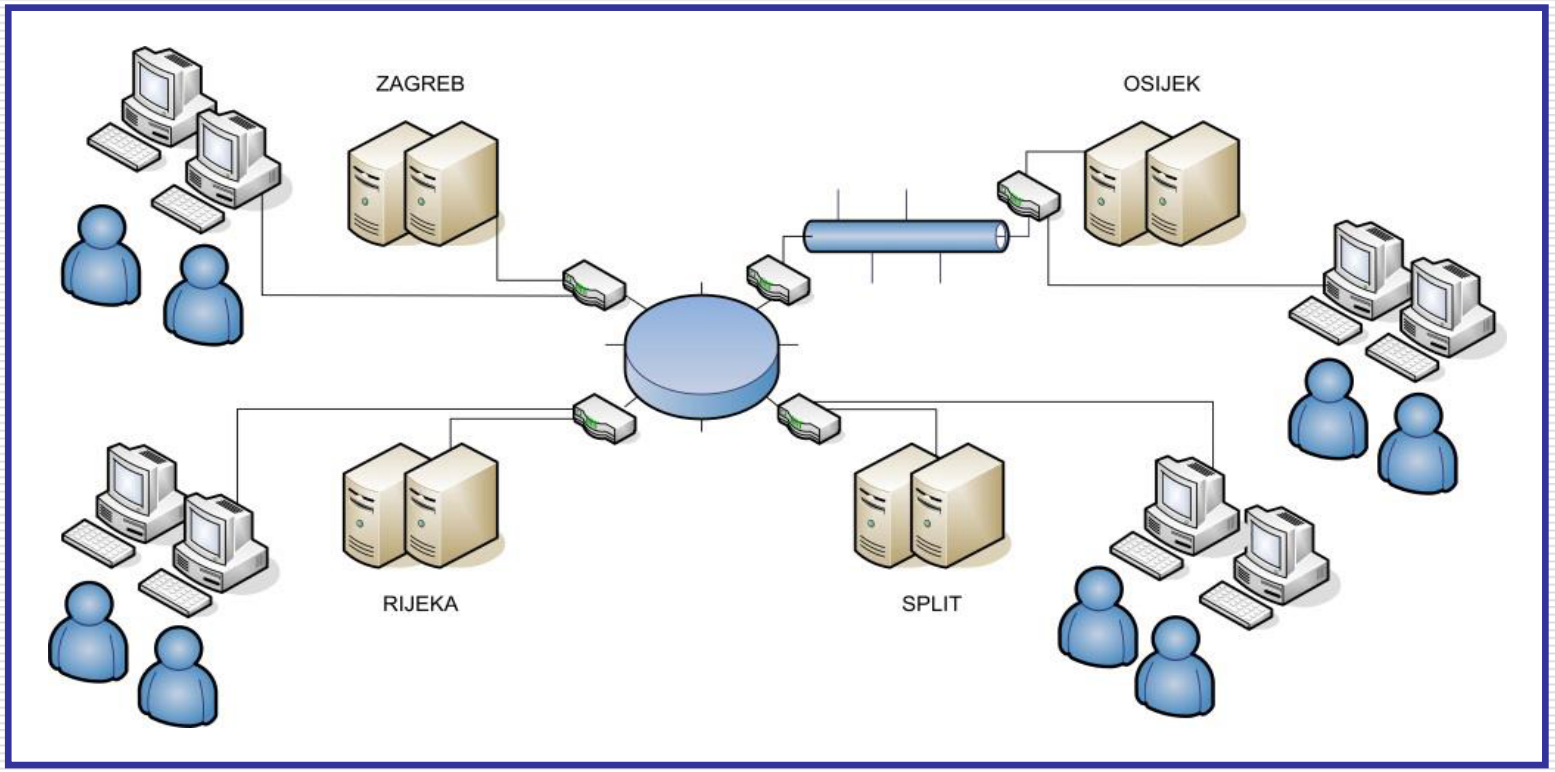
CLUSTER iAS-a NA HYPER-V ZA KAPACITET I RASPOLOŽIVOST

Dubravko Miljković
HEP-SIT
Zagreb

Opis dosadašnje infrastrukture aplikacijskih servera

Centri dosadašnjeg geografski disperziranog clustera aplikacijskih servera

Opis dosadašnje infrastrukture aplikacijskih servera



Mrežna povezanost centara u dosadašnjem rješenju

Opis dosadašnje infrastrukture aplikacijskih servera

- Inicijalno 2 servera po centru – 8 servera
- Kasnije instalacija narasla na 16 servera
  - Zagreb 7 servera
  - Split 5 servera
  - Rijeka 2 servera
  - Osijek 2 servera
- Glomazno rješenje, ali izrazito redundantno
- 10 godina pouzdanog rada

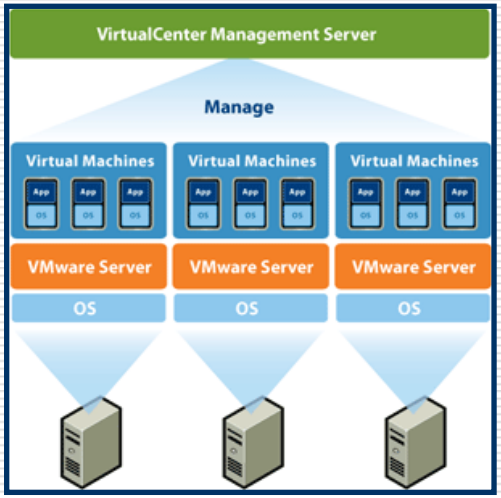
NLB cluster internet aplikacijskih servera (Oracle iAS-a)

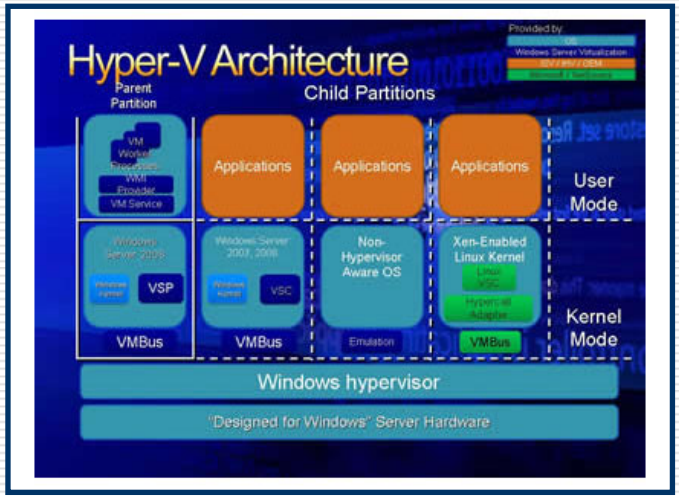
- Dvije Hyper-V platforme (Windows 2012)
  - szg01hvias01
  - szg01hvias02
- 12 iAS-a instaliranih na virtualnim serverima
  - Windows 2008 SP2 operacijski sustav
  - dimenzionirani za do 250 istovremenih korisnika
  - szg01vias01 – szg01vias12

Zašto virtualizacija

- Prvenstveno licenčne prednosti
- Manji broj fizičkih servera
- Manji troškovi održavanja servera
  - Prašinu možete usisati i sami
- Da li u sustavima poput HEP-a ovakve ušteda ima smisla?
  - Svaka ušteda je dobra, ali ih ipak treba staviti u perspektivu

Zašto Hyper-V





- I VMWare i Hyper-V su zreli proizvodi
- iAS-i su instalirani na Windowsima
- Povoljnije licenciranje Hyper-V za Windows 2012 Datacentar Edition
  - Unlimited virtualization rights

Oracle licenciranje

- Dva načina licenciranja
  - Per named user
  - Per CPU
- Virtualizacija
  - Ako se ne koristi Oracle VM, proizvod mora biti licenciran za sve CPU (procesore) na serveru
  - Povoljno npr. kad se koristi jedan proizvod na serveru na svim procesorima
- Alternativa novom rješenju?
  - Povaditi/disable procesore iz servera postojećeg geografski disperziranog clustera

Dvije Hyper-V platforme Hardware

- HP DL380 G8
  - 2 x šesterojezgreni Intel Xeon E5-2640 na 2.5 GHz
  - 96 GB RAM
  - 4 x 146 GB HDD 15k (po 2 diska u RAID 1 polju za OS)
  - 8 x 600 GB HDD 10k (po 4 diska u RAID 5 polju za virtualne iAS-e)
  - 6 mrežnih 1 Gbit Ethernet kartica

Zašto stari iAS 1.0.2?

- Forms/Reports 6
  - Starija tehnologija
  - Prokušano i pouzdano
  - Radi čak i pdf font embedding
  - Portfolio od 50+ aplikacija
- Forms 10i Weblogic 11/12 g
  - Novija tehnologija, ali
  - Zanemariva dobit, velik posao konverzije i testiranja
- Krajnji korisnik praktički ne vidi nikakvu razliku
  - Aktivni grafovi (2D grafovi postaju 3D pod mišem)

Zašto baš 250 istovremenih korisnika?

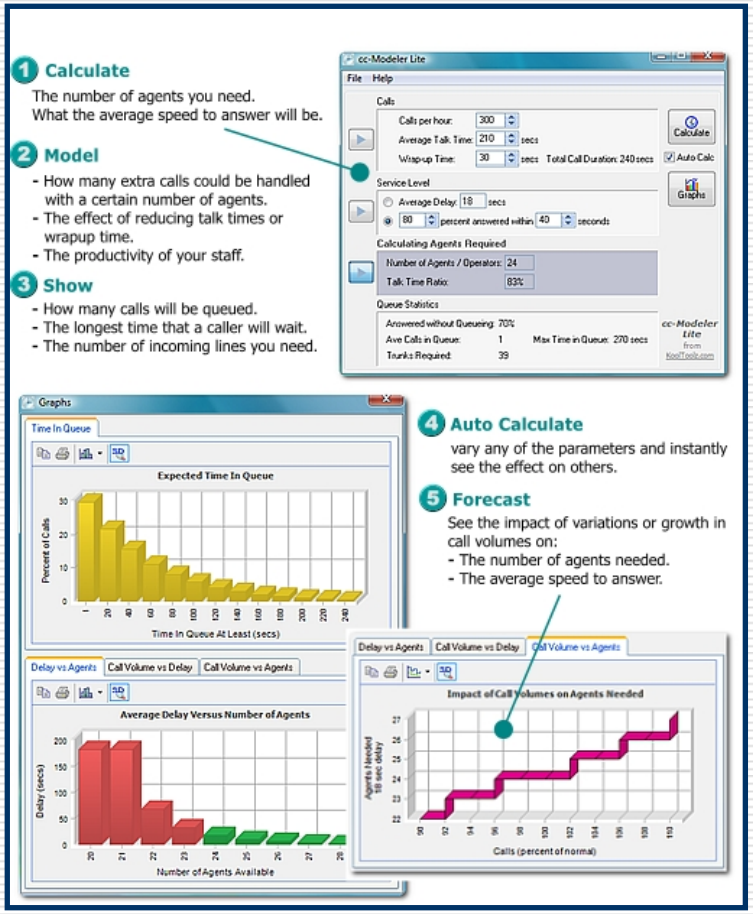
- Forms procesa moglo bi biti i više
  - Windows non-interactive desktop heap size
- In-Process Reports
  - Samo jedan report istovremeno (na Forms/Reports 6)
- Stand Alone Reports Server
  - Proizvoljni broj istovremenih reporta (specificirano sa brojem reports engines)
- Oko 200 po starom programu

Queueing theory za reports server

- In-process reporti
  - Samo jedan engine

    bez uobičajenih parametara
    - initEngine
    - minEngine
    - maxEngine
  - Do 150 istovremenih korisnika nema problema
  - 150 - 200 manje čekanje
  - 200 - 250 veliko čekanje
  - 250+ nikad dočekati

