

Povećanje raspoloživosti sustava distribuiranim repliciranjem podataka

Increasing system availability
through distributed data replication

SAŽETAK: Opisan je višekorisnički sustav s višestrukim bazama podataka i bitnim uvjetom da kvar bilo koje komponente sustava ne uzrokuje prekid rada sustava na vrijeme dulje od zadanoг maksimalnog intervala neraspoloživosti. Za zadani maksimalni interval neraspoloživosti od 10 minuta, rješenje je ostvareno distribuiranim sustavom s višestrukim serverima i distribucijom kopija podataka, zasnovanom na lokalnoj mreži prema standardu IEEE 802.3. U radu su također opisana rješenja problema koherentnosti podataka u sustavu, te problema dinamičkog prespajanja korisničkih terminala prilikom pojave kvara. U ostvarenju sustava upotrebljene su isključivo komercijalno dostupne sklopovske i programske komponente.

ABSTRACT: This paper describes the multiuser multidatabase system with essential request that the failure of any system's component could not disable it for a period longer than proposed maximal unavailability interval. For determined maximal unavailability interval of 10 minutes, the solution comprises distributed multiple server system with data replication, based on IEEE 802.3 standard local area network. The paper also discusses the solution to data coherence, as well as dynamic reconnection of user terminals in case of failure. The system was realised using only off-the-shelf hardware and software components.

UVOD

Svi tehnički uređaji, pa tako i elektronički uređaji na kojima se zasnivaju današnji informacijski sustavi, podložni su kvarenju. Bez obzira na kvalitetu materijala i izrade ne može se jamčiti rad bez kvarenja komponenti sustava. Veća kvaliteta komponenti ogleda se u manjoј vjerojatnosti kvarenja, odnosno u većim razmacima između pojava kvarova. Uvijek se, međutim, moraju postaviti pitanja: što će se dogoditi kad neki uređaj ipak otkaže, kakav će to utjecaj imati na čitav sustav i koliko dugo će te taj utjecaj trajati. U vezi s trajanjem utjecaja kvara komponente na rad sustava definira se i vrijeme neraspoloživosti sustava kao vrijeme u kojem osnovne funkcije sustava nisu na raspolaganju korisnicima, kao posljedica kvara komponente sustava.

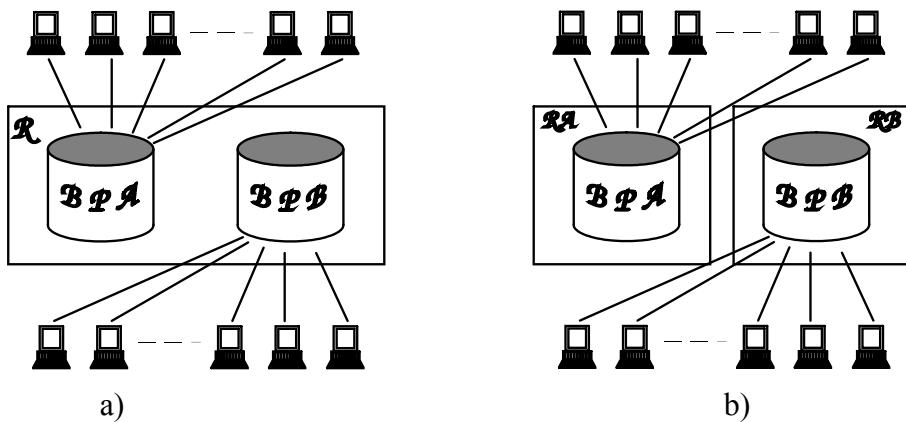
Kod informacijskih sustava kojima je osnovna funkcija podrška poslovanju poduzeća vrijeme neraspoloživosti sustava može se dovesti u direktnu vezu sa financijskim uspjehom poduzeća. Stoga je razumljiva želja da se ovo vrijeme smanji na najmanju isplativu mjeru.

Definiranje problema

U poduzeću X dnevni promet robe zahtijeva obradu nekoliko stotina transakcija dnevno. Zatoj u obradi transakcija duži od 15 minuta izazvao bi velike poremećaje u distribuciji robe a time i gubitke. Prilikom izrade projekta informatizacije distribucijskog centra centralno pitanje bilo je kako osigurati da kvar bilo koje komponente sustava ne uzrokuje neraspoloživost sustava u trajanju dužem od 10 minuta.

