

Kotitehtävät vko 4

Puskurit. Transaktioidenhallinta. Suorituskyky. Monitorointi ja jäljitys. Hälytykset. IBM DB2:een tutustuminen:

- a) Miten kuormitustestin /benchmark testin **koejärjestely** tulee suunnitella. Tätä on käsitelty esim DB2-manuaalissa [Troubleshooting and Tuning Database Performance](#) (Chapter 1)

Vastaus:

Järjestelyistä on kerrottu kattavasti osoitteessa: [Troubleshooting and Tuning Database Performance](#) (Chapter 1) s. 5-10 (löytyy myös IBM:n osoitteesta:

http://public.dhe.ibm.com/ps/products/db2/info/vr97/pdf/en_US/DB2PerfTuneTroubleshoot-db2d3e971.pdf). Lyhyesti esitettynä, on oleellista, että järjestelyt vastaavat todellista tilannetta

tietokantapalvelimen kannalta. Jos testauksen aikana käynnistyy esimerkiksi muita ohjelmia, joilla ei ole mitään tekemistä tietokantapalvelimen kanssa, on tietysti pyrittävä välttymään tältä. Testi on oltava myös sellainen, että se mittaa jotain tärkeää ominaisuutta tietokannan käytönkannalta. Testi on oltava myös sellainen, että sen pystyy toistamaan täsmälleen samanlaisena uudelleen. Tietokantapalvelimen konfiguraatio on siis oltava yksityiskohtia myöten selvillä.

- b) Suunnittele miten kannattaisi tietokantapalvelimen suorituskykyä mitata.

Vastaus:

Asiaa on käsitelty kattavasti osoitteissa: Ks. [http://technet.microsoft.com/en-us/library/ms188320\(v=sql.105\).aspx](http://technet.microsoft.com/en-us/library/ms188320(v=sql.105).aspx), [http://technet.microsoft.com/en-us/library/ms190382\(v=sql.105\).aspx](http://technet.microsoft.com/en-us/library/ms190382(v=sql.105).aspx). Tietokantapalvelimen muistikäyttö, levyliikenne ja prosessorin kuorma ovat tärkeitä asioita arvioitaessa palvelimen suorituskykyä. Myös palvelintietokoneen verkkokortin tai korttien liikennekuorma kiinnostaa samasta syystä. Tässäkin mittauksessa on tärkeää, että muut tietokantapalvelimen toimintaan liittymättömät ohjelmat eivät vääristä mittaustuloksia. Eli ne ajetaan alas palvelintietokoneelta tai siirretään muille palvelimille jos mahdollista. Siten tehtäviin mittauksiin voi ainakin tältä osin luottaa. Peräti itse ylläpitäjäkin ottaa yhteyden mitattavaan palvelintietokoneeseen omalta työasemaltaan remote desktop connection:in avulla. Siten minimoidaan muistin ja prosessorin kuormittaminen tältäkin osin palvelintietokoneella. Kaikkia osa-alueita ei myöskään kannata mitata kerralla vaan valita vain osa niistä siten, että itse counter:it eivät muodostu liian suureksi kuormaksi palvelimelle. Loput osa-alueethan voi mitata myöhemmin.

Miten mittaisit koko järjestelmän suorituskykyä?

Vastaus:

Ks. osoitteesta: [http://technet.microsoft.com/en-us/library/ms188320\(v=sql.105\).](http://technet.microsoft.com/en-us/library/ms188320(v=sql.105).)

- c) Mitä Buffer Pool:illa tarkoitetaan tietokantapalvelimella?

Vastaus:

Buffer Pool on tietokantapalvelimen käynnistyksen yhteydessä palvelintietokoneen muistista tietokantapalvelimen käyttöön varattu muistialue. Sinne luetaan levyllä olevat tietokannan datat, joita käyttäjät tarvitsevat. Esimerkiksi, jos käyttäjä haluaa muuttaa tietokannassa olevan henkilön tietoja, tarkistaa tietokantapalvelin, että mistä kyseisen henkilön tiedot löytyvät. Ovatko ne jo valmiiksi luettu Buffer Pool:in aikaisemmin vai onko ne haettava levyltä. Tämä tietysti vie enemmän aikaa kuin suoraan muistista lukeminen. Kun Buffer Pool:ssa olevaa tietoa muutetaan merkitään se muuttuneeksi, jotta tietokantapalvelin tietää, että levyllä oleva vastaava tieto on vanhentunut ja se pitää päivittää buffer pool:ssa olevalla päivitetyllä tiedolla. Asiasta on kerrottu tarkemmin SQL Server:in osalta osoitteessa: [http://technet.microsoft.com/en-us/library/aa337525\(v=sql.105\).aspx](http://technet.microsoft.com/en-us/library/aa337525(v=sql.105).aspx)

d) Mikä on keskeisin asia miksi transaktioidenhallintaa tarvitaan?