  (ovisi o reportima, naš slučaj)



iAS na Windows 2008

- iAS 1.0.2.2.2a namijenjen za Windows 2000
- Ne da se instalirati na Windows 2008
  - Čak ni primjenom uobičajenih trikova kao Compatibility Mode itd.
- Windows Server 2008 – 32 bit
- Kopirana jedna postojeća instalacija s Windows 2000 Server
  - Preseljena struktura direktorija
  - Izmijenjene sve Oracle datoteke s nazivima stroja
  - Izmijenjen Windows registry
  - I još ponešto sitno

Broj nodova - 12

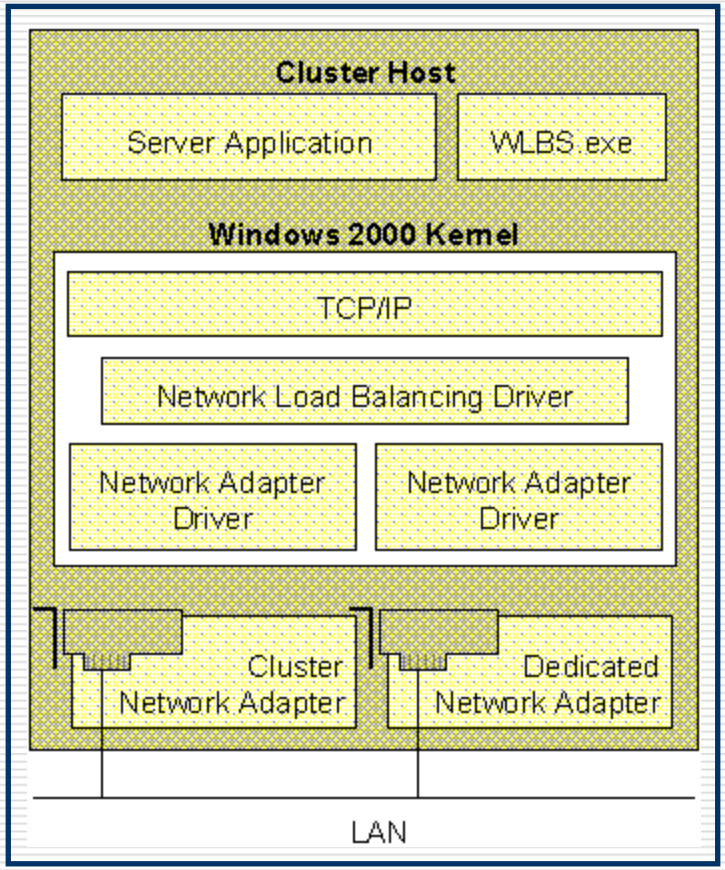
- 2000 istovremenih korisnika
  - 167 po nodu - optimalna performansa
- Max. 3000 istovremenih korisnika
  - 250 po nodu – malo sporiji in-process reporti
- Ispad jednog noda neznatno utječe na cjelokupni kapacitet
  - 8.3% korisnika puca sesija
  - U tipičnoj slučaju samo jedan korisnik u uredu doživljava problem, drugima sustav radi
  - Pomišlja da je problem u njegovom PC-ju
- Ispad jedne Hyper-V platforme sa 6 nodova još uvijek osigurava uslugu za 1500 korisnika

Visoka raspoloživost
High Availability (HA)

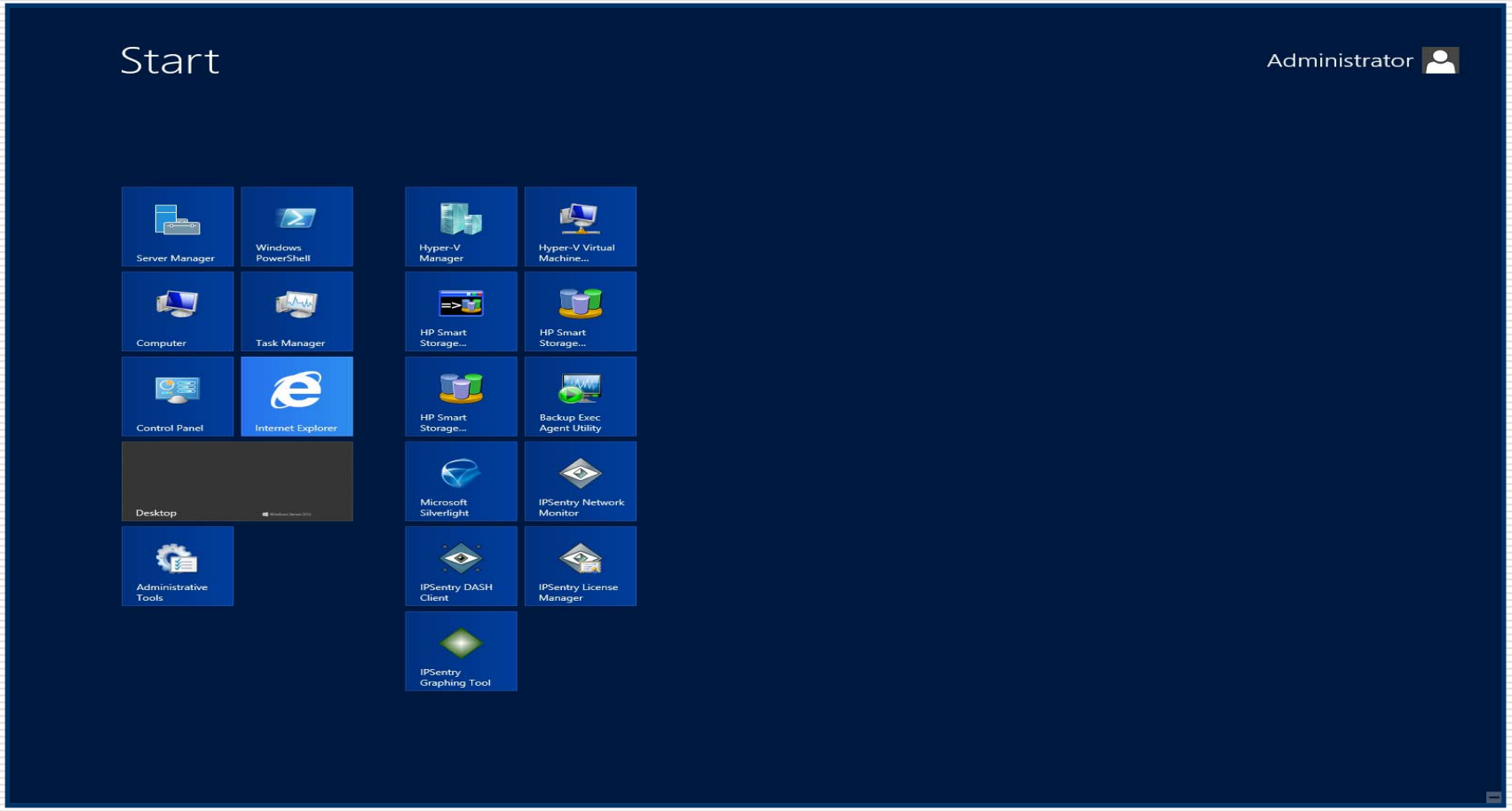
- Uporaba Microsoft NLB (Network Load Balancing)
  - Svi nodovi ravnopravno pristupaju mreži (bus struktura)
  - Ne postoji jedna točka prekida (Single Point of Failure)
  - Korisnici se raspodjeljuju na temelju broja konekcija stohastičkom procedurom
  - Kod velikog broja korisnika opterećenje je ravnomjerno raspoređeno na sve nodove (zakon velikih brojeva)

Visoka raspoloživost
High Availability (HA)

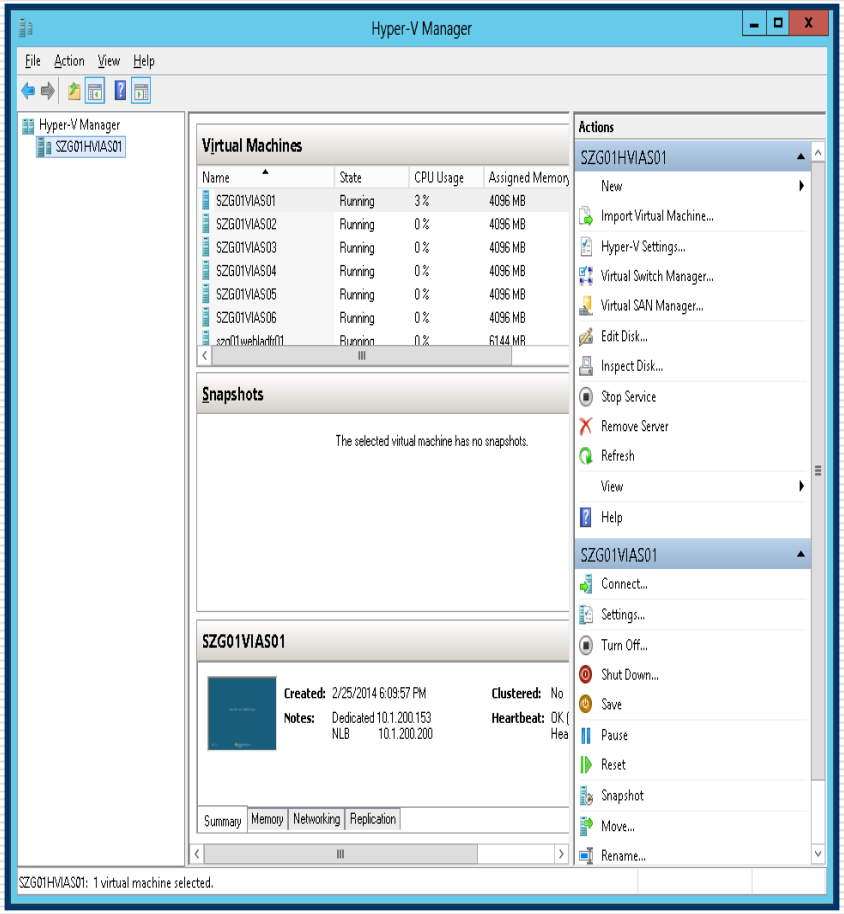
- NLB sa dvije virtualne kartice po virtualnom stroju
  - Unicast mode
- NLB (Cluster) kartica
  - Dolazni promet
- Dedicated kartica
  - Izlazni promet sa virtualnog stroja
- Microsoft NLB je dobro i jeftino rješenja, čak i za goleme promete

