КРОК
№1 в России по созданию ИТ-инфраструктур

ДЕДУПЛИКАЦИЯ. НЕТ ГРОМОЗДКИМ ЛЕНТОЧНЫМ БИБЛИОТЕКАМ.

КРОК | EMC² where information lives

EMC VELOCITY² SIGNATURE SOLUTION CENTER

Сергей Верчёнов
ИНЖЕНЕР КОМПАНИИ КРОК

А ЧТО У ВАС С РЕЗЕРВНЫМ КОПИРОВАНИЕМ?

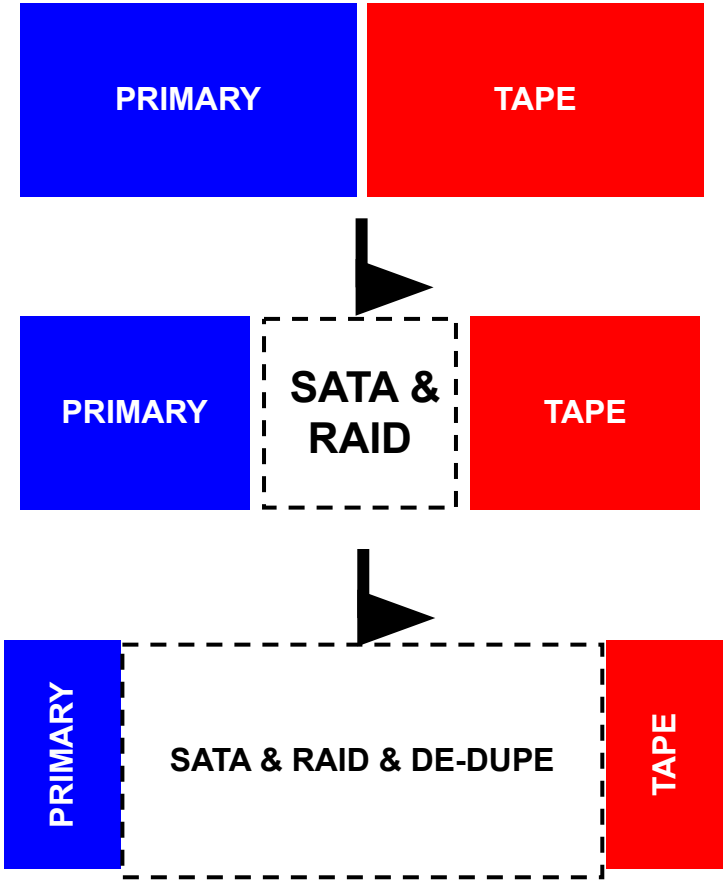
- Копирование всех данных на ленту
- Отправка лент в хранилище [?]

Обычное резервное копирование

- В неделю копируется до 200% объема данных (full+ incr, часто больше)
- Скорость резервного копирования далека от идеальной
- Скорость восстановления «такова, какова она есть»
- Риск потери данных из-за невозможности восстановления [>10%]

ОЧЕВИДНЫЙ ВЫХОД – ПЕРЕХОД К КОПИРОВАНИЮ НА ДИСК

- Дисковые системы быстрее.. (Особенно при восстановлении)
- Дисковые системы надежнее.. (RAID, Кэш с батарейкой, дублирование компонентов и т.п.)
- Дисковые системы проще в управлении..
- Дисковые системы проще и безопаснее при развертывании удаленного хранилища резервных копий..
- Дисковые системы стремительно развиваются и дешевеют..



КЛЮЧ К ЭФФЕКТИВНОМУ ХРАНЕНИЮ – ДЕДУПЛИКАЦИЯ ДАННЫХ

Факты:

“Развертывание систем с дедупликацией прогрессирует необычно быстро для достаточно новой технологии.”

Влияние на рынок:

“Gartner рассматривает эту технологию как революционную, потому что она радикально снижает стоимость резервного копирования и восстановления с использованием дисков...”

“…игнорировать ее невозможно.”

Рекомендации:

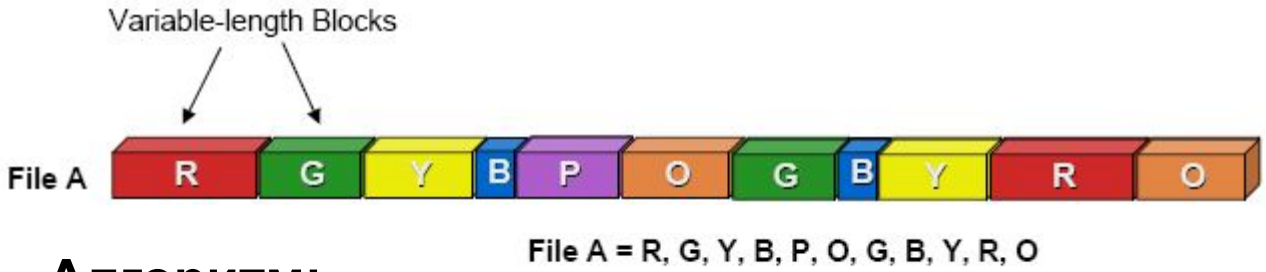
“…убедитесь, что вашу организацию устраивает надежность и зрелость предлагаемого вендором решения.”