Vastaus:

Transaktioidenhallinnassa halutaan varmistaa, että operaatio onnistuu virheettömästi, tietokannan eheys pysyy kunnossa ja virhetilanteissa voidaan peruuttaa operaation sql kieliset lauseet virheettömästi siten että tietokannan data palautuu aikaisempaan tilaansa. Kyse on siis siitä, että halutaan ns. ACID ominaisuudet. Hyvä käytännön esimerkki tästä voisi olla esimerkiksi maksun maksaminen nettiyhteyttä tms. käyttäen. Kun asiakas haluaa maksaa vaikkapa sähkölaskunsa on hänen tililtään vähennettävä maksun summa ja heti perään lisättävä sähkölaitoksen pankkitilille. Tämä vaatii siis kaksi update lausetta. Niiden ajon ollessa kesken, ei tietenkään ole hyvä asia, jos vaikkapa sähkölaitoksen controlleri pääsee katsomaan tilannetta, jolloin asiakkaan tililtä on kyllä vähennetty maksun summa, mutta sitä ei ole vielä ehditty lisätä sähkölaitoksen tilille. Tilisiirron operaatio pitää siis eristää muista operaatioista siten, että ne eivät pääse näkemään kesken olevia operaatioita. Myös virhetilanteissa on pystyttävä peruuttamaan kesken jäänyt tilisiirto virheettömästi alkuperäiseen tilaansa. Tämän vuoksi esimerkiksi asiakkaan operaatio eristetään omaksi transaktiokseen ja sähkölaitoksen controllerin operaatio omaksi transaktiokseen. Niistä ajetaan ensimmäisenä se joka ehtii ensimmäisenä ajoon. Toisen on odotettava, jos se tarvitsee samoja resursseja, kuten esimerkiksi sähkölaitoksen controlleri halusi nähdä asiakkaamme tilitiedot samaan aikaan kuin asiakas oli tekemässä maksua. Kun tilisiirto on ehditty tehdä loppuun asti virheettömästi, pääsevät seuraavat transaktiot käyttämään vapautuneita resursseja. Ks. <http://msdn.microsoft.com/en-us/library/ms190612%28v=sql.105%29.aspx> ja <http://msdn.microsoft.com/en-us/library/ms188929.aspx>.

e) Selitä lyhyesti mihin yleensä tarvitaan tietokantaympäristössä **monitorointeja** ja mitä tarkoituksia on erilaisia jäljityksillä (**tracing**). Voit tarkastella tätä yleisesti tai tietyn DBMS-tuotteen kannalta.

Vastaus:

Monitoroinneista on kerrottu ehkä selkeimmin SQL Server:in sivuilla: <http://msdn.microsoft.com/en-us/library/ms191199%28v=sql.105%29.aspx> ja <http://msdn.microsoft.com/en-us/library/bb510705%28v=sql.105%29.aspx>. Tietokantapalvelimella halutaan siis monitoroida tietokantainstanssiin yhteydessä olevien käyttäjien yhtyeiksi, prosessien numeroita ja niiden tilaan. Myös käyttäjien suorittamat sql kieliset lauseet ovat kiinnostuksen kohteena tässä ja objektien (taulut, proseuurit, indeksit, triggerit jne) lukitukset ja millaisia nämä lukot ovat (read vai write lukkoja esimerkiksi). Transaktiot voivat päättyä myös "pattitilanteeseen" eli deadlock tilanteeseen. Tietokantapalvelin ratkaisee tämän yleensä siten, että transaktiosta rollback:taan se jos menetetään vähinten tehtyä työtä, tai se jolla on eniten työtä vielä tekemättä. Tässä on tietysti seurattava sitä, ettei käy niin, että jotain transaktiota ei saada koskaan ajettua loppuun. Myös tätä halutaan monitoroida.

IBM DB2:n osalta monitoroinnista on kerrottu osoitteessa: <http://my.haaga-helia.fi/~dbms/ict4tn006/DB2PerfTuneTroubleshoot-db2d3e972.pdf> kappaleessa 2 ja <http://pic.dhe.ibm.com/infocenter/db2luw/v10r1/index.jsp?topic=%2Fcom.ibm.db2.luw.admin.dm.doc%2Fdoc%2Ft0057364.html>. Tavoitteet DB2:n monitoroinnissa ovat samat kuin minkä tahansa tietokantapalvelimen, esimerkiksi SQL Server:in.

f) Mitkä hälytykset (error levels) kannattaa laittaa esimerkiksi SQL Server:ssä tulemaan dba:n sähköpostiin. Ainakin jos ylläpidettäviä tietokantoja on vain yksi.

Vastaus:

Microsoft Server:in tärkeimmät/vakavimmat hälytykset ovat välillä 19-24. Niistä on kerrottu kattavasti osoitteissa: [http://technet.microsoft.com/en-us/library/ms180982\(v=sql.105\).aspx](http://technet.microsoft.com/en-us/library/ms180982(v=sql.105).aspx), [http://technet.microsoft.com/en-us/library/ms175076\(v=sql.105\).aspx](http://technet.microsoft.com/en-us/library/ms175076(v=sql.105).aspx). Itse error level:ien numerot ja niiden merkitys on kerrottu osoitteessa: <http://technet.microsoft.com/en-us/library/ms164086.aspx>.

IBM DB2:n dokumentaatiossa on ehkä selkeimmin kerottu trace:istä:

<http://publib.boulder.ibm.com/infocenter/db2luw/v9r5/index.jsp?topic=%2Fcom.ibm.db2.luw.admin.trb.doc%2Fdoc%2Fc0020800.html>. Trace:en avulla voi siis tallentaa tietoa tietokantapalvelimessa suoritettavista operaatioista ja tapahtumista. Edellä esiteltyjä monitoroitavia asioita voi siis automatisoida seurantaan ja tallentaa syntyvät "raportit" trace:n avulla.

Microsoft:in SQL Server trace:stä on kerrottu osoitteessa: <http://msdn.microsoft.com/en-us/library/hh245121.aspx> ja <http://msdn.microsoft.com/en-us/library/ms188662.aspx>