

СЪВРЕМЕННА ИНФРАСТРУКТУРА ЗА ЗАПАЗВАНЕ НА ДАННИ¹⁾

Асен Захариев

Висше училище по телекомуникации и пощи

Резюме. Текстът представя факторите, които се взимат под влияние за изграждането на инфраструктура за запазване и съхранение на информация. Дадена е основна информация, която ще доведе до опазването на бизнеса в случай на бедствие.

Keywords: backup, storage, archive, information, disaster

Въведение. За да се ориентира в околния свят и да може да определя действията си, човек непрекъснато се нуждае от информация – прогноза за времето, програмата на телевизията, цените на стоките, работното време на магазините, седмичната програма, печелившите числа от тотото. Информацията придобива първостепенна роля във всички сфери на живота – управление, образование, научни изследвания, здравеопазване, администрация, развлечение и др. Думата „информация“ произлиза от латинската дума *informacio*. Това прави информацията едно от най-ценните неща в днешни дни.

Запазването, или бекъп (от английски – *backup*), е термин, използван в ИТ средите, отнасящ се до запазване и архивиране на информация/данни. Както и до тяхното възстановяване в случай на някакво „бедствие“. Това може да е повреда на файл, някакъв вид природно бедствие или най-срещаното – човешка грешка. Много от нас не осъзнават нуждата от запазване/архивиране, докато не се случи най-лошото.

Тук също трябва да изясним каква е разликата между запазване (*Backup*) и архивиране (*Archive*). Запазването (*Backup*) е действие, което цели бързото възстановяване на изгубените данни. Тези запазени копия се държат на една ръка разстояние. Архивът, от своя страна, са данни, които не са използвани толкова често като запазените. Те се държат в отдалечени съоръжения, специализирани за тази цел.

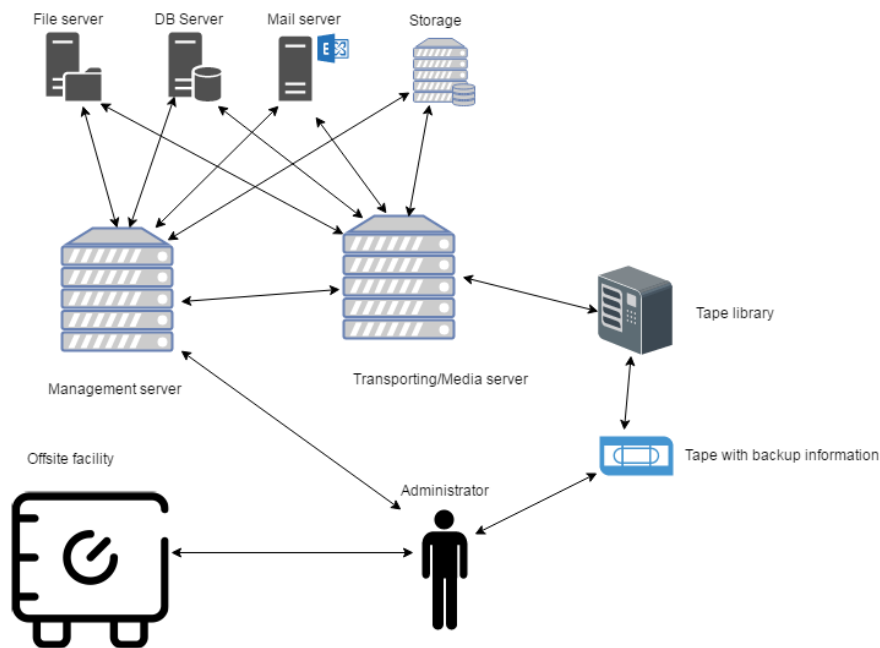
Преди да бъде имплементирано решение за съхранение на данните, трябва да се зададат няколко основни въпроса: коя е важната за нас информация; колко да бъде далечна във времето точката на възстановяване; имаме ли план

за сигурност; каква ще е протекцията на версиите; каква е сегашната инфраструктура и каква – нуждата от допълнителна техника; и не на последно място – колко финансови средства можем да отделим за изпълнението на този план.

