

Poređenje performansi rada RAID-1 i pojedinačnog diska u LINUX okruženju

B. Đorđević, and V. Timčenko

Sadržaj — Rad predstavlja ispitivanje hardverskih RAID rešenja pod Linux operativnim sistemom. Rad uključuje komparaciju performansi jednog od najpopularnijih RAID nivoa, RAID-1, u odnosu na pojedinačni disk, pod Linux kernel verzijom 2.6. Performanse se mere korišćenjem Postmark benchmark programa.

Ključne reči — RAID, journaling tehnika, performanse, sistem datoteka, ext3.

I. UVOD

RAID je kombinacija više manjih, jeftinih diskova uređenih u niz diskovih polja. Osnovna karakteristika RAID rešenja su poboljšanje performansi kao i veći kapacitet u odnosu na jedan disk. Osnova poboljšanja performansi kod RAID je tehnika deljenja podataka između različitih diskova kao i paralelno čitanje i upisivanje na više diskova [1].

Osim toga, RAID se odlikuje visokom pouzdanošću a fault-tolerance se postiže preko redundantnog skladištenja podataka primenom tehnika mirroring ili tehnike parnosti diskova. RAID se definiše kroz šest osnovnih nivoa arhitekture, počev od RAID-1 do RAID-6, pri čemu svaka od ovih arhitektura obezbeđuje fault-tolerance, poboljšanje performansi kao i niz drugih mogućnosti. Ovi nivoi su osnova za mnoge druge ugnežđene RAIDX+Y nivoe koji se mogu realizovati. Ugnežđeni RAID1+0 nivo je trenutno najpopularniji među RAID konfiguracijama. Postoje dva pristupa realizaciji RAID arhitekture: hardverska (Hw-RAID) i softverska (Sw-RAID). Svaka od ovih realizacija nudi specifične mogućnosti i performanse.

U poslednjih deset godina došlo je velikih poboljšanja u arhitekturi računara a s aspekta RAID tehnologije to se ogleda u povećanoj brzini okretanja diskova, od 10K-15K okretaja, skraćenim vremenom pozicioniranja, poboljšanim mehaničkim kašnjenjima i boljim tehnikama keš baferisanja. Osim toga, na tržištu se mogu naći visoko kvalitetni disk interfejsi, naročito su cenjeni SAS i FC, kojima se postiže prenos podataka od 3 Gb/sec ili 4 Gb/sec. Takođe su poboljšane brze systemske magistrale kao što su paralelna PCI i serijska PCI-express. Paralelna PCI magistrala je 64-bit systemska magistrala zasnovana na taktu od 64MHz csa teorijski mogućim protokom od

533 MB/sec. PCI express je serijska magistrala komutacionog tipa sa visoko kvalitetnim gigabitnim linkovima.

Linux je moderan, sofisticiran i moćan operativni sistem. Novije verzije Linux kernela uključuju podršku za rad sa visoko performansnim journaling sistemima datoteka, poput 32bitnih ext3 i ReiserFS, i 64bitnih XFS i JFS sistema datoteka razvijenih u Silicon Graphics i IBM laboratorijama respektivno[2], [3], [4], [5].

II. CILJ I MOTIV RADA

Cilj ovog rada je da se opiše rad i performanse RAID nivoa-1, kao i da se uporede njihove brzine rada u odnosu na rešenja koja nisu zasnovana na RAID tehnologiji, tj na pojedinačni disk. U odnosu na rešenja koja ne podrazumevaju upotrebu RAID nivoa, RAID nudi visoku pouzdanost i hot-swap mogućnosti, tj. "vruću zamenu" diskova. Osim toga, mogućnost paralelnog upisa i čitanja sa više diskova dodatno poboljšava kvalitet performansi RAID rešenja. RAID tehnologija se zasniva na implementaciji dve tehnike obrade podataka "Parity" i "Mirror" (check/generate) koje se primenjuju na više diskova istovremeno i ne mogu se naći u slučaju ostalih, Non-RAID tehnologija. Naše rešenje se fokusira na evaluaciji performansi RAID i Non-RAID konfiguracija pod Linux operativnim sistemom. Najpre smo izabrali server na kojem smo instalirali RAID-SAS kontroler zatim smo odabrali RAID konfiguraciju i podesili parametre na željene vrednosti. Posebna pažnja je posvećena testiranju različitih RAID nivoa u fer-play uslovima a koji podrazumevaju testiranje pod istim ili sličnim uslovima, u istom hardverskom okruženju i pod istim operativnim sistemom. Iste veličina fajl sistema za sve testirane nivoe omogućava jednake uslove pri korišćenju slobodnog prostora na diskovima (oko 50GB).

