

# Информационни Системи и Облачни Технологии

## Изготвили проекта:

Георги Христанов  
Иван Адамов  
Лъчезар Младенов

## Съдържание:

1. Същност на Cloud computing
2. Налагане и тенденции
3. Архитектури
  - 3.1. Видове архитектури
  - 3.2. Видове облаци
4. Processing и Организация
  - 4.1. Big Data
  - 4.2. Map-Reduce
  - 4.3. Hadoop
5. Сигурност
  - 5.1. Сигурността като облачна услуга
  - 5.2. Сигурност на информацията
6. Връзка с други технологии
  - 6.1. Мобилни интерактивни приложения
  - 6.2. Анализ на данни
  - 6.3. Приложения за настолни системи
7. Заключение и изводи

# 1. Същност на Cloud computing

При разработката на всяка дадена информационна система основен момент е частта с планирането на технологиите, които ще бъдат необходими за внедряването ѝ за използване. Почти винаги не е нужно да се използва пълния капацитет на дадена технология или нейното актуализиране, за да работи дадена система след време. За да се спести изразходването на ненужни ресурси, има Cloud услуги - услуги, които спестяват нуждата от лично притежание на технология, като предоставят за ползване частична или пълна нейна функционалност срещу определени условия. Това дава възможност да се стартира бизнес с минимална инвестиция и спестяването на излишни разходи. В днешно време почти всичко работи в cloud среда - онлайн хостинг за сайтове, хранилища за запазване на данни, използване на платформи за интегриране на приложения и дори услуги, които да се използват за взаимодействие с информационни системи. Cloud услугите спестяват доста ресурси и дават възможност на бизнеса да расте, докато достигне ниво, при което се нуждае от собствено място за съхранение на данни, платформи и технологии, но вече дори и растежа се осъществява в cloud-а. Излиза в пъти по-евтино поради липсата на нужда от персонал, на когото да се заплаща, за да поддържа системата, защото цялата поддръжка я осигурява доставчика на услуги. Доста популяризиран е бизнеса с доставянето на такива услуги за други бизнеси, които имат нужда от тях.

## 2. Налагне и тенденции

Пионери са Amazon, сега всяка голяма компания има собствен cloud - Microsoft(Azure), IBM(Cloud), Apple(iCloud), Oracle(Public Cloud), Google(App Engine), Amazon(Web Services) и много други. Преди около 10-на години Amazon тръгва като онлайн магазин за книги, превръща се в най-големия онлайн магазин за книги и в днешно време е магазин за всичко(дори плодове и зеленчуци). За да се достигне този ръст, е била нужна технология за тези онлайн магазини, която изисква да има някаква технология, но не цял сървър. Поради огромния си успех решават да продават такива услуги на клиенти - достъп до технологии без нужда от закупуване на собствени сървъри. Към днешна дата, cloud computing-а е дотолкова внедрен, че Apple, например, синхронизира и качва всички файлове на потребителите на нейните смартфони iPhone в собствения си cloud. Тенденцията в сферата е големите фирми да предоставят cloud услуги, които по-малките фирми да използват и предоставят.

В софтуерната сфера, все повече нараства търсенето на хората, които разбират cloud архитектурите и могат да работят в cloud среда. Съответно нарастват и изискванията към софтуерните инженери да разбират начина, по който работи cloud-а, а именно виртуализацията на машини с цел достъпа до определен брой ресурси от различни места. Това води до драстично увеличение в броя на търсенето на подобни системни програмисти. Поради трудността си, много по-лесно се намират кадри за създаване на информационна система, която работи с cloud услуги, отколкото система, която развива такива.

## 3. Архитектури

### 3.1. Видове архитектури

#### 3.1.1. Публични

Това са места, където всяко нещо е достъпно до всеки. Повечето информационни системи, които са бизнес насочени са такива.

#### 3.1.2. Вътрешни

Това са места, където само определени организации имат достъп до ресурсите. Тук вече може да са и сървъри, и софтуер. Обикновено от такива имат нужда банки и правителства или вътрешни фирмени системи, но особено при първите, това не е навлязло поради характера на организацията и страха от пробив в сигурността.

#### 3.1.3. Хибридни

Сбор от публични и вътрешни части на системата. Пример е система на фирма, която има част за разпределяне на задачи за служителите и онлайн каталог на продуктите.

