



Oracle eða Informix? Val á gagnagrunnskerfi
fyrir Vatnamælingar Orkustofnunar og
Landsvirkjun

Tómas Jóhannesson

Greinargerð TJ-91-02

ORACLE EÐA INFORMIX?

VAL Á GAGNASAFNSKERFI FYRIR VATNAMÆLINGAR ORKUSTOFNUNAR OG LANDSVIRKJUNAR

1. INNGANGUR

Verið er að þróa nýtt úrvinnslukerfi fyrir Vatnamælingar Orkustofnunar og Landsvirkjunar (sjá greinargerð TJ/ÁS/KE/-ÁSn-91/01). Velja þarf venslagagnasafnskerfi fyrir nýja vatnamælingakerfið og lýsir eftirfarandi greinargerð samanburði á kerfunum Oracle og Informix sem koma helst til greina í þessu sambandi. Oracle og Informix eru einu gagnasafnskerfin fyrir UNIX tölvur sem komin eru með umboðsaðila og nokkurn notendahóp hér á landi og voru önnur kerfi því ekki athuguð.

Markmiðið með prófununum sem hér er lýst er í fyrsta lagi að athuga hvort gagnasafnskerfin séu frambærileg og henti fyrir þróun og uppsetningu nýja vatnamælingakerfisins. Þar koma til álita notendaskilkerfanna, nettengingar, þróunarumhverfi o.fl. Í öðru lagi var tilgangurinn að bera saman vinnsluhraða og plássnýtingu Oracle og Informix. Í þriðja lagi var lagt mat á vinnsluhraða og plássþörf þegar tímaraðir eru geymdar á "normaliseruðu" formi í venslagagnasafnskerfi. Þessi möguleiki hefur þegar verið athugaður af Verkfræðistofnunni Streng fyrir Landsvirkjun og reyndist óhóflega frekur á diskapláss og vinnslutíma. Ekki var því lögð mikil vinna í tilraunir með tímaraðir á "normaliseruðu" formi í samanburðinum sem lýst er hér að neðan.

Af Oracle er þegar komin talsverð reynsla á Orkustofnun og gengu prófanir á Informix að nokkru leyti út á að athuga hvort Informix gæti leyst ýmis verk sem nú eru unnin í Oracle gagnasafnskerfi Orkustofnunar. Ef Informix verður fyrir valinu fyrir vatnamælingakerfið er líklegt að notkun þess á Orkustofnun verði meiri en í vatnamælinga-

kerfinu einu og þarf kerfið því að geta unnið vandræðalaust þau verkefni sem þegar eru unnin í Oracle.

2. SAMANDREGNAR NIÐURSTÖÐUR

Niðurstöður samanburðarins, sem lýst er í allmiklum smáatriðum hér að neðan, eru eftirfarandi.

1. Uppbygging Oracle og Informix er svipuð og kerfin eru sambærileg hvað varðar notkun frá þriðjukynslóðarforritunarmálum, og nettengimöguleika.
2. Informix getur unnið með helstu töflur í gagnasafni Orkustofnunar vandræðalaust.
3. Oracle er á stærðarþrepinu 4 sinnum hraðvirkara en Informix í fyrirspurnum.
4. Plássþörf kerfanna er sambærileg.
5. Þróunarumhverfi Informix (4GL) hentar betur til kerfisþróunar en þróunarumhverfi Oracle (SQL*Menu, SQL*Forms og SQL*ReportWriter). Þróunarumhverfi Oracle má bæta með hugbúnaði frá þriðja aðila.
6. Oracle er mun dýrara en Informix. Stofnkostnaður (fjölnotendaleyfi á OS, vinnustöðvaleyfi fyrir 1 notanda á LV og Egilsstöðum) er um 1.4 milljónir kr. fyrir Informix en 2.6 milljónir fyrir Oracle. Með viðhaldi í 3 ár er kostnaðurinn um 2.1 milljón kr. fyrir Informix en 4.3 milljónir fyrir Oracle. Sá Oracle búnaður sem þegar er til á Orkustofnun lækkar þennan kostnað töluvert fyrir Oracle, en Oracle er

engu að síður nokkru dýrara en Informix. Þessar tölur taka ekki tillit til kostnaðar við kaup á þróunarhugbúnaði frá þriðja aðila fyrir Oracle en val á slíkum hugbúnaði er nú til athugunar.

7. Kerfin henta bæði illa til geymslu tímaraða á "normaliseruðu" formi, sérstaklega Informix. Plássþörf tímaraðanna á þessu formi er 4-10 sinnum meiri en í hefðbundum stýrikerfisskrám. Vinnsluhraði við fyrirspurnir er um 20-50 falt hægari (Oracle) og um 200-500 falt hægari (Informix) en hægt er að ná með þriðjukynslóðarforritunarmálum og stýrikerfisskrám.

3. UPPSETNING GAGNASAFNSKERFANNA

Oracle (útgáfa 6.0.27) keyrir á HP9000/825 tölvu Orkustofnunar. Í sambandi við þessar prófanir setti Snorri Bergmann frá Verkfræðistofunni Streng upp Informix (útgáfu 4.10) á sömu tölvu.