III. METODOLOGIJA TESTIRANJA

Postoji nekoliko mogućih scenarija za određivanje performansi sistema datoteka. Testiranje se može obaviti pomoću svetski priznatog benchmark softvera, koji simulira različite vrste opterećenja, poput opterećenja Internet Service Provider-a ili NetNews servera. Drugi način uključuje korišćenje specijalnih testova, specijalno dizajniranih u te svrhe, poput testova sekvencijalnog i slučajnog čitanja i pisanja, kreiranja datoteka i simulacije rada u aplikaciji.

Za potrebe ovog rada korišćen je PostMark [7] softver koji simulira opterećenje Internet Mail servera. PostMark kreira veliki inicijalni skup (pool) slučajno generisanih

Borislav Đorđević, Institut Mihajlo Pupin, Volgina 15, 11050 Beograd, Srbija; (e mail: bora@impcomputers.com)

Valentina Timčenko, Institut Mihajlo Pupin, Volgina 15, 11050 Beograd, Srbija; (e mail: valentina.timcenko@institutepupin.com)

datoteka na bilo kom mestu u fajl sistemu. Nad tim skupom se dalje vrše operacije kreiranja, čitanja, upisa i brisanja datoteka i određuje vreme potrebno za izvršavanje tih operacija. Redosled izvođenja operacija je slučajno čime se dobija na verodostojnosti simulacije. Broj datoteka, opseg njihove veličine i broj transakcija su u potpunosti konfigurabilni, a radi eliminisanja cache efekata preporučuje se kreiranje inicijalnog skupa sa što većim brojem datoteka (bar 10000) i izvršenje što većeg broja transakcija. Konfiguraciju za testiranje performansi sistema datoteka odlikuju sledeći fundamentalni parametri: matična ploča, vrsta i radni takt procesora, količina i vrsta drugostepene keš memorije, količina i vrsta operativne (RAM) memorije, tip i model disk kontrolera, tip i model diska.

Performanse sistema datoteka su testirane na sledećoj konfiguraciji:

TABELA 1: KARAKTERISTIKE TESTING SISTEMA.

Server	HP Proliant ML350 G5
RAM	4 GB
Procesori	Intel(R)Xeon(R)
CPU Model	Quad-core 410@2.33GHz
Nmbr. of CPU cores	4
CPU brzina	2333MHz
L2 keš	2 x 6 MB
Controllere	
RAID	HP SmartArray E200i (SAS)
RAID keš memorija	128MB
Disk (jedan port)	HP SAS 10K, 146GB, 2.5" t
Operativni sistem	Linux Fedora 8, 2.6.23.1-42

Glavne karakteristike diskova upotrebljenih u testu prikazane su u tabeli 2.

TABELA 2: KARAKTERISTIKE DISKA HP SAS 10K

kapacitet	146GB
interfejs	SAS plug
srednje vreme pozicioniranja	4ms
rotaciona brzina	10,000 rpm
Maks. brzina bafera	3Gb/sec

Testiranje je obavljeno na distribuciji Linux-a, Red Hat Fedora 8 sa stabilnom verzijom kernela 2.6.23.1-42.

Sistemi datoteka su kreirani u logičkim particijama na sledeći način:

Filesystem	Size	Mounted on
LogVol00	90-530GB	/ root FS
LogVol01	2GB	swap
LogVol02	50GB	/test testing FS

Sistem datoteka Log Vol02 je korišćen za testiranje performansi

IV. REZULTATI TESTIRANJA

Izvršena su tri različita testa performansi nad različitim skupovima slučajno generisanih datoteka. Testovi su

obavljeni, na ext2 sistemu datoteka, na sve tri journaling opcije ext3 sistema datoteka i na Reiser sistemu datoteka.