Изграждане на съвременна инфраструктура за запазване на данни

Съвременната инфраструктура за запазване на данни може да се обобщи до следната схема, дадена на фиг. 1 по-долу. Необходимите компоненти, управлявани от администратора (Administrator) на фигурата, са:

- сървъри за данни, които ще бъдат запазвани (File server, DB Server, Mail Server, Storage);
- сървър за управление на процеса на запазване на данните (Management server);
- транспортиращ сървър (Transporting/Media server);
- лентово устройство за съхранение (Tape library) и касета (Tape with backup information);
- отдалечена сграда, в която се съхраняват архивите/данните – обикновено намираща се далеч от седалището на фирмата. (offsite facility).



Фигура 1. Общ вид на съставните компоненти на съвременната инфраструктура за запазване на данни

Важна информация може да бъде всичко! Преди ИТ инженерът да започне с изпълнението на зададената му задача, е задължително той да се запознае с дейността на организацията, за която ще работи. Трябва да се съобразят фактори, като дали информацията може да се изнася извън държавата или Европейския съюз. Дали е задължена от съответните законодателни органи да пази за определен брой години тази информация. След като изпълним това действие, ние си отговаряме на първия въпрос.

За пример ще вземем адвокатска кантора. За такова предприятие е важно освен да държи досие за всеки служител и клиенти, така и своите документи за минали, сегашни и бъдещи дела. Информацията ще е от тип файлове – текстове, сканирани документи и няколко тип „бази данни“. Други фирма, която можем да ползваме за пример, е производител на храни. При тях освен споменатото по-горе имаме наличност и специално създадени софтуери за бизнес анализ (маркетинг), както и такива, които управляват производствените ленти. Там освен основните бази от данни (SQL и Oracle) имаме и немалко налични виртуални машини. А в някои случаи дори и на cloud услуги. След като сме наясно с този фактор, вече можем да преминем към следващата точка.

Колко да бъде далечна точката на възстановяване и времето? Recovery point objective и Recovery time objective – най-конструктивните фактори са тези два. Върху тях ще положим основите на нашата инфраструктура. Нека да ги разгледаме поотделно. Точка на възстановяването е отражението вследствие възстановяването на изгубените поради бедствие услуги или данни. Желателно е това време да е възможно най-близко до момента преди загубата. Ако използваме примера с адвокатската кантора – миналите дела можем да зачислим към архивирането. При тях няма нови промени, данните не се променят и датата, от която можем да ги възстановим, може да е далечна. Например дата, след която делото е приключило. Но ако това е активно дело, тази точка на възстановяване може да бъде от няколко часа до ден. Другият ни пример е програмата, управляваща поточната линия. При нея няма изменения. Веднъж загубена, тази програмата може да се възстанови напълно функционираща от момента, в който е инсталирана.

Точка за време на възстановяването е услуга. Буквално се разбира време, за което дадените загуби ще бъдат възстановени. Дейността от нашия пример с адвокатската кантора ще иска по-голямо време за възстановяване на минали дела. Докато програмата, управляваща производствената линия, трябва да се възстанови възможно най-бързо поради финансовите загуби, породени от липсата на производство.

Освен това за да бъдат запазени и/или архивирани данните, е важно да се вземе под внимание и сигурността на вече направените копия. Трябва да се вземе под внимание дали обслужваме отдалечени офиси, при кои-

то процесът протича през незащитена интернет връзка. Избраният от нас софтуер трябва да позволява криптиране на данните. Предпочитан метод за криптиране е AES-256 (Advanced Encryption Standard с 256 битов ключ). Това е първият и единственият публично достъпен шифроващ метод, одобрен от Националната агенция за сигурност на САЩ. Тук данните се криптират, преди да бъдат изпратени по мрежата, а когато се доставят, биват декриптирани.

