

Распределенные системы и технологии.

Облачные системы

Дмитрий Юрьевич Чалый
декан факультета ИВТ,
зав. кафедрой информационных и сетевых технологий



22 мая 2016 г.

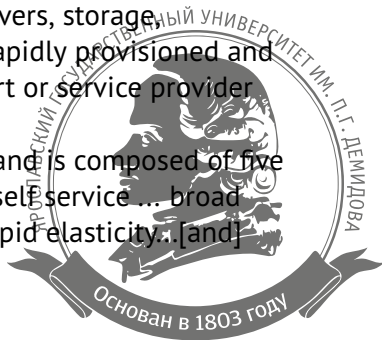
Что такое облако?

Аренда по требованию без интереса к физическим деталям.

NIST:

«Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

This cloud model promotes availability and is composed of five essential characteristics: ...on-demand, self-service ... broad network access ... resource pooling ... rapid elasticity...[and] measured service.»



- HaaS (Hardware as a Service): вы арендуете железо.
- IaaS (Infrastructure as a Service): вы арендуете виртуальную вычислительную инфраструктуру/сервис хранения данных.
- PaaS (Platform as a Service): вы арендуете инфраструктуру, интегрированную с программной платформой.
- SaaS (Software as a Service): вы арендуете доступ к сервису.



IaaS: Infrastructure as a Service

Вычислительная инфраструктура: серверы, хранилища данных, виртуальные сети.

Что делает провайдер:

- закупка/мониторинг/замена оборудования;
- виртуализация;
- инфраструктура.

За что отвечаете вы:

- администрирование (ОС, БД, приложений);
- разработка приложений;
- мониторинг, резервное копирование;
- установка рабочих мест пользователей.

Примеры:

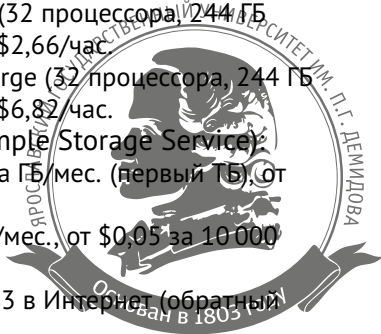
- Amazon Web Services: EC2 (виртуальные машины), S3 (хранилище данных), VPC (виртуальная сеть).
- Microsoft Azure: виртуальные машины, хранилище Azure, виртуальные сети.



IaaS на примере Amazon

Amazon Web Services: <https://aws.amazon.com>

- Виртуальные машины, сервис EC2 (Elastic Compute Cloud):
 - Общее назначение: t2.nano (1 процессор, 0,5 ГБ ОЗУ, Linux), \$0,0065/час.
 - Общее назначение: m3.2xlarge (8 процессоров, 30 ГБ ОЗУ, 2×80 ГБ SSD диск, Linux), \$0,532/час.
 - Оптимизация вычислений: c3.xlarge (32 процессора, 60 ГБ ОЗУ, 2×320 ГБ SSD диск, Linux), \$1,68/час.
 - Оптимизация памяти: r3.8xlarge (32 процессора, 244 ГБ ОЗУ, 2×320 ГБ SSD диск, Linux), \$2,66/час.
 - Оптимизация хранилища: i2.8xlarge (32 процессора, 244 ГБ ОЗУ, 8×800 ГБ SSD диск, Linux), \$6,82/час.
- Хранилище данных, сервис S3 (Simple Storage Service):
 - Стандартное хранилище: \$0,03 за ГБ/мес. (первый ТБ), от \$0,004 за 10 000 запросов.
 - Хранилище Glacier: \$0,007 за ГБ/мес., от \$0,05 за 10 000 запросов.
 - \$0,09 за ГБ трафика из Amazon S3 в Интернет (обратный трафик бесплатный).



PaaS: Platform as a Service

Инструменты и сервисы для разработки и внедрения приложений.

Что делает провайдер:

- закупка/мониторинг/замена оборудования;
- виртуализация;
- инфраструктура;
- администрирование (ОС, БД, средств разработки);

За что отвечаете вы:

- администрирование приложений;
- разработка приложений;
- мониторинг, резервное копирование;
- установка рабочих мест пользователей.

Примеры:

- Google App Engine: Python, Java, Go.
- Amazon Web Services: Amazon Lambda (не совсем так).
- Microsoft Azure: виртуальные машины, хранилище Azure, виртуальные сети.



Использование сервисов и приложений по требованию.
За что отвечаете вы:

- установка рабочих мест пользователей.