A. Test1

U prvom testu (testu malih i srednjih datoteka) je izvršeno 50.000 transakcija nad skupom od 2000 slučajno generisanih datoteka čije se veličine kreću u opsegu 1KB-100KB, što rezultuje čitanjem i pisanjem približno 1.5GB podataka. Ova suma prevazilazi količinu sistemske memorije i generalno eliminiše efekte keširanja diskova.

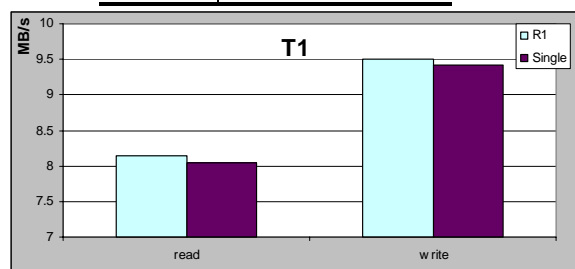
PostMark konfiguracija:

- set size 1000 100000
- set number 2000
- set transactions 50000

Rezultati testa dati su u tabeli 3, a grafički prikazani na slici 1.

TABELA 3: REZULTATI PRVOG TESTA

MB/s	RAID-1	Single
read	8.14	8.05
write	9.51	9.42



Sl. 1. Grafički prikaz performansi (prvi test).

U ovom testu malih datoteka, veliki broj I/O operacija uključujući i metadata operacije i filedata operacije. Zato se očekuje da na performanse imaju kombinovan uticaj i journaling tehnika i keširanje datoteka (file caching). U ovom testu, RAID1 je nešto brži od jednog diska (oko 15%).

B. Test2

U drugom testu (ultra male datoteke) je izvršeno 50.000 transakcija nad velikim skupom slučajno generisanih datoteka, 30000 datoteka, čije se veličine kreću u opsegu 1bajt-1KB, što rezultuje čitanjem i pisanjem približno oko 25MB podataka. Ovakva konfiguracija generiše veliki broj zahteva za ažuriranje meta-data oblasti, odnosno inode tabele.

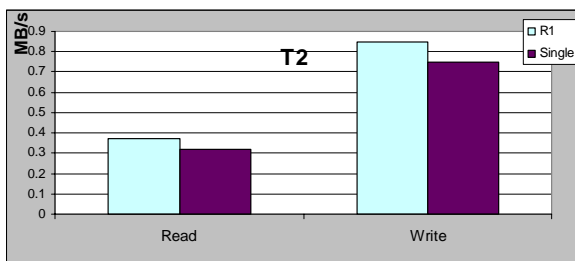
PostMark konfiguracija:

- set size 1 1000
- set number 30000
- set transactions 50000

Rezultati testa dati su u tabeli 4, a grafički prikazani na slici 2.

TABELA 4: REZULTATI DRUGOG TESTA

MB/s	RAID-1	Single
read	0.37	0.32
write	0.85	0.75



Sl. 2. Grafički prikaz performansi (drugi test).

Ovaj test uključuje ogroman broj veoma malih datoteka, pa samim tim i ogroman broj metadata operacija. Zato se očekuje da journaling i komponente file-keša, kao što su metadata keš i direktorijumski keš imaju dominantan uticaj na performanse. U ovom testu, RAID1 je nešto brži od jednog diska (oko 15%)

C. Test3

U trećem testu (širok dijapazon malih i srednjih datoteka) je izvršeno 50.000 transakcija nad skupom od 2000 slučajno generisanih datoteka čija je maksimalna veličina povećana na 300KB, što rezultuje čitanjem i pisanjem približno 4,5GB podataka. Ovaj test je vrlo intenzivan - ukupna količina podataka za čitanje i upis je znatno veća od količine sistemske memorije i u potpunosti eliminiše efekte svih mehanizama keširanja.