Друг метод е тези данни да бъдат допълнително криптирани, докато биват записвани на съответните устройства, който ще разгледаме малко по-късно. Криптирането става, докато информацията бива записвана върху носител, и само и единствено носител с дешифриращия ключ може да ги прочете. Има и физическа сигурност на носителите. Трябва да се знае кои хора упълномощаваме да управляват данните адекватно и отговорно. Устройствата, които записват и съхраняват данните, трябва да са изолирани от останала част от сградата, както и от останалата ИТ техника. Това изискване е задължително поради важността на информацията. Веднъж изнесени от сградата на клиента, те трябва да бъдат транспортирани според норми, които предпазват както от природни бедствия (пожар, електрическо/магнитно влияние), така и от злонамерени посегателства (кражба). Копията/архивите трябва да се държат на достатъчно, но не много отдалечено място. Фактори, като огнеупорни сейфове/куфари, влажност и адекватен одит, също трябва да се вземат под внимание.

Стратегия за запазване на данни. Версия на даденото копие може да се определи като дата или ден от седмицата, когато запазването на данните се е случило. Нека да го изясним с прост пример. Имаме сървър с два диска. Единият ни диск е системният, още познат ни като английската буква С или сисвол (от английски – sysvol, system volume). На другия диск имаме записани критични данни, нека за по-кратко да го наречем Дата. Искаме да имаме повече версии на Дата заради неговата роля. За този Дата диск ще правим всеки ден бекъп. А **сисвол** ще бъде запазван само веднъж в седмицата поради малкото промени, които се правят на него, и лесното му ръчно изграждане от началото. При **Дата** ще имаме следните версии – *ДатаПонеделник*, *ДатаВторник* ... *ДатаНеделя*. А при сисвол ще имаме само *СисволСъбота*. Всяка версия има и своята протекция. Протекцията е период от време, за което дадената версия се пази от презаписване. След тази дата на протекция дадената версия може да бъде изтрита/презаписана. Можем да пазим версиите от понеделник до петък за 14 дни, а съботно/неделните – за 24 дни. Така, ако следваме горния пример:

За *ДатаПонеделник* с дата 1 март ще имат протекция до 15 март. След това тази версия може да бъде изтрита или презаписана. Съответно *ДатаВторник* ще бъде възможно да бъде изтрита на 16 март, и така нататък. При следващия

понеделник на 8 март ще имаме две версии – тази от 1 и от 8 март. Докато СисволСъбота и ДатаСъбота от 6 март ще бъдат с протекция 4 седмици. Неофициална статистика показва, че 90% от възстановяването на дадена информация се случва до 10 дни, след като последно е работено по данните. Най-често потребителите виждат, че нещо е повредено или са направили грешка, в понеделник или петък.

След като сме поставили нашите основи на стратегията, е време да се премине към практическата част от заданието. Тук е моментът да помислим и от инфраструктурна гледна точка. Ако тепърва се изгражда инфраструктурата – това е най-лесният и удобен вариант. При тази ситуация можем лесно да пресметнем какво ни е нужно, за да постигнем нашата цел. Примерно каква връзка ни трябва между устройствата за записване и сървърите с информацията, която искаме да запазим/архивираме. Кое поколение и модел устройства ни трябва за нашите нужди, както и много други фактори.

Предизвикателството е, когато вече имаме дадената инфраструктура и трябва да се съобразим с нейното състояние, за да вкараме нашето решение за запазване/архивиране на информация. Първо определяме така наречения *bottleneck*. Това е факторът, който може буквално да спъне нашата работа. Дали ще е капацитетът на мрежата, скоростта на запис/четене на устройствата, или ще е чисто или просто фирмена политика, ние трябва да сме наясно с тези фактори.