Примеры:

- Salesforce.
- Google Apps.
- ...и многое другое!



Новый вектор: data-intensive computing

- Computation-intensive computing: мало данных, но много вычислений ($P? = NP?$);
- Data-intensive computing: много данных, но вычисления относительно простые (линейные или квазилинейные алгоритмы):
 - в датацентрах хранится колоссальный объем данных;
 - перемещают не данные, а программы;
 - важно не время работы, а скорость ввода/вывода.



- Общедоступное, публичное облако (public cloud);
- Частное облако (private cloud).

Что проще: арендовать существующие сервисы или создать свое облако?



Приватное или публичное облако?

Пример датацентра:

- 5120 серверов класса IBM x3550 (1U, 2× Intel Dual-Core 2,66 Гц, 8ГБ ОЗУ);
- CapEx: \$101,8 млн. (в ценах 2007 года, сейчас выше);
- OpEx: \$5,9 млн./год;

Аренда таких же мощностей в Amazon сейчас:

- 5120 on-demand инстансов c4.xlarge (\$0,209 за час) круглый год:

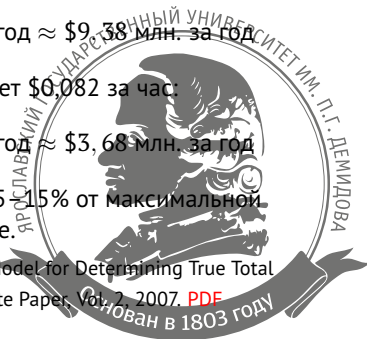
$$5120 \times \$0,209 \text{ за час} \times 8766 \text{ часов в год} \approx \$9,38 \text{ млн. за год}$$

- При резервировании на три года цена будет \$0,082 за час:

$$5120 \times \$0,082 \text{ за час} \times 8766 \text{ часов в год} \approx \$3,68 \text{ млн. за год}$$

- Загрузка среднего датацентра составляет 5-15% от максимальной производительности, оптимизация на споте.

J. Koomey, K. Brill, P. Turner, J. Stanley, B. Taylor. A Simple Model for Determining True Total Cost of Ownership for Data Centers // Uptime Institute White Paper, Vol. 2, 2007. [PDF](#)



- **C**ommon Infrastructure;
- **L**ocation-independence;
- **O**nline connectivity;
- **U**tility pricing;
- on-**D**emand Resources.

Weinman J. Clouconomics: A Rigorous Approach to Cloud Benefit Quantification
// J. of Software Technology, Vol. 14, № 4, Oct. 2011



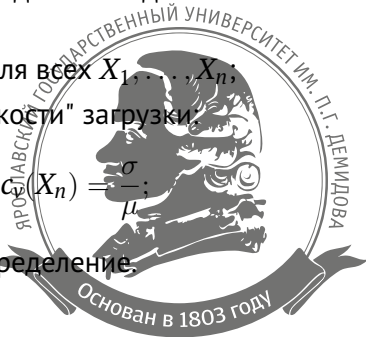
- 1 Экономика масштаба (economy of scale).
- 2 "Статистика масштаба" ("statistics of scale"): больше загрузка ресурсов облачного провайдера.

Пусть X_1, X_2, \dots, X_n — независимые случайные величины:

- среднеквадратичное отклонение σ одинаково для всех X_1, \dots, X_n ;
- мат. ожидание $\mu > 0$ и одинаково для всех X_1, \dots, X_n ;
- коэффициент вариации, мера "гладкости" загрузки:

$$c_v(X_1) = c_v(X_2) = \dots = c_v(X_n) = \frac{\sigma}{\mu};$$

- X_1, \dots, X_n могут иметь разное распределение.



Common Infrastructure

- 1 Экономика масштаба (economy of scale).
- 2 "Статистика масштаба" ("statistics of scale"): больше загрузка ресурсов облачного провайдера.

Пусть X_1, X_2, \dots, X_n – независимые случайные величины:



$$M(X_1 + X_2 + \dots + X_n) = n\mu;$$



$$D(X_1 + X_2 + \dots + X_n) = n\sigma^2;$$

- и среднеквадратическое отклонение совокупного спроса равно

$$\sqrt{n} \times \sigma;$$

- следовательно коэффициент вариации

$$\frac{\sqrt{n}\sigma}{n\mu} = \frac{\sigma}{\sqrt{n}\mu} = \frac{1}{\sqrt{n}} c_v(X_i)$$



Common Infrastructure

- 1 Экономика масштаба (economy of scale).
- 2 "Статистика масштаба" ("statistics of scale"): больше загрузка ресурсов облачного провайдера.

