Money.pl

Linux Contextualization

Michał Jura
Dział IT i Rozwoju

PLNOG 7, Kraków, 28 Września 2011 r.

AGENDA

Money.pl

1. Początki naszej infrastruktury
2. Metody wirtualizacji
3. VServer case study
4. Jak tego używamy
5. Linux Containers
6. Podsumowanie i wnioski

TŁO HISTORYCZNE
1/11

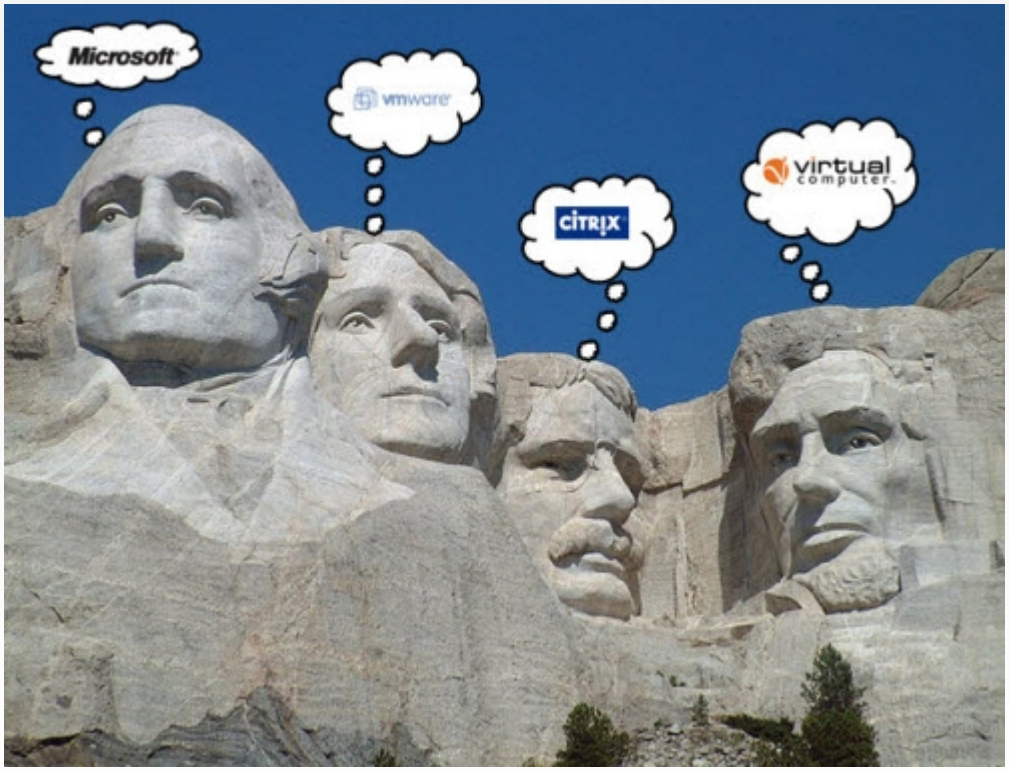
Money.pl

Nasze cele na początku 2008 roku:

- Wdrożenie nowych serwerów
- Uruchomienie nowych projektów
- Przygotowanie do zmiany Data Center
- Zwiększenie poziomu bezpieczeństwa