Вече разполагащи с цялостната визия за бизнеса на клиента, неговите желания и състоянието на инфраструктурата можем да започнем работа. Тук ще отчетени и финансовата гледна точка, докато проучваме нужната ни техника. Ще определим така наречения *backup window*. Това е диапазонът, в който ни е позволено да извършваме действието запазване/архивиране. Това време обичайно е след като е приключила дейността на клиента – най-често „прозорец“ от 19:00 ч. до 08:00 ч. Но ако даденият бизнес работи в няколко часови зони и имаме по-малко време, да кажем 4 или 6 часа? И за двата случая имаме предоставени два типа устройства за записване. Те могат да бъдат лентово или дисково базирани носители. Като, разбира се, могат да се комбинират, за да предоставят по-добри резултати.

Лентовите устройства са по-известни с името касети. Те представляват намотани на ролка магнитни ленти. Макар и принципът да е остарял, предимството при тях е липсата на каквато и да е електроника, водеща до неизправности. При правилно експлоатиране и запазване такава касета има възможност да стигне до десетки години годност. Друг важен фактор, който трябва да добавим, е цената. На най-новите и масово използвани касети от шесто поколение можем да поберем 6.25 терабайта (при компресия от 2,5:1) обем от информация, и то за около 40 долара на касета. Ако погледнем твърдите дискове, за същите пари можем да се

сдобием само с 1 терабайт. За съжаление, финансовата гледна точка се променя, когато трябва да се закупят устройствата за записване и четене на тези касети. Така едно от тези устройства в самостоятелна конфигурация може да излезе около 2000 долара. При по-голяма конфигурация, наричана библиотека, цената може да достигне до 40 000 долара и повече. Устройства от този тип могат да бъдат използвани при един добър диапазон от свободно време, в което можем да извършим запазването на информация, както и след това използваните касети да бъдат изпратени на специализирано място за съхранение. Така се подsigуряваме, че дори и ако офисът/дата центърът или мястото, на което оригинално е съхранявана информация, бъде разрушено, ще можем да подновим действието на бизнеса в рамките на дни, дори и на часове.

Дисково базираните устройства са другата крайност. Там имаме голямо количество електроника, която може да се повреди по една или друга причина. Най-големият фактор, който надделява пред лентовите носители, е бързината на запис и четене и отстраняването на механичното движение. Тук записът на информация се извършва почти мигновено върху диска, независимо дали е твърд диск (HHD), или полупроводниково дисково устройство (SSD). Този тип устройства постигат още по-ефективно действие, използвайки RAID (Redundant Array of Independent Disks) технологията. Това е масив от независими дискове, които работят като един, позволявайки сигурност и бързо действие за записване и четене на информация. Придружени със специален софтуер за дедупликация и операционна система, тези дисково базирани устройства могат да запазят стотици гигабайтове информация за броени часове.

Хибридният вариант или сценарий, при който може да се ползват и двете технологии, най-вече се среща при гореспоменатия малък времеви диапазон. Тъй като дисковите устройства ни предлагат по-бърз запис и четене, първоначално можем да запазим информацията на тях, след което да бъде преместена върху лентови носители дори извън зададеното ни време. За този вариант ни е нужна самостоятелна мрежа между двете устройства. Така оставяме незасегнатата мрежата, която се ползва през работно време.

Сравнителен анализ на съвременните комерсиални приложения за запазване на данни

След като достигнем финансовата страна на инфраструктурата и платформите за сравнение на данни, трябва да вземем всичките гореизброени фактори и да ги приложим в зависимост от нашите нужди. Благодарение на бързо развиващия се ИТ сектор можем да използваме програми с отворен код или някои от корпоративните решения. Тук трябва да калкулираме различни многобройни фактори за създаването на такава инфраструктура.



Фигура 3.1. Графика с четирите квадранта на Гартнер, показваща лидерите, претендентите, визионерите и играчите в специализирана ниша от сектора „Запазване на информация“

Като заключение можем да кажем, че за да изградим такава инфраструктура, трябва да се поставим на мястото на бизнеса, да разберем условията и нуждите му. Трябва да вникнем в детайли в това, което ни се предоставя от модерния ИТ свят, и да се възползваме максимално от дадените ни похвати и инструменти.