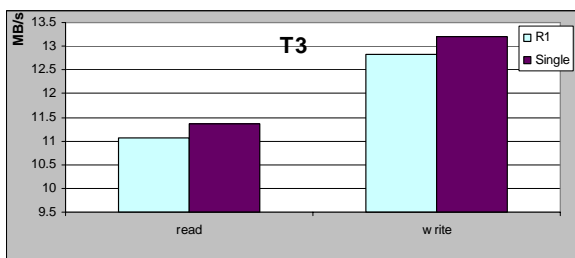
PostMark konfiguracija:

- set size 1000 300000
- set number 2000
- set transactions 50000

Rezultati testa dati su u tabeli 6, a grafički prikazani na slici 3.

TABELA 6: REZULTATI TREĆEG TESTA ZA EXT3

MB/s	RAID-1	Single
read	11.06	11.36
write	12.84	13.19



Sl. 3. Grafički prikaz performansi (treći test).

U okviru testa malih i srednjih datoteka, sa povećanjem veličine test datoteka, filedata transferi dominiraju tako da filedata keširanje ima dominantan uticaj na performanse. U ovom testu, RAID1 je za nijansu sporiji od jednog diska (2.5%).

V. ZAKLJUČAK

U ovom radu smo sumirali rezultate PostMark testova za RAID1 nivo i za slučaj pojedinačnog diska. Testovi su pokazali da je postoji mala razlika od oko 1% između RAID1 i sistema sa jednim diskom. RAID1 je brži za ultra male i male datoteke, dok je sistem sa jednim

diskom brži za veće datoteke. RAID1 je veoma prijatno iznenađenje jer u većini slučajeva ima pad u performansama od svega 1%. To zapravo znači da je RAID1 veoma prihvatljivo rešenje u slučaju hardverskog RAID kontrolera sa SAS interfejsom kojim se omogućavaju višestruki brzi linkovi za različite diskove. U tom slučaju duplo upisivanje može da se obavlja paralelno a RAID keš dodatno poboljšava performanse pri upisu. U posmatranom slučaju, zanimljivo je posmatrati i uporediti RAID1 i sistem sa samo jednim diskom na RAID kontroleru. Pokazuje se da su im performanse slične osim u pojedinim slučajevima kada je RAID1 nešto brži. Naime, inteligentni RAID kontroleri mogu da ubrzaju performance slučajnog (random) čitanja na osnovu metode uparenih diskova u paraleli. Najveći problem sa RAID1 konfiguracijom je 50% memorije koja mora da se odvoji za skladištenje podataka, ali uzimajući u obzir da je cena diskova svakim danom sve niža možemo očekivati da ovo neće biti veliki nedostatak ove konfiguracije.

LITERATURA

- [1] B. Baude "RAID on Linux on POWER", IBM eServer Solutions Enablement, nov2005
- [2] Tweedie S., "EXT3, Journaling Filesystem" 20 July, 2000
- [3] Bill von Hagen, "Exploring the ext3 Filesystem", April 5, 2002
- [4] Robbins D., "Introducing ext3", Gentoo Technologies, Inc., Updated October 9, 2005
- [5] Robbins D., "Surprises in ext3", Gentoo Technologies, Inc. 1 December 2001
- [6] The Software-RAID HOWTO, Jakob Østergaard and Emilio Bueso April 2004
- [7] Hardware RAID vs. Software RAID: Which Implementation is Best for my Application?, Company: Adaptec Published: January 2008
- [8] M. Seltzer, G. Ganger, M. McKusick, K. Smith, C. Soules, C. Stein, "Journaling versus Soft Updates: Asynchronous Meta-data Protection in File Systems", USENIX Conf. Proc., pp. 71-84, San Diego, CA, June 2000.
- [9] A. Silberschatz, P. Galvin, Operating System Concepts. Addison-Wesley, 2007
- [10] V. Danen "Set up Logical Volume Manager in Linux", Mar 09 2007
- [11] J. Katcher, "PostMark: A New File System Benchmark", Technical Report TR3022. Network Appliance Inc, Oct. 1997..

ABSTRACT

Abstract: — This paper represents an (evaluation) of hardware RAID solutions under the Linux Operating System. Paper includes a performance comparison of one most popular RAID levels such as RAID-1 and single disk, under Linux kernel version 2.6. The performance is measured using Postmark benchmark software application..

EXAMINATION OF HARDWARE RAID SOLUTIONS UNDER LINUX: RAID-1 V SINGLE DISK

Dordević, B., Timčenko, V.