ДОКЛАДЫ АН ВШ РФ
2021 апрель–июнь № 2 (51)

ТЕХНИЧЕСКИЕ НАУКИ

УДК 519.8

ВЫБОР ОПТИМАЛЬНОГО РАЗМЕЩЕНИЯ ДАННЫХ ПРИ ИСПОЛЬЗОВАНИИ ОБЛАЧНОЙ ИНФРАСТРУКТУРЫ

В.И. Мейкшан[1], Н.Б. Тесля[2]
[1] *Сибирский государственный университет телекоммуникаций и информатики*
[2] *Сибирский университет потребительской кооперации*

Преимущества использования облачных технологий очевидны, их применение расширяется, что определяет неуклонный рост спроса. Особенную актуальность облачные вычисления приобрели для крупных компаний, работающих в сфере интернет-услуг, ритейла, логистики и генерирующих большие объемы деловой и прочей информации. Применение облачных технологий позволяет организовать совместное потребление ресурсов, решить проблемы хранения и передачи весомых объемов данных. Потребительская кооперация России относится к таким крупным, территориально распределенным организациям, активно формирующим собственную цифровую экосистему. Вопрос хранения и обработки данных для организаций потребительской кооперации очень актуален. При этом цены у поставщиков облачных услуг существенно различны, и требуется решать задачу минимизации затрат на хранение и передачу значительных объемов данных. Для выбора оптимальной схемы хранения данных у нескольких поставщиков облачных сервисов при разных технико-экономических параметрах пакета предоставляемых услуг (максимальный объем хранилища, стоимость выделяемых ресурсов) рассматривается применение метода линейного программирования. Математическая модель включает в себя уравнение расходов по хранению и передаче данных с рядом ограничений по объему хранилища, объему данных и их сохранности. В качестве программного инструмента, позволяющего выполнить численные расчеты, выбран Microsoft Excel в сочетании с надстройкой «Поиск решения». В соответствии с математической моделью установлены условие минимизации величины затрат на облачное хранение и необходимые ограничения. Исходные данные задачи заданы для трех центров формирования данных, определенных размеров хранилищ у пяти провайдеров облачных услуг и условных цен на услуги хранения и передачи информации. Расчеты затрат выполнены в нескольких вариантах: без оптимизации, с решением оптимизационной задачи, а также при повышении цен у поставщиков облачных сервисов. Результаты расчетов подтверждают необходимость решения задачи минимизации затрат на облачные услуги для корпоративных клиентов. Представленную модель можно расширить под любые условия затрат, а также для разных сфер применения облачных сервисов.

Ключевые слова: облачные технологии, облачные вычисления, облачный сервис, облачное хранилище, хранение данных, математическая модель оптимизации затрат, линейное программирование, симплекс-метод, потребительская кооперация, потребительские общества.

DOI: 10.17212/1727-2769-2021-2-34-42

Введение

Согласно исследованиям делового портала TAdviser [1] в 2020 г. на российском рынке наблюдался взрывной спрос на облачные сервисы, и он продолжает расти. Крупнейшими поставщиками облачных услуг для корпоративной сферы являются всемирно известные компании Amazon, Microsoft, Google, а также рос-

© 2021 В.И. Мейкшан, Н.Б. Тесля

сийские провайдеры МТС, КРОК и другие. В России продукция лидеров «облачного» рынка наиболее востребована в сфере интернет-услуг, телекоммуникаций, в розничной торговле, в секторе транспорта и логистики, в банковской сфере.

В сферу деятельности российской потребительской кооперации, представляющей собой социально-ориентированную экосистему с чрезвычайно разветвленной структурой, также входят торговля, закупки, логистика, заготовки, общепит и др. Следовательно, в применении облачных технологий для потребительских обществ и их союзов, представленных в 71 регионе России, заложен высокий потенциал для развития.

Облачные вычисления являются основой концепции совместного потребления информационных технологий и ресурсов. Обычно использованием облачных хранилищ решают проблему дистанционного хранения и обработки больших объемов данных без приобретения и обслуживания дорогостоящего оборудования, лицензий на программное обеспечение.