Prvi zadatak u postizanju tog cilja bio je da se identificiraju kritične komponente sustava kako bi se njima posvetila posebna pažnja. Lako se može zaključiti da će periferijska oprema (terminali, štampači) imati minimalan utjecaj na raspoloživost sustava jer kvar pojedine periferijske komponente samo smanjuje funkcionalnost sustava ali ne uzrokuje njegov zastoj. Također se ove komponente mogu imati u dovoljnoj rezervi i brzo zamijeniti da se ponovo uspostavi puna funkcionalnost sustava. Kritične komponente sustava svakako su mjesta na kojima se pohranjuju i obrađuju podaci, dakle centralne procesorske jedinice sa svojim diskovima.

Analizom poslovanja distribucijskog centra ustanovljeno je da se čitavo poslovanje može organizirati oko dvije nezavisne baze podataka. Fizička osnovica informacijskog sustava može se u tom slučaju ostvariti na jedan od dva načina, kako je prikazano slikom 1.



Slika 1.: Konfiguracije sustava

Konfiguracija a), u kojoj se obje baze podataka nalaze na istom računalu, očigledno nije pogodna kao rješenje jer će u slučaju kvara računala sustav biti neraspoloživ sve do popravka računala. U slučaju kvara diska ovo vrijeme se još produžava zbog potrebne restauracije baza podataka sa zaštitnih kopija. Raspoloživost se može povećati dodavanjem rezervnog računala koje će, u slučaju kvara na prvom računalu, preuzeti sav posao. Time se, međutim, neopravdano udvostručuju troškovi opreme. Ako se izaberu kvalitetna računala tada će rezervno računalo uglavnom stajati neiskorišteno.

Konfiguracija b) je nešto povoljnija jer u slučaju kvara jednog od računala ostaje u funkciji polovina sustava. Ovo drugo računalo može preuzeti i poslove sa neispravnog računala i sustav će nastaviti s radom uz nešto slabiji odziv. Ovo rješenje izgledalo je perspektivno uz uvjet da se riješi način prijenosa baze podataka s neispravnog računala na ispravno, održavanja koherentnosti podataka i prespajanje korisničkih terminala na drugo računalo. S financijske strane je ovo rješenje također povoljnije jer su oba računala iskorištena cijelo vrijeme u normalnom radu.

Rješenje problema

Prijenos baze podataka s jednog računala na drugo

Prilikom pojave kvara na računalu, podaci pohranjeni na njemu bit će privremeno ili trajno nedostupni. Snimanje zaštitnih kopija podataka uobičajeni je način osiguranja od gubljenja podataka. Zaštitne kopije obično se snimaju na jeftinijem mediju od diska, najčešće magnetskim trakama. U slučaju kvara računala baza podataka može se rekonstruirati sa zaštitne kopije na drugom računalu. Međutim, rekonstrukcija baze podataka sa trake nije dovoljno brza da bi zadovoljila postavljenu granicu vremena raspoloživosti. Kod najbržih suvremenih traka (DAT trake) brzina prijenosa podataka je oko 10 MB/min što bi uz očekivanu veličinu baze podataka od 200 MB tražilo više od 20 min za restauriranje baze podataka. Drugi mediji za masovno pohranjivanje podataka, poput optičkih diskova, zadovoljavaju u pogledu brzine restauriranja baze podataka, ali im je cijena visoka. Kao optimalno rješenje u ovoj situaciji pokazali su se magnetski diskovi kao medij za pohranjivanje podataka i lokalna mreža Ethernet kao sredstvo za prijenos podataka iz jednog računala u drugo.

Održavanje koherentnosti podataka

Rezervna kopija baze podataka na drugo računalo obnavlja se jednom dnevno. To znači da u slučaju kvara jednog od računala na drugom računalu imamo bazu podataka sa stanjem iz prethodnog dana, što nije zadovoljavajuće. Da bi se na toj rezervnoj bazi nastavio rad ona se mora dovesti u stanje koherentnosti sa bazom podataka na računalu koje je upravo otkazalo.