### 3.2. Видове облаци

#### 3.2.1. IaaS(Infrastructure as a Service)

Предоставят за ползване само хардуер(наемат се виртуални машини), използва се предимно за сървъри, където да се пазят данни. От потребителя зависи качването на операционна система. По всяко време може да се разшири потреблението. Най-трудна е разработката в тези облаци и е най-рядко срещана. То се осъществява чрез даване на повече виртуални машини, защото не може да се задели повече памет, след като вече е достигнат даден лимит.

#### 3.2.2. PaaS(Platform as a Service)

Предоставят за ползване операционна система(предимно Linux), платформи(JRM, .NET), фреймуъркци(Spring, Microsoft MVC framework), езици за сървърна част(Java, C#, PHP, JavaScript(Node.js), Ruby, Python), СУБД-та(релационни, нерелационни и blob хранилища), уеб услуги и API-та. Дават възможност за създаване на приложения за крайни потребители. Такива са Heroku Cloud, Parse.com, MongoLab и други. Обикновено са чатове, системи, базирани на известия, CMS, системи за лог-ване, CDN-и.

#### 3.2.3. SaaS(Software as a Service)

Предоставят за ползване набор от софтуерни приложения - програми. Предназначени са за крайни потребители. За използването им не са нужни

задълбочени познания по информационни технологии. Примери: Google Docs, Dropbox, Wordpress. Тук разработката е най-лесна и съответно най-разпространена.

## 4. Map-Reduce

### 4.1. Big Data

Big Data е широко понятие за огромни обеми от данни (терабайти / петабайти). При такъв обем, традиционната обработка на данни не само е неадекватна, но и невъзможна за една машина. Типични компоненти на големи big data системи са:

- Дистрибутирани бази данни (Cassandra, Hbase, Hive)
- Дистрибутирани фреймуърци за обработка (Apache Hadoop)
- Дистрибутирани системи за обработка (Map-Reduce)
- Дистрибутирани файлови системи (Hadoop Distributed File System - HDFS)

### 4.2. Map-Reduce

#### 4.2.1. Какво е Map-Reduce?

Най-просто казано, това е изчислителна концепция, която когато се имплементира, дава възможност за обработка на големи обеми от данни. Map-reduce ползва паралелни изчисления върху големи клъстери (хиляди node-ове). Разчита се на дистрибутирана инфраструктура – например Apache Hadoop. Входните и изходните данни са разположени на дистрибутирана файлова система, даваща възможност за разбиването на файл на части и записването им на отделни места. Фреймуърците се грижат за поддръждането, пускането и следенето на отделните задачи, както и повторното пускане на неуспешно изпълнените задачи.

#### 4.2.2. Как работи?

При Map-reduce се изпълняват следните стъпки:

- 1) Входните данни се разделят на независими парчета.
- 2) Всяко едно от парчетата минава през “Map” функцията, която групира данните в чифтове ключ–стойност, като това става паралелно.
- 3) “Reduce” функцията произвежда изходният резултат.

Изключително важно е „Map“ функцията да раздробява на множество парчетата, като в противен случай се губи паралелността. Това е факт, тъй като всички данни които имат еднакъв ключ се обработват от един node. Съответно, ако „Map“ функцията винаги слага един и същ ключ, то нямаме паралелност.

„Map“-а също така трябва да разделя на логически уникални групи. Всяка една от групите в последствие се обработва от „Reduce“ функцията. По принцип, основната обработка е в „Reduce“-а, тъй като той взема всички ключове, обработва ги по някакъв начин и връща съответен резултат.

Процедурата може да бъде пускана по няколко пъти. Например: Ако разполагаме с всички градове в България и тяхното население, но искаме да изведем само тези в северна България с над 100 000 души, то първо трябва да изведем градовете в северна България, а след това да видим кои от тях разполагат с население над 100 000.

## 4.3. Hadoop

### 4.3.1. Какво е Hadoop?

Изключително рядко в компютърният свят, се случва всички софтуерни гиганти, включително Microsoft, IBM и Oracle, да кажат - „Това е!“. Такъв обаче е случаят с дистрибутираният фреймуърк за обработка – Hadoop. Истината е, че всяка една от гореспоменатите компании е работила по собствени проекти, но в крайна сметка те са били изоставени и изместени от Hadoop.