4. KERFISREKSTUR

Uppbygging Oracle og Informix gagnasafnskerfanna er að flestu leyti mjög svipuð. Kerfunum er úthlutað diskplássi undir gögn í stórum skráum sem hægt er að bæta við eftir þörfum. Sérstakir "prósessar" sjá um að skrifa í gagnatöflur, læsa færslum o.fl. og er hægt að fylgjast með ástandi kerfanna með til þess ætluðum eftirlitsforritum. Rekstur kerfanna er því í aðalatriðum eins og bjóða bæði kerfin upp á svipuð hjálpartæki að þessu leyti.

Oracle hefur eigið notendabókhald og krefst aðgangsorða í hvert sinn sem forrit eru keyrð (nema farið sé í kringum aðgangsorðakerfið með svokölluðum OPS notendum). Informix lætur sér duga notendanöfn stýrikerfisins og einfaldar það notkun kerfisins mikið.

Informix virðist betur aðhæft UNIX stýrikerfinu en Oracle. Það á auðveldara með að skrifa skrár sem önnur UNIX forrit

þurfa að lesa (t.d. awk), einfaldara er að eiga við pípur, bakgrunnsvinnu o.m.fl. Hins vegar virðist það ekki geta unnið með terminfo skjálýsingar sem notaðar eru á HP tölvum Orkustofnunar og er það nokkur galli.

Svokallað "data dictionary" virðist við fyrstu sýn aðgengilegra og betur skipulagt í Oracle en í Informix. Sjónarhorn ("view") í "data dictionary" Oracle kerfisins eru fleiri en í Informix og auðveldara er að fá fram ýmsar upplýsingar um notkun kerfisins á plássi o.fl.þ.h.

Handbækur kerfanna eru mjög ítarlegar og skýra sig að mestu leyti sjálfar. Handbækur Oracle bötundu mjög í útgáfu 6 og taka handbókum Informix nokkuð fram að skipulagi, indexar nýttast betur o.fl.

5. NOTKUN Á GAGNASAFNSKERFUNUM FRÁ ÞRIÐJUKYNSLÓÐARFORRITUM

Bæði Oracle og Informix bjóða upp á tengingu við þriðjukynslóðarforrit með forþýðendum fyrir C og ýmis önnur forritunarmál. Forþýðandi Oracle fyrir C nefnist Pro*C og hefur hann verið notaður dálítið á Orkustofnun. Forþýðandi Informix fyrir C nefnist ESQL og var hann prófaður með því að breyta nokkuð stóru forriti sem vinnur með efnafræðigögn Orkustofnunar þannig að hægt væri að þýða það fyrir Informix. Aðgangur forritunarmála að SQL fyrirspurnum er mjög vel staðlaður og þurfti því lítið að breyta forritinu, sem skrifað er fyrir Oracle, til þess að það gengi í Informix. Forritið vann eðlilega og virðist sem Pro*C og ESQL séu að flestu leyti sambærileg til forritapróunar. ESQL er örlítið takmarkaðra en Pro*C að því leyti að það getur ekki skilað gögnum í vectora í einni aðgerð ("array bind" í "fetch"). Hins vegar er notkun á ESQL auðveldari en á Pro*C vegna þess að esqt skipunin sér um ýmsa hluti sem notandi Pro*C verður að gera sjálfur með hjálp "make" skráa.

6. NETTENGINGAR

Bæði Oracle og Informix geta sótt gögn yfir á aðrar tölvur með nettengingum. Nettengin Oracle heitir SQL*Net og hefur hún verið í notkun á Orkustofnun um nokkurra ára skeið og reynst vel. Nettengibúnaður Informix var ekki prófaður en er eftir því sem næst verður komist sambærilegur við nettengibúnað Oracle fyrir þau verk sem hér um ræðir (Oracle getur tengst fleiri tölvutegundum en Informix og veitt aðgang að gögnum í ýmsum öðrum gagnsafnskerfum). Þó virðist sem nettenging Informix breyti ekki stafrófi í strengjum sem fluttir eru á milli tölva sem nota mismunandi stafróf, t.d. þegar PC tölva sem notar IBM Code Page 861 stafróf les gögn úr töflum á UNIX tölvu sem notar ISO-8859/1 stafróf. Þetta atriði skiptir þó væntanlega ekki höfudmáli í nýja vatnamælingakerfinu.

7. NOTENDASKIL OG ÞRÓUNARUMHVERFI

Notendaskil Oracle og Informix eru mjög frábrugðin hvort öðru. Bæði kerfin geta svarað á skjá SQL fyrirspurnum sem notandi slær inn frá hnappaborði (SQL*Plus í Oracle, ISQL í Informix). Í Oracle eru notuð þrjú sjálfstæð forrit, SQL*Menu, SQL*Forms og SQL*ReportWriter, til þess að setja upp valmyndir, skráningar- og leitarmyndir og svara fyrirspurnum. Í Informix eru öll þessi verk unnin með einu forriti sem nefnist 4GL þó einfaldari verk sé stundum hægt að vinna með öðrum forritum kerfisins.

Notendaskil Oracle eru sundurleit og oft þunglamaleg í notkun. SQL*Menu og SQL*ReportWriter hafa ekki verið notuð á Orkustofnun en þessi forrit voru prófuð fyrir nokkru (eldri útgáfur en nú eru seldar) og reyndust illa. SQL*ReportWriter hefur jafnframt fengið slæma dóma í tölvutímaritum. SQL*Forms (útgáfa 2.3) hefur verið notað um nokkurra ára skeið á Orkustofnun og reynst misjafnlega. Forritið reynist vel til þess að setja upp einfaldar skjámyndir en afleitlega til þess að gera flóknari hluti. Viðhald og breytingar á flóknum skjámynd-

um í SQL*Forms (útgáfa 2.3) er ótrúlega tafsamnt. Helstu gallar SQL*Forms eiga að hafa verið lagaðir í nýjustu útgáfu þess sem enn hefur ekki verið sett upp á Orkustofnun.