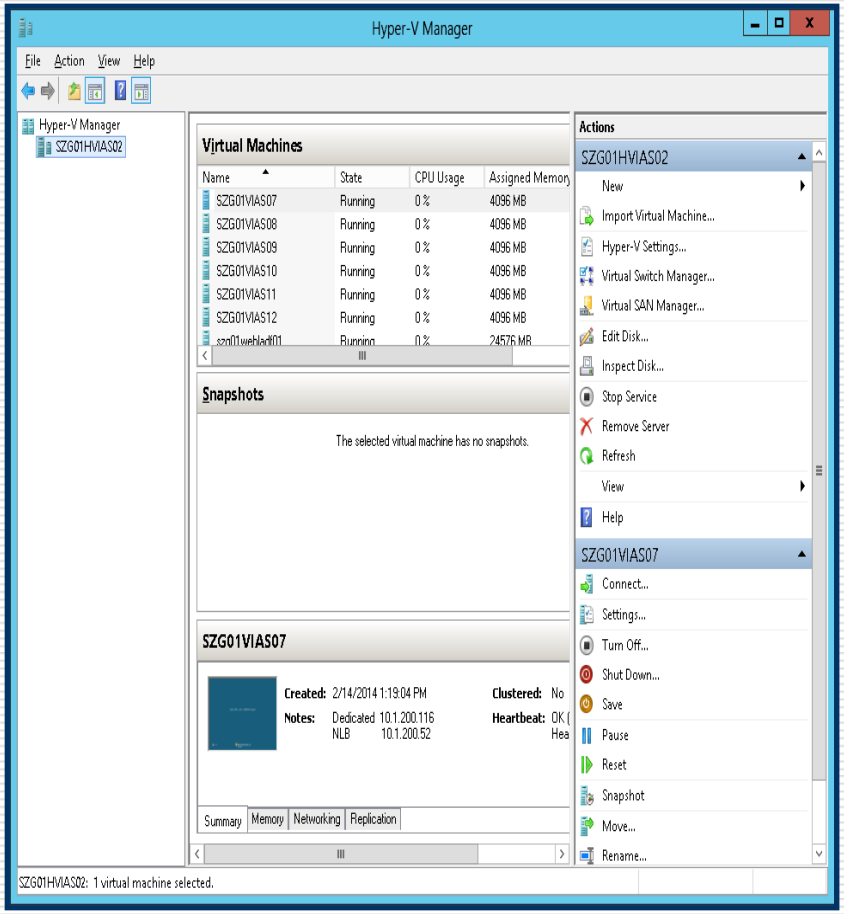


Hyper-V platforma
Windows 2012 Server sa Hyper-V rolom



NLB - po 6 virtualnih strojeva na svakoj Hyper-V platformi: Hyper-V Instanca 1 i Hyper-V Instanca 2



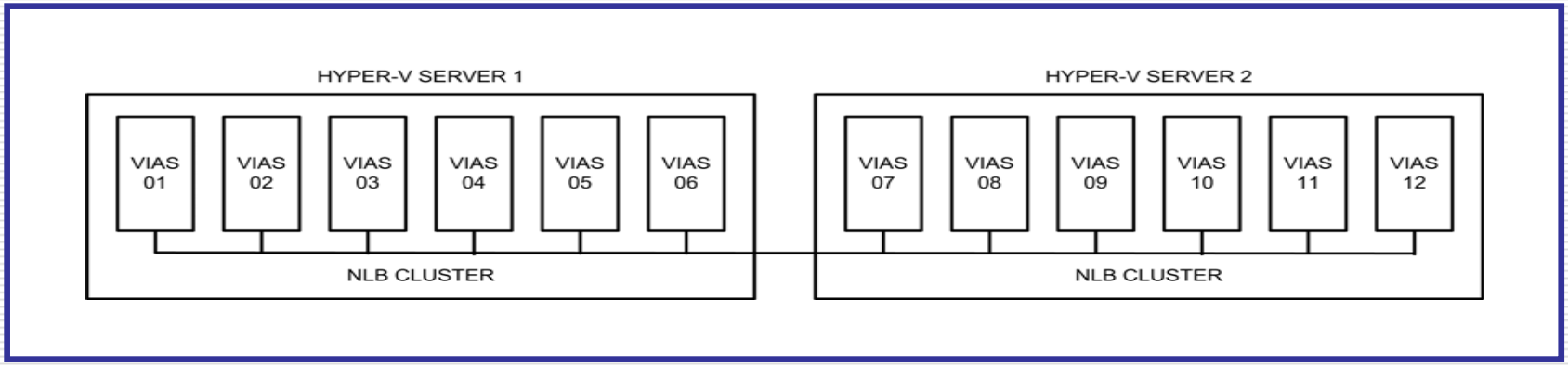


NLB cluster internet aplikacijskih servera (Oracle iAS-a)

- Ukupno opterećenje clustera je oko 1800-2000 istovremenih korisnika
- Maksimalno opterećenje je oko 3000 istovremenih korisnika
  - Uz smanjenu performansu
    - Dulje vrijeme uspostave sesije
    - Dulje čekanje na in-process reporte (jedan engine)
- Privremeno opterećenje 6 iAS-a na jednoj Hyper-V platformi: 1500
- Ako padne jedna Hyper-V platforma 1500 korisnika može se odmah rekonektirati

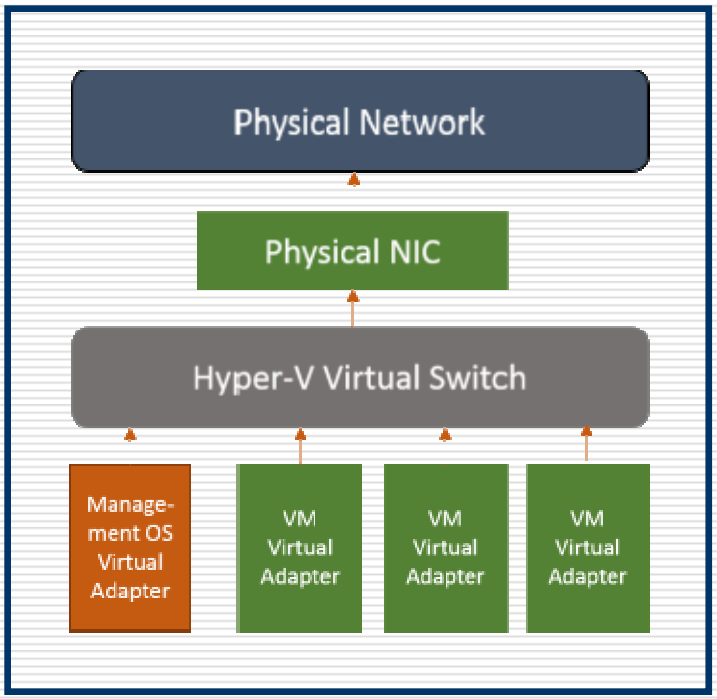
Visoka raspoloživost
High Availability (HA)

- Dvije Hyper-V platforme
- NLB uključuje sve virtualne iAS-e na obje platforme

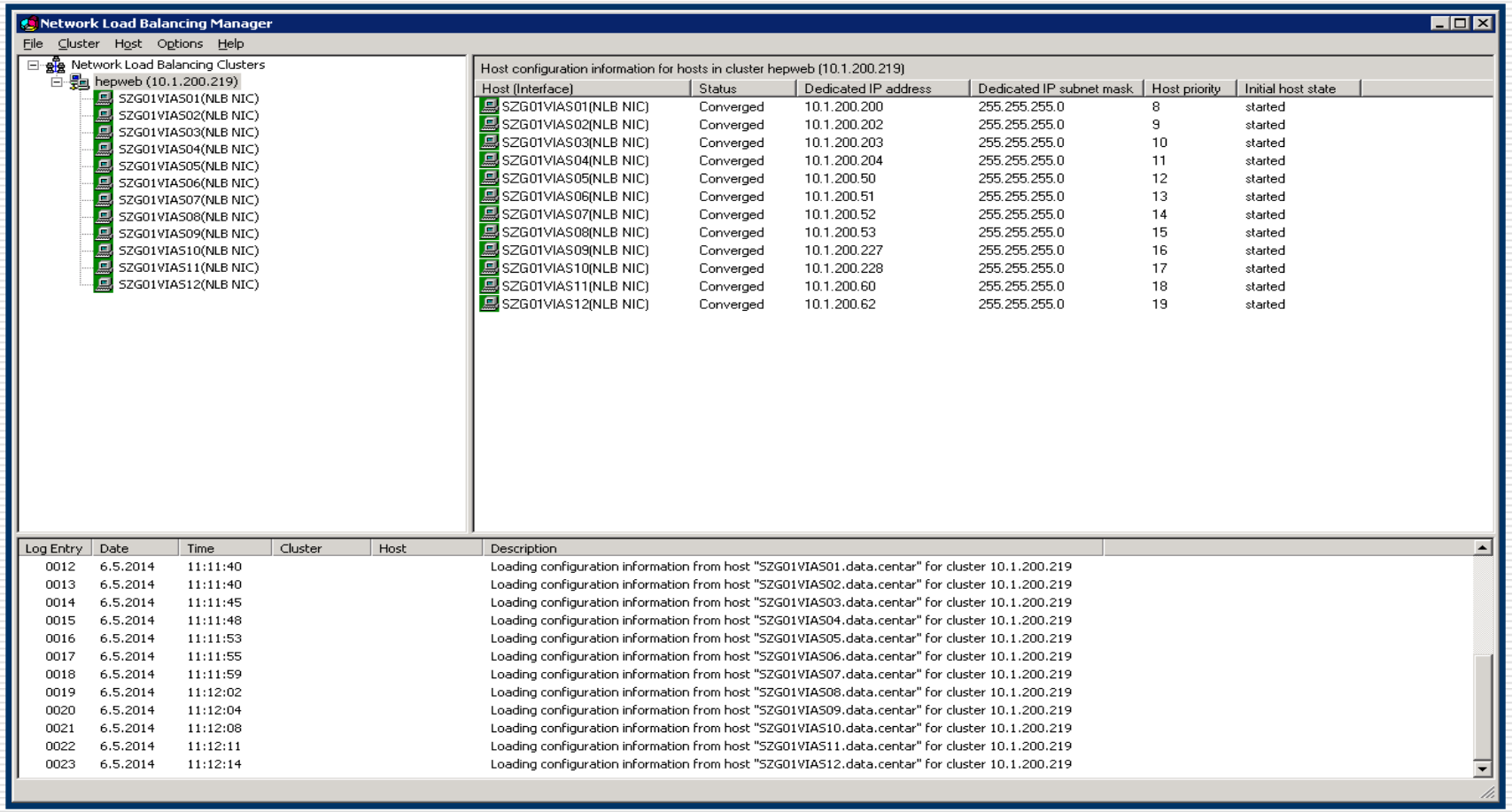


Hyper-V platforma virtual switch

- External virtual switch
  - Osigurava komunikaciju između
    - Fizičke mreže
    - Management operating sustava
    - VM - virtualnih strojeva
  - Spojen na fizičke adaptere
  - Network teaming
- VM – virtualni strojevi
  - Komuniciraju svaki preko dva Virtual NIC-a

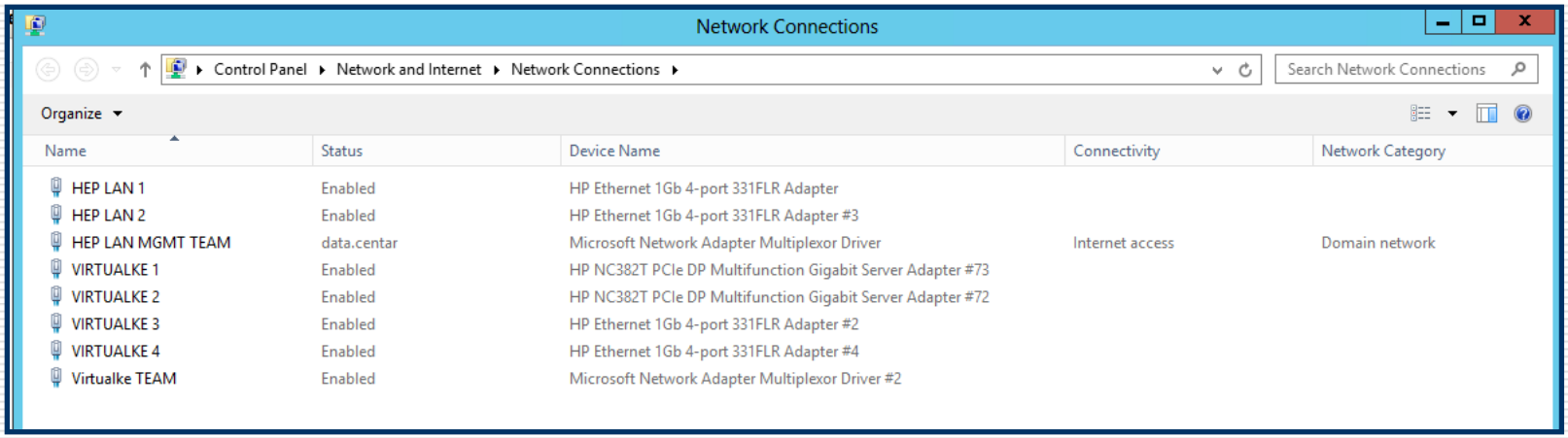


Visoka raspoloživost
High Availability (HA)