METODY WIRTUALIZACJI
1/11

Money.pl

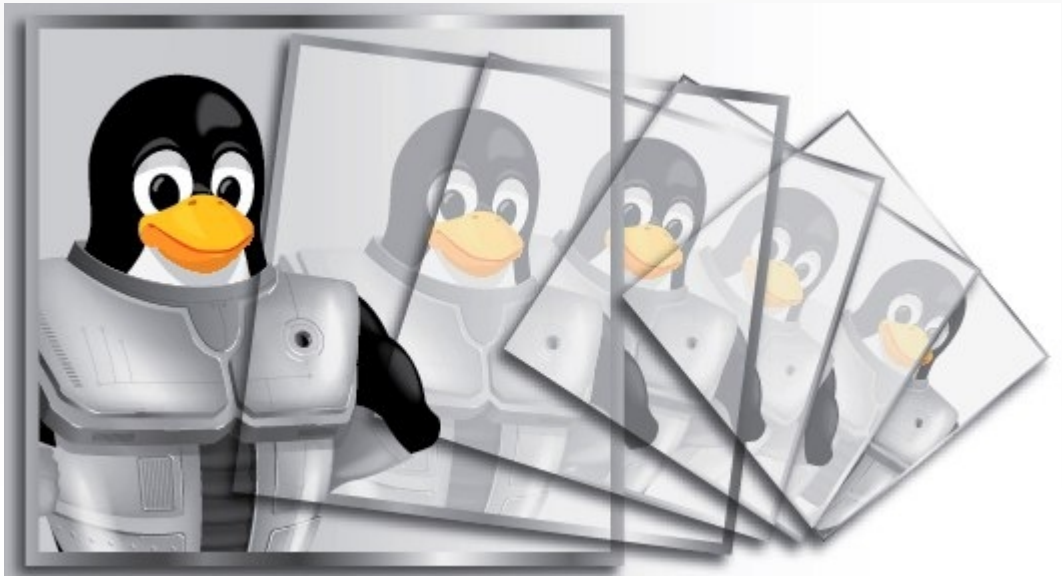


- Emulation
  - Qemu
- Paravirtualization
  - Xen
- Native Virtualization
  - KVM, VMware
- Operating System-Level Virtualization
  - **OpenVZ, FreeBSD jail, Solaris Zones, LXC**

NASZE OCZEKIWANIA
1/11

Money.pl

- Niewielki narzut na wirtualizację
- Elastyczność w konfiguracji
- Różne przypadki użycia
- Bezpieczeństwo
- Open Source



Linux-VServer

Money.pl

- Strona projektu: **linux-vserver.org**
- Seperacja procesów w obrębie jądra Linux
- Wykorzystuje znane mechanizmy systemu Linux
  - Linux Capability System
  - Resource Limits
  - File Attributes
  - Change Root Environment
- Współdzielone zasoby dla wszystkich instancji wirtualnych

Linux Capability System

Money.pl

- Począwszy od jądra 2.6.24 pełna obsługa POSIX Capabilities
- Obecnie zdefiniowane 34 uprawnienia z możliwością wprowadzenia dodatkowych
- Bezproblemowa koegzystencja tradycyjnych aplikacji (SUID-0) oraz tych korzystających z nowego mechanizmu
- Wszystkie informację są indywidualne dla każdego wątku w systemie ustawiane trzy bitową maską: I(inheritable), P(permitted) i E(effective)

man capabilities

Money.pl

Zdolność	Opis
[0] CAP_CHOWN	modyfikacja właściciela pliku lub grupy
[5] CAP_KILL	możliwość wysyłania sygnałów
[6] CAP_SETGID	zezwala na setgid(), setgroups()
[7] CAP_SETUID	zezwala na setuid()
[11] CAP_NET_BRODCAST	wysyłanie brodcastów i nasłuchiwanie multicastów
[12] CAP_NET_ADMIN	konfiguracja sieci w systemie
[13] CAP_SYS_MODULE	usuwanie i ładowanie modułów do jądra