Informix 4GL er hægt að nota til þess að setja upp valmyndir, skráningarmyndir og útprintanir af flestum gerðum. Þessir möguleikar eru vel samhæfðir og taka hliðstæðum forritum í Oracle langt fram. Til þess að Oracle verði sambærilegt við Informix að þessu leyti þarf væntanlega að finna notendaskil fyrir Oracle frá þriðja aðila (mörg slík forrit eru til) sem bíður upp á þá kosti sem Informix 4GL hefur.

8. SAMANBURÐUR Á VINNSLUHRAÐA OG PLÁSSPÖRF

8.1 Uppsetning

HP9000/825 tölvan er ekki afkastamikil miðað við ýmsar vinnustöðvar sem nú eru komnar á markaðinn. Tölvurnar hér að neðan segja því ekki mikið um vinnsluhraða gagnasafnskerfanna á t.d. SUN SPARC, HP9000/700 eða IBM RS6000 vinnustöðvum. Samanburður á vinnsluhraða kerfanna á HP9000/825 tölvunni hefur hins vegar almennt gildi þar sem ætla má að afköst þeirra breytist í svipuðum hlutföllum ef prófanirnar yrðu gerðar á afkastameiri tölvum (keyrslurnar reynast CPU takmarkaðar). Samanburður á plássþörf kerfanna hefur að sjálfsögðu fullt gildi á HP9000/825 tölvunni. Prófanirnar fóru fram þegar lítil önnur vinnsla var á tölvunni og mæla því vinnsluhraða sem á við einn notanda. Prófanir snerust eingöngu um lestur á gögnum enda hefur vinnslutími við aðra vinnu litla þýðingu fyrir vatnamælingakerfið sem hér um ræðir.

Oracle kerfi Orkustofnunar hefur verið í rekstri um nokkurt skeið og er því komin reynsla á uppsetningu þess á HP9000/825 tölvunni. Snorri Bergmann og sérfræðingur Informix frá Bandaríkjunum yfirfóru uppsetningu Informix kerfisins og hafa staðfest að hún sé í lagi. Bæði kerfin geyma gögn í nokkrum stórum stýrikerfisskrám (ekki er

notaður möguleiki á "raw device" uppsetningu sem bæði Oracle og Informix bjóða upp á) og eru sett upp með ámóta stóra minnisbuffera (tæp 2 Mb). Uppsetning kerfanna er þannig að flestu leyti sambærileg þó vafalaust megi bæta hana fyrir þau bæði. Niðurstöðurnar hér að neðan eiga því að vera sanngjarn, en mjög takmarkaður, samanburður á afköstum kerfanna.

Töflur til prófananna voru flestar fluttar úr Oracle gagnasafni Orkustofnunar inn í Informix. Þessar töflur geyma raunveruleg gögn sem notuð eru í daglegri vinnslu í gagnasafnskerfi Orkustofnunar. Fyrirspurnirnar, sem prófaðar eru hér að neðan, eru í flestum tilfellum einfaldaðar útgáfur af fyrirspurnum sem notaðar eru í daglegum rekstri gagnasafns Orkustofnunar. Töflurnar í Oracle hafa orðið til smám saman og hafa því ekki jafn "góða" uppbyggingu þar og þær fá þegar þær eru lesnar inn í einu lagi eins og gert var þegar töflurnar voru búnar til í Informix. Þessi munur ætti að vera Oracle í óhag í samanburðinum hér að neðan.

Nokkrar töflur sem eru einfölduð útgáfa af bókhaldstöflum Orkustofnunar voru einnig prófaðar. Gögn til þessara prófana voru lesin úr "óformateruðum" stýrikerfisskrám sem notaðar eru í bókhaldskerfi Orkustofnunar. Fyrirspurnir sem nota þessar töflur er hliðstæðar við fyrirspurnir sem notaðar eru í C-forritum í bókhaldskerfi Orkustofnunar.

Engar tilraunir hafa verið gerðar með uppsetningu gagna í töflunum til þess að fá fram hámarks vinnsluhraða eða plássnýtingu heldur eru notuð sjálfgefin gildi sem gagnasafnskerfin velja ef annað er ekki tekið fram.

8.2 Borholugagnatafla Orkustofnunar

Fyrsta tilraunin sem gerð var fólst í að lesa borholumælingatöflu Orkustofnunar inn í Informix og bera saman afköst Oracle og Informix í fyrirspurnum sem vinna með borholugögnin. Ástæða þess að borholumælingataflan varð fyrir valinu var sú að hún er sem stendur stærsta tafla í gagnasafni Orkustofnunar (gögn og index samtals um 9Mb)

og ætti því að reyna á hæfni kerfanna við að fást við raunveruleg gögn sem eru mun meiri að umfangi en pláss er fyrir í minnisbufferum. Einnig hefur þessi tafla að stofni til sömu uppbyggingu og komið hefur til tals að nota við varðveislu á tímaröðum (þ.e. auðkennisnúmer, dýpi, mæligildi) ef dýpisdálkur töflunnar er túlkaður sem tími. Að lokum hefur komið í ljós galli á "optimizer" í Oraclekerfi Orkustofnunar í ákveðnum fyrirspurnum sem nota þessa töflu og er forvitnilegt að sjá hvort "optimizer" Informix reynist betur í slíkum fyrirspurnum. Index var skilgreindur á fyrstu tvo dálka töflunnar.