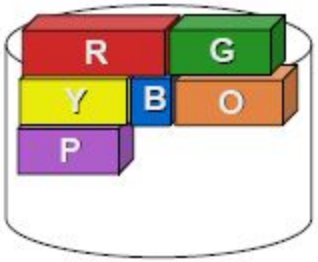
“К 2012 г. дедупликация будет использоваться в 75% всех СРК.”

Gartner

Source: Dave Russell, Gartner.

КАК РАБОТАЕТ ДЕДУПЛИКАЦИЯ?





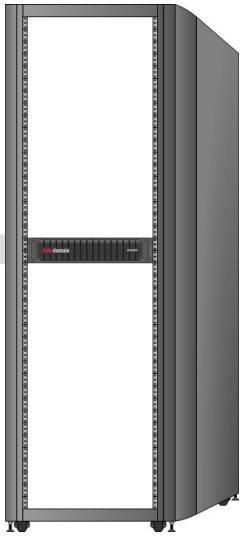
Алгоритм:

- Разбиение потока данных на блоки переменной длины
- Вычисление сигнатуры (хэша) для каждого блока
- Сохранение только уникальных блоков в сжатом виде

Результат: возможность упаковать 10 ТВ резервных копий в 1 ТВ дискового пространства = снижение стоимости хранения, возможен отказ от лент, эффективная репликация и т.п.

ЧТО ЖЕ ДАЕТ ДЕДУПЛИКАЦИЯ?

- Позволяет отказаться от хранения данных на лентах
- Уменьшение загрузки сети
- Уменьшает объем (10-30 раз) и окно бэкапа



- «Тормозит» процесс роста данных
- Упрощает управление данными, позволяет хранить дольше
- Упрощает восстановление в случае сбоя
- Снижает стоимость хранения данных

ФАКТОРЫ, ВЛИЯЮЩИЕ НА ДЕДУПЛИКАЦИЮ

Типы данных

Больше создаваемых пользователями файлов, неструктурированные данные* = больше коэффициент дедупликации

*Зашифрованные и сжатые данные – не идеальный вариант для дедупликации

Степень изменения данных

Меньше изменений = больше коэффициент дедупликации

Период хранения

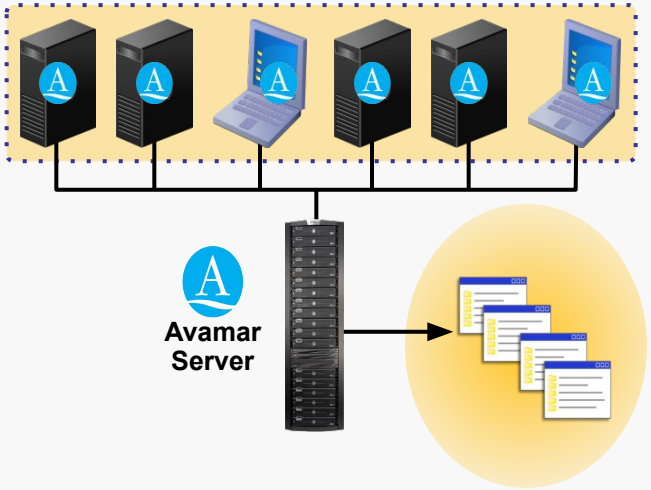
Дольше период хранения данных = больше коэффициент дедупликации

Соотношение уровней копирования

Больше полных бэкапов = больше коэффициент дедупликации

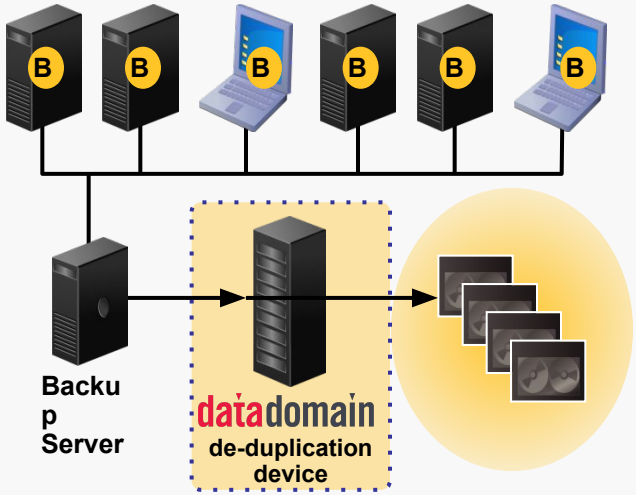
ДЕДУПЛИКАЦИЯ EMC: ВАРИАНТЫ

Дедупликация на источнике (EMC Avamar)



- Уменьшение объема хранения до 50 раз
- Сокращение трафика до 500 раз
- Сокращение времени резервного копирования до 10 раз
- Скорость восстановления – десятки МБ/с

Дедупликация на устройстве (EMC Data Domain)

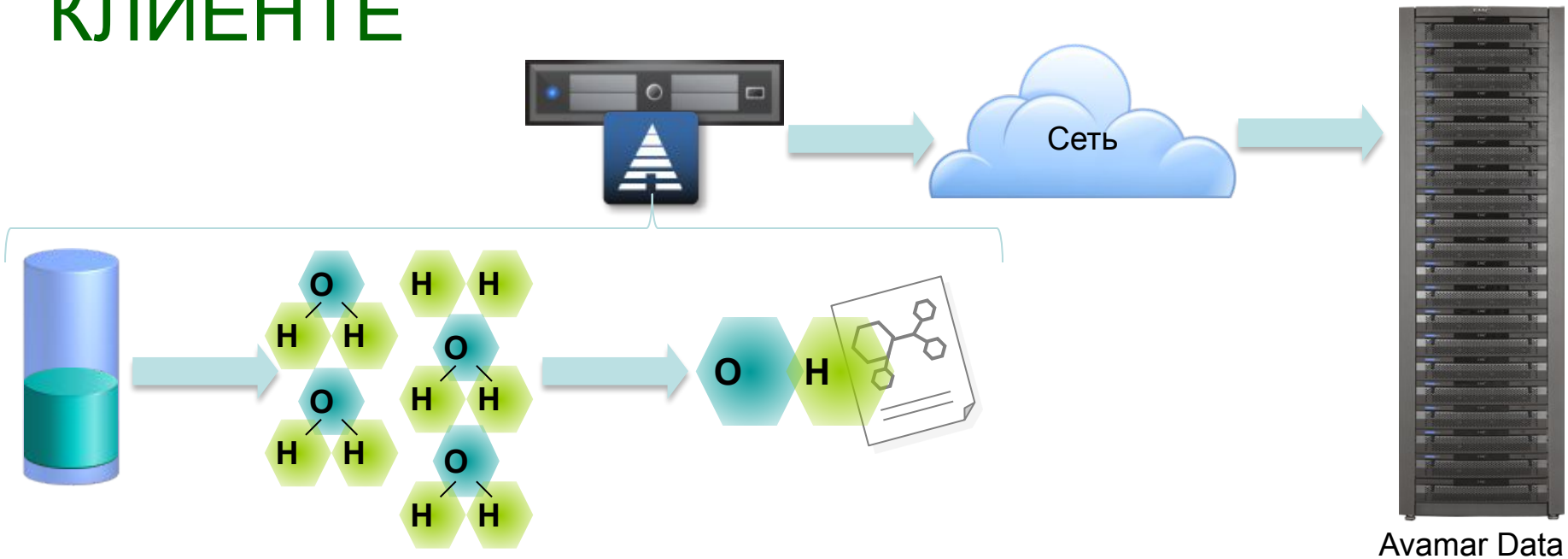


- Уменьшение объема хранения до 20 раз
- Интеграция с существующими системами резервного копирования
- Скорость восстановления - сотни МБ/с

Область дедупликации

AVAMAR

AVAMAR: ДЕДУПЛИКАЦИЯ НА КЛИЕНТЕ

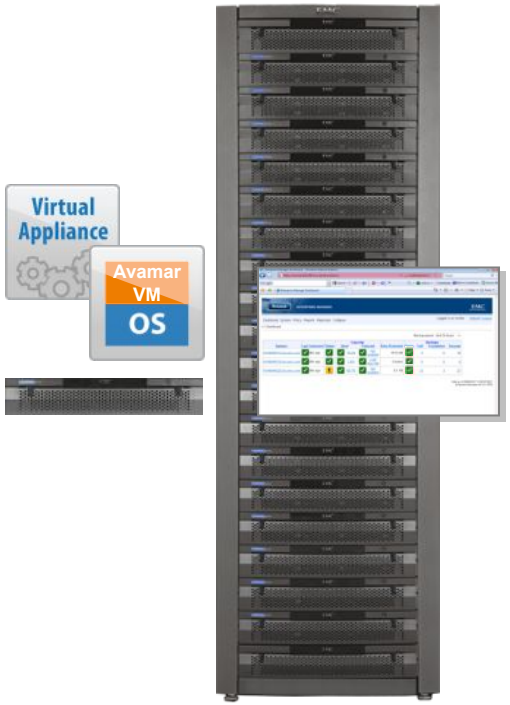


На уровне источника (дедупликация осуществляется именно тут):

1. Разбиение данных на блоки переменной длины
2. Определение уникальности блока – находится ли он уже в хранилище?
3. Копирование только уникальных блоков
4. Пересылка сжатых и шифрованных данных малого объема

На уровне хранилища: Координированная дедупликация между сайтами, серверами и во времени

ПОЧЕМУ AVAMAR?



- Целостное и автономное решение для бэкапа, программно-аппаратный комплекс
- Сокращает окно бэкапа, уменьшает нагрузку на сеть
- Высокая доступность данных и отказоучтойчивость всех компонент решения
- Гибкая конфигурация «под вашу» среду:
  - Avamar Data Store
  - Avamar Virtual Edition

Восстановление из полного бэкапа за один шаг

Существенно выше надежность бэкапа (по сравнению с записью на ленту)

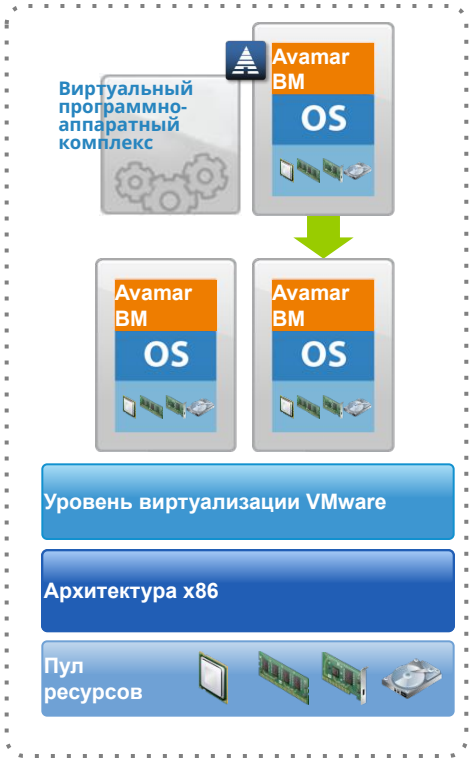
Увеличение отдачи от вложенных средств

СЕРВЕРНАЯ ЧАСТЬ: AVAMAR DATA STORE

- Avamar Data Store
  - Многоузловые конфигурации масштабируются от 4 ТБ до 52 ТБ дедублицированной емкости
  - Это эквивалентно 1.1 ПБ обычного хранения на дисках или лентах*
  - Требования к емкости снижаются в 20–40 раз
  - Доступность и надежность RAIN, RAID
- Avamar Data Store Single Node
  - Бывают емкостью 1 ТБ, 2 ТБ, 3.3 ТБ для хранения дедублицированных данных
  - Это эквивалентно 70 ТБ обычного хранения на дисках или лентах *
  - Простое удаленное развертывание
  - Дают возможность быстрого восстановления без зависимости от канала WAN

** Предположение: 100 % данных - файлы Microsoft Office, еженедельные полные и дневные инкрементальные копии, без компрессии, скорость изменения данных 10 % в день, срок хранения копий 90 дней*

AVAMAR VIRTUAL EDITION FOR VMWARE



Сервер Avamar в качестве виртуальной машины

- Первый в отрасли виртуальный комплекс для дедупликации резервных копий, операционного и аварийного восстановления
- Использование существующих серверов ESX и ресурсов хранения
- Репликация устраняет необходимость перевозки лент
- Поддержка vMotion для гибкого внедрения
- Два виртуальных программно-аппаратных комплекса Avamar Virtual Edition для VMware на один сервер VMware ESX

vmware PARTNER
TECHNOLOGY ALLIANCE

ПРЕИМУЩЕСТВА EMC AVAMAR

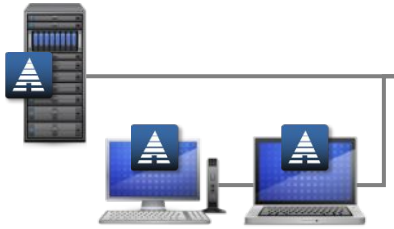
- Снижение нагрузки при резервном копировании
  - Уменьшение времени резервного копирования
  - Сокращение нагрузки на серверы
  - Сокращение нагрузки на сеть
- Снижение стоимости резервного копирования
  - Копия занимает значительно меньше места
  - Меньше требования к пространству хранилища
- Централизованное удаленное копирование
  - Использование низкоскоростных каналов передачи данных
  - Централизация средств резервного копирования для удаленных офисов
  - На удаленных площадках больше не требуется инфраструктура и персонал для резервного копирования

AVAMAR: ГДЕ ИСПОЛЬЗУЕМ?

Бэкап на Avamar Data Store в Основном ЦОД

Рабочие станции /ноутбуки

SMALL REMOTE OFFICE



•Минимизация WAN трафика, централизованное управление

•Полный бэкап NAS-устройств по NDMP

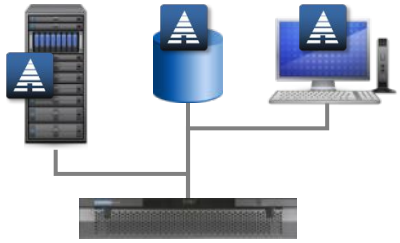
Основной ЦОД

Резервный ЦОД

•Репликация минимальных объемов данных

Бэкап на локальный Avamar Data Store, репликация на Avamar Data Store в Основном ЦОД

LARGE REMOTE OFFICE



•Минимизация WAN трафика, централизованное управление

Бэкап на Avamar Virtual Edition, репликация на Avamar Data Store в Основном ЦОД

Виртуальная инфраструктураVMware

•Бэкап Guest или Image уровня – уменьшение влияния на ESX

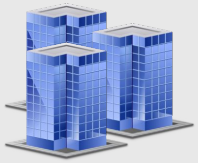
AVAMAR: ГДЕ ИСПОЛЬЗУЕМ?

Интеграция с приложениями:

- Oracle
- Microsoft SQL Server
- Microsoft Office SharePoint
- Microsoft Exchange Server
- IBM DB2
- Lotus Notes
- NDMP

Схемы резервного копирования:

Virtual



Remote Offices

Desktop/Laptop

NAS Storage

EMC Unified Management

EMC Global Services for Avamar

Инфраструктура:

EMC Avamar Software

EMC Avamar Data Store

EMC Avamar Virtual Edition for VMware

Операционные системы:

- Microsoft
- VMware
- Microsoft Server 2008 Hyper-V
- AIX
- HP
- Solaris
- Red Hat
- SuSE
- Novell
- MacOS

datadomain

DATA DOMAIN: ДЕДУПЛИКАЦИЯ НА ЦЕЛЕВОМ УСТРОЙСТВЕ

Control

Backup and Archive Applications

EMC²
symantec.
commvault
Tivoli software
BakBone SOFTWARE
vizioncore Enhancing Virtual Infrastructure

CIFS, NFS, NDMP, DD Boost

Ethernet

Virtual Tape Library (VTL) over Fibre Channel

Сеть

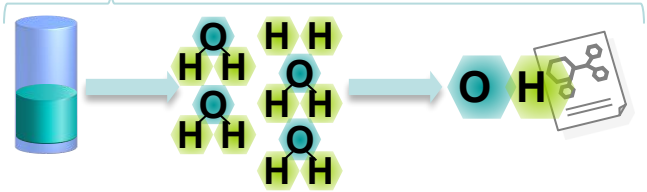
Target

DD appliance

Replication

Disaster Recovery

DD appliance

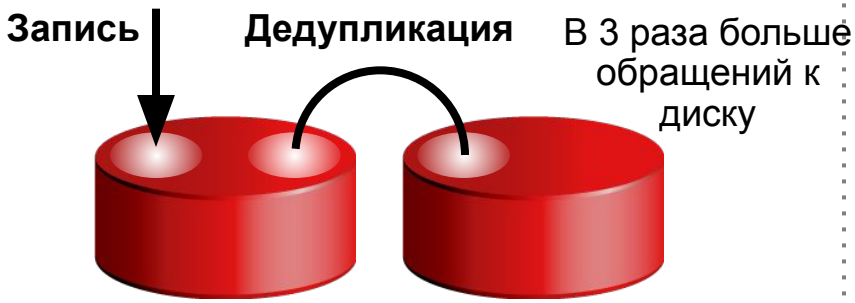


Параметры DD890:

- 2U
- 2 to 10 ports
- 10 and 1 Gigabit Ethernet; 8 Gb/s Fibre Channel
- RAID 6
- Up to 285 TB usable capacity with shelves
- 2 TB or 1 TB 7.2K rpm SATA HDD in shelf
- File system
- NVRAM
- N+1 fans and redundant, hot-plug power supplies

МЕТОДОЛОГИЯ: ДЕДУПЛИКАЦИЯ INLINE VS. POST-PROCESS

Post Process: Дедупликация после записи на диск

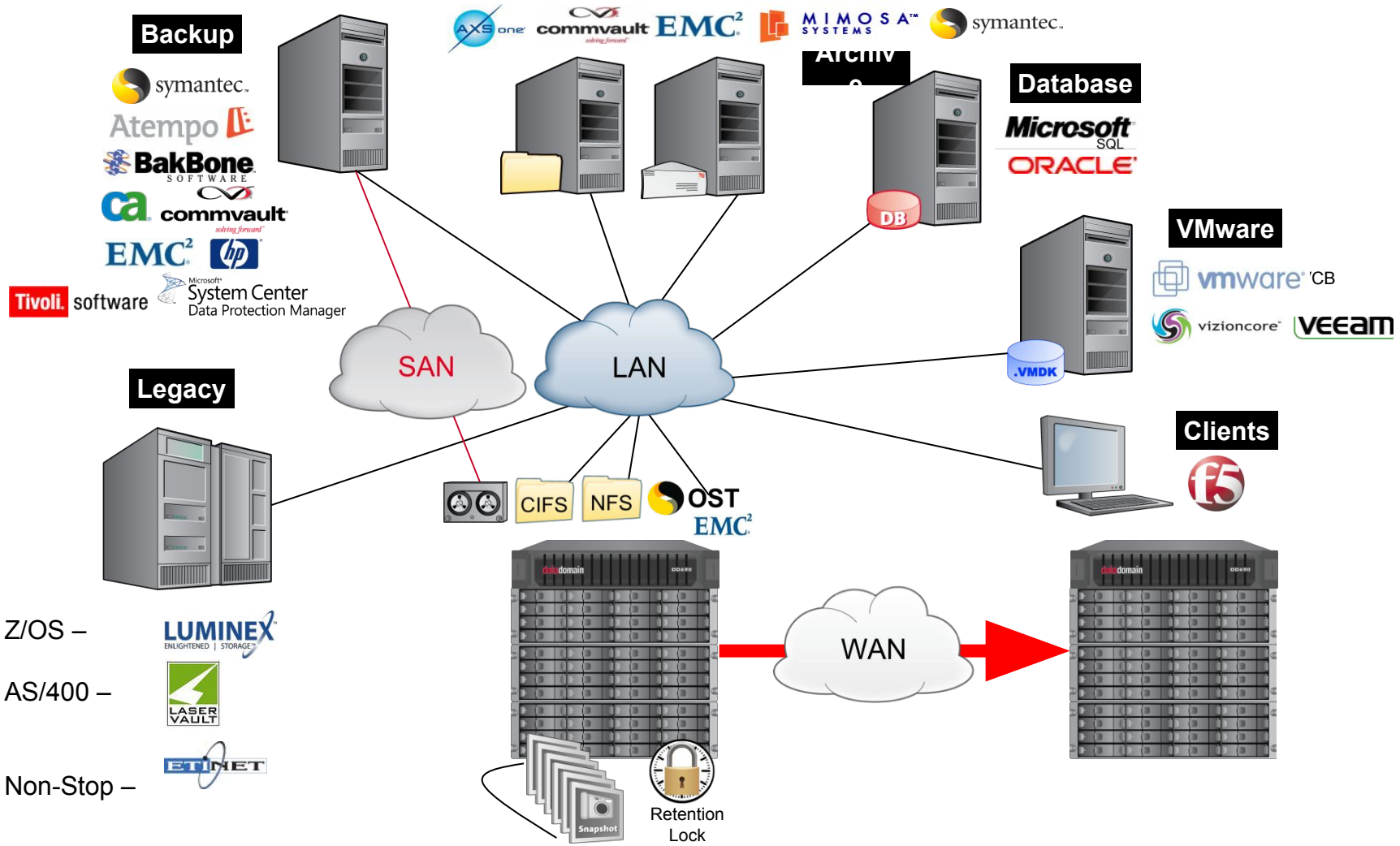


- Дедупликация затрагивает процессы:
  - Копирование на ленту: медленное решение для потоковой передачи данных на ленту
  - Восстановление: Увеличение параметров SLA
  - Репликация: Медленная передача данных на резервную площадку
  - Дедупликация «наслаивается» на процесс копирования/восстановления
- Затраты на персонал для устранения этих проблем

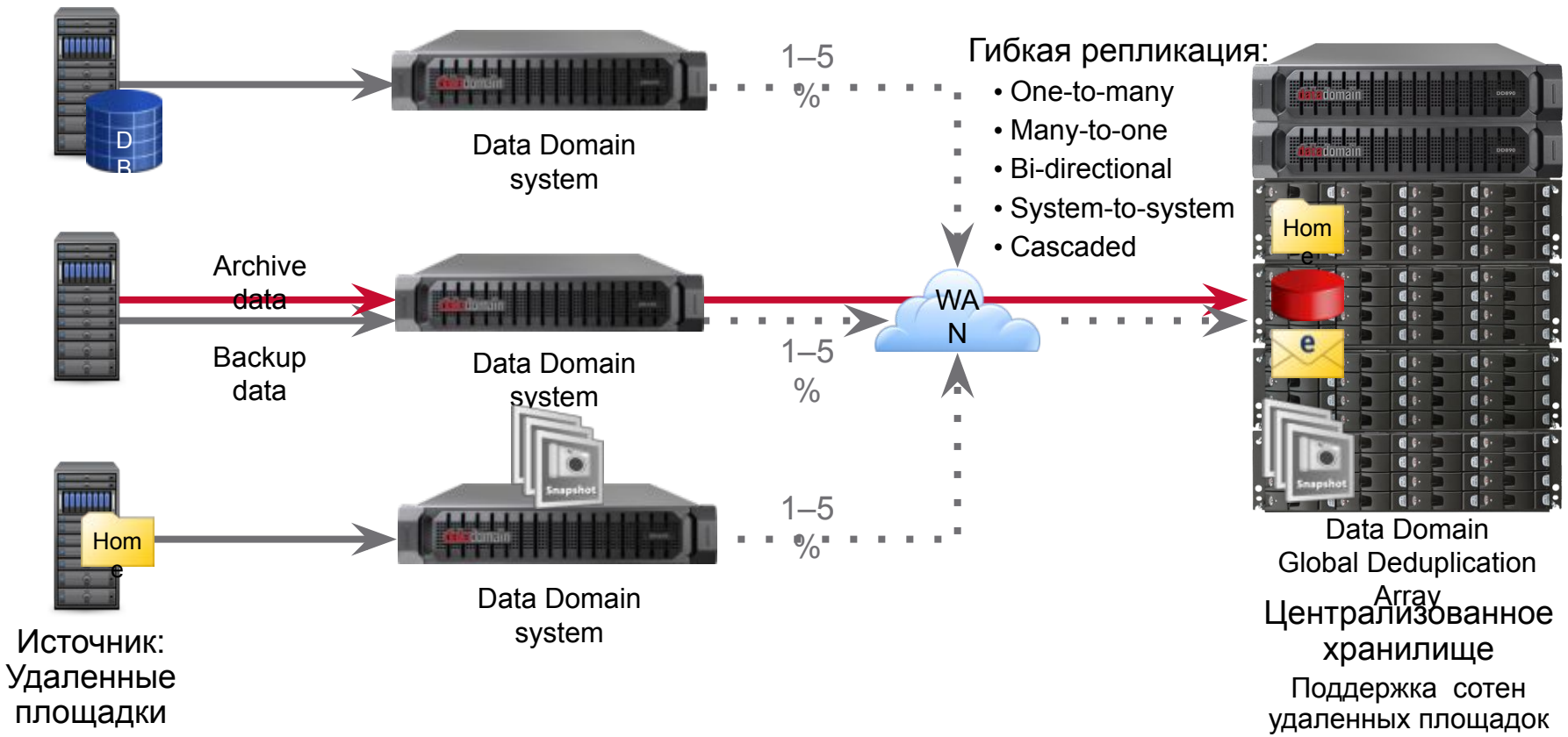
Inline: Дедупликация до записи на диск

- Дедупликация практически не влияет на другие процессы
  - Предсказуемость
  - Простота

DATA DOMAIN: ГДЕ ИСПОЛЬЗУЕМ?



ЭФФЕКТИВНАЯ РЕПЛИКАЦИЯ ДАННЫХ

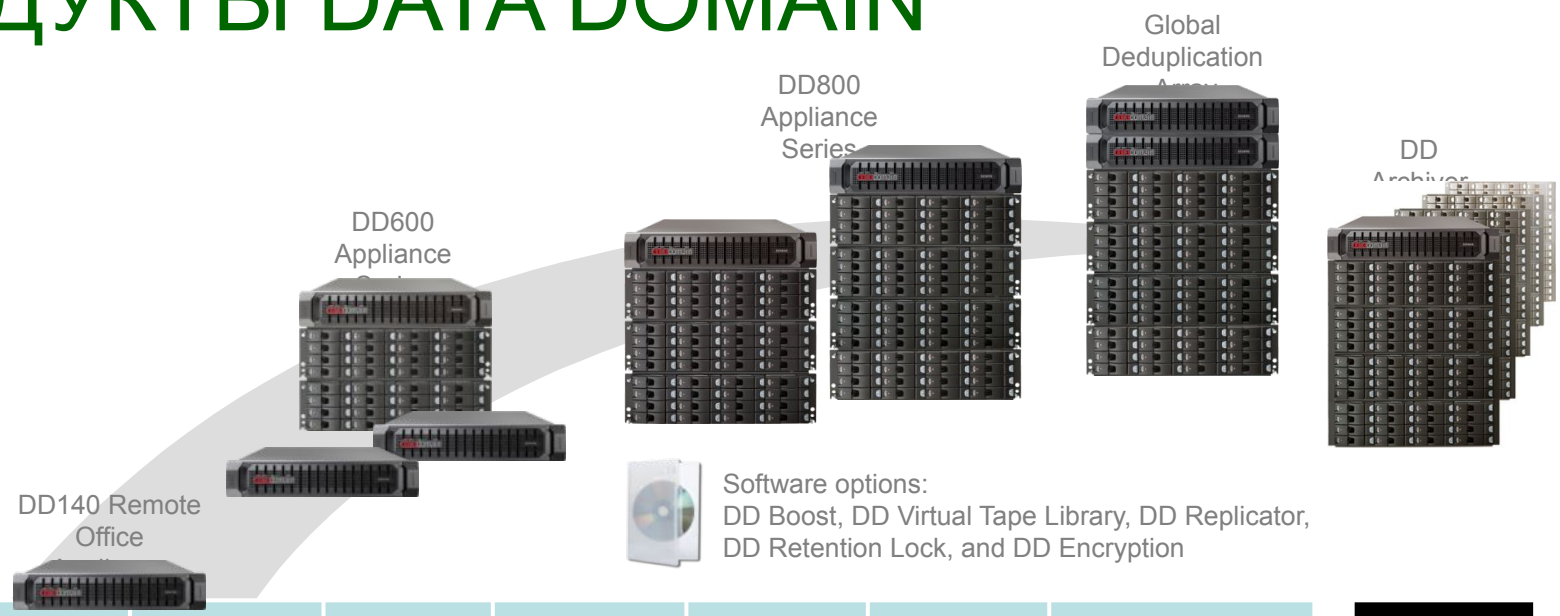


95–99% уменьшение загрузки WAN-каналов

ФУНКЦИОНАЛ DD BOOST (OST)

- Осуществляет «перенос» части процесса дедупликации на сервер резервного копирования
- Поддержка основных продуктов на рынке
  - EMC NetWorker
  - Symantec NetBackup and Backup Exec
- Ускоряет процесс копирования вплоть до 50 процентов
- Позволит копировать больше с имеющимся оборудованием
  - 20–40 % снижения общего воздействия на сервер бэкапа
  - 80–99 % снижения воздействия на сеть
- Позволит управлять репликацией Data Domain из приложения резервного копирования

ПРОДУКТЫ DATA DOMAIN



	DD140	DD610	DD630	DD670	DD860	DD890	Global Deduplication Array	DD Archiver
Speed (DD Boost)	490 GB/hr	1.3 TB/hr	2.1 TB/hr	5.4 TB/hr	9.8 TB/hr	14.7 TB/hr	26.3 TB/hr	9.8 TB/hr
Speed (other)	450 GB/hr	675 GB/hr	1.1 TB/hr	3.6 TB/hr	5.1 TB/hr	8.1 TB/hr	10.7 TB/hr	4.3 TB/hr
Logical capacity	9–43 TB	40–195 TB	84–420 TB	0.6–2.7 PB	1.4–7.1 PB	2.9–14.2 PB	5.7–28.5 PB	5.7–28.5 PB
Raw capacity	1.5 TB	Up to 6 TB	Up to 12 TB	Up to 76 TB	Up to 192 TB	Up to 384 TB	Up to 768 TB	Up to 768 TB
Usable capacity	0.86 TB	Up to 3.98 TB	Up to 8.4 TB	Up to 55.9 TB	Up to 142 TB	Up to 285 TB	Up to 570 TB	Up to 570 TB

ПРЕИМУЩЕСТВА EMC DATA DOMAIN

- Дедупликация **в реальном времени**
  - Минимизация требуемого пространства
  - Эффективная репликация
  - Предсказуемая производительность и простота управления
- Правильная **архитектура**
  - Максимальная производительность при минимуме дисков
  - Сквозная проверка целостности
  - Открытость
- **Зрелость** решения
  - Более 10000 систем установлено
  - Средний уровень дедупликации 17.4:1
  - В среднем одна открытая эскалация в службе техподдержки
  - Система, которая «просто работает»

ДЕМОНСТРАЦИЯ DATA DOMAIN

datadomain

КОГДА ЧТО ИСПОЛЬЗОВАТЬ?

AVAMAR

- Требуется защита виртуальных серверов, удаленных офисов, рабочих станций, серверов объемом до 500 ГБ
  - EMC Avamar

datadomain

- Требуется защита больших объемов данных с восстановлением по SAN, интеграция различных технологий, поддержка лент или дисковых систем
  - Традиционное ПО (например, EMC NetWorker)
  - ..и EMC Data Domain

СПАСИБО ЗА ВНИМАНИЕ!

Сергей Верчёнов

ИНЖЕНЕР КОМПАНИИ КРОК

Тел.:(945)974-2274 (4626)
E-mail: serverchenov@croc.ru