Hadoop не е една технология – тя е фамилия от технологии, които се ползват заедно. Два основни елемента предоставени „out of the box“ са:

- Дистрибутирана файлова система – това представлява инфраструктура, в която може да бъдат записвани големи файлове, както и голям брой файлове. Hadoop автоматично решава кое къде да бъде сложено, като се постига redundancy – едно нещо се записва на няколко node-а, за да се предотврати загуба на данни.
- Map-Reduce

### 4.3.2. Hadoop Ecosystem

- 1) Файлова система:
  - Redundant (3 копия)
  - Файловете се разделят на блокове от по 64/128mb
  - Може да се скалира до хиляди node-ове
- 2) MapReduce API
- 3) Hadoop библиотеки
  - Pig, Hive, HBase и др.

## 5. Сигурност

### 5.1. Сигурността като облачна услуга

Сигурността като услуга (security as a service), известна като съкращението SecaaS (или SaaS) е модел на облачна технология, при който доставчикът осигурява управлявани услуги за сигурност на потребителите в Интернет. SecaaS е базиран на облачната технология, софтуер като услуга SaaS (software as a service), но се ограничава до специализирани услуги за информационна сигурност на данните и приложенията на потребителите. Също така улеснява осигуряването на управлявани услуги за сигурност в облака, като носи редица ползи на организациите:

Намаляване на разходите за организациите - SecaaS решенията се предоставят на потребителите под наем или на база закупен лиценз;

Лесно управление и аутсорсинг на административни функции – доставчиците на услуги предоставят общо управление на облачни услуги за сигурност, обща политика на сигурност и общо администриране;

Адаптивност и мащабируемост - услугите включват пакет за сигурност Security Suite под формата на anti-virus/malware/spyware;

Непрекъснато обновяване - SecaaS услуги гарантират, че софтуерът за сигурност се поддържа с най-актуалната антивирусна защита.

### 5.2. Сигурност на информацията при облачните услуги

Важно значение при ползването на облачни услуги има осигуряването на защита на информацията, което е свързано с идентифицирането на заплахите и набеязване на възможните решения за осигуряване на информационна сигурност на данните и приложенията на потребителите в облачна среда, насочване на усилията за преодоляването им в конкретни направления. Според редица организации част от основните заплахи за информационната сигурност на потребителите в облачна среда са свързани с: оторизацията на потребителя, шифриране на данните, физическата безопасност при работата на оборудването, „вреден“ трафик и др.

При избора на правилен доставчик на облачни услуги е необходимо организациите да се ориентират от наличието на сертификати, данни от трети лица, информация за използваните механизми на защита и др. В помощ биха могли да бъдат ръководствата за безопасност на информацията при ползване на облачни услуги на организации като: Cloud Security Alliance (CSA), Cloud Computing Information Assurance Framework (ENISA) и др.

### 5.2.1. Оторизация на потребителя

При ползване на облачни услуги най-широко разпространен способ за оторизация на потребителите е защита с парола. Доставчиците осигуряват на потребителите по-висока надеждност, като прилагат и средства като сертификати и електронни подписи. Прилагат се стандартите LDAP (Lightweight Directory Access Protocol), SAML (Security Assertion Markup Language) и др., които осигуряват взаимодействие на доставчика със системата за идентификация на потребителите и определяне на даваните пълномощия. Благодарение на това доставчикът разполага с актуална информация за оторизираните потребители.