Пусть X_1, X_2, \dots, X_n – **позитивно коррелированные** случайные величины:



$$M(X_1 + X_2 + \dots + X_n) = n\mu;$$



$$D(X_1 + X_2 + \dots + X_n) = n^2 \sigma^2;$$

- таким образом,

$$c_v(X_1 + X_2 + \dots + X_n) = \frac{n\sigma}{n\mu} = c_v(X_1).$$



Задержки в распространении сигнала имеют свою цену:

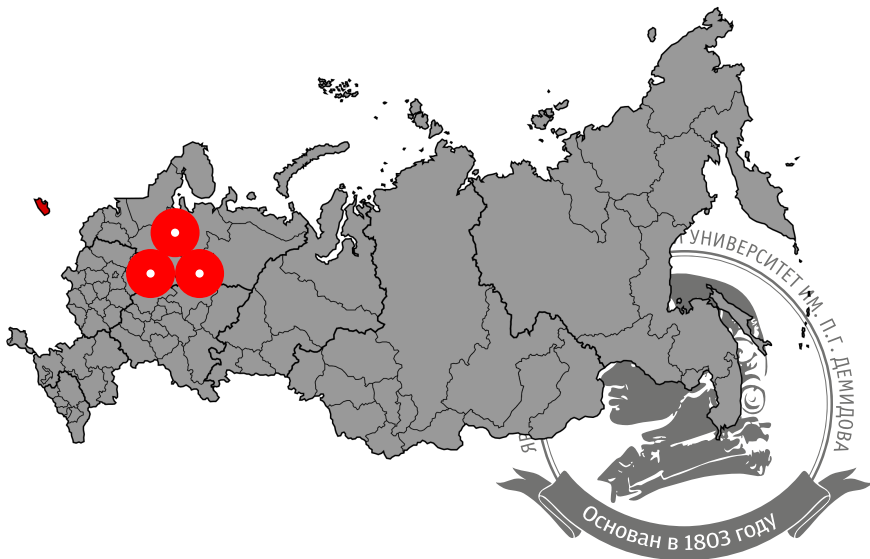
- Сигнал ограничен скоростью света (≈ 300 км/мс);
- **Например:**
 - 0,5 секунды задержки могут привести к потере 20% трафика у Google;
 - каждые 100 мс стоят Amazon 1% продаж.
- Современные веб-сервисы являются сложными.



Location Independence



Location Independence



Location Independence

Пусть

- r — радиус покрытия каждого датацентра;
- k — коэффициент, который зависит от выбора способа покрытия.

Тогда A — площадь покрытия равна

$$A = kn\pi r^2,$$

и если A фиксировано, то

$$r \propto \frac{1}{\sqrt{n}}.$$



- Если мы все наши приложения в одном облаке, то и задержки на координацию сервисов будут малы;
- мы можем использовать распределенные системы доставки контента для минимизации задержек;
- если мы используем сторонний сервис, который расположен в том же облаке, мы получаем минимальные задержки.



- Без онлайн-доступа ничего не работает.



Использование облака, как ресурса:

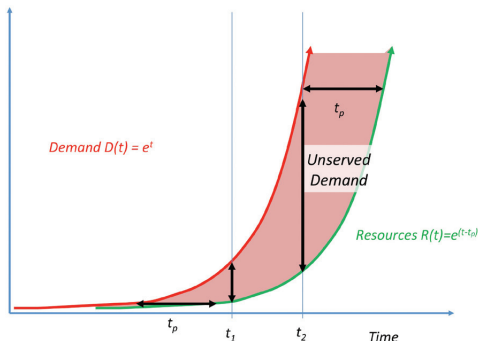
- Kia Rio (1350 руб./сутки аренда, 600 тыс. руб. покупка, в кредит 18 000 руб./мес. на 36 мес.);
- если цена облачного ресурса ниже, чем покупка своего, то всегда надо предпочесть облачный;
- использование вычислительных ресурсов может иметь пики.



On-demand Resources

Пусть $D(t)$ – спрос на ресурс, а $R(t)$ – имеющийся объем ресурсов:

- если $D'(t) = 0$, то нам надо закупить столько ресурсов, сколько требуется;
- если $D'(t) = c$, то надо организовать своевременную поставку ресурсов;
- если $D(t) = e^t$, а $R(t) = e^{t-k}$, то у нас проблемы:



- 1 IAAS для бизнеса по кирпичикам.
- 2 Экономическое обоснование миграции в облако для модели IaaS.