Существует множество схем ценообразования для облачных услуг, которые предлагают провайдеры. Чаще всего основу тарифа составляют стоимость хранения данных (на почасовой основе или за месяц) и исходящий трафик (скачивание данных). При этом увеличение объема данных приводит к увеличению стоимости облачных услуг. Однако при наличии довольно большого рынка облачных услуг и различия в стоимости этих услуг потребителям целесообразно решать задачу минимизации затрат на облачное хранение-передачу данных.

Хранение и обмен большими объемами информации актуальны для потребительской кооперации, особенно при взаимодействии Центросоюза, главного координирующего органа, с районными потребсоюзами, география которых представлена от Мурманска до Камчатки. Последние годы огромное внимание уделяется созданию цифровой кооперативной экосистемы. Активно проводимые процессы цифровизации способствуют существенному увеличению объема информации, поэтому вопрос снижения издержек по операциям с данными очень актуален.

Исходя из вышеизложенного, рассмотрим математическую модель и численные результаты ее реализации по оптимизации затрат при хранении данных с помощью облачных сервисов. Один из подходов к решению рассматриваемой задачи предложен ранее в работах [3–5] для создания киберинфраструктуры, обеспечивающей хранение и интеграцию данных для системы государственного мониторинга водных ресурсов [2].

1. Постановка задачи

Целью исследования является реализация математической модели, основанной на методе линейного программирования, для минимизации затрат при хранении и передаче облачных данных.

Установим релевантные факторы и условия для создания математической модели оптимизации затрат. Предполагается, что задан конечный набор N гетерогенных источников данных, причем ожидаемый объем данных от i -го источника $(i=\overline{1,N})$ равен ν_i. Поступающие данные планируется накапливать в облачных хранилищах нескольких провайдеров. Из совокупности M поставщиков облачных сервисов каждый j-й провайдер $(j=\overline{1,M})$ может предоставить хранилище данных стоимостью C_j за единицу данных при максимальном объеме хранилища, не превышающем V_i. Затраты на передачу единицы данных от i-го источника

$(i=\overline{1,N})$ до хранилища для *j*-го провайдера $(j=\overline{1,M})$ оцениваются величиной a_{ij}. Задача состоит в том, чтобы найти оптимальное распределение $\{x_{ij} : i=\overline{1,N}; j=\overline{1,M}\}$, где x_{ij} – объем данных, которые относятся к *i*-му источнику и будут размещаться в хранилище *j*-го провайдера.

Для решения поставленной задачи используем метод линейного программирования. Он широко применяется для решения управленческих задач, которые связаны с оптимизацией, минимизацией затрат и получением экономии.

Общая стоимость затрат (S) включает в себя расходы на хранение (S_1) и затраты на передачу (S_2) данных:

$$S = S_1 + S_2 = \sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}C_j + \sum_{i=1}^{N}\sum_{j=1}^{M} x_{ij}a_{ij} .$$

Отсюда вытекает следующая математическая формулировка оптимизационной задачи:

$$S = \sum_{i=1}^{N}\sum_{j=1}^{M} (C_j + a_{ij})x_{ij} \to \min. \quad (1)$$

Определим ряд дополнительных ограничений:

1) объем данных от i-го источника, размещаемых у *j*-го провайдера, не может быть отрицательным, т. е.

$$x_{ij} \geq 0 \quad (i=\overline{1,N}; j=\overline{1,M}); \quad (2)$$

2) все данные, которые порождаются i-м источником, должны сохраняться в полном объеме, т. е.

$$\sum_{j=1}^{M} x_{ij} = \nu_i \quad (i=\overline{1,N}); \quad (3)$$

3) общий объем данных, которые можно разместить у *j*-го провайдера, не должен превышать емкости его хранилища, т. е.

$$\sum_{i=1}^{N} x_{ij} = V_j \ (j=\overline{1,M}) . \quad (4)$$

Аналогичная по смыслу задача сформулирована в [3–5] как задача бинарного линейного программирования. В этом случае блок данных от каждого источника считается неделимым элементом, и схема их распределения становится менее гибкой. Кроме того, при такой постановке получается задача, которая относится к NP-полным задачам, и ее решение сопряжено с существенными вычислительными сложностями. К примеру, стандартные алгоритмы линейного программирования приходится использовать в сочетании с методом ветвей и границ.