Money.pl

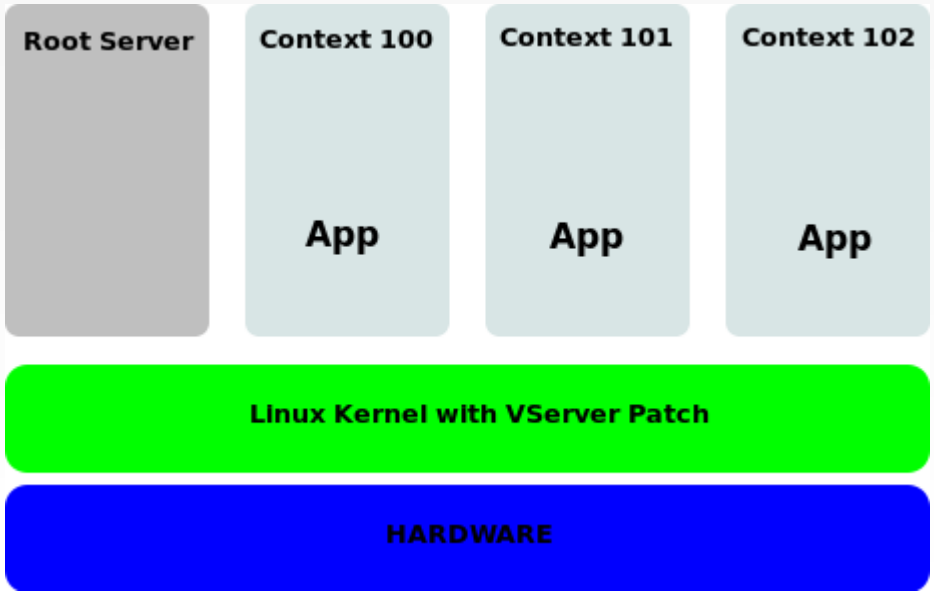
Modyfikacje jądra Linux dla VServer

- **Context Separation**
  - rozszerza funkcjonalność jądra o użycie kontekstów
  - separuje procesy serwerów wirtualnych
  - umożliwia użycie dwóch identycznych uid przez różne instancje
- **Network Separation**
  - nie wprowadza żadnych zmian w celu zwiększania narzutu
  - dowolność przypisywania adresów IP
- **Chroot Barier**
  - ulepszony mechanizm chroot chroniący przed wyjściem z niego lub nieautoryzowaną modyfikacją
- **Resource Isolation**
- **Filesystem XID Tagging**

VSERVER ARCHITEKTURA

Money.pl

- Wpływ na wydajność Root Servera
  - w okolicy 0%
- Narzut na wirtualizację
  - mniej niż 2%
- Ilość linii kodu VServer Patch
  - około 1112 linii



VSERVER ADMINISTRACJA

Money.pl

- Tworzenie nowej instancji VServer

  vserver vserver1 build -m debootstrap –context 100 \
  --hostname vserver1.mny –interface eth0:192.168.1.10/24 \
  -- -d squeeze -m http://ftp.pl.debian.org/debian

- Zarządzanie wirtualnymi serwerami

  vserver vserver1 start,stop,suexec,restart,enter,status,delete

- Dodatkowe narzędzia w pakiecie **util-vserver**

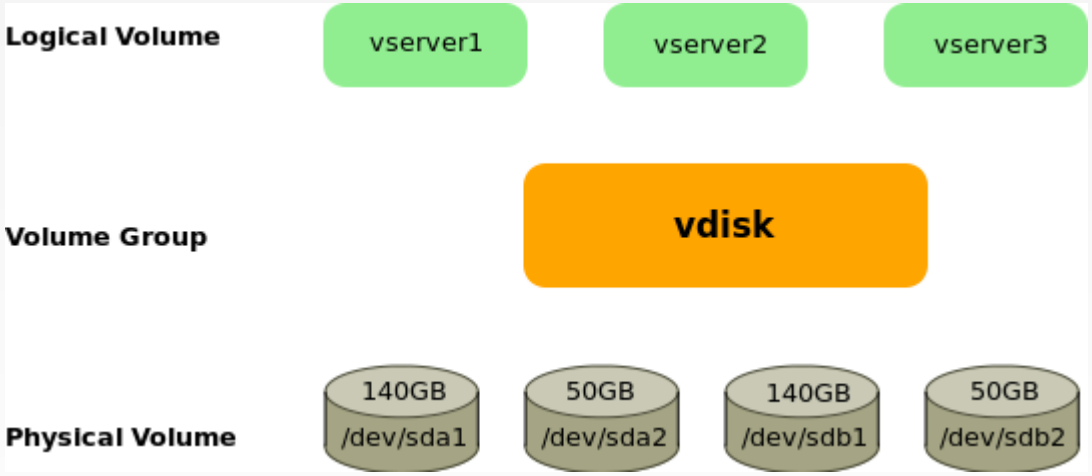
  vps, vlimit, vkill, vpstree, vnamespace, vapt-get, vrpm, vemerge