Bazu podataka možemo promatrati kao stroj s konačnim brojem stanja u kojemu transakcije predstavljaju ulazne pobude. Stanje baze podataka određeno je slijedom transakcija koje se nad njom obavljuju:

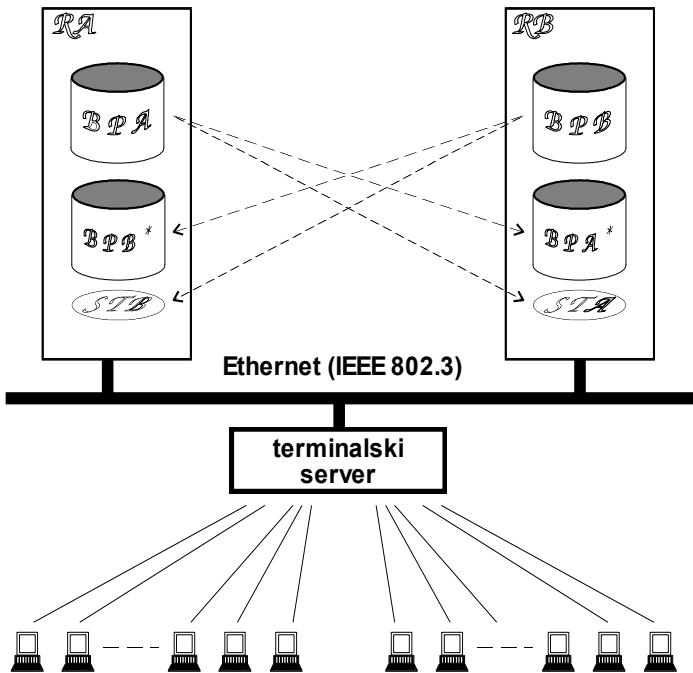
$$\mathcal{B}_i = \mathcal{B}_{i-1} \xrightarrow{\mathcal{T}_i}$$

gdje je \mathcal{B}_i stanje baze podataka nakon izvođenja transakcije \mathcal{T}_i , a $\xrightarrow{\cdot}$ predstavlja operaciju izvođenja transakcije nad bazom podataka. Svojstvo stroja sa konačnim brojem stanja je ponovljivost: ako je neki slijed ulaznih pobuda doveo stroj iz stanja S_i u stanje S_j , svaki put kad se taj slijed pobuda dovede na ulaz stroja u stanju S_i stroj će preći u stanje S_j . Primijenimo li to na bazu podataka, to znači da pamćenjem slijeda trasnsakcija koje se izvode nad bazom od posljednjeg uspostavljanja koherencije kopije i originala, možemo kopiju dovesti u koherentno stanje s originalom i nakon kvara na jednom od računala. Slijed transakcija se mora, naravno, također pamtitи na rezervnom računalu.

Prespajanje terminala

U slučaju otkaza jednog od računala, svi korisnici s tog računala nastavljaju raditi na drugom računalu. To znači da njihove terminale treba odspojiti s pokvarenog računala i prespojiti na ispravno računalo. Kad bi se to radilo fizičkim prespajanjem kabela, potrebno vrijeme za prespajanje bilo bi predugo. Rješenje ovog problema nađeno je u primjeni terminalskega servera koji omogućuje logičko povezivanje terminala s računalom posredstvom Ethernet mreže i odgovarajućeg komunikacijskog protokola. Prespajanje terminala izvodi se u tom slučaju jednostavnom promjenom parametra u tablici povezivanja terminal servera.

Integriranje rješenja



Slika 2.: Konačno rješenje

Slika 2. šematski pokazuje integrirano rješenje sustava. Na svakom od računala, $\mathcal{R}\mathcal{A}$ i $\mathcal{R}\mathcal{B}$, nalazi se po jedna radna baza podataka, $\mathcal{B}\mathcal{P}\mathcal{A}$ i $\mathcal{B}\mathcal{P}\mathcal{B}$, te po jedna zaštitna kopija baze podataka, $\mathcal{B}\mathcal{P}\mathcal{A}^*$ i $\mathcal{B}\mathcal{P}\mathcal{B}^*$. Na svakom od računala također se nalazi i datoteka u kojoj se pamti slijed transakcija nad bazom s drugog računala, $\mathcal{S}\mathcal{T}\mathcal{A}$ i $\mathcal{S}\mathcal{T}\mathcal{B}$.

U normalnom radu radne funkcije sustava dopunjene su dvjema funkcijama koje doprinose povećanju raspoloživosti.