В отличие от этого задача (1)–(4) принадлежит к классу транспортных задач линейного программирования. В арсенале современной вычислительной математики имеются эффективные и быстродействующие методы точного решения подобных задач, причем многие из этих методов доведены до программной реализации и входят в состав различных пакетов прикладных программ.

2. Результаты исследования

В качестве программного обеспечения для реализации модели оптимизации затрат выбрано офисное приложение MS Excel с использованием надстройки «Поиск решения».

Считаем, что информация, предназначенная для облачного хранения, формируется в трех региональных центрах (источниках данных, которые расположены в разных регионах России): центр А, центр Б и центр В. Минимальная единица тарификации имеет определенный физический объем, обычно 1 ГБ. Для примера объемы данных для облачного хранения распределяются между источниками следующим образом: центр А – 300 ГБ, центр Б – 500 ГБ, центр В – 200 ГБ.

Поскольку именно затраты на передачу данных будут существенно различаться у разных провайдеров для выбранных центров, находящихся в разных регионах страны, то стоимость хранения у всех поставщиков облачных хранилищ для расчетов установим постоянную и одинаковую, равную одной условной денежной единице (у. д. е.) за единицу тарификации в месяц. Стоимость передачи данных за единицу у разных провайдеров (их имена представлены исключительно для удобства восприятия информации) для цели представления методологического подхода взяты произвольно (табл. 1). Подобная информация предоставляется поставщиками услуг.

Таблица 1 / Table 1

Стоимость передачи данных у различных провайдеров для региональных центров
The cost of data transmission from different providers for regional centers

Центр	Объем данных, ГБ	Стоимость передачи единицы данных у поставщиков, у. д. е. за 1 ГБ				
		Amazon	Google	Microsoft	МТС	КРОК
А	300	17	7	11	15	14
Б	500	12	9	16	12	12
В	200	14	8	9	7	15

Возможное решение поставленной задачи выбора поставщиков облачных услуг для хранения данных без оптимизации затрат представлено в табл. 2. При этом условия, заданные математической моделью, выполняются. Так, максимальный объем размещения данных в хранилищах не превышает максимальные объемы облачных хранилищ у поставщиков.

Для оптимизации затрат в MS Excel используем функцию «Поиск решения» для линейной задачи симплекс-методом. Симплекс-метод применяется для итерационного решения системы уравнений в области допустимых решений до тех пор, пока целевая функция не достигнет оптимального значения. Выбираем условие минимизации целевой функции (суммарной величины затрат на облачное хранение). Задаем ограничения, определенные в математической модели затрат: неотрицательные значения объемов данных, сохранение объемов данных в полном объеме, отсутствие превышения объема данных емкости хранилища поставщика облачных услуг.

В результате найденного решения объемы хранящихся данных будут автоматически распределены между поставщиками облачных хранилищ в соответствии с ограничениями, как представлено в табл. 3.

Таблица 2 / Table 2

Распределение данных между поставщиками и расчет затрат центров на облачный сервис без оптимизации затрат

Distribution of data between suppliers and calculation of the cost of cloud service centers without cost optimization

Центр	Объем данных, ГБ	Поставщики облачных технологий				
		Amazon	Google	Microsoft	МТС	КРОК
А	300	290	0	0	0	10
Б	500	0	150	350	0	0
В	200	0	0	0	100	100
Итого у поставщика, ГБ		**290**	**150**	**350**	**100**	**110**
Максимальный объем хранилища поставщика, ГБ		290	165	470	162	270
Стоимость хранения 1 ГБ, у. д. е.		1	1	1	1	1
Центр	Всего ГБ	Стоимость передачи данных, у. д. е. за 1 ГБ				
А	300	17	7	11	15	14
Б	500	12	9	16	12	12
В	200	14	8	9	7	15
Расчет затрат на облачные услуги, у. д. е.						
Затраты на передачу	14 220	4930	1350	5600	700	1640
Затраты на хранение	1000	290	150	350	100	110
Всего затрат	**15 220**	5220	1500	5950	800	1750