Приложение 1. Сравнение на характеристиките на основните доставчици и техните платформи за запазване на данни

Компания и продукт	Категория	Плюсове	Минуси
Actifio/Actifio	Визия за бъдещето	Позволява по-бързо запазване и възстановяване, отколкото традиционните методи. Ползван от много корпорации, този продукт позволява по-добро структуриране на архивирани данни. Actifio предоставя обширен списък на продуктите, с които работи.	Възможност за по-голям разход от гледна точка на заемано място. Това се дължи на липсата на механизъм за компресиране на данните, които се намират върху физическите устройства За да се достигне максимално използване на този софтуер, потребителите трябва да взаимодействат с други софтуерни администратори. Actifio не предлага възстановяването на операционни системи, без те да са предварително инсталирани.
Arcserve/ Unified Data Protection	Специализирана ниша	Освен основните традиционни методи за запазване на данни Arcserve започна да предлага така наречения snapshot метод. Първоначално е интегриран с NetApp и Nimble устройства. DRaaS (възстановяване от бедствие) функцията на Arcserve предлага предимство за готовност на виртуални машини, които могат да заменят оригиналните при тоталната им загуба. Гъвкавата възможност да заплатите само за това, което искате да ползвате. Лицензирането може да бъде въз основа на нужните функции, използван брой процесори, заето място или използвани операционни системи.	Все още не е доказано решение за корпорации. Предимно се среща в по-малки или средни организации. В момента се разработва интеграция с лентови устройства, които не са включени в досегашните версии на продукта. Липсата на автоматично разпределяне на ресурсите за запазване или възстановяване на информацията, което води до повече ръчни настройки.
Commvault/ Commvault	Лидери	Commvault предлага мащабност – единичен пакет за дадена специфика до цялостно решение дори и за най-големите корпорации. По-голяма гъвкавост при ценообразуването. Предлага съвместимост с най-големите производители на дискови устройства за съхранение.	Първоначалното стартиране, конфигуриране, както и работата с Commvault може да бъдат не толкова интуитивни в сравнение с други конкурентни продукти.

		Commvault може лесно да се интегрира с водещите на пазара доставчици на публични облаци – Microsoft Azure и Amazon Web Services	Препоръчително е преди работа да се изкарат специализирани подготвителни курсове. В EMEA и APAC регионите нивото на обслужване може да бъде под стандартта заради липсата на партньори в тези региони.
EMC/ Data Protection Suite (combination of Avamar and Networker)	Лидери	Увеличаването на обслужваните продукти от DPS допринася за неговото развитие. Във версия 9 на Networker са включени голям брой подобрения относно скоростта, с която се извършва запазването и възстановяването на данни, както и на по-лесно управление от интерфейса.	Остава да бъде проблем в портфолиото на EMC многобройните интерфейси за управление, както и тяхната поддръжка. Трябва да се вземе под внимание, че без използването на дисковото устройство на EMC – Data Domain, производителността на Networker и Avamar намалява значително. Голям въпрос е какво ще стане с предлаганите продажби и поддръжката от страна на EMC покрай придобиването им от Dell.
HPE/ Data Protector	Претенденти	Предлага голям брой поддръжани от софтуера продукти с лесен за конфигуриране и използване интерфейс. Backup Navigator and Storage Optimizer предлагат анализ и оценка на вашата инфраструктура.	Не е предназначен за големи инфраструктурни среди. След разделянето на Hewlett-Packard на две компании софтуерът е без ясна визия за бъдещето, което води до съмнения колко добре ще бъде предлаган на пазара и поддържан в бъдеще.
IBM/ Spectrum Protect	Лидери	Поддържа голям диапазон от платформи (OS, бази, дискови устройства) заедно със способността да запазва информация върху дискови или лентови устройства, както и облачни доставчици. Подобрявайки своите алгоритми за дедупликация, както и предоставянето на метода incremental-forever, IBM съчетава минимални	Труден за навигиране и работа интерфейс. Ниската осведоменост за функциите и подобренията кара много клиенти да прибягват до по-стари версии на продукта. Някои от новите функции и подобрения изискват повишаване на верси-