Keyrðar voru tvær fyrirspurnir sem eru dæmigerðar fyrir þær fyrirspurnir sem svara þarf fyrir töflur eins og hér um ræðir. Fyrri fyrirspurnin (1) tók saman fjölda mæligilda, meðaldýpi og meðaltal mæligilda fyrir 10 mælingar með "group by" aðgerð. Mælingarnar voru tilteknaðar með "between n1 and n2" og aðgangur að þeim því hraður með indexinum. Hin fyrirspurnin (2) sótti sömu upplýsingar fyrir allar mælingar frá ákveðnum mælistað með "join" yfir í einfaldaða töflu sem geymdir almennar upplýsingar um hverja mælingu, þ.m.t. númer mælistaðar. Í þeirri töflu voru indexar á dálkum sem geyma auðkennisnúmer mælingar og númer mælistaðar. Að auki var prófuð fyrirspurn (3) sem þurfti að lesa alla töfluna frá upphafi til enda til þess að athuga hversu hraður aðgangur kerfanna að miklu gagnamagni væri. Sú fyrirspurn bað um meðaltal allra mæligilda í töflunni. Meðaltalið er í þessu tilfalli merkingarlaust sem slíkt en fyrirspurnin hefur sama gildi til samanburðar á kerfunum þrátt fyrir það.

Að lokum voru prófaðar tvær fyrirspurnir sem tengjast áður nefndum galla í "optimizer" Oracle. Fyrirspurnirnar biðja um fjölda mæligilda í ákveðinni mælingu (þ.e. fjölda færslna með tiltekið auðkennisnúmer). Önnur fyrirspurnin (4) notaði gagnatöfluna beint. Hin (5) var eins nema hvað notað var sjónarhorn ("view") sem tók saman fjölda mæligilda fyrir allar mælingar töflunnar með "group by" aðgerð. Báðar fyrirspurnirnar eiga að vera auðveldar viðfangs, þar

sem indexinn veitir hraðan aðgang að mælingunni sem spurt er um hverju sinni. Gallinn á "optimizer" Oraclekerfisins á Orkustofnun lýsir sér í því að seinni fyrirspurnin er meira en 100 falt tímafrekari en sú fyrri. Þetta stafar af því að notkun kerfisins á indexum í fyrirspurnum sem nota sjónarhorn er ófullkomin. Niðurstöður samamburðarins koma fram í töflu 1 hér að neðan.

Eins og sjá má þá tekur borholumælingataflan nokkru meira pláss (32%) í Oracle en í Informix. Þetta stafar að hluta til af því að taflan og sérstaklega indexinn í Oracle hafa byggst upp á löngum tíma og taka því heldur meira pláss en ef gögnin hefðu verið lesin inn í einu lagi eins og gert var þegar þau voru lesin inn í Informix. Einnig stafar mismunurinn af því að Oracle þarf nokkru meira pláss fyrir hverja tölu sem geymd er (6.6 batur að meðaltali) en Informix (5.3 batur að meðaltali) þegar tekið er tillit til pláss sem ekki nýtist beint ("overhead"). Þegar tekið er tillit til index þurfa kerfin 43 batur (Oracle) og 32 batur (Informix) til þess að geyma hverja færslu. Munurinn í plássnotkun kerfanna er ekkert sem máli skiptir þegar litið er til þess að gögnin sem slík þurfa ekki nema u.þ.b. $277255 * 8 = 2218$ Kb pláss (þ.e. 8 batur pr. færslu) ef þau eru geymd á hefðbundinn hátt á disk (óformateruð, hver mæling í sérstakri skrá, ekki þarf að geyma auðkennisnúmer, index er óþarfur). Gögnin taka því um 4-5 falt meira pláss á "normaliseruðu" formi en strangt tiltekið er nauðsynlegt, hvort sem þau eru geymd í Oracle eða Informix. Ef litið er til jafnsafnaðra tímaráða þá verður þessi munur enn meiri, eða 8-10 faldur (þá er hægt að komast hjá að geyma tíma fyrir hvert mæligildi). Þessi niðurstaða er til marks um það að venslagagnasafnskerfi henti illa til varðveisla tímaráða með ofangreindum hætti hvert sem kerfið er og er í samræmi við athuganir Verkfræðistofunnar Strengs á þessu atriði.

Þegar litið er til keyrslutímans þá kemur í ljós að bæði kerfin svara fljótt og vel fyrirspurn sem getur gengið beint að tiltölulega litlu gagnamagni eftir index (fyrirspurn 1).