Таблица 3 / Table 3

Результат оптимизации затрат на использование облачных сервисов с минимизацией искомой функции

The result of optimizing the cost of using cloud services while minimizing the desired function

Центр	Объем данных, ГБ	Поставщики облачных хранилищ				
		Amazon	Google	Microsoft	МТС	КРОК
А	300	0	165	135	0	0
Б	500	290	0	0	0	210
В	200	0	0	38	162	0
Итого у поставщика, ГБ		**290**	**165**	**173**	**162**	**210**
Максимальный объем хранилища поставщика, ГБ		290	165	470	162	270
Стоимость хранения 1 ГБ, у. д. е.		1	1	1	1	1
Центр	Объем данных, ГБ	Стоимость передачи данных, у. д. е. за 1 ГБ				
А	300	17	7	11	15	14
Б	500	12	9	16	12	12
В	200	14	8	9	7	15
Расчет затрат на облачные услуги, у. д. е.						
Затраты на передачу	10 116	3480	1155	1827	1134	2520
Затраты на хранение	1000	290	165	173	162	210
Всего затрат	**11 116**	3770	1320	2000	1296	2730

Для подтверждения работоспособности реализации математической модели в MS Excel выполним оптимизацию затрат при увеличении стоимости передачи

данных. Для этого увеличим стоимость передачи данных для четырех из пяти поставщиков облачных хранилищ. Стоимость передачи данных для хранилища с самым большим объемом (470 ГБ) оставим без изменений (табл. 4).

Таблица 4 / Table 4

Оптимизация затрат после увеличения стоимости передачи данных у поставщиков облачных технологий

Cost optimization after increasing the cost of data transmission from cloud providers

Центр	Объем данных, ГБ	Поставщики облачных технологий				
		Amazon	Google	Microsoft	МТС	КРОК
А	300	0	0	300	0	0
Б	500	203	165	132	0	0
В	200	0	0	38	162	0
Итого у поставщика, ГБ		**203**	**165**	**470**	**162**	**0**
Максимальный объем хранилища поставщика, ГБ		290	165	470	162	270
Стоимость хранения 1 ГБ, у. д. е.		1	1	1	1	1
Центр	Объем данных, ГБ	Стоимость передачи данных, у. д. е. за 1 ГБ				
А	300	34	14	11	30	28
Б	500	24	18	16	24	24
В	200	28	16	9	14	30
Расчет затрат на облачные услуги, у. д. е.						
Затраты на передачу	15 864	4872	2970	5754	2268	0
Затраты на хранение	1000	203	165	470	162	0
Всего затрат	**16 864**	5075	3135	6224	2430	0

3. Обсуждение результатов

Полученные результаты показывают, что без оптимизации затраты на облачные услуги могут оказаться достаточно высоки. Так, при обычном подходе, согласно расчетам (см. табл. 2), суммарные затраты на хранение и передачу данных из облачного хранилища составят 15 220 у. д. е. Применение симплекс-метода для минимизации суммарных затрат на облачные услуги провайдеров (см. табл. 3) показало, что по сравнению с первым вариантом (см. табл. 2) получена экономия финансовых затрат почти на 27 %. При этом величина суммарных затрат на облачное хранение в результате оптимизации составила 11 116 у. д. е.

После увеличения у провайдеров стоимости передачи данных в два раза наблюдается перераспределение данных по хранилищам поставщиков (см. табл. 4). Как и предполагалось, наибольший объем данных передан поставщику, у которого стоимость передачи данных не изменилась. Благодаря использованию метода суммарная стоимость затрат после распределения составила 16 864 у. д. е., т. е. затраты увеличились по сравнению со вторым вариантом (см. табл. 3) на 52 % при стопроцентном увеличении тарифов большинства провайдеров. При этом отметим, что в случае хранения данных по увеличенным тарифам в первоначальном распределении (см. табл. 2) затраты на облачные услуги составят 19 405 у. д. е., что приведет к удорожанию облачных услуг на 75 %.