- Schematy serwerów dla różnych dystrybucji

STORAGE WIRTUALIZACJA

Money.pl

- Wirtualizacja zasobów dyskowych przez LVM2



- Zwiększanie partycji online

```
lvresize -L +10G /dev/vdisk/vserver1
resize2fs /dev/vdisk/vserver1
```

- Snapshoty

```
lvcreate -s -L +10G -n vserver1-`date +%F` vdisk/vserver1
```

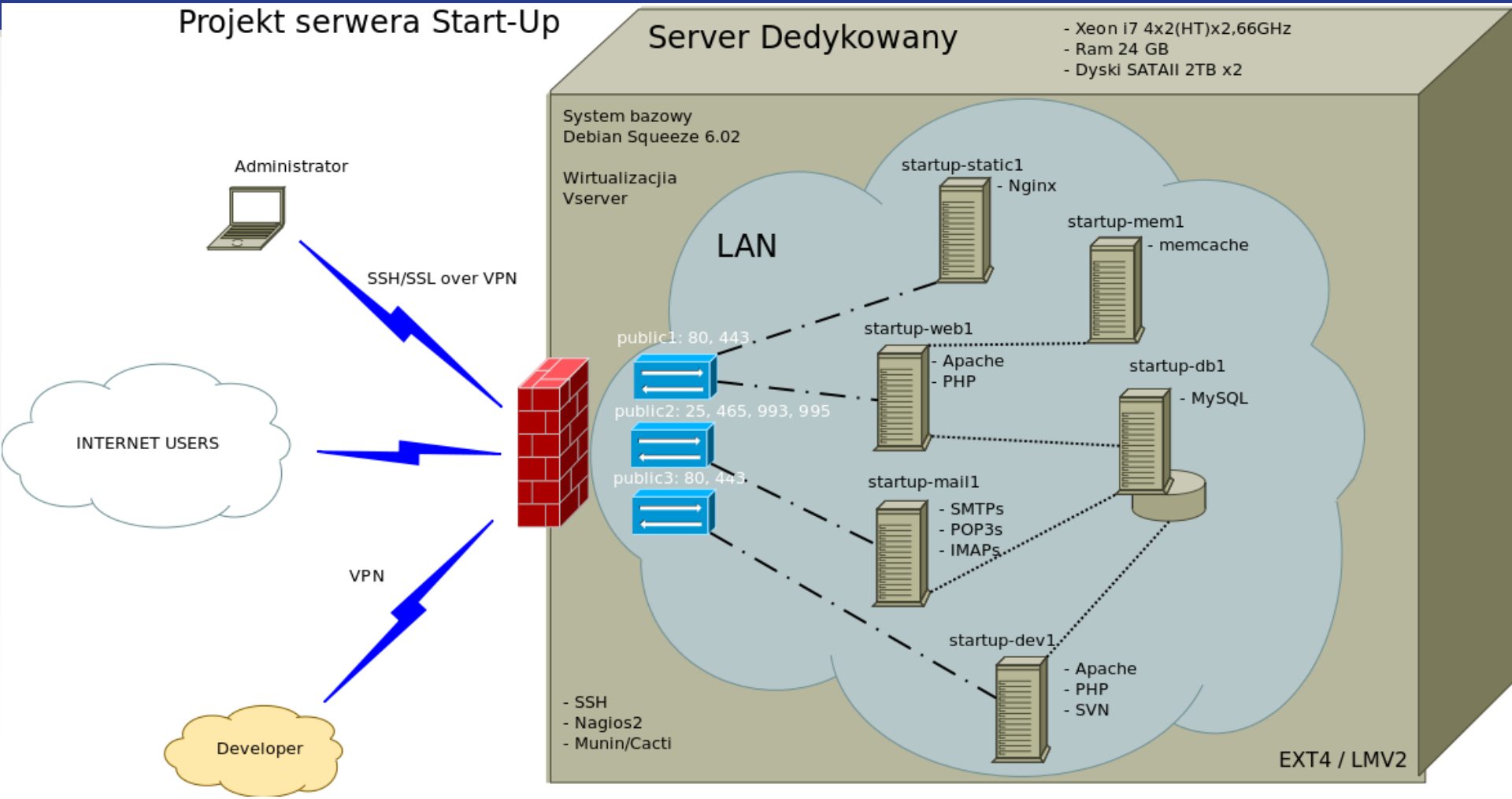
START-UP INFRASTRUKTURA

Money.pl

- Specyfika projektów Start-Up:
  - serwer http treść statyczna
  - serwer http treść dynamiczna
  - serwer aplikacji
  - baza pamięciowa
  - baza relacyjna
  - serwer poczty
- Zwykle jeden lub dwa serwery dedykowane
- Gotowość na szybki wzrost ruchu i pojawienie się wąskich gardeł
- Partycjonowanie serwerów **One Service, One Context**