Hins vegar hefur Oracle verulega yfirburði (um 3 falt skemmri keyrslutími) þegar leita þarf í hóflegu gagnamagni í tveimur indexuðum töflum með "join" (fyrirspurn 2) og enn meiri yfirburði (á stærðarþrepinu 10 falt skemmri keyrslutími) þegar stór tafla (með litlum færslum) er lesin frá upphafi til enda (fyrirspurn 3). Ekki er ljóst af hverju þetta stafar þar sem Oracle geymir töluleg gögn á sínu eigin formi og þarf að eyða CPU tíma í að breyta þeim yfir á innra form tölvunnar þegar fyrirspurnum er svarað, en Informix geymir gögnin á formi tölvunnar og þarf því ekki að breyta um form. Síðasttalda fyrirspurnin felur í sér lestur á 5,5 Mb (Oracle) og 4,4 Mb (Informix) af gögnum af disk tölvunnar (ekki þarf að lesa indexinn í þessu tilfelli). Gagnaflæðið í þessari fyrirspurn svarar því til um 50 Kb/s fyrir Oracle en um 3-4 Kb/s fyrir Informix. Gagnaflæðið hjá Oracle telst hvergi nærri gott þar sem tölvan getur auðveldlega lesið um og yfir 250 Kb/s af gögnum af disk. Gagnaflæðið hjá Informix í þessari fyrirspurn er meira en 10 falt hægara en hjá Oracle og undir 2% af þeim hraða sem tölvan getur tæknilega staðið undir. Þetta hæga gagnaflæði er ekki eðlilegt og virðist stafa af því að Informix ræður illa við svona stóra töflu þegar færslurnar eru jafn litlar og raun ber vitni (12 batur). Afköstin (reyndar bæði hjá Oracle og Informix) eru mun meiri þegar færslur eru stærri eins og kemur fram hér að neðan.

Eins og áður en nefnt þá taka tímaráðir á bilinu 4 til 10 falt meira pláss á "normaliseruðu" formi í Oracle eða Informix en á hefðbundnu formi í stýrikerfisráðum. Þegar við bætist að vinnsluhraði kerfanna er einungis um 1/5 (Oracle) eða minna en 1/50 (Informix) af hámarksafköstum tölvunnar, þá fæst að lestur á tímaröðum sem geymdar eru með þessum hætti er um 20-50 falt hægari (Oracle) eða um 200-500 falt hægari (Informix) en fengist með "gamaldags" stýrikerfisráðum (eða í "raw" breytum í Oracle og "blob" breytum í Informix). Þetta gerir það að verkum að geymsla tímaráða með þessum hætti, hvort sem er með Oracle eða Informix, er ekki raunhæfur möguleiki í

TAFLA 1: Samanburður á Oracle og Informix, Borholugögn.

	Oracle	Informix
Færslufjöldi	277255	277255
Pláss, gögn, o.fl. (Kb)	5500	4418
Pláss, index (Kb)	6300	4510
Pláss alls (kb)	11800	8928
Tími fyrirsp. 1 (liðinn/CPU) (s)	5/3	5/3
Tími fyrirsp. 2 (liðinn/CPU) (s)	10/6	30/20
Tími fyrirsp. 3 (liðinn/CPU) (s)	117/90	1200/750
Tími fyrirsp. 4 (liðinn/CPU) (s)	<1/<1	<1/<1
Tími fyrirsp. 5 (liðinn/CPU) (s)	150/100	>1800/600

nýja vatnamælingakerfinu. Þó geymsla tímaraða á "normaliseruðu" formi sé óraunhæfur möguleiki í báðum kerfunum, hefur samanburðurinn hér að ofan engu að síður nokkurt gildi sem samanburður á afköstum kerfanna tveggja í tímafrekum fyrirspurnum úr stórum töflum.

Eins og sjá má á niðurstöðunum úr fyrirspurnum (4) og (5) þá stendur "optimizer" Informix sig ekki vel við lestur úr sjónarhorninu úr borholumælingatöflunni. Reyndar er Informix meira en 1000 sinnum hægvirka en þörf krefur en Oracle "aðeins" 100 sinnum. Bæði kerfin virðast því viðkvæm fyrir framsetningu fyrirspurna.

8.3 Jarðefnafræðitafla Orkustofnunar

Næsta tilraun notaði efnafræðitöflu Jarðefnafræðistofu Orkustofnunar. Þessi tafla er ein af stærstu töflum í gagnasafni Orkustofnunar og hefur mjög marga dálka. Dálkarnir geyma bæði heilar tölur, rauntölur og strengi og eru 85 alls. Taflan er gisin, þ.e. margir dálkannar eru auðir vegna þess að ekki eru öll efnanna mæld fyrir öll sýni.

Keyrðar voru tvær fyrirspurnir. Önnur (1) bað um summu (eða meðaltal) 8 efna sem eru greind í mörgum eða flestum sýnum töflunnar en hin (2) bað um summu (eða meðaltal) 8 efna sem eru ekki greind nema í tiltölulega fáum sýnum. Fyrirspurnirnar gáfu sömu niðurstöðu í báðum gagnasafnskerfunum eins og vera ber. Niðurstöður samanburðarins koma fram í töflu 2 hér að neðan.

Eins og sjá má þá tekur taflan sem næst helmingi meira pláss í Informix en í Oracle. Þetta stafar af því að Oracle geymir ekki gildi sem eru óskilgreind (NULL). Það leiðir til mikils sparnaðar á plássi fyrir efnafræðigögnin vegna þess hversu algengt er að dálkar séu óskilgreindir í efnafræðitöflunni. Munurinn á tímanum sem fer í fyrirspurnirnar er enn meiri en munurinn á plássinu eða um 4 til 5 falt lengri fyrir Informix, bæði hvað varðar liðinn tíma og CPU tíma. Fyrri fyrirspurnin tekur af einhverjum ástæðum hartnær helmingi lengri tíma í Informix en sú síðari þó báðar þurfi að lesa töfluna frá upphafi til enda.

Fyrirspurnirnar taka báðar til svo margra færslna að útilokað á að vera að geymsla á færslum í bufferum kerfanna hafi áhrif á niðurstöðuna hér að ofan. Þessi möguleiki var engu að síður athugaður með því að taka Oracle kerfið niður og keyra fyrirspurn 1 strax eftir að kerfið kom upp aftur. Þá tók fyrirspurnin 12/10 sek. sem er í fullu samræmi við fyrri niðurstöður.

Lestrarafköst gagnasafnskerfanna við lestur úr efnafræðitöflunni eru betri en afköst við lestur úr borholumælingatöflunni (145-180 Kb/s fyrir Oracle en 60-90 Kb/s fyrir Informix). Þessi afköst Oracle eru um 1/2 til 2/3 af þeim hraða sem tölvan getur tæknilega staðið undir en afköst Informix eru um 1/3 til 1/4. Samkvæmt þessu má gera ráð fyrir að léleg afköst við lestur úr borholumælingatöflunni, sérstaklega hjá Informix (2%), stafi af óvenjulega litlum færslum og "overhead" kerfanna í tengslum

TAFLA 2: Samanburður á Oracle og Informix, Efnafræðigögn.

	Oracle	Informix
Færslufjöldi	7810	7810
Pláss, gögn, o.fl. (Kb)	1600	3136
Pláss, index (Kb)	152	96
Pláss alls (kb)	1752	3232
Tími fyrirsp. 1 (liðinn/CPU) (s)	11/10	55/50
Tími fyrirsp. 2 (liðinn/CPU) (s)	9/8	35/30

við það.

Þessi afköst fást þegar ein tafla er lesin frá upphafi til enda en slíkur lestur er langhrað-virkasti lestur sem gagnasafnskerfin bjóða upp á. Þegar mikið gagnamagn er lesið úr mörgum töflum samtímis með "join" má gera ráð fyrir mun minni afköstum (sjá prófanir sem nota bókhaldsgögn hér að neðan). Í mörgum tilfellum verða afköst þá 1/5 eða minna en þau afköst sem nást þegar lesið er úr einni töflu.

8.4 Bókhaldstöflur

Þegar ljóst er að tímaraðir verða ekki geymdar í venslagagnasafnskerfum í nýja vatnamælingakerfinu, stendur eftir að gagnasafnskerfi þarf einkum til þess að geyma ýmsar rekstrarupplýsingar, staðsetningar o.f.l.þ.h. Prófanirnar hér að ofan segja ekki mikið um vinnslugetu fyrir slík gögn og voru því að síðustu gerðar nokkrar tilraunir með 2 töflur með bókhaldupplýsingum en þær gefa betri mynd af þeirri vinnslu sem verður í nýja vatnamælingakerfinu að þessu leyti.

Bókhaldsgögnin voru lesin úr stýrikerfisskrám í bókhaldskerfi Orkustofnunar inn í 2 töflur. Tafla I (10 dálkar) geymir færslur úr viðskiptamannabókhaldi stofnunarinnar og tafla II (5 dálkar) geymir sundurliðun færslanna í töflu I á viðfangsefnisnúmer í verk-efnabókhaldi. Keyrðar voru 5 fyrirspurnir sem eru hliðstæðar við fyrirspurnir sem notaðar eru í C-forriti sem er hluti af bókhaldskerfi Orkustofnunar. Niðurstöður prófana á þessum töflum í Oracle og Informix má bera saman við tímann sem forrit þetta þarf til að vinna með sömu gögn í stýrikerfisskrám og fá þannig samanburð við þau af-

köst sem hefðbundin forritun í þriðjukynslóðarforritunarmálum gefur.

Fyrirspurn (1) leitar að skráningarvillum í töflu I með því að skoða allar færslurnar, tafla II er ekki lesin. Fyrirspurnir (2) og (3) leita að ólöglegum tilvísunum milli taflna I og II með "subquery" og "join". Fyrirspurn (4) leitar að misræmi milli bókaðrar upphæðar í töflu I og sundurliðunarinnar í töflu II með "join" og "group by". Fyrirspurn (5) leitar að misræmi í millifærslum í töflu I með "group by", tafla II er ekki lesin. Niðurstöður þessara prófana koma fram í töflu 3 hér að neðan.

Til samanburðar þá taka þessi gögn alls 342 Kb í stýrikerfisskrám bókhaldskerfisins (þar er ekki geymdur index). C-forrit sem vinnur nokkru meira verk en allar fyrirspurnirnar í töflunni hér að ofan tekur 4 sek. í keyrslu, og notar 2.2 sek. af CPU tíma (1.0 "user", 1.2 "system"). Þetta forrit flettir upp í 2 hliðlægum töflum fyrir hverja færslu í töflum I og II auk þess sem gert var í prófununum á Oracle og Informix. Forritið þarf jafnframt að lesa gögn hliðlægu taflanna úr stýrikerfisskrám auk taflna I og II. Samtals les forritið rúmlega 800 Kb af gögnum.