		изисквания откъм хардуер и разхода по имплементирането в големи инфраструктурни среди. Лесно до 5 петабайта от информация могат да бъдат предпазени от една инстанция/сървър на Spectrum Protect.	ята на сегашната инфраструктура. Например на по-висока версия на база от данни, което е обвързано с допълнително заплащане и поддръжка, а и невинаги тези нови версии са съвместими със софтуера, който ги ползва.
Unitrends/Unitrends	Специализирана ниша	Unitrends предлага облачно запазване, както и „възстановяване от бедствие“. Доставчици като Amazon Web Services, Microsoft Azure, Google и OpenStack ползват услугите на този вендор. Подобрен алгоритъм за дедупликация.	Предоставените дискови устройства за запазване на данните са с малък капацитет – по-малко от 50 терабайта. Съвместимостта с дискови устройства се ограничава само до EMC VNX и NetApp. Използването на Unitrends, като част от облачно решение, е по-предпочитано за архивиране на данни, отколкото за краткосрочното им запазване.
Veeam/ Veeam	Лидери	Специализиран и надежден софтуер за запазване на VMware и Hyper-V решения. Без допълнително инсталирани приложения предоставя частично и лесно възстановяване.	Липса за обслужване на физически, както и облачни решения. Изисква се от администраторите да имат много странични знания, за да предоставят оптимизирана инфраструктура. Липсата на съвместимост с облачни доставчици.
Veritas Technology/ NetBackup	Лидери	Широк диапазон на обслужвани операционни системи и приложения. Предлаганата опция Accelerator за оптимизиране на времето за запазване, както и OpenStorage Technology (OST) за по-лесно управление на дисковите устройства прави NetBackup да се отличава пред конкуренцията. Централизиран интерфейс, позволяващ управлението и разгъването на много от свойствата на софтуера.	Все още в разработка е интегрирането с доставчици на облачни услуги. Вътрешен одит на софтуера, инициализиран още по времето на Symantec и продължаващ и до днес от Veritas, кара клиентите да търсят други решения за своите нужди.

NOTES / БЕЛЕЖКИ

1. Авторът е студент II курс, специалност „Компютърно администриране на софтуерни приложения“.

Консултант при разработване на настоящия текст е гл. ас. д-р инж. Страхил Соколов, катедра „Информационни технологии“, Факултет по телекомуникации и мениджмънт, Висше училище по телекомуникации и пощи.
E-mail: strahil.sokolov@gmail.com

2. How to build a storage and backup strategy for your small business: <http://www.cio.com/article/2378019/small-business/how-to-build-a-storage-and-backup-strategy-for-your-small-business.html>
3. Backup infrastructure presents services prospects: <http://searchitchannel.techtarget.com/tip/Backup-infrastructure-presents-services-prospects>
4. Five ways to create a more efficient backup infrastructure: <http://www.computerweekly.com/news/1326870/Five-ways-to-create-a-more-efficient-backup-infrastructure>
5. Backup for dummies – Acronis Special Edition by Joel Berman 2014: <https://www.ahsay.com/blog/tag/backup-infrastructure-as-a-service/>
6. HP Data Protector 8.xx and 9.xx. HP internal books
7. Gartner

THE MODERN BACKUP INFRASTRUCTURE

Abstract. The text presents the factors to be taken into effect for the construction of infrastructure for the preservation and storage of information. Basic information that will help protecting of the business in the event of a disaster is given.

✉ **Mr. Asen Zahariev**

University of Telecommunications and Post

Sofia, Bulgaria

E-mail: zaharievasen@gmail.com