Таким образом, полученные результаты свидетельствует о том, что модель оптимизации затрат при принятых ограничениях сформирована корректно. Следовательно, для решения задачи минимизации затрат облачного хранения метод линейного программирования применять целесообразно, особенно при наличии многих переменных (стоимости услуг, центров данных, числа провайдеров).

Заключение

Область применения и возможности облачных технологий очень велики. Хранение и обмен данными с использованием облачных сервисов актуальны для корпоративных клиентов, особенно для структурно разветвленных и территориально распределенных. Преимущества облачных технологий необходимо в полной мере применять для потребительской кооперации, так как ее объем информационных данных, охватывающих различные экономические и социальные сферы (торговлю, образование, науку, здравоохранение и т. д.) весьма велик, а центры формирования этих данных расположены по всей территории России.

При использовании облачных сервисов необходимо грамотно подходить к выбору не только надежных провайдеров, но и к расчетам затрат на использование их услуг. Необходимо определиться с объемами хранения, поставщиками, тарифами облачных хранилищ. Эффективность модели особенно очевидна при выборе нескольких провайдеров с разными тарифами и при наличии многих центров генерации данных. Применение модели позволит без особого труда оптимизировать затраты на использование облачных хранилищ с помощью общераспространенного офисного приложения.

ЛИТЕРАТУРА

1. TAdviser. Государство. Бизнес. ИТ: web-сайт. – URL: http://www.tadviser.ru (дата обращения: 29.06.2021).
2. Information-centric systems for supporting decision-making in watershed resource development / M. Mocanu, L. Vacariu, R. Drobot, M. Muste // 19th International Conference on Control Systems and Computer Science (CSCS). – Bucharest, Romania, 2013. – P. 611–616.
3. **Negru C., Pop F., Cristea V.** Cost optimization for data storage in public clouds: a user perspective // Proceedings of 13th International Conference on Informatics in Economy. – Bucharest, Romania, 2014.
4. Budget constrained selection of cloud storage services for advanced processing in datacenters / C. Negru, F. Pop, O.C. Marcu, M. Mocanu, V. Cristea // 14th International Conference-Networking in Education and Research (RoEduNet NER). – Craiova, Romania, 2015. – P. 158–162.
5. Cost-aware cloud storage service allocation for distributed data gathering / C. Negru, F. Pop, M. Mocanu, V. Cristea, A. Hangan, L. Vacariu // IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR). – Cluj Napoca, Romania, 2016. – P. 1–5.

SELECTION OF OPTIMAL DATA PLACEMENT USING CLOUD INFRASTRUCTURE

Meikshan V.I.[1], Teslya N.B.[2]
[1] *Siberian States University of Telecommunications and Information Sciences, Novosibirsk, Russia*
[2] *Siberian University of Consumer Cooperation, Novosibirsk, Russia*

Benefits of using cloud technology are obvious, their application is expanding, as a result, it determines the steady growth of demand. Cloud computing has acquired particular relevance for large companies connected with Internet services, retailing, logistics that generate large volume of business and other information. The use of cloud technologies allows organizing the joint consumption of resources, solving the problems of storing and transferring significant amounts of data. Russian consumer cooperation refers to large territory distributed organizations actively forming their own digital ecosystem. The issue of data storing and processing for consumer cooperation organizations is very relevant. At the same time, the prices of cloud service providers are

significantly different and require solving the problem of minimizing the cost of storing and transferring significant amounts of data. The application of the linear programming method is considered to select the optimal data storage scheme for several cloud service providers having different technical and economic parameters of the package (maximum amount of storage, cost of allocated resources). Mathematical model includes the equation of costs for data storing and transferring and restrictions on the amount of storage, the amount of data and its safety. Software tool that allows to perform numerical calculations is selected Microsoft Excel in combination with the "search for solutions" add-on. In accordance with the mathematical model, the conditions for minimizing the amount of cloud storage costs and the necessary restrictions are established. Initial data are set for three data forming centers, storages of certain size for five cloud service providers and nominal price for information storage and transmission. Calculations of expenses are performed in several variants: without optimization, with the solution of the optimization problem, with price increase by cloud service providers. Results of the calculations confirm the necessity to solve the problem of minimizing the cost of cloud services for corporate clients. The presented model can be expanded for any cost conditions as well as for different areas of cloud applications.