### 5.2.2. Шифриране на данните

Дори данните да са шифрирани, то при тяхната обработка те трябва да бъдат разшифровани. Целта на Private Core е да отстрани необходимостта от разшифроване на данните или в крайна сметка да изпълни тази операция по най-безопасен начин. Private Core разработва собствен безопасен хипервизор, който позволява на клиента да управлява защитени виртуални машини, публична инфраструктура, използвайки технологията Intel TXT (Trusted Execution Technology - технология на доверения достъп на изпълнение). Безопасността се осигурява чрез шифриране на всички данни в паметта на физическата машина, като се оставя много малка сигнатура с възможност за атака. Данните се разшифроват и съхраняват в нешифрован вид само в кеш-паметта на процесора, който поддържа технологията на доверените изчисления. Така ако апаратната мощност на публичната платформа поддържа такива технологии, то организациите разполагащи своите данни в облака не зависят повече от мерките по безопасност на физическите сървъри – всички операции се изпълняват на виртуалните машини в зашифрован вид. Подобна схема на организация на безопасност на сървърите има и торент-трекера The Pirate Bay. Допълнително преимущество е това, че изпълнението на жизнено важни операции с помощта на виртуални машини може да бъде пренесено на няколко отделни физически сървъри в различни райони на света. В случай на отказ на 1 от сървърите работата може да бъде пренесена бързо на резервен. По такъв начин системата става достатъчно устойчива към хакерски атаки. Въпреки това както и във всяка друга криптографска система сигурността зависи от качеството на кода на хипервизора, както и от липсата на грешки в системата за управление на крипто-ключовете. Вреден трафик Разполагането на услугите при доставчика усложнява задачата по осигуряване на безопасност при трафика между приложенията и виртуалната машина. Ако хакерите, намерят уязвимост в един виртуален компютър, могат с лекота да получат достъп и до останалите. Отсъствието на контрол при трафика може да доведе до изтичане на данни, заразяване с вируси и др. Възможен е вариант на разположение на всички виртуални машини, принадлежащи на 1 собственик на 1 физически сървър, но такава стъпка намалява ефективността от използването на ресурсите на сървъра. Много разработчици на антивирусни системи предлагат свои версии продукти за защита на виртуалните среди.

## 6. Връзка с други технологии

### 6.1. Мобилни интерактивни приложения

Това са приложения, реагиращи на информация, постъпваща от потребители или сензори, в режим на реално време, а също така приложения, които трябва да бъдат постоянно достъпни и използват големи масиви от данни, които е най-удобно да се съхраняват в изчислителен център. Добър пример за интерактивни приложения са услуги, съчетаващи 2 или повече източника на данни и други услуги. Облачната изчислителна среда предоставя отлична архитектура за такива приложения, особено в случаите, когато се ползват постоянно включени към мрежата мобилни устройства.

### 6.2. Анализ на данни

Извличането на знания от данни (на английски: *Data mining*) е процес на откриване на смислени корелации, зависимости, повтарящи се образци (на английски: *patterns*), тенденции и аномалии в големи масиви от данни, съхранявани в складове чрез използване на техники и алгоритми от областта на машинното обучение, разпознаването на образи, статистиката, невронните мрежи и визуализацията на данни. За разлика от извличането на данни посредством заявки към база данни, извличането на знания от данните генерира формулирана, не очевидна, предварително известна, потенциално полезна информация.

Това е една дейност, която изисква много ресурси и може удобно да бъде изпълнена в "облак".

### 6.3. Приложения за настолни системи

Системи, изискващи много ресурси. Такива са например приложенията за обработка на графика и тримерна анимация. Всякакви такива приложения може да не се инсталират на настолна система, а да се пускат за изпълнение в частен облак, работещ в крупен център за обработка на данни.



## 7. Заключение и изводи

Основното нещо, което можем да заключим е, че cloud технологиите променят бизнеса. Те спестяват излишни ресурси като време, пари, човешки ресурси. Осигуряват оптимизация на поддръжката, backup на даните и редица други неща, които, ако трябва разработчикът на информационната система да ги осигури, не биха били разработени и поддържани толкова качествено. Cloud услугите дават възможности на разработчиците на системи да вършат по-лесно работата си и да се фокусират върху конкретната задача и максималната ефективност на продукта, който създават. Самата архитектура на cloud-a позволява оптимално използване на ресурсите.

Като изводи можем да обособим следните:

- 1) Бъдещето е cloud - глобалната мрежа върви към преминаване в общ cloud. Това ще доведе до централизиране на услугите, подобряването им, но и по-малка вариативност, подобно на .NET платформата.
- 2) Бизнесът ще използва cloud - просто поради по-ниските разходи и лесното достигане на стойност на една система.
- 3) Разработката и подобряването на cloud ще бъде приоритет на големите компании и ще повиши цената на софтуерните инженери в тази област.
- 4) Хардуерът ще спре да се разширява така драстично поради липсата на нужда - по-лесно е да се създаде още една виртуална машина, отколкото да се сменя целия сървър.
- 5) Общият cloud ще означава повече информация и повече търсене, което ще бъде основният приоритет на големите компании - изкуствен интелект за оптимизиране на търсенето.