START-UP INFRASTRUKTURA

Money.pl



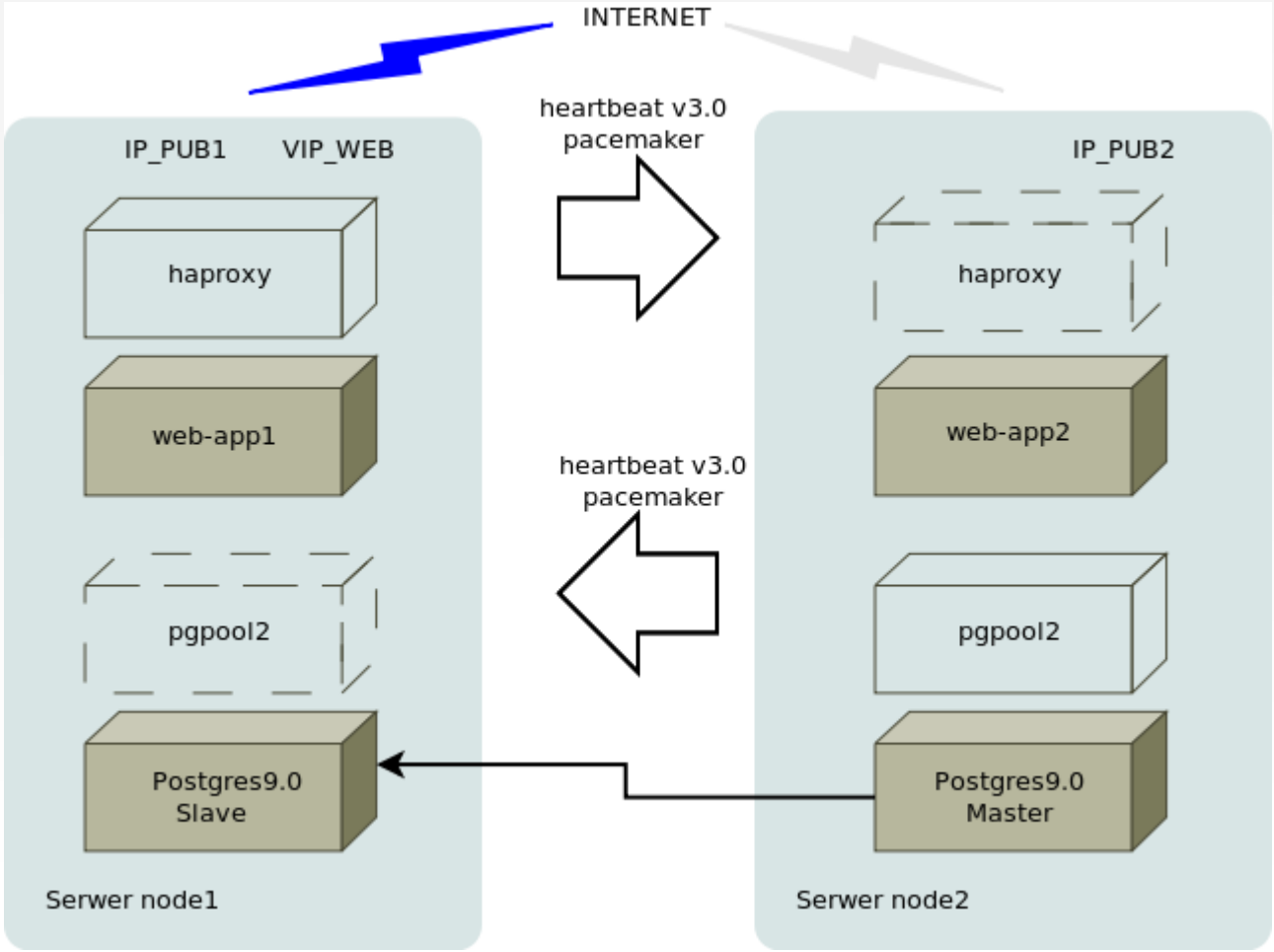
PRZYPADKI UŻYCIA

Money.pl

- Backup
  - dla większości wirtualnych serwerów wystarczy zwykły **rsync** w locie
  - dla serwerów z bazą **LVM2** snapshot
  - dla produkcji agent **bacula-fd** na Root Server
- Monitoring
  - **Nagios3** na Root Server + VServer Plugin
  - **Collectd** na Root Server + VServer Plugin
- Konfiguracja
  - własna implementacja **chef**
- Każdy programista posiada dostęp do swojego projektu