Key words: cloud technologies, cloud calculation, cloud service, cloud storage, data storage, mathematical model, optimization of expenses, linear programming, simplex method, consumer cooperation, consumer organizations.

DOI: 10.17212/1727-2769-2021-2-34-42

REFERENCES

1. TAdviser. Goverment. Business. IT: website. (In Russian). Available at: http://www.tadviser.ru (accessed 29.06.2021).
2. Mocanu M., Vacariu L., Drobot R., Muste M. Information-centric systems for supporting decision-making in watershed resource development. *19th International Conference on Control Systems and Computer Science (CSCS)*, Bucharest, Romania, 2013, pp. 611–616.
3. Negru C., Pop F., Cristea V. Cost optimization for data storage in public clouds: a user perspective. *Proceedings of 13th International Conference on Informatics in Economy*, Bucharest, Romania, 2014.
4. Negru C., Pop F., Marcu O.C., Mocanu M., Cristea V. Budget constrained selection of cloud storage services for advanced processing in datacenters. *14th International Conference-Networking in Education and Research (RoEduNet NER)*, Craiova, Romania, 2015, pp. 158–162.
5. Negru C., Pop F., Mocanu M., Cristea V., Hangan A., Vacariu L. Cost-aware cloud storage service allocation for distributed data gathering. *IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*, Cluj Napoca, Romania, 2016, pp. 1–5.

СВЕДЕНИЯ ОБ АВТОРАХ

Мейкшан Владимир Иванович – родился в 1949 году, д-р техн. наук, профессор, профессор кафедры автоматической электросвязи Сибирского государственного университета телекоммуникаций и информатики. Область научных интересов: качество функционирования и надежность мультисервисных сетей связи, распределенные информационные системы, управление сетями связи. Опубликовано более 180 научных и учебно-методических работ. (Адрес: 630102, г. Новосибирск, ул. Кирова, 86. E-mail: meikshanvi@gmail.com).

Meikshan Vladimir Ivanovich (b. 1949) – Doctor of Sciences (Eng.), Professor, Professor of the Automatic Electrical Communication Department in the Siberian States University of Telecommunications and Information Sciences. His research interests are currently focused on on dependability and performance of multiservice networks, distributed information systems and telecommunication networks management. He is author of over 180 scientific papers. (Address: 86, Kirova St., Novosibirsk, 630102, Russia. E-mail: meikshanvi@gmail.com).

Тесля Наталья Борисовна – родилась в 1971 году, канд. техн. наук, доцент, заведующий кафедрой информатики, Сибирский университет потребительской кооперации. Область научных интересов: информационные системы и технологии. Опубликовано 58 научных работ. (Адрес: 630073, Россия, г. Новосибирск, проспект К. Маркса, 26. E-mail: teslnb@sibupk.onmicrosoft.com).

Teslya Natalia Borisovna (b. 1971) – Candidate of Technical Sciences, Associate Professor, Head of the Department of Computer Science, Siberian University of Consumer Cooperation. His research interests are currently focused on Information systems and technologies. He is author of 58 scientific papers. (Address: 26, K. Marx Av., Novosibirsk, 630073, Russia. E-mail: teslnb@sibupk.onmicrosoft.com).

Статья поступила 30 мая 2021 г.
Received May 30, 2021

To Reference:

Meikshan V.I., Teslya N.B. Vybor optimal'nogo razmeshcheniya dannykh pri ispol'zovanii oblachnoi infrastruktury [Selection of optimal data placement using a cloud infrastructure]. *Doklady Akademii nauk vysshei shkoly Rossiiskoi Federatsii = Proceedings of the Russian higher school Academy of sciences*, 2021, no. 2 (51), pp. 34–42. DOI: 10.17212/1727-2769-2021-2-34-42.