Í sambandi við plásstölurnar í töflu 3 ber að nefna að bókhaldstöflurnar í Oracle voru vegna mistaka búar til þannig að þær stækkuðu í þrepum upp á 100K þegar gögnin voru lesin inn. Plássið sem töflurnar þurfa í Oracle getur því verið allt að 100K meira fyrir hvora töflu og hvorn index en strangt tiltekið er nauðsynlegt. Þrátt fyrir að tekið sé tillit til þessa, virðast töflurnar taka talsvert meira pláss í Oracle en í Informix.

TAFLA 3: Samanburður á Oracle og Informix, Bókhaldsgögn.

	Oracle	Informix
Færsluföldi, Tafla I	5072	5072
Færsluföldi, Tafla II	10467	10467
Pláss, gögn, o.fl. (Kb)	700	460
Pláss, index (Kb)	350	164
Pláss alls (kb)	1150	624
Tími fyrirsp. 1 (liðinn/CPU) (s)	7/5	25/21
Tími fyrirsp. 2 (liðinn/CPU) (s)	28/23	180/140
Tími fyrirsp. 3 (liðinn/CPU) (s)	31/28	220/140
Tími fyrirsp. 4 (liðinn/CPU) (s)	65/51	150/140
Tími fyrirsp. 5 (liðinn/CPU) (s)	2/<1	5/2
Tími samtals (liðinn/CPU) (s)	133/108	580/443

Vinnsla bókhaldsgagnanna í Informix er mun seinlegri en í Oracle hvort sem miðað er við liðinn tíma eða CPU tíma, þrátt fyrir að gögnin taki meira pláss í Oracle. Informix er allt frá 2.3 sinnum upp í rúmlega 7 sinnum hægvirkara en Oracle þegar miðað er við liðinn tíma og hlutföllin eru svipuð þegar miðað er við CPU tíma. Samtals tóku allar fyrirspurnirnar 4.4 sinnum lengri tíma fyrir Informix og notuðu 4.1 sinnum meiri CPU tíma. Þessar tölur eru svipaðar og fyrir efnafræðigögnin.

Bæði kerfin eru ótrúlega miklu seinvirkari en C-forrit, eða 33 sinnum fyrir Oracle og 145 sinnum fyrir Informix ef miðað er við liðinn tíma (þrátt fyrir að C-forritið vinni mun meira verk). Í þessu sambandi má hafa í huga að vinnsla á bókhaldsgögnum er einmitt sú vinna sem venslagagnasafnskerfi eru sniðin að og eiga því bæði kerfin að vera á heimavelli í þessum prófunum. C-forritið þarf lítinn CPU tíma miðað við gagnasafnskerfin sem sýnir að "raunveruleg" vinna sem fyrirspurnirnar krefjast (samlagning, frádráttur, samanburður á gildum, o.s.fr.) hlýtur að vera mjög lítill hluti af CPU tímanum sem gagnasafnskerfin nota. Þetta ásamt því að notaður CPU tími allra fyrirspurnanna er tiltölulega nálægt liðnum tíma, bendir til þess að gagnasafnskerfin noti mikinn meirihluta tímans sem þau þurfa til þess að svara fyrirspurnum í CPU "overhead". Nefna má að HP9000/825 tölvan getur framkvæmt um 5 milljónir aðgerða (heiltölureikniaðgerðir,

samanburður, o.s.fr.) á sek. Á 25 CPU sek. (Oracle fyrirspurnir (2) og (3)) getur tölvan því framkvæmt um 125 milljónir aðgerða, sem svarar til um 8000 aðgerða á hverja færslu í töflum I og II. Á 140 CPU sek. (Informix fyrirspurnir (2), (3) og (4)) svara afköst tölvunnar til um 50000 aðgerða á hverja færslu í töflum I og II. Ljóst er að einungis þarf lítið brot af þessum aðgerðum til þess að vinna "raunverulega" vinnu í sambandi við fyrirspurnirnar.

Samkvæmt ofansögðu er líklegt að hlutfallslegur munur á afköstum Oracle og Informix í fyrirspurnum á öðrum og hraðvirkari tölvum sé svipaður og á HP9000/825 tölvunni. Þetta stafar af því að gagnasafnskerfin virðast eiga hlutfallslega auðvelt með að lesa gögn af diskum en eyða mestöllum tíma sínum í CPU "overhead" í sambandi við úrvinnslu gagnanna eftir að þau eru komin í minni tölvunnar. Hraðvirkar vinnustöðvar geta verið með 5-10 sinnum öflugra miðverk (CPU) en þær lesa alla jafna ekki mikið hraðar af diskum en HP9000/825 tölvan. Það er ekki líklegt að diskurinn verði flöskuháls í prófununum hér að ofan þó svo vinnslugeta miðverksins margfaldist 5-10 sinnum. Því má ætla að Oracle sé á stærðarþreppinu 4 sinnum hraðvirkara en Informix í fyrirspurnum af þessum toga á mörgum tölvutegundum tiltölulega óhád afköstum tölvanna.

Rétt er að nefna að samanburðurinn hér að ofan tekur aðeins til fyrirspurna frá einum

notanda (en ekki skráningar) og segir því ekkert um afköst kerfanna í svokallaðri OLTP vinnslu sem oft er miðað við í samanburði á gagnasafnskerfum.

