

Juha-Pekka Mielonen

# Raportointiratkaisujen kehitys ja ylläpito Microsoft Business Intelligencessä

Metropolia Ammattikorkeakoulu  
Insinööri (AMK)  
Tietotekniikka  
Insinöörityö  
Päivämäärä 15.5.2012

Tekijä(t) Otsikko	Juha-Pekka Mielonen Raportointiratkaisujen kehitys ja ylläpito Microsoft Business Intelligencessä
Sivumäärä Aika	41 sivua 15.5.2012
Tutkinto	Insinööri (AMK)
Koulutusohjelma	Tietotekniikka
Suuntautumisvaihtoehto	Tietoverkot
Ohjaaja(t)	Yliopettaja Janne Salonen
<p>Insinööriyön tavoitteena oli tutustua Microsoft Business Intelligencen toteutukseen, selvittää sen käyttökohteita ja tutkia toimintaa harjoitustehtävien avulla.</p> <p>Työn pohjana toimi Microsoft SQL Server 2008, Business Intelligence Development and Maintenance -kurssin (70-448) materiaali ja harjoitukset.</p> <p>Työssä käydään ensin läpi fyysinen ja virtuaalinen toteutusympäristö. Lisäksi käydään läpi BI:n olennaista aluetta, ETL:ää ja tietovaraston keskeistä mallia. Samalla tutustutaan Microsoft SQL -kehitystyökaluihin Microsoft Server Management Studio ja Business Intelligence Development Studio. Tämän jälkeen tutustutaan Business Intelligencen sisäisiin komponentteihin.</p> <p>Insinööriyössä tutustuttiin syvällisesti Microsoft Business Intelligence -tuotteen kehitysympäristöihin ja ylläpitotoimiin. BI havaittiin joustavaksi työkaluksi pienistä projekteista lähtien aina suurien yritysten raportointityökaluksi asti. Lopuksi käydään läpi kurssiin kuuluvat harjoitustehtävät.</p>	
Avainsanat	Business Intelligence, SQL, datamigraatio, 70-448

Author(s) Title	Juha-Pekka Mielonen Reporting Solution Development and Maintenance in Microsoft Business Intelligence
Number of Pages Date	41 pages 15 May 2012
Degree	Bachelor of Engineering
Degree Programme	Information Technology
Specialisation option	Data networks
Instructor(s)	Janne Salonen, Senior Lecturer
<p>The purpose of the final project was to explore the Microsoft Business Intelligence solution as well as to clarify the business needs of BI and examine the solution through the course material.</p> <p>This final project was based on the self-paced kit called Microsoft SQL Server 2008, Business Intelligence Development and Maintenance (70-448).</p> <p>The first part of the study covers the physical and virtual environment this course platform was implemented in. The following chapter describes the ETL process and dimensional model which both are essential for Business Intelligence. The Microsoft Server Management Studio and Business Intelligence Development Studio development tools were also discussed in this project. After these subjects the internal components are covered.</p> <p>This thesis covers profoundly the development and maintenance of Microsoft Business Intelligence software. BI was found fit to be an agile tool from the small projects to corporate size reporting systems. Finally the exercises for the course are explained in detail.</p>	
Keywords	Business Intelligence, SQL, data migration, 70-448

## Sisällys

1	Johdanto	1
2	Ympäristö	2
2.1	Virtuaalialusta, VMware	2
2.1.1	Windows 2008 Server R2	3
2.1.2	SQL Server 2008 R2 + Adventureworks2008R2 -tietokannat	4
3	ETL ja ulotteinen mallinnus (Dimensional modeling)	5
3.1	ETL - Extract, Transform, Load	5
3.2	Dimensional modeling (DM) - ulotteinen mallinnus	7
4	Microsoft SQL -kehitystyökalut	8
4.1	SQL Server Management Studio (SSMS)	8
4.1.1	Object Explorer	9
4.1.2	Solutions Explorer	9
4.1.3	Template Explorer	10
4.2	Business Intelligence Development Studio (BIDS)	11
5	BIDS-komponentit	12
5.1	SQL Server Integration Services (SSIS)	12
5.2	Server Analysis Services (SSAS)	15
5.3	SQL Server Reporting Services (SSRS)	17
6	Kurssin harjoitukset	19
6.1	Luku 1: SSIS-paketin kehitys	19
6.2	Luku 3: SSIS-pakettien konfigurointi ja asentaminen	22
6.3	Luku 4: SSIS-pakettien hallinnointi, suojaus ja suorittaminen	24
6.4	Luku 5: SSAS-kuution kehitys	25
6.5	Luku 6: SSAS-kuutioiden laajentaminen	27
6.6	Luku 7: SSAS-tietovaraston hallinta, prosessointi ja asennus	29
6.7	Luku 8: SSAS:n hallinta ja turvaaminen	30
6.8	Luku 9: SSAS-datalouhintatyöskentely	32
6.9	Luku 10: SSRS-raporttien kehitys	34

6.10	Luku 11: Raporttien laajentaminen ja käyttöönotto	36
6.11	Luku 12: Asennettujen raporttien ja datalähteiden ajastus ja turvaaminen	37
6.12	Luku 13: SSRS-palvelimen konfigurointi ja hallinnointi	38
7	Yhteenveto	39
	Lähteet	41

## Lyhenteet ja käsitteet

BI	Business Intelligence. Microsoftin ohjelmistotuote liiketoimintatiedon hallintaan.
BIDS	Business Intelligence Development Studio. Microsoftin kehitysympäristö Business Intelligence -tuotteelle.
Cube	Kuutio. SSAS:n datamallin perusmoduli.
DB schema	Tietokantakaavio. Tietokannan malli.
DMX	Data Mining Extensions. SSAS:n tukema kyselykieli datalouhintