High-availability

проф. д-р инж. Христо Вълчанов

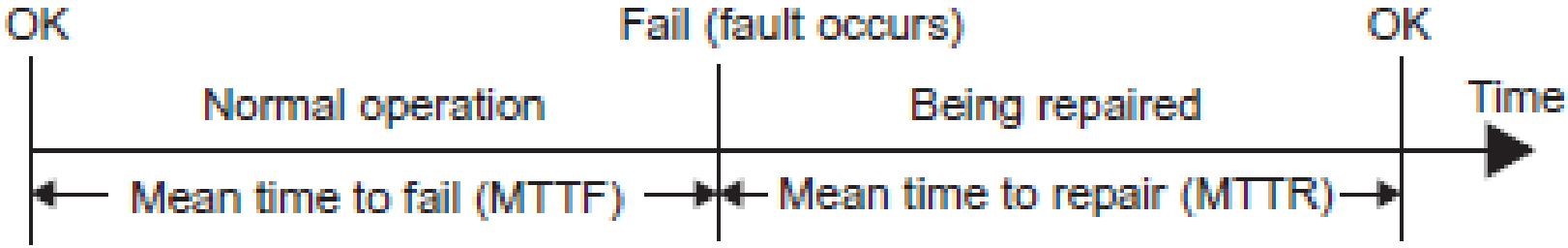
http://cs.tu-varna.bg

Наличност на услугите

- Наличност (Availability) – отношението на времето през което системата е работила (uptime) към общото време в рамките на период (година).
- Типично се указва като брой 9-ки в съотношението

Nines	Application	Up time %	Actual down time
2	Office equipment	99%	3 days, 15.6 hours
3	Most IT infrastructure	99.9%	8.76 hours
4	Internet infrastructure	99.99%	52 minutes 34 sec
5	PSTN and other business critical	99.999%	5 minutes, 16 sec
6	Carrier class core/edge	99.9999%	32.56 seconds

Оценка на наличност



- Реално се използват статистически данни за средното време за отказ (Mean Time to Failure – MTTF)
- Mean Time to Repair (MTTR)

- **Availability = MTTF / (MTTF + MTTR)**

Оценка на наличност

- Ако система предоставя 50 000 часа MTTF и средно 15 мин за възстановяване или подмяна, наличността на системата е пет 9-ки: 99.9995%.

- Увеличаване на MTTF
- Намаляване на MTTR

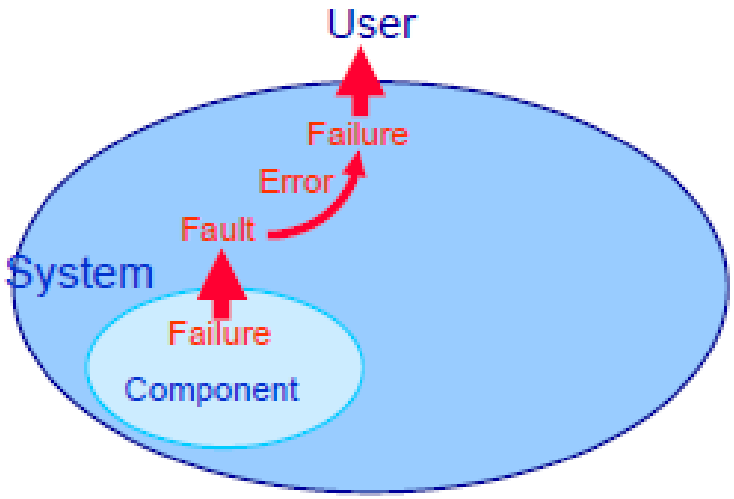
Време на отказ (downtime)

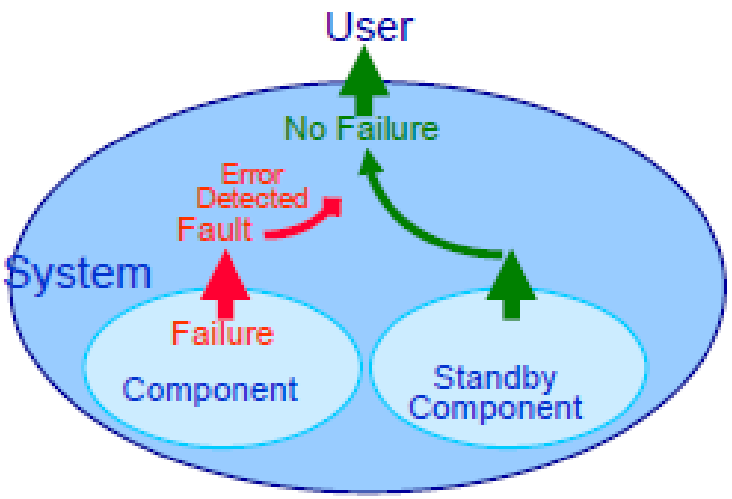
- Планирано време
  - Типично за поддръжка, обновяване
- Непланирано време
  - Отказ на хардуер
  - Грешно натискане на бутон
  - Природни бедствия

Откази

За възстановяване при отказ е необходимо резервиране на компоненти

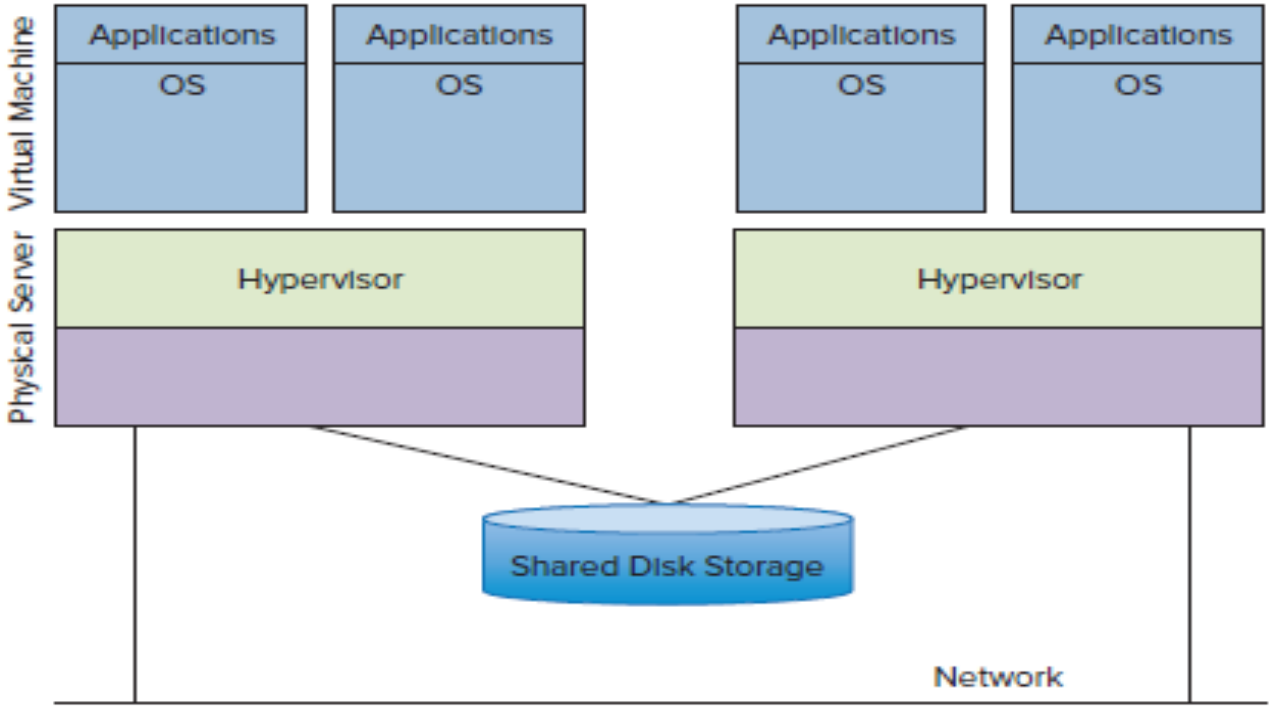
Чрез тях системата може да открие отказ и да се преконфигурира, използвайки резервирания компонент





Клъстери

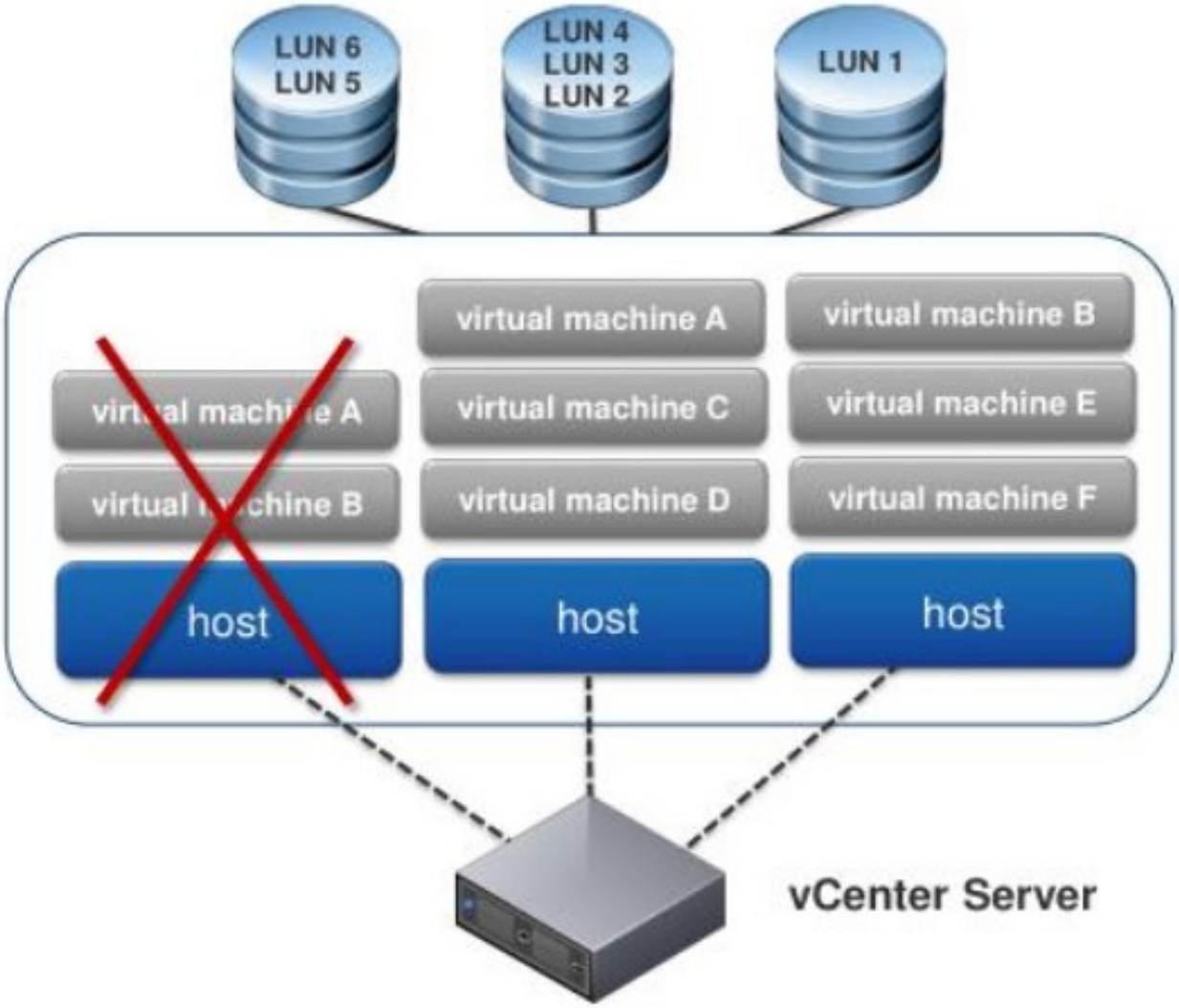
- Свързване на няколко сървърни машини
- Споделени ресурси за съхранение
- Клъстерен софтуер



High availability

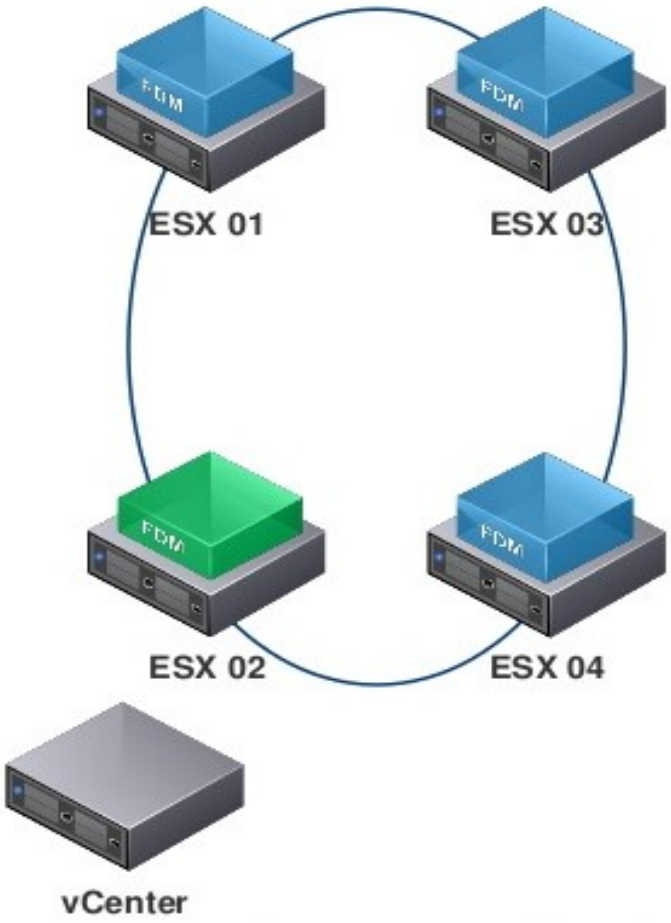
- Предоставя автоматично рестартиране на ВМ при отказ на физически сървър
- Предоставя висока наличност докато редуцира необходимостта от пасивен чакащ хардуер (standby)
- VMware използва vCenter

High availability



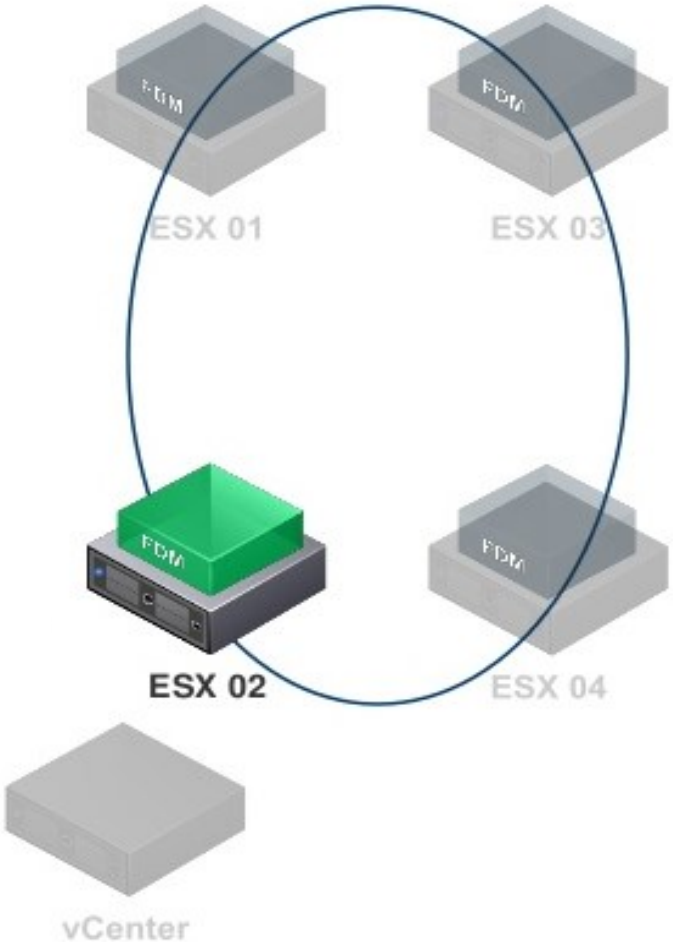
High availability

- На всеки хост се изпълнява VMware HA агент - Fault Domain Manager (FDM)
- FDM наблюдава съобщенията (heartbeats) между хостовете
- Само един агент е главен, останалите са подчинени
- Съобщенията се изпращат всяка секунда
- Липса на съобщение е индикация за отказ на хост



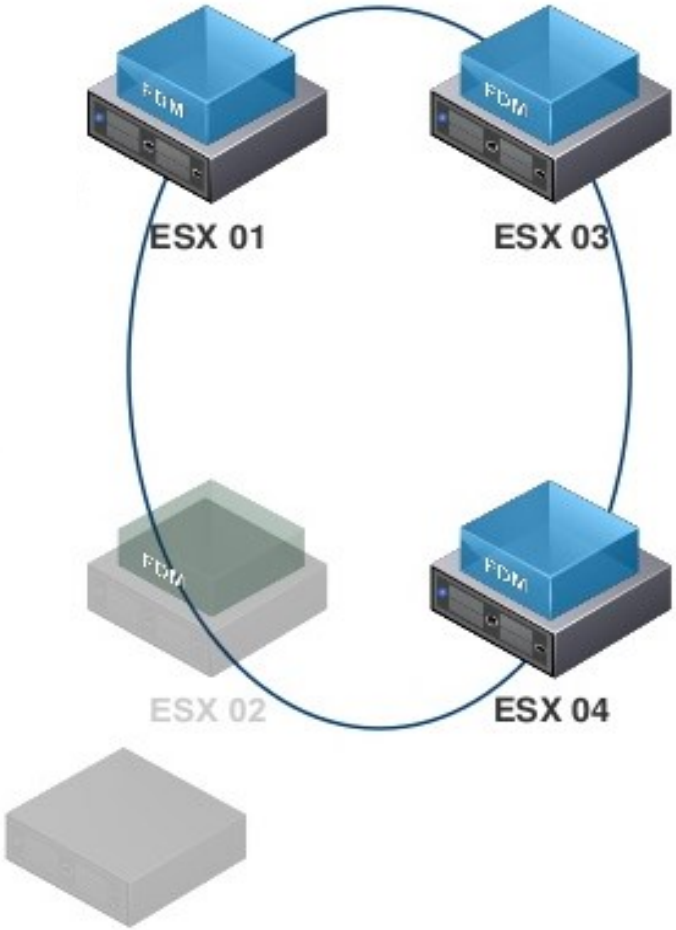
Роля на главен FDM

- Наблюдава ESX хостовете и наличността на ВМ
- Ако подчинен хост откаже, рестартира защитените машини от него
- Ако защитена ВМ откаже, рестартира я



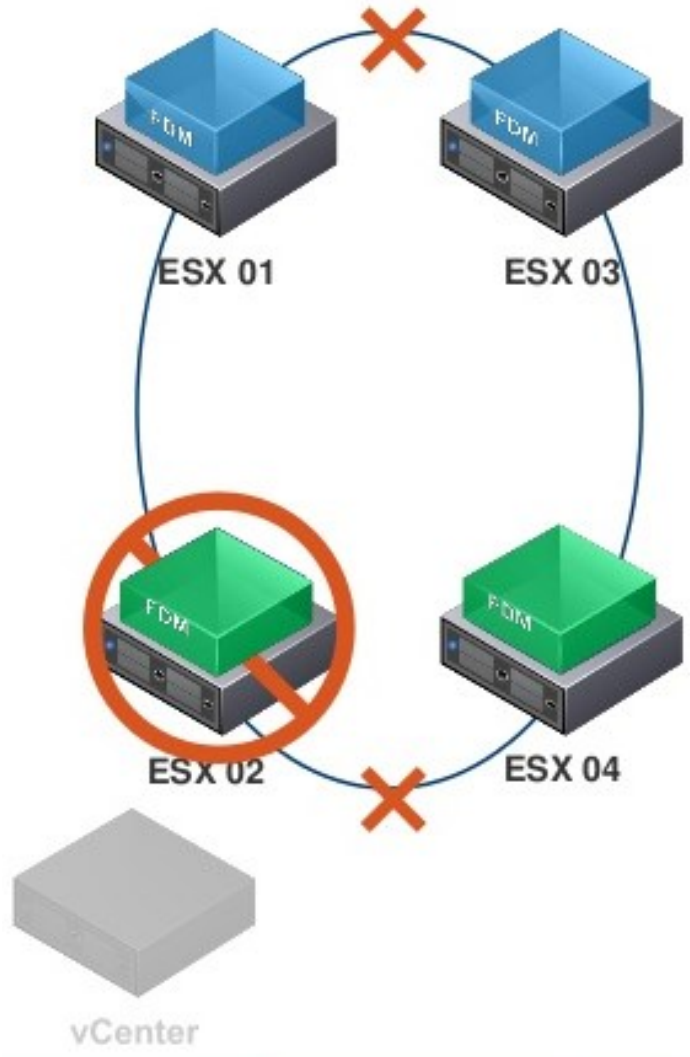
Роля на подчинен FDM

- Наблюдава състоянието на локалните ВМ и изпраща информация при значителни промени към главния FDM
- Следи наличието (health) на главния. Ако той откаже, започва се избор на нов главен



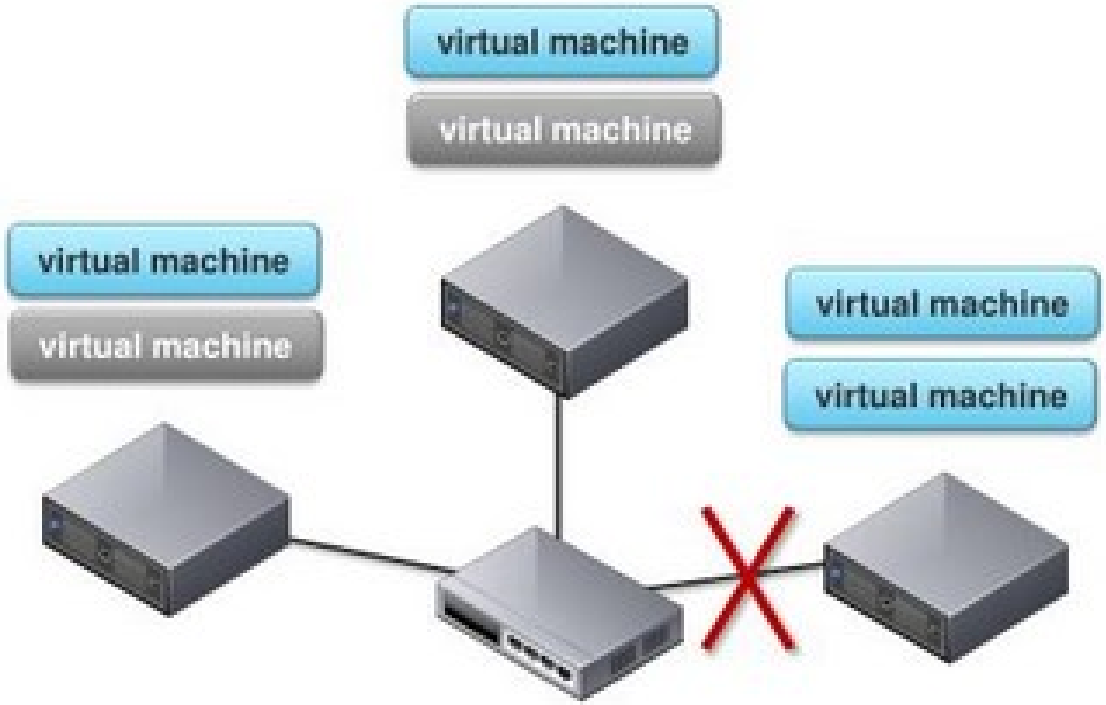
Избор на главен FDM

- Избор настъпва при:
  - vSphere е разрешен
  - Отказ на главния, изключен или в режим на maintenance
- Алгоритъм:
  - Хостът с достъп до datastore с най-голям номер печели
  - Ако има няколко, този с лексикографски по-висок ID печели (“host-99” печели пред “host-100”
- Използва се UDP
- След избора всички комуникации са през SSL

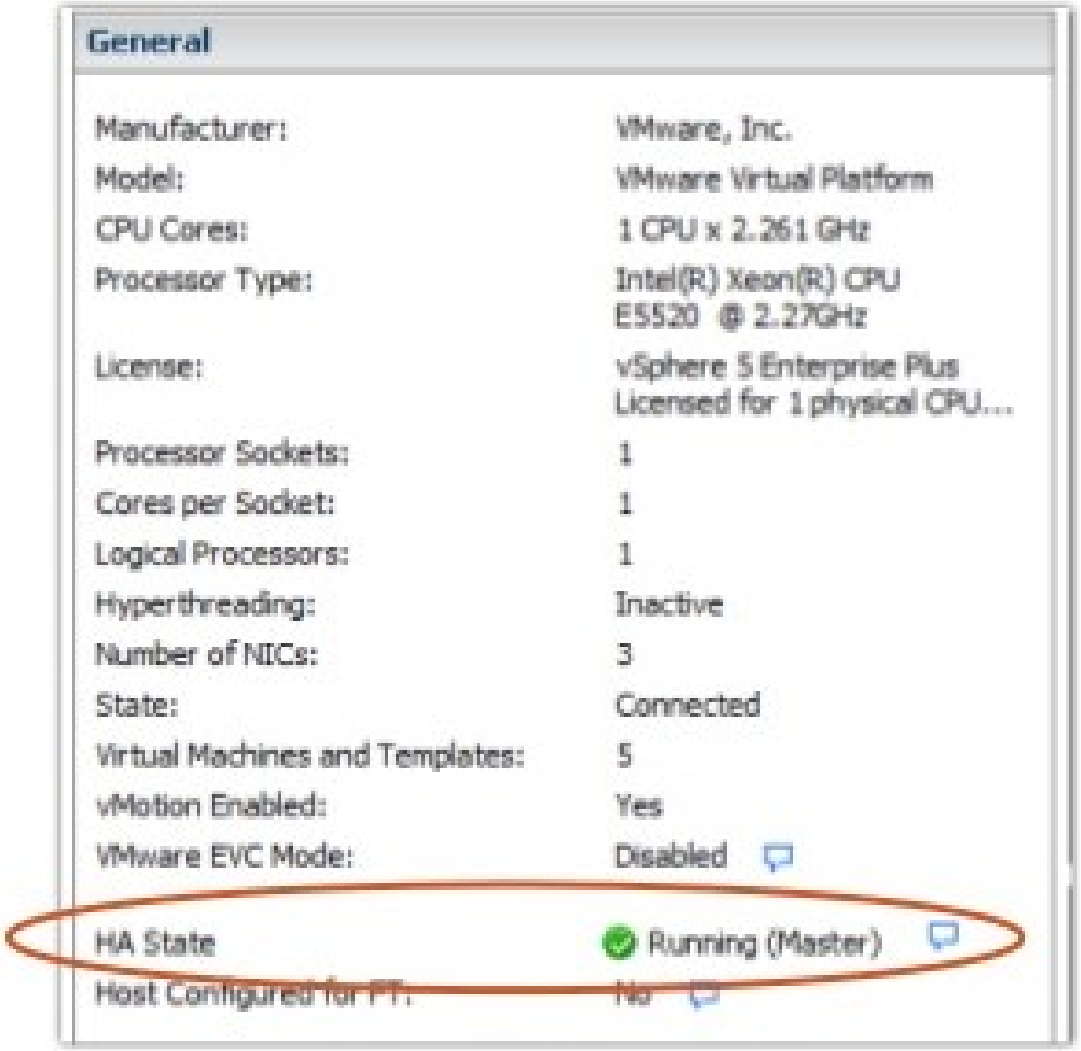


Изолиране на хост

- Ако хост в клъстера изгуби heartbeats но функционира, той се изолира от клъстера
- VMware HA изчаква 12 сек преди да определи, че хостът е изолиран



HA състояния



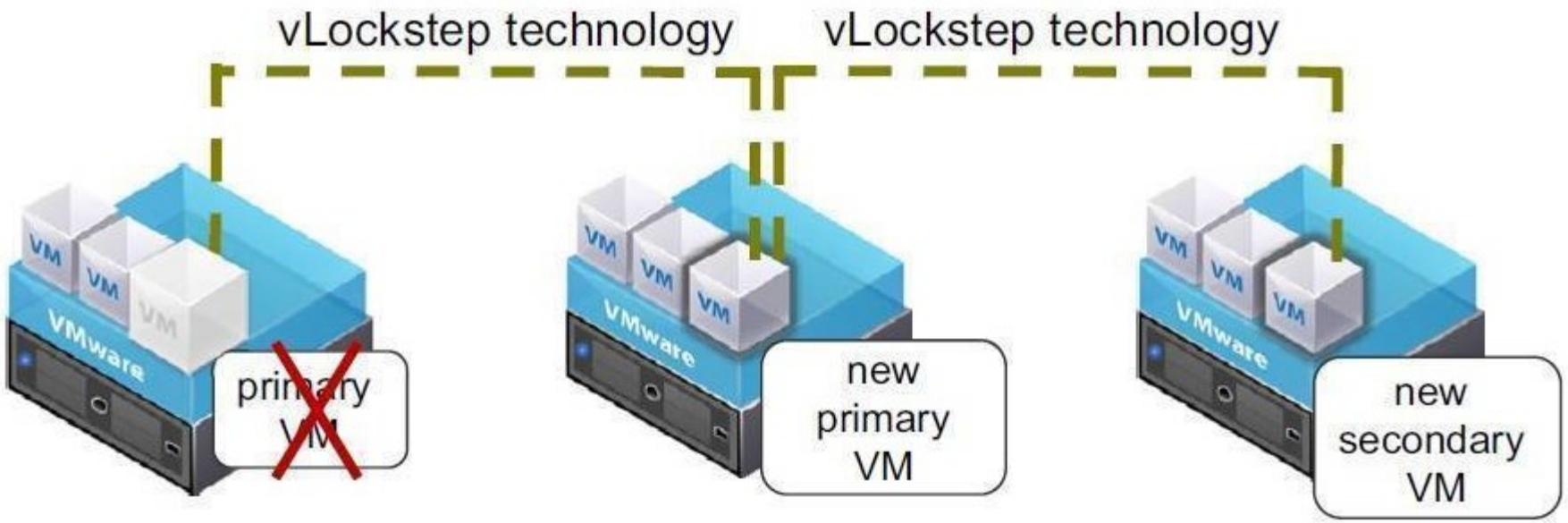
General

Manufacturer:	VMware, Inc.
Model:	VMware Virtual Platform
CPU Cores:	1 CPU x 2.261 GHz
Processor Type:	Intel(R) Xeon(R) CPU E5520 @ 2.27GHz
License:	vSphere 5 Enterprise Plus Licensed for 1 physical CPU...
Processor Sockets:	1
Cores per Socket:	1
Logical Processors:	1
Hyperthreading:	Inactive
Number of NICs:	3
State:	Connected
Virtual Machines and Templates:	5
vMotion Enabled:	Yes
VMware EVC Mode:	Disabled
HA State	Running (Master)
Host Configured for FT:	No

Отказоустойчивост (FaultTolerance)

- FaultTolerance (FT) – предоставя по-високо ниво на активност и наличност от HA
- Предоставя нулево downtime и нулева загуба на приложения
- Подходяща за критични приложения

Отказоустойчивост (FaultTolerance)

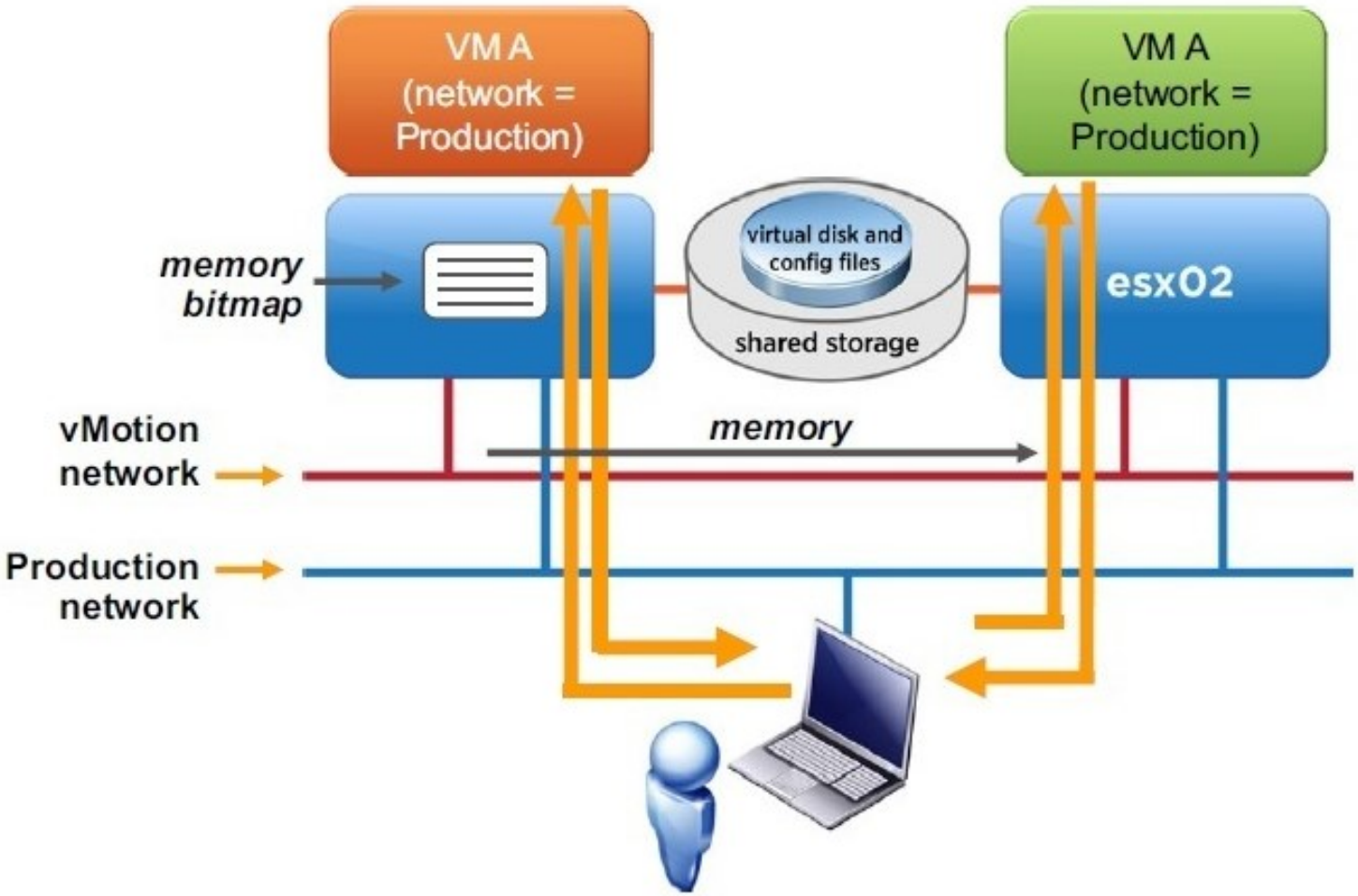


Мигриране на ВМ

- Необходимост от спиране на сървър за поддръжка
- Обновяване на софтуер
- Замяна на хардуер

➢ Не трябва да се спира работата на ВМ - **vMotion**

Мигриране на ВМ



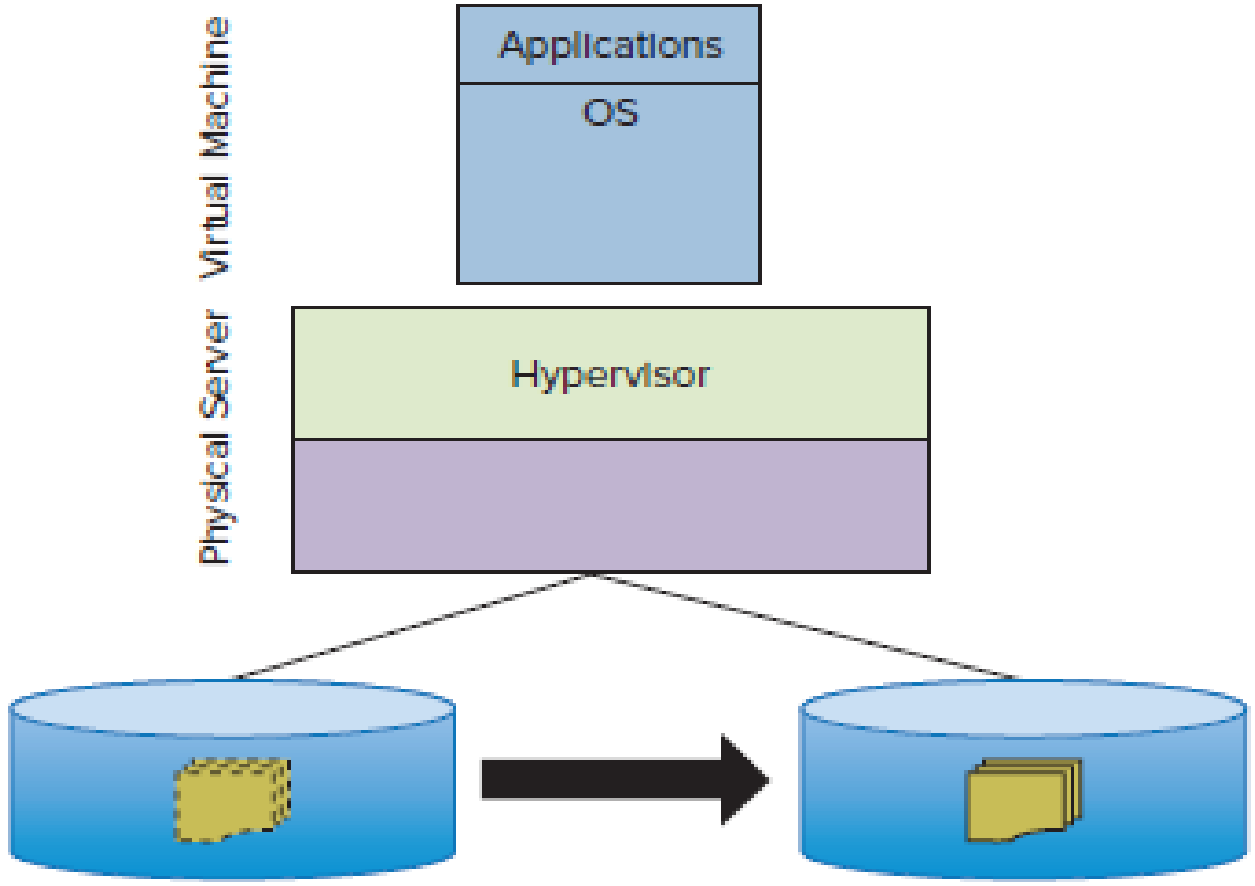
Мигриране на ВМ

1. Инициира се мигриране от администратора
2. Състоянието на паметта на ВМ се копира в новия хост. Списък от модифицираните страници се съхранява в битмап на първия хост
3. След пълното копиране, ВМ от първия хост остава в паметта, но не обслужва клиентски заявки. Файлът с битмап паметта и състоянието на устройствата на ВМ се прехвърлят на втория хост
4. Втория хост прочита адресите от битмап паметта и изисква съдържанието на тези адреси от първия хост
5. След прехвърлянето на съдържанието на паметта, ВМ се стартира
6. След успешното стартиране, паметта на ВМ от първия хост се маркира като свободна

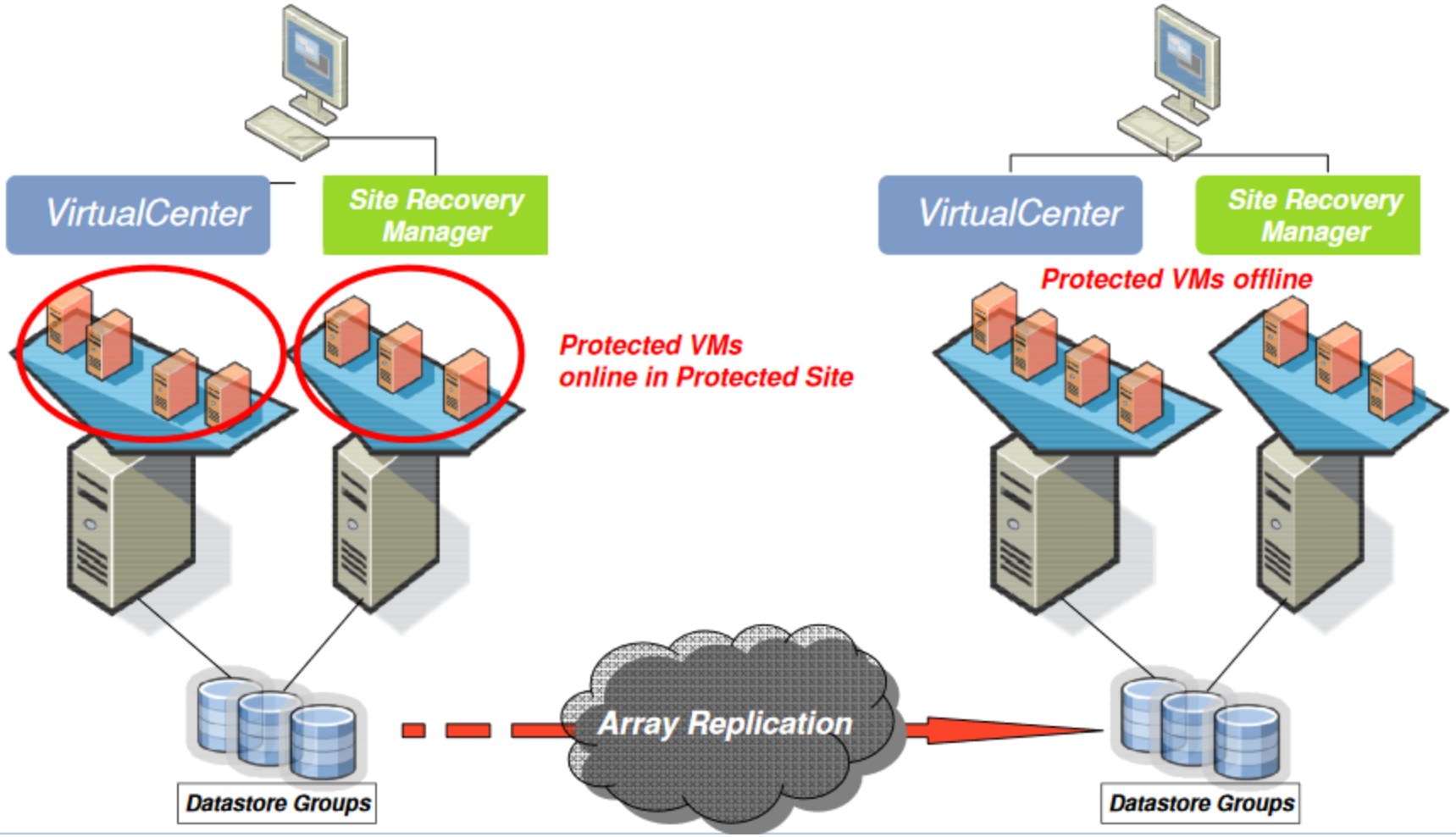
Мигриране на сторидж

- Физическо преместване на файлове от един диск на друг
- Преместване на ВМ от един дисков масив в друг
- Подмяна на дисков масив
- Не се спира работата на ВМ

Мигриране на сторидж



Защита на центове за данни



Клониране на ВМ

- Разгръщане на ВМ на база точно копие на съществуваща такава
- Възможност за създаване на резервни копия

Видове клониране

- Пълно – създава се независима ВМ, без връзка с оригиналната ВМ. Не споделя нейните дискове. Повече време за създаване.
- Свързано – създава се от снимка (snapshot) от оригинала. Всички файлове при оригинала в момента на снимката са налични за копието. Копието трябва да има достъп до оригинала.

Използване на темплейти

- Темплейт – основно копие на ВМ, включващо гост ОС, приложения и специфична конфигурация
- Удобно за разгръщане на множество ВМ, които са еднородни
- Не може да се стартира за разлика от клонирана ВМ

Формати на темплейти

- Open Virtualization Format (OVF) – позволява обмен на виртуални устройства (appliance) между различни продукти и платформи. Пакет, включваш множество файлове (метаданни и елементи, описващи ВМ и информация за разгръщането)
- Open Virtual Appliance (OVA) – OVF пакет в единичен архивен файл, типично TAR формат.

Въпроси?