопрос используемой ими вычислительной инфраструктуры станет не так уж важен.

Что до тех, кто финансирует исследования, то как государственные, так и частные организации изучали и применяли облачные решения в инфраструктуре и проектах, в которые вкладывались. Достигнуть этого можно разными путями: например, через развертывание облачной инфраструктуры с использованием общедоступных НРС-центров или предоставление ваучеров на использование ресурсов облачных вычислений от коммерческих поставщиков вместо того, чтобы выделять время общедоступных высокопроизводительных вычислительных центров.

Еще один довод в пользу широкого использования и внедрения облачных технологий: они предоставляют прекрасную возможность для демократизации развития научных исследований, так как отменяют необходимость создавать и обслуживать собственные НРС-центры. Это особенно важно для финансирующих научных изыскания организаций, исследователей и лабораторий, не имеющих доступа к таким центрам, а также тех, кто получил финансирование под конкретный ограниченный во времени проект и не хочет вкладывать средства в покупку и обслуживание оборудования, которое устареет или не будет использоваться в будущем. Потенциал этой технологии огромен: она устраняет препятствия для исследований в самых неблагополучных странах и уголках планеты, т. е. там, где инфраструктура для проведения научно-исследовательских и опытно-конструкторских работ развита недостаточно.

Что немаловажно, существует возможность объединить облачные вычисления с другими технологиями, такими как большие данные и машинное обучение, для повышения продуктивности научных исследований. Подобная интеграция позволит использовать комбинацию таких функций, как хранение данных или реализация моделей в облачных системах, что обеспечит прямой доступ к готовым решениям, уже реализованным поставщиком и полностью адаптированным к среде облачных вычислений. Тем самым будет снижена техническая нагрузка и устранены возможные проблемы при адаптации или настройке сервисов. Интеграция функций упрощает анализ, интерпретацию и представление данных, хранящихся в облаке, и создает большой потенциал для получения новых, порой неожиданных результатов и достижений. Подход, при котором происходит эффективная комбинация этих методов, получил название «облачного датамайнинга»⁴.

Еще одно преимущество комбинации всех этих технологий состоит в том, что она предлагает способ реализовать сценарий «нулевой загрузки», решая одну из насущных проблем, связанных с доступностью исследовательских данных, – чрезмерное количество и объем запросов на загрузку, для обслуживания которых требуются огромные хранилища данных. Для некоторых ученых, лабораторий и областей исследований этот подход стал практически культурной революцией: отказ от загрузки данных для локальных вычислений и переход к выполнению вычислений в облаке с загрузкой лишь окончательных результатов.

Наконец, если рассматривать технологии с точки зрения рынка труда, то предложения о работе на руководящих должностях в сфере научных вычислений отражают существующее движение к облачным технологиям. Причем особенно подчеркивается, что соискатели должны обладать знаниями, позволяющими управлять частной и публичной облачной инфраструктурой. На этом этапе тем, кто трудится в сфере науки и исследований, необходимо как можно скорее получить полное представление об экосистеме облачных вычислений, о сложностях работы с ней и возможностях, которые она предлагает для расширения научных изысканий, оптимизации расходов и воздействия на окружающую среду.

⁴ Barua H.B., Mondal K.C. (2019) A comprehensive survey on cloud data mining (CDM) frameworks and algorithms. ACM Comput Surv 52(5): 104. <https://doi.org/10.1145/3349265>.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.