8.5 Niðurstöður

Niðurstöður prófananna hér að ofan eru eftirfarandi:

1. Oracle er á stærðarþrepinu 4 sinnum hraðvirkara en Informix í fyrirspurnum.
2. Plássþörf kerfanna er í aðalatriðum sambærileg. Oracle er heldur plássfrekara í töflum þar sem flestir dálkar eru notaðir en Informix getur þurft meira pláss í gisnum töflum.
3. "Optimizer" beggja kerfanna eru viðkvæmir fyrir framsetningu fyrirspurna þannig að fyrirspurnir sem kerfin ráða illa við geta tekið meira en 100 falt lengri tíma en öðruvísi orðaðar en jafngildar fyrirspurnir. Informix virðist ef eitthvað er viðkvæmara fyrir þessu en Oracle þrátt fyrir svokallaðan "cost based optimizer" sinn.
4. Bæði kerfin eru 10 til 100 sinnum hægvirvari en forrit sem skrifuð eru í hefðbundnum þriðjukynslóðarforritunarmálum.
5. Kerfin henta bæði illa til geymslu tímaraða á "normaliseruðu" formi, sérstaklega Informix. Plássþörf tímaraðanna á þessu formi er 4-10 sinnum meiri en í hefðbundum stýrikerfisskrám. Vinnsluhraði við fyrirspurnir er um 20-50 falt hægari (Oracle) og um 200-500 falt hægari (Informix) en hægt er að ná með þriðjukynslóðarforritunarmálum og stýrikerfisskrám.

9. KOSTNAÐUR

Kostnaður við að koma upp gagnasafnskerfunum á fjölnotendatölvu á OS og á vinnustöðvum með einum notanda á Egilstöðum og á Landsvirkjun kemur fram í töflum hér að neðan.

Orkustofnun hefur afsláttarsamning við Oracle og er tekið tillit til þess í töflunum. Verð til Landsvirkjunar er tekið beint úr verðlistum og gæti því lækkað ef samningar nást um það við sölumenn.

Fjölnotendaleyfi fyrir Oracle er þegar til á Orkustofnun en kostnaður við það er engu að síður hafður með til samanburðar. Til greina kemur að nota ekki notendaskil Oracle (SQL*Menu, SQL*Forms og SQL*ReportWriter) en kaupa þess í stað búnað frá þriðja aðila. Þá fellur burt liðurinn "töl" í tölunni fyrir Oracle en í staðinn kemur kostnaður við kaup á öðrum hugbúnaði. Gera má ráð fyrir að hugbúnaður frá þriðja aðila verði dýrari en tólin frá Oracle.

Í sambandi við kostnað við Informix þá er miðað við svokölluð "run time" leyfi á vinnustöðvunum. Ef keypt eru full leyfi fyrir þær þá hækkar kostnaðurinn verulega ("run time" leyfi kosta á bilin 33% til 75% af fullu verði). Á fjölnotendatölvu OS er miðað við 2-16 notanda "run time" fyrir einingarnar 4GL og ESQL en 1 notanda þróunarleyfi fyrir þessar einingar sem felur í sér DBG auk 4GL og ESQL.

Verð á einingum Oracle er gefið upp í dönskum kr. án virðisaukaskatts. Verð á einingum Informix er gefið upp í þús. íslenskra kr. án virðisaukaskatts.

TAFLA 4: Kostnaður, Oracle.

Fjölnotendatölva á OS (2-16 not.)	Kostn. (DKR)
Grunnkerfi	123.900
SQL*Plus	30.975
Nettenging	43.365
Tól	92.925
Alls: 291.165	
Alls með 66.7% afslætti:	96.958
Viðhald 3 ár:	43.675
Vinnust. á Egilsst. (1.not)	
Grunnkerfi	42.840
SQL*Plus	10.710
Nettenging	14.994
Tól	32.130
Alls: 100.674	
Alls með 66.7% afslætti:	33.524
Viðhald 3 ár:	25.143
Vinnust. á Landsv. (1 not.)	
Grunnkerfi	42.840
SQL*Plus	10.710
Nettenging	14.994
Tól	32.130
Alls án afsláttar:	100.674
Viðhald 3 ár:	75.506
Samtals (DKR):	231.156
Samtals með 3 ára viðhaldi:	375.480
Samtals með vask (þ. ÍKR):	2634
Samtals með 3 ára viðhaldi:	4279

TAFLA 5: Kostnaður, Informix.

Fjölnotendatölva á OS	Kostn. (þ. ÍKR)
Grunnkerfi (2-16 not.)	677
Nettenging (2-16 not.)	170
4GL, ESQ (rt, 2-16 not.)	147
4GL, DBG, ESQ (1 not.)	334
Alls: 1328	
Alls með 55% afslætti:	598
Viðhald 3 ár:	313
Vinnust. á Egilsst. (rt, 1 not.)	
Grunnkerfi	250
Nettenging	63
4GL, ESQ	72
Alls: 385	
Alls með 55% afslætti:	173
Viðhald 3 ár:	68
Vinnust. á Landsv. (rt, 1 not.)	
Grunnkerfi	250
Nettenging	63
4GL, ESQ	72
Alls án afsláttar:	385
Viðhald 3 ár:	152
Samtals:	1156
Samtals með 3 ára viðhaldi:	1689
Samtals með vask (þ. ÍKR):	1439
Samtals með 3 ára viðhaldi:	2103