- 1) Svaka transakcija zapisuje se kao transakcijski zapis u datoteku za pamćenje slijeda transakcija STx. Transakcijski zapis je relativno malen i njegovo upisivanje ne optereće značajno sustav tako da se ne primjećuje smanjenje performansi sustava u posluživanju korisnika.
- 2) Jednom dnevno, na kraju radnog vremena, kopira se baza BPx u zaštitnu kopiju BPx*. Brzina prijenosa podataka ovisi o propusnosti mreže i diskovnih podsustava obaju računala. U ispitivanju sa konkretnom opremom postignute su brzine od 20 - 25 MB/min što omogućuje da se čitava baza kopira za 8-10 min. No, samo vrijeme kopiranja baze nije kritično za raspoloživost sustava.

Prilikom pojave kvara odvija se slijedeći scenarij (pretpostavimo da se pokvari računalo $\mathcal{R}\mathcal{A}$):

- 1) Administrator sustava pokreće na računalu RB postupak dovođenja baze podataka BPA* u stanje koherencije s bazom podataka BPA. Postupak se sastoji od primjenjivanja slijeda transakcija STA na bazu podataka BPA*. Nakon provedenog postupaka baza BPA* sadržavat će potpuno identične podatke kao i baza BPA na računalu RA koja je zbog kvara trenutno nedostupna.
- 2) Za vrijeme dok na računalu RB traje postupak dovođenja baze podataka BPA* u stanje koherentnosti s bazom podataka BPA, administrator sustava mijenja tablicu logičkih veza

u terminalskom serveru. Svi terminali, do tada povezani s računalom RA, prespajaju se na računalo RB. Ovaj postupak može se obaviti jednom jedinom naredbom sa konzolnog terminala terminalskog servera.

3) Korisnici se prijavljuju na računalo RB i nastavljaju s radom.

Dobra organizacija i pažljivo konfiguriranje sustava omogućuje da ovaj postupak protekne glatko i bez problema. Korisnicima se čak omogućuje da zadrže isto ime i lozinku za prijavljivanje. Tako korisnici ne moraju ništa osjetiti osim da je sustav bio u zastoju nekoliko minuta. Prilikom ispitivanja sustava, sa ispitnim bazama podataka veličine 150 MB, vrijeme neraspoloživosti sustava nije prelazilo 5 minuta.

Zaključak

Opisani sustav povećanja raspoloživosti nastao je iz potrebe da se omogući uspostavljanje funkcije sustava u razumno kratkom roku nakon kvara kritične komponente, sa prihvatljivim povećanjem cijene sustava. Povećanje cijene ispitanih sustava bilo je oko 10% i sastojalo se uglavnom u povećanju kapaciteta diskova. Nadalje, sve komponente sustava su komercijalno dostupne te nije bilo posebnih zahvata niti razvoja specijalne opreme. Jedino je unutar aplikacijskih programa bilo potrebno ugraditi podršku za opisane postupke.

Opisani sustav izведен je na dva računala ali se jednostavno može primijeniti i na većem broju računala i baza podataka. Također se može postići zaštita od višestrukih kvarova tako da se distribuira više od jedne kopije baze podataka. Ova proširenja ideje treba podvrći ispitivanju radi ocjene ekonomičnosti primjene.

LITERATURA

1. COMPAQ SYSTEMPRO/LT Family of personal Computer Servers Features/ Specifications, Second Edition, March 1992, Compaq Computer Corporation
2. SCO UNIX 3.2v4. Administrators Guide, The Santa Cruz Operation Inc., 1992
3. SCO LLI Drivers Package Version 3.0.0. Release and Installation Notes, March 1992., The Santa Cruz Operation Inc., 1992
4. SCO TCP/IP Runtime Version 1.1.3. Administrators Guide, The Santa Cruz Operation Inc., 1992
5. SCO NFS NetWork File System Version 1.1.0. System Administrators Guide, The Santa Cruz Operation Inc., 1992
6. Specilaix Modular Terminal Server, Guide to Installation and Operation, Second Edition, March 1992, Specialix International Plc.
7. ES3210 Installation Manual, Second Edition, March 1992, Racal Datacom

mr. Dragutin Vuković, dipl.inž.

MicroLAB d.o.o.

Savska cesta 41/VII, PP 17

41000 ZAGREB

HRVATSKA

e-mail: drago.vukovic@etf.uni-zg.ac.mail.hr