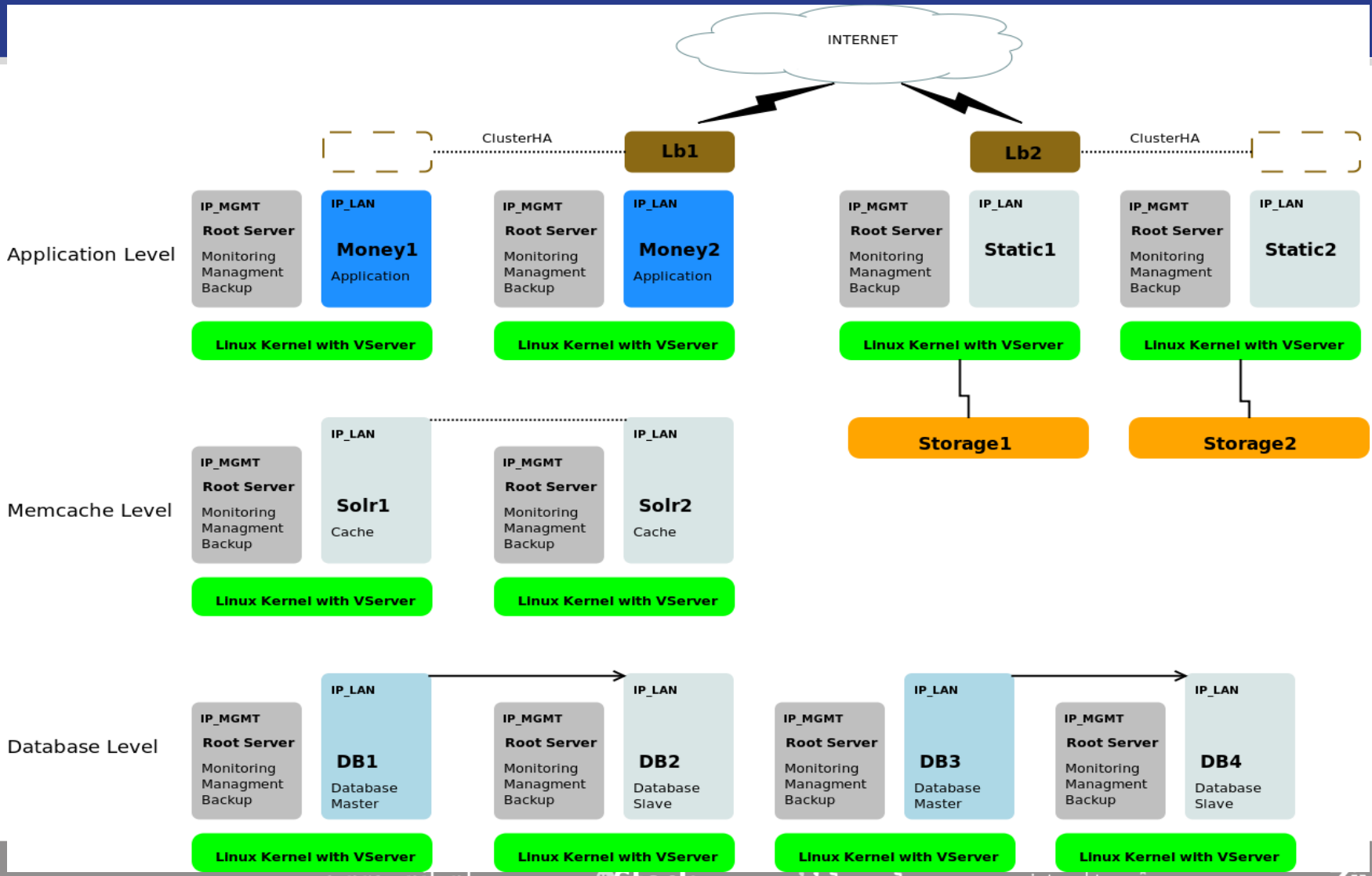
WDRAŻANIE NIEZAWODNOŚCI

Money.pl



TOPOLOGIA
11/11

Money.pl



- Strona projektu: **lxc.sourceforge.net**
- Mechanizm zarządzanie przestrzenią użytkowników określany mianem chroot na sterydach
- Linux Containers **lxc** wykorzystuje:
  - zarządzanie zasobami przez system plików **cgroup** w jądrze Linux od wersji 2.6.29
  - ulepszona izolacja zasobów dzięki funkcjonalności **namespaces** w jądrze Linux od wersji 2.6.26
- Projekt oficjalnie wspierany przez IBM

Money.pl

- Montowanie systemu plików zarządzania grupami

  mkdir -p /cgroup

  mount none -t cgroup /cgroup

- Konfiguracja sieci za pomocą **bridge**

  brctl addbr br0

  ifconfig br0 192.168.2.1

- Tworzenie i zarządzanie kontenerem

  lxc-create -n vserver1 -f /etc/lxc/vserver1.conf

  lxc-start, lxc-stop, lxc-freeze, lxc-monitor

PODSUMOWANIE
8/11

Money.pl

- Super lekka wirtualizacja bez emulacji urządzeń, narzut mniej niż 2%
- Lepsza konsolidacja i utylizacja serwerów
- Doskonałe rozwiązanie dla hostingu i aplikacji internetowych
- Separacja i organizacja usług przez odrębne instancje oraz katalogi
- Możliwość implementacji dodatkowej podsieci do zarządzania serwerami
- Dowolność przy partycjonowaniu serwerów

PYTANIA ....

Money.pl