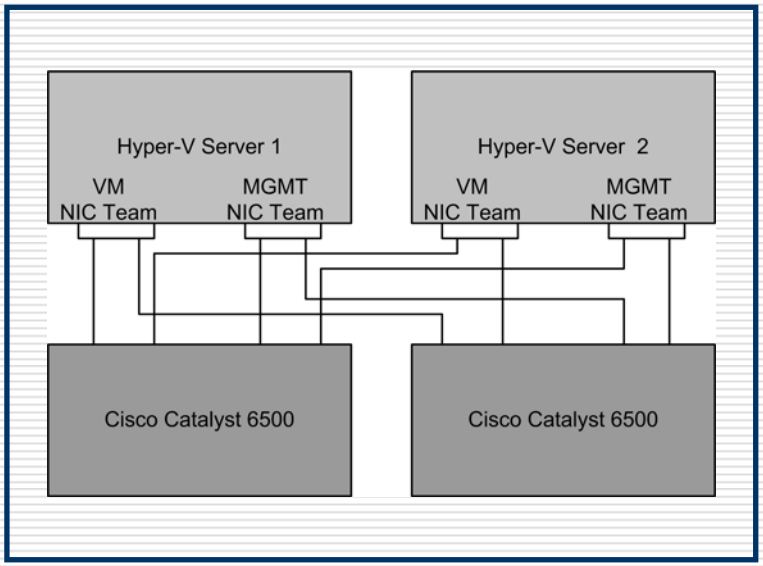
Visoka raspoloživost
High Availability (HA)

- MGMT Team
  - Dvije fizičke kartice
- VM Team
  - Četiri fizičke kartice



Visoka raspoloživost
High Availability (HA)
2 x Cisco Catalyst 6500 Switch

- Fizičke mrežne kartice su u teaming-u
  - Polovica fizičkih kartica spojeno na jedan switch, druga polovica na drugi switch



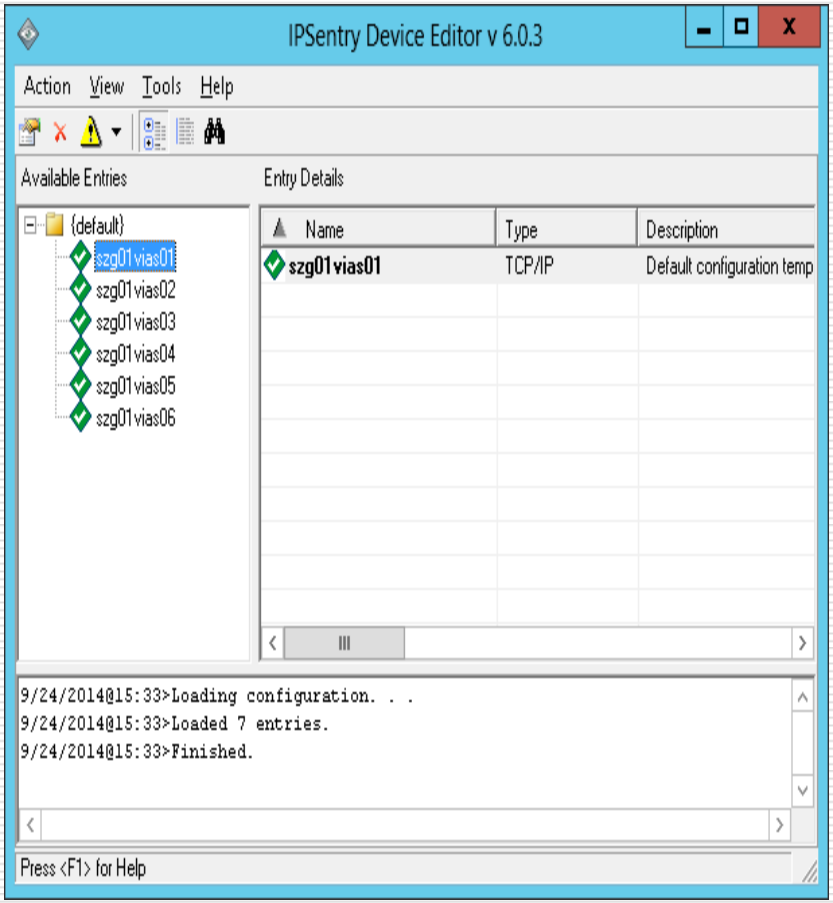
Affinity reset

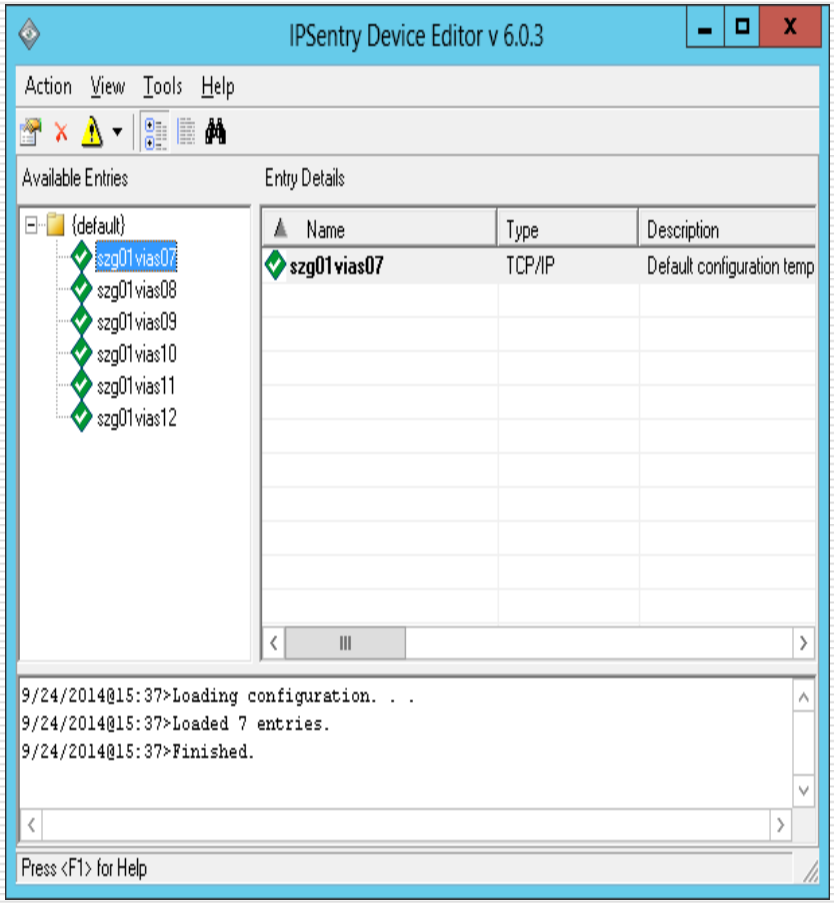
- NLB affinity: single
  - Zasnovan na IP adresi korisnika
  - Korisnik se tijekom sesije spaja na isti nod (i iAS)
- Affinity se gubi nakon zaustavljanja NLB-a
- Wlbs (Windows Load Balancing Service)
  - Start
  - Stop
- Procedura (tijekom noći)
  - wlbs stop
  - Pauza cca. 1 min
  - wlbs start
  - Nakon affinity reset-a svaki korisnik nanovo uspostavlja vezu s nekim od nodova (slučajan izbor)

Nadzor nodova (OS + iAS) - IPSentry

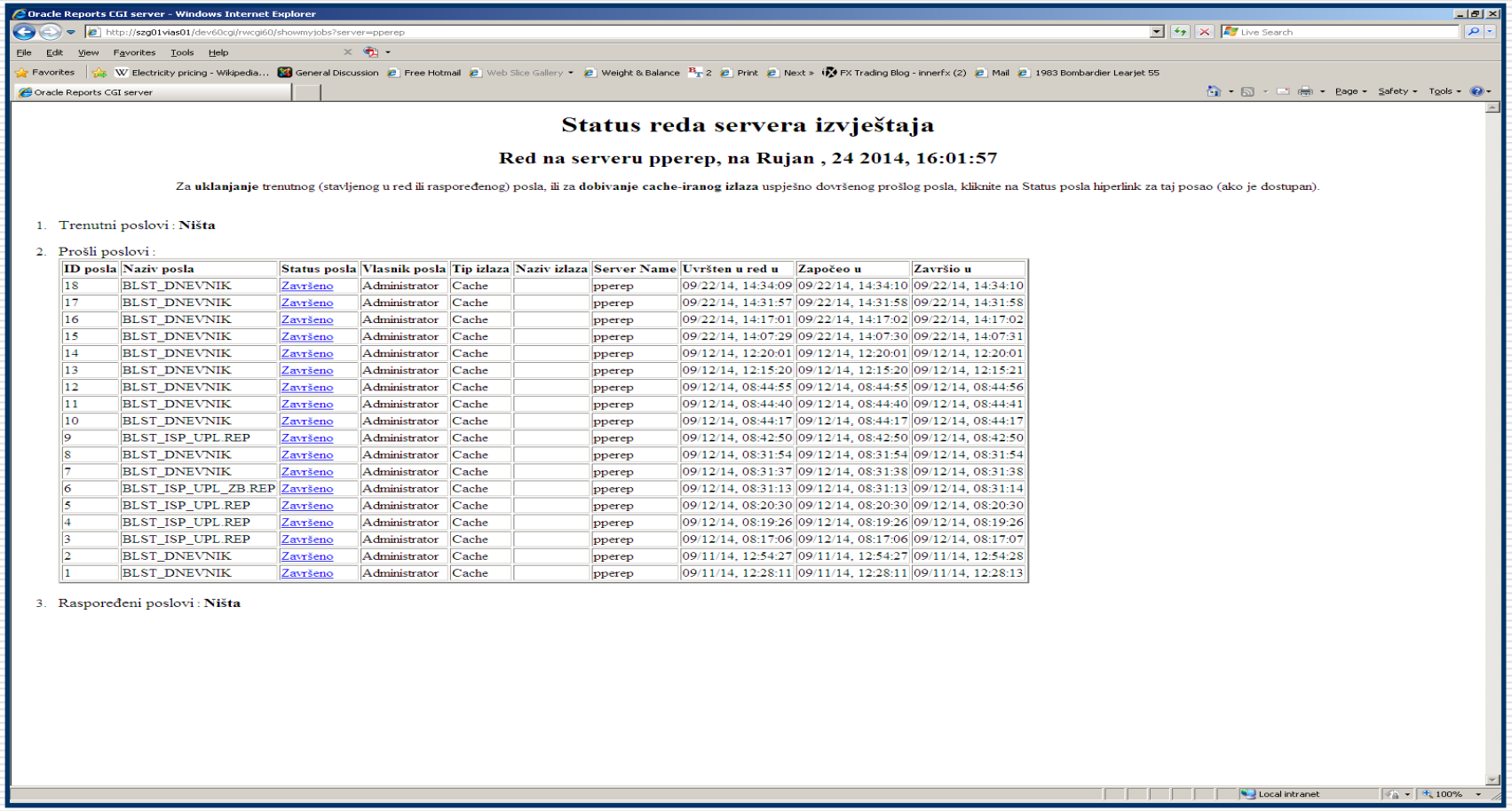
- IPSentry software
  - Uključivanje funkcionalnog i isključivanje nefunkcionalnog noda iz NLB-a
    - wlbs start
    - wlbs stop
  - Inicijalno stanje noda u NLB-u je neaktivno
    - Uključivanje tek kad se utvrdi ispravnost noda
  - Nadzor i restart komponente ili cijelog noda (ovisno o potrebi)
    - Web i Forms (Apache)
    - Reports serveri

Nadzor iAS-a - IPSentry





Status reports servera
nadzor brojnih reports servera



Status reda servera izvještaja

Red na serveru pperep, na Rujan , 24 2014, 16:01:57

Za **uklanjanje** trenutnog (stavljenog u red ili raspoređenog) posla, ili za **dobivanje cache-iranog izlaza** uspješno dovršenog prošlog posla, kliknite na Status posla hiperlink za taj posao (ako je dostupan).

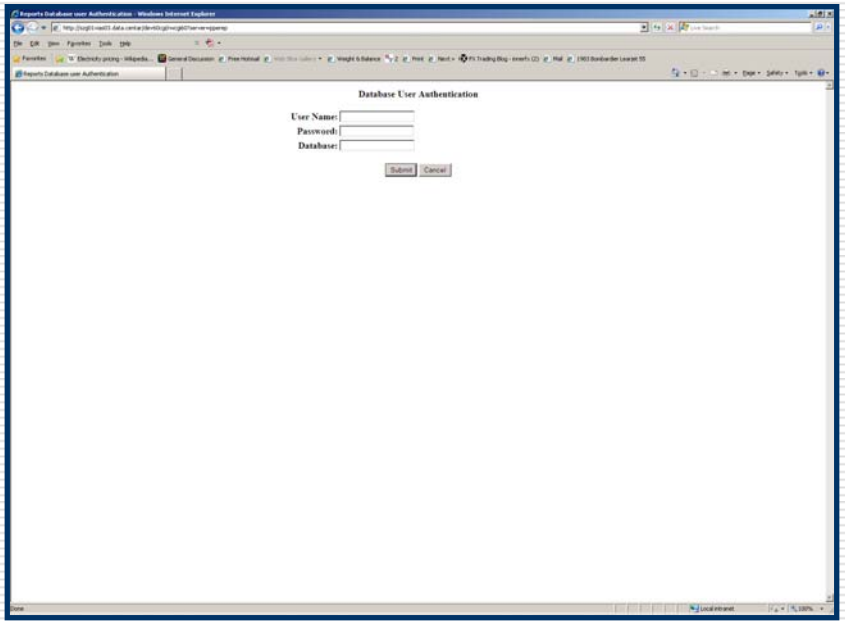
1. Trenutni poslovi : **Ništa**
2. Prošli poslovi :

ID posla	Naziv posla	Status posla	Vlasnik posla	Tip izlaza	Naziv izlaza	Server Name	Uvršten u red u	Započeo u	Završio u
18	BLST_DNEVNIK	Završeno	Administrator	Cache		pperep	09/22/14, 14:34:09	09/22/14, 14:34:10	09/22/14, 14:34:10
17	BLST_DNEVNIK	Završeno	Administrator	Cache		pperep	09/22/14, 14:31:57	09/22/14, 14:31:58	09/22/14, 14:31:58
16	BLST_DNEVNIK	Završeno	Administrator	Cache		pperep	09/22/14, 14:17:01	09/22/14, 14:17:02	09/22/14, 14:17:02
15	BLST_DNEVNIK	Završeno	Administrator	Cache		pperep	09/22/14, 14:07:29	09/22/14, 14:07:30	09/22/14, 14:07:31
14	BLST_DNEVNIK	Završeno	Administrator	Cache		pperep	09/12/14, 12:20:01	09/12/14, 12:20:01	09/12/14, 12:20:01
13	BLST_DNEVNIK	Završeno	Administrator	Cache		pperep	09/12/14, 12:15:20	09/12/14, 12:15:20	09/12/14, 12:15:21
12	BLST_DNEVNIK	Završeno	Administrator	Cache		pperep	09/12/14, 08:44:55	09/12/14, 08:44:55	09/12/14, 08:44:56
11	BLST_DNEVNIK	Završeno	Administrator	Cache		pperep	09/12/14, 08:44:40	09/12/14, 08:44:40	09/12/14, 08:44:41
10	BLST_DNEVNIK	Završeno	Administrator	Cache		pperep	09/12/14, 08:44:17	09/12/14, 08:44:17	09/12/14, 08:44:17
9	BLST_ISP_UPL REP	Završeno	Administrator	Cache		pperep	09/12/14, 08:42:50	09/12/14, 08:42:50	09/12/14, 08:42:50
8	BLST_DNEVNIK	Završeno	Administrator	Cache		pperep	09/12/14, 08:31:54	09/12/14, 08:31:54	09/12/14, 08:31:54
7	BLST_DNEVNIK	Završeno	Administrator	Cache		pperep	09/12/14, 08:31:37	09/12/14, 08:31:38	09/12/14, 08:31:38
6	BLST_ISP_UPL_ZB REP	Završeno	Administrator	Cache		pperep	09/12/14, 08:31:13	09/12/14, 08:31:13	09/12/14, 08:31:14
5	BLST_ISP_UPL REP	Završeno	Administrator	Cache		pperep	09/12/14, 08:20:30	09/12/14, 08:20:30	09/12/14, 08:20:30
4	BLST_ISP_UPL REP	Završeno	Administrator	Cache		pperep	09/12/14, 08:19:26	09/12/14, 08:19:26	09/12/14, 08:19:26
3	BLST_ISP_UPL REP	Završeno	Administrator	Cache		pperep	09/12/14, 08:17:06	09/12/14, 08:17:06	09/12/14, 08:17:07
2	BLST_DNEVNIK	Završeno	Administrator	Cache		pperep	09/11/14, 12:54:27	09/11/14, 12:54:27	09/11/14, 12:54:28
1	BLST_DNEVNIK	Završeno	Administrator	Cache		pperep	09/11/14, 12:28:11	09/11/14, 12:28:11	09/11/14, 12:28:13

3. Raspoređeni poslovi : **Ništa**

Status reports servera
query string

- Funkcionalni query
  (obična provjera Windows servisa nije dovoljna):
  - http://szg01vias01/dev60cgi/rwcgi60?server=pperep
- Check response:
  - Npr. string "Database"
- Korektivna akcija
  - Restart reports servera
  - Mail notifikacija u slučaju neuspjelog restarta
    - Moguće kill-ati reports process i potom ga startati
    - rwmts60.exe



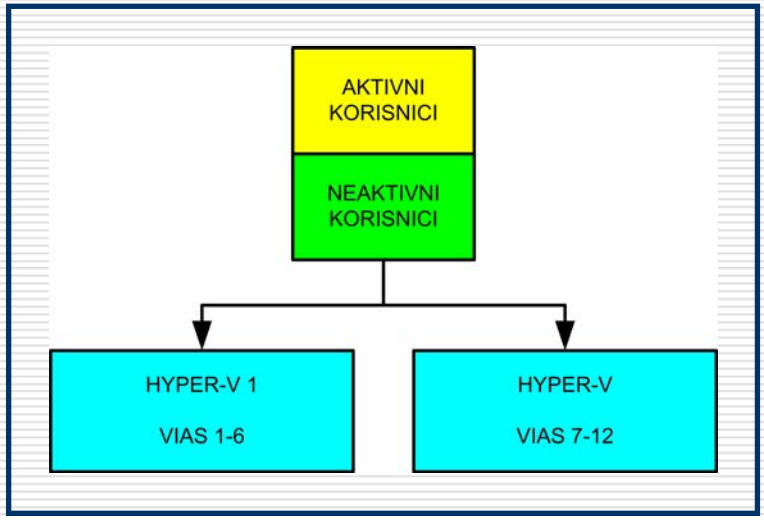
Restart pojedinog noda

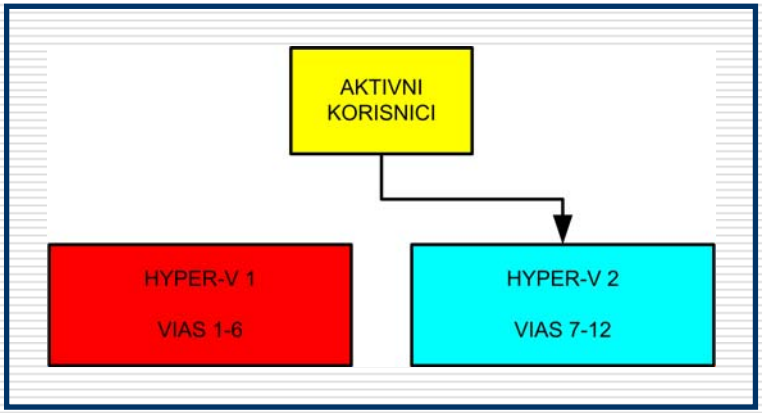
- Distribucija opterećenja
  - Ravnomjerno na 12 nodova (Load weigh - Equal)
  - Kod pada jednog noda preostaje 11 nodova (91.6%)
- Tranzijentni kvar
  - Nod se nakon automatskog restarta ponovo uključuje u cluster
  - Email notifikacija
- Permanentni kvar
  - Nod se ne uključuje u cluster (ozbiljniji kvar)
  - Email notifikacija

Pad jedne virtualne platforme

- Korisnici spojeni na iAS-e instalirane na dotičnoj platformi
  - Pucaju sesije
  - Korisnici se trebaju se rekonektirati
- Platforma se restarta kroz nekoliko minuta
- Dio korisnika je spojen na aplikaciju, i aktivan je
  - Rekonektiraju se odmah
- Dio korisnika je spojen na aplikaciju, ali nije aktivan
  - Ne rekonektiraju se odmah

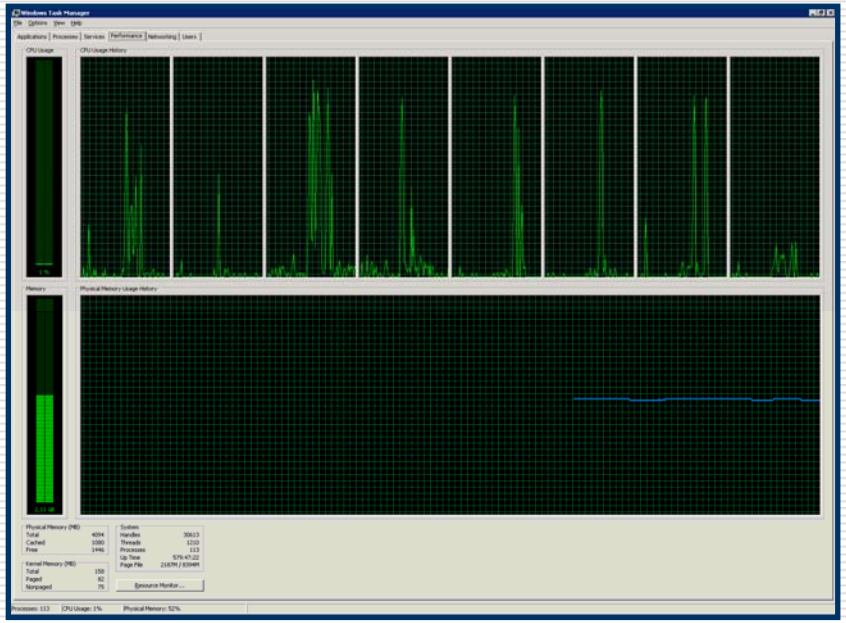
Pad jedne virtualne platforme

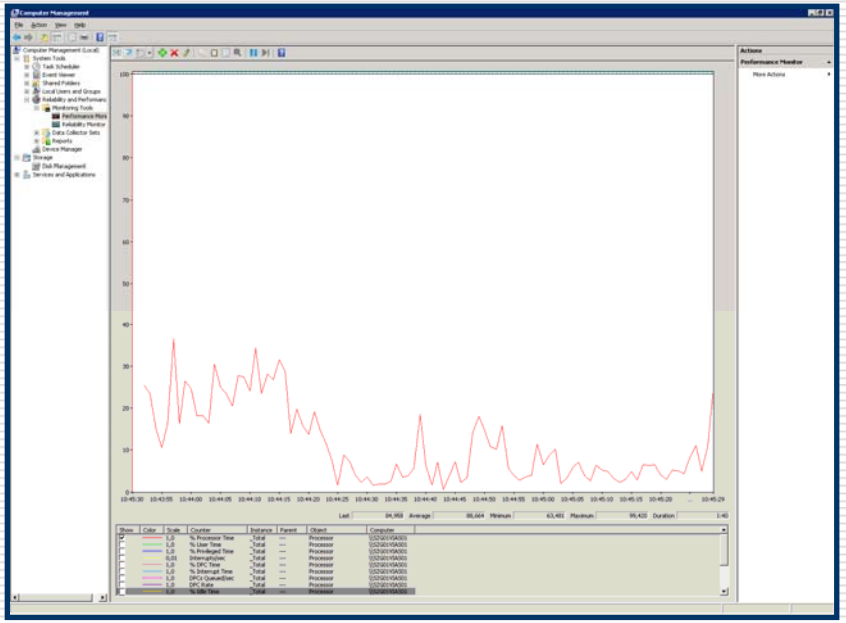




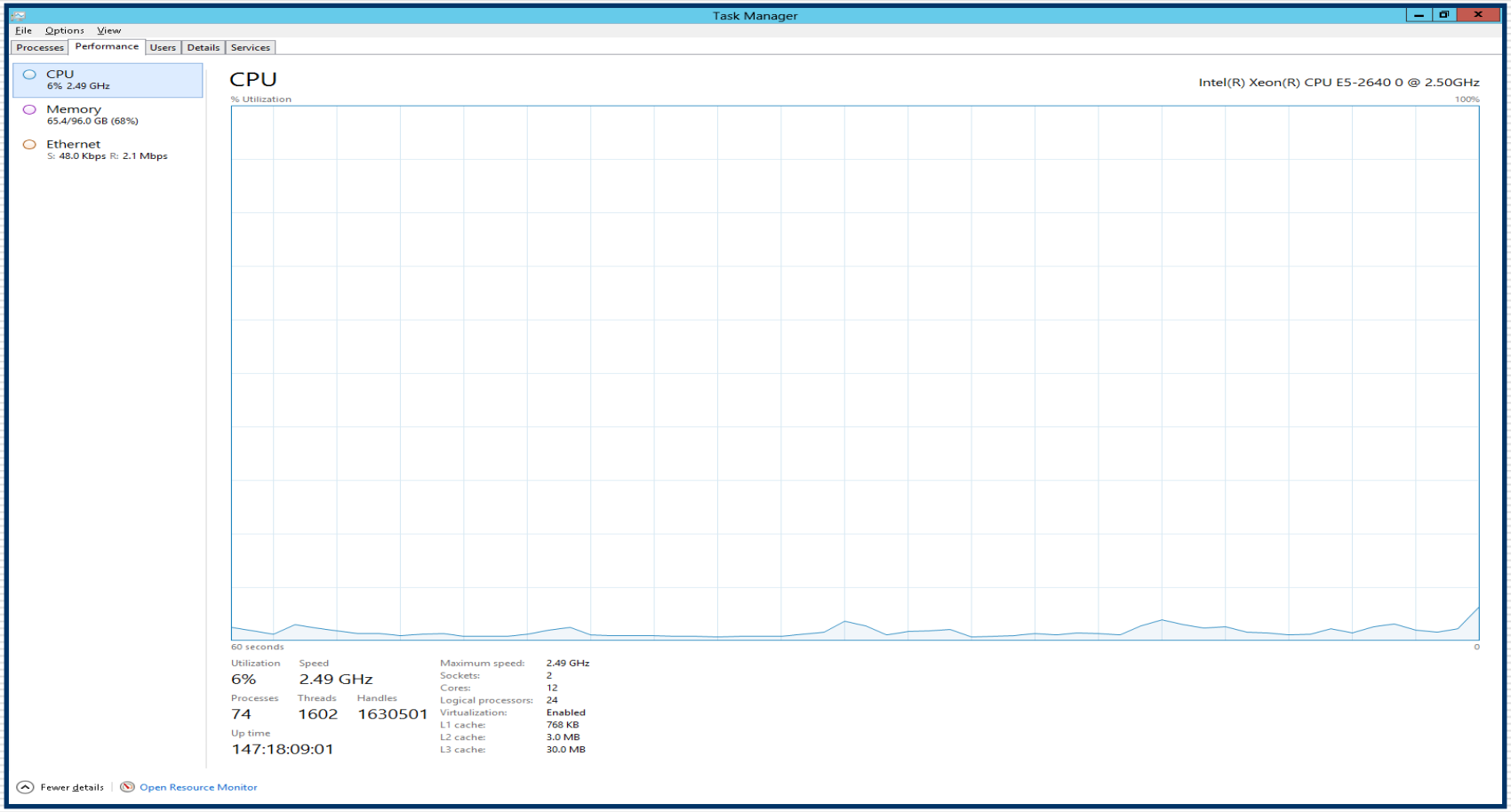
- Dio korisnika je loggiran, ali neaktivan
- Neaktivni korisnici neće se odmah spojiti na drugu virtualnu platformu
- Jedna virtualna platforma zadovoljava međuperiod dok se druga platforma restartom ne osposobi

Opterećenje virtualnog noda - iAS

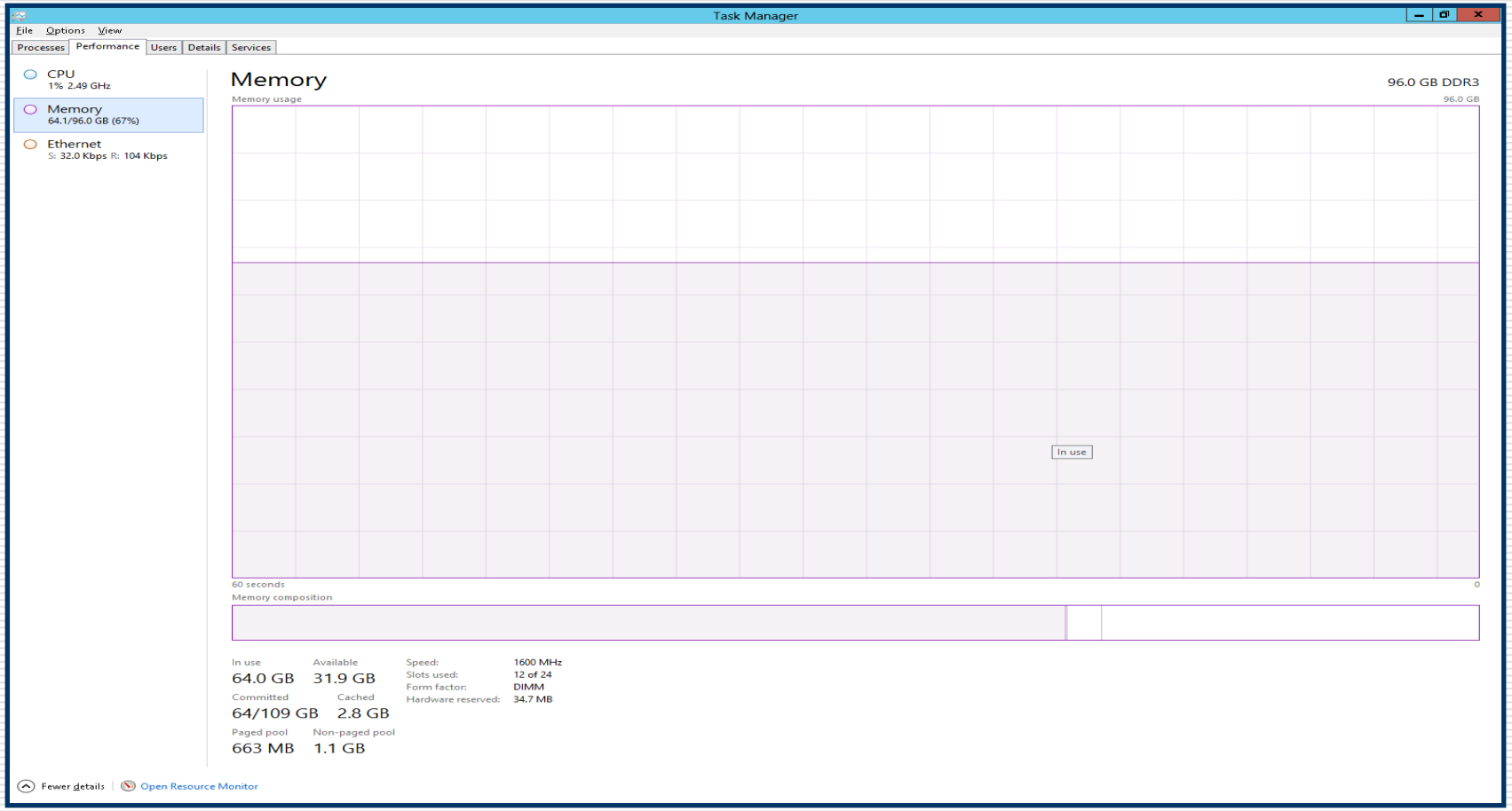




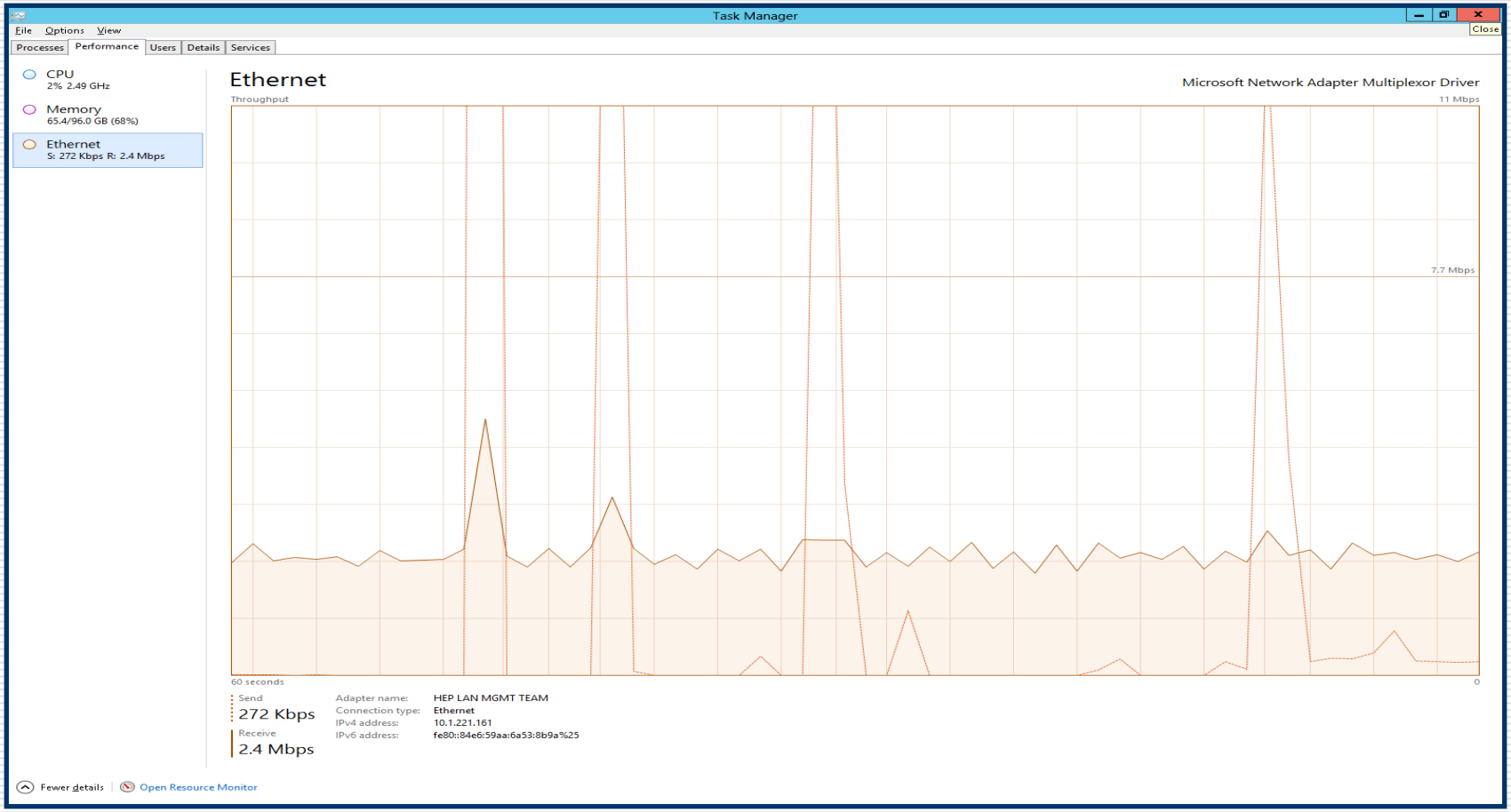
Opterećenje Hyper-V platforme - CPU



Opterećenje Hyper-V platforme - RAM

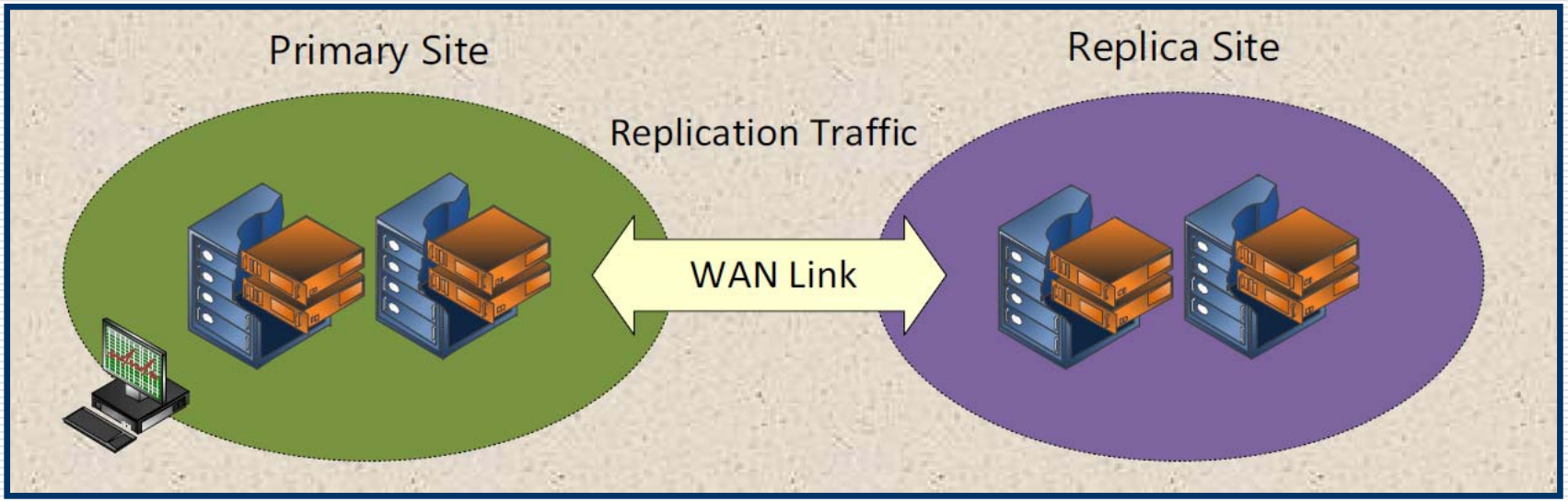


Opterećenje Hyper-V platforme - Ethernet



Disaster Recovery (DR) rješenje

- Primary site - Zagreb - 2 servera: HA
- Replica site - Split - 1 server



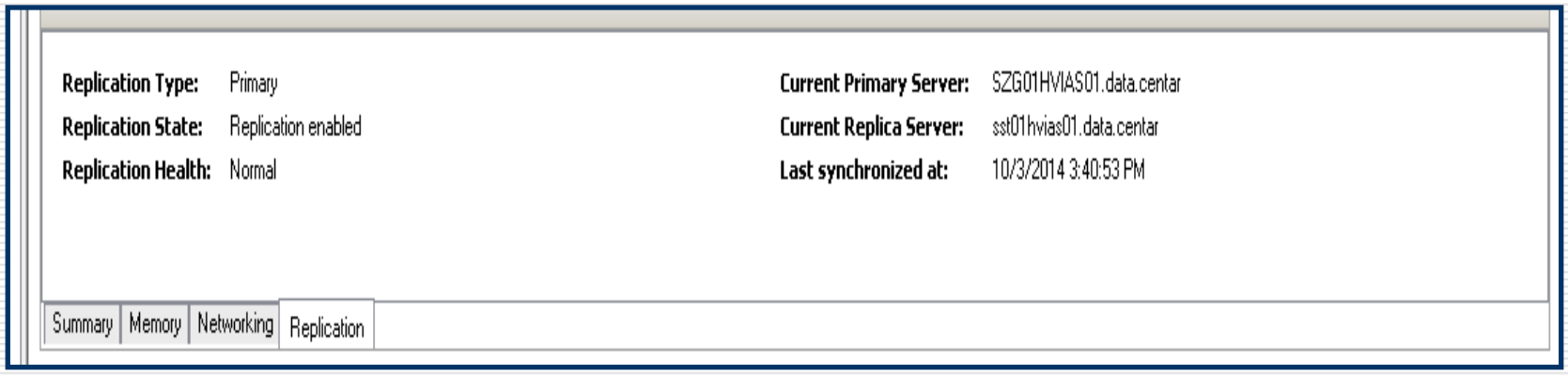
Disaster Recovery (DR) rješenje

Hardware na sekundarnoj lokaciji Split

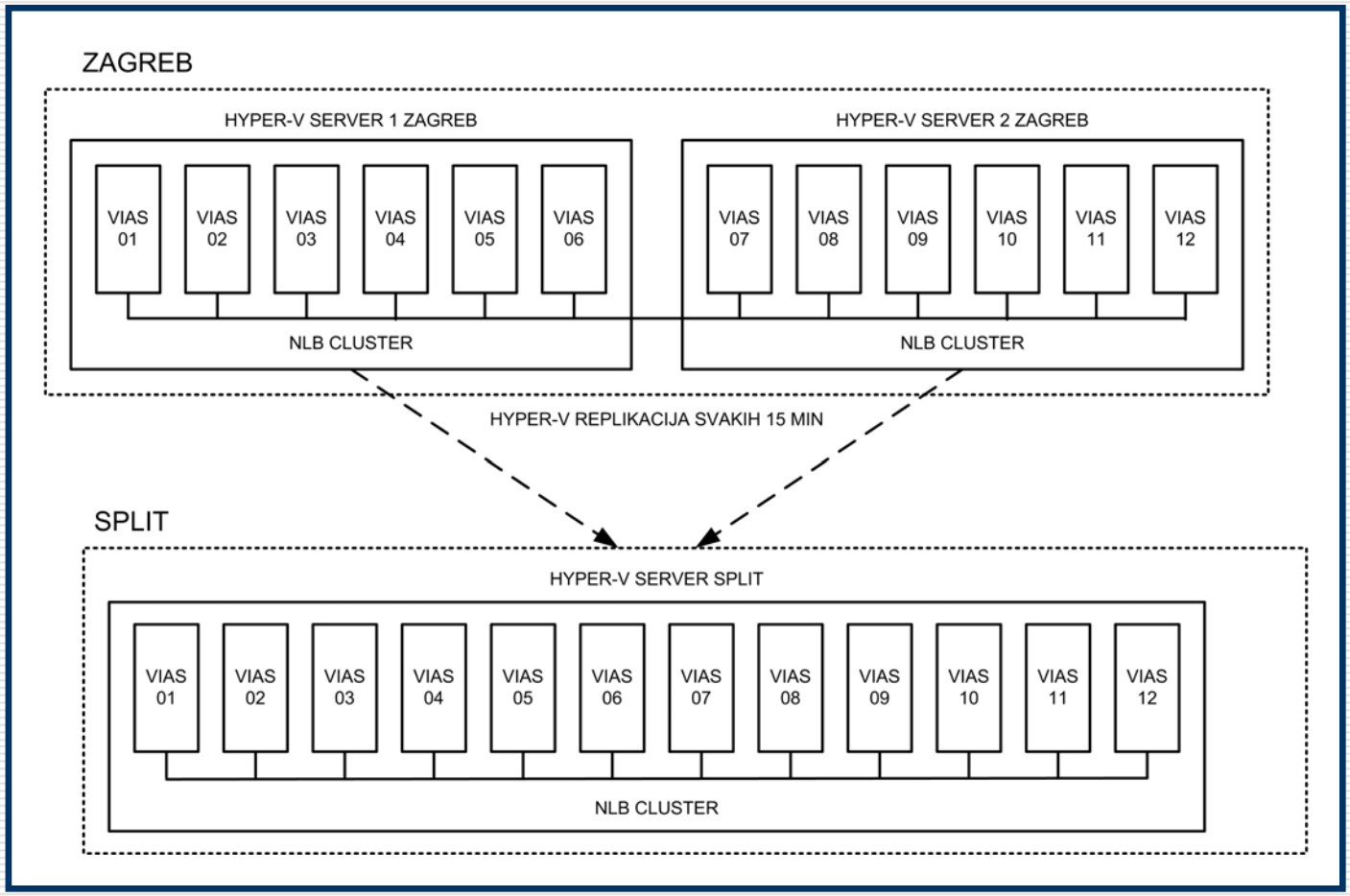
- Server
  - SST01HVIAS01
  - HP Proliant DL380G7
- CPU
  - 2 x Intel Xeon 6 jezgri
- RAM
  - 48 GB
- Diskovi
  - 6x600 GB
  - 3x146 GB
- 6 mrežnih 1 Gbit Ethernet kartica

Disaster Recovery (DR) rješenje

- Replicira svih dvanaest virtualnih servera na jednu Hyper-V platformu u Splitu
  - svakih 15 minuta
  - VM u Splitu spuštene (Off)
- U slučaju potrebe cijeli NLB cluster iAS-a radio bi iz lokacije Split sa nešto nižom performansom
- Ručna intervencija u Splitu – podizanje VM



Disaster Recovery (DR) rješenje

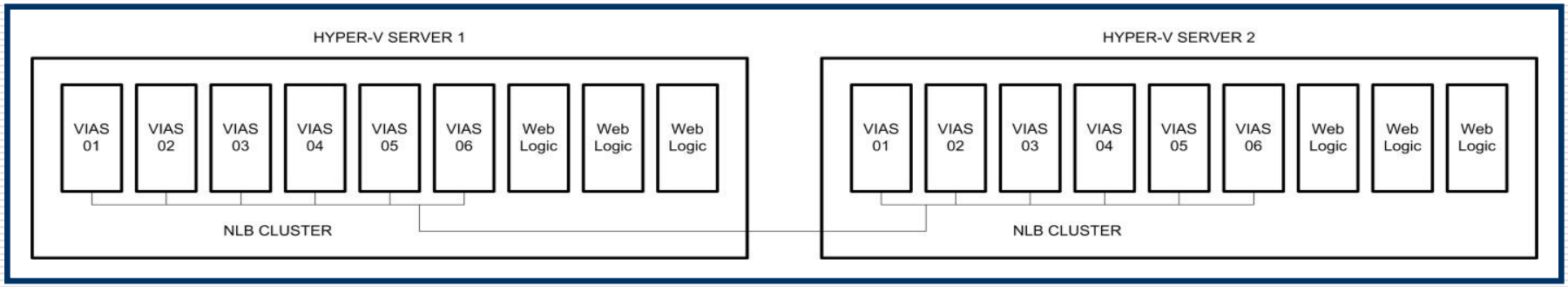


Disaster Recovery (DR) rješenje

- Drugi IP adresni range u Splitu
  - Hyper-V replike s uključenom alternate IP address
- Zagreb
  - IP adrese 10.1.200.53-10.1.200.204
- Split
  - IP adrese u rangu 10.129.64.44-10.129.64.49 za replike novih virtualnih strojeva na Hyper-V platformi u SIT Split
- Nakon podizanja VM na Hyper-V platformi u Splitu potrebna ručna izmjena adrese clustera u DNS-u HEP-a

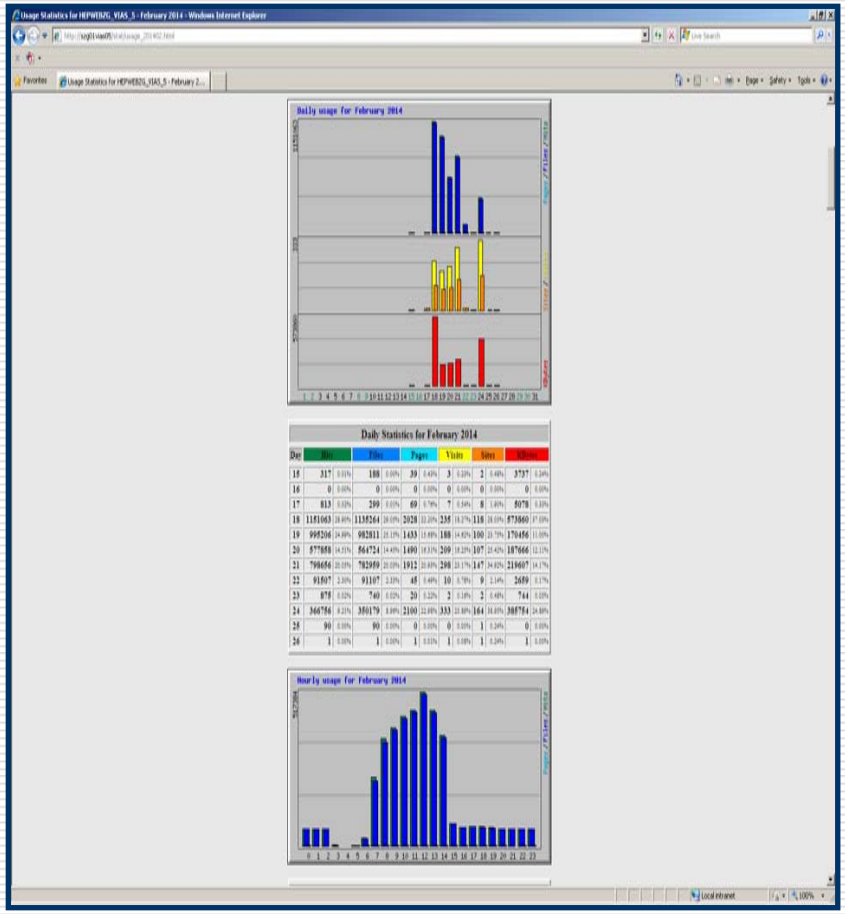
Ostali aplikacijski serveri instalirani na Hyper-V platformama

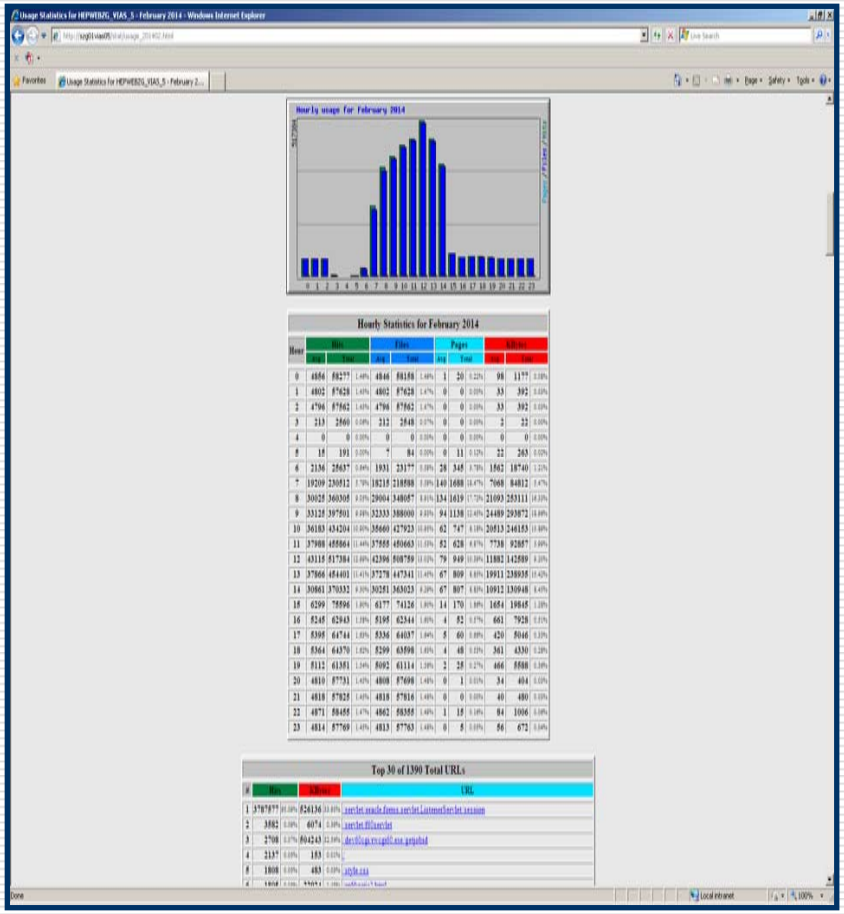
- Da bi se maksimalno iskoristile prednosti Oracle per CPU licenciranje
  - Tri dodatna WebLogic servera na svakoj Hyper-V platformi



Statistika uporabe – Webalizer

Potrebna integracija logova sa svih nodova





Replikacija i update aplikacija

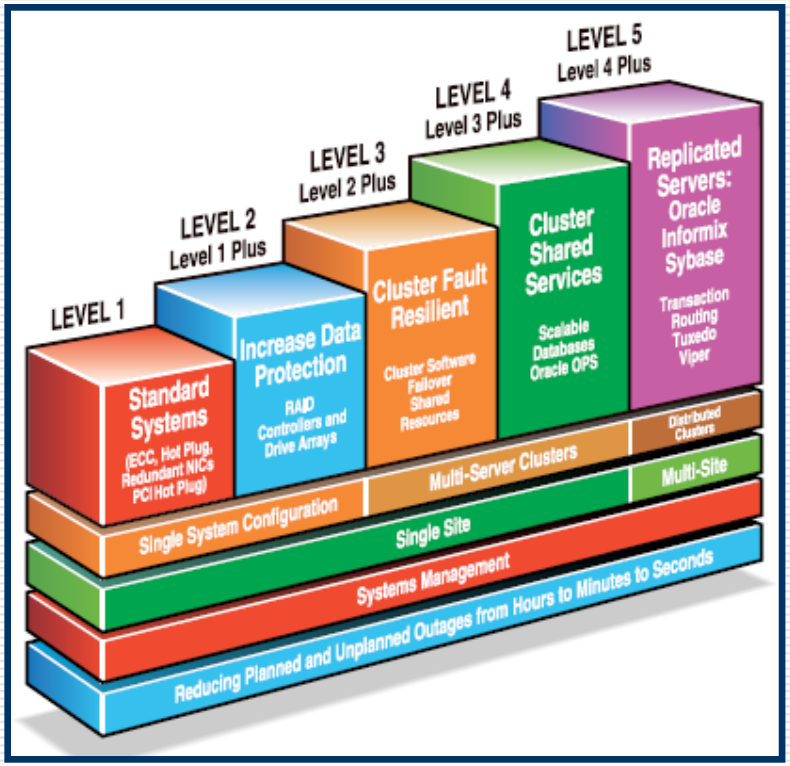
- Replikacija na sve nodove clustera
  - Dvorazinsko: Hyper-V platforma > VM na platformi
  - Kopiranje unutar 2 minute
- Update koda zajedničkih aplikacija
  - Forms - *.fmx
  - Menus - *.mmx
  - Reports - *.rep
  - Tijekom noći - vrlo mali broj ili bez korisnika
- Kill svih ifweb60.exe na svim nodovima
  - ifweb60 proces "drži" formu (lock)
- Lock riješeno na verzijama 10g+, ali
  - Neučinkovito koristi RAM
  - Stoga nije pogodno za produkciju (samo razvoj)

Raspoloživost

Hyper-V NLB cluster

Geografski disperzirani cluster

Raspoloživost	Godišnje vrijeme van pogona*
99%	87.6 sati
99.9%	8.76 sati
99.99%	52.5 minuta
99.999%	5.25 minuta



*Kvarovi i planirano održavanje

Airbus A320

Cessna 172

Primjer usporedbe dva tehnička sustava s različitom konstrukcijom, redundancijom i mogućnostima

- Da li je A320 siguran?
  - Accident rate 0.08/1.000.000 flights
- Da li je Cessna 172 sigurna?
  - Accident rate 5/1.000.000 flight hours
- Airbus A320 oko 50 puta sigurniji od Cessne 172
- Na malom uzorku nema razlike
  - Kroz kraći period oba zrakoplova funkcioniraju podjednako pouzdano
- Na velikom uzorku razlika je ogromna
  - Dulji period i velika flota

Iskustva

- Da li je cluster iAS-a na Hyper-V instaliran na samo dva fizička servera pouzdan kao geografski disperzirani cluster?
  - Sigurno ne!
  - Geografski disperzirani cluster ima raspoloživost i do 99.999% (**overkill???**)
- Da li zadovoljava trenutno stanje
  - Da, pogotovo ako smo zadovoljni s raspoloživosti reda 99.9%
- Da li se ruši?
  - Zasad ne!
  - Poneki otkaz Hyper-V management konzole

Iskustva

- Hyper-V umjesto geografski disperziranog clustera?
  - Zanimljivo za male i srednje poslovne sustave gdje je potreban veliki kapacitet
  - Nema dovoljno resursa za izgradnju geografski disperziranog clustera (npr. bez regionalnih centara)
- Štednja na licencama?
  - Zanimljivo opet za male i srednje poslovne sustave
  - Manje smisla u velikim poslovnim sustavima (zanemariva relativna ušteda u usporedbi s velikim drugim troškovima)
- Centralizacija sustava
  - U realnost i centralizirana administracija

Zaključak

- Pogodno za visok kapacitet
  - Tisuće istovremenih korisnika
- Pogodno za manje i srednje velike poslovne sustave bez više regionalnih centara
- Per CPU Oracle licenciranje
  - Troškovi licenciranja su stvarno niži
- Malo manje pogodno za visoku raspoloživost
  - No free lunch – jeftinije licenciranje, niža raspoloživost
  - Raspoloživost još uvije dovoljna za većinu namjena, reda 99.9% (mogući ispadi reda nekoliko sati godišnje)
  - Ipak otprilike 10-100 puta manja raspoloživost od geografski disperziranog clustera (99.99% – 99.999%)

HVALA NA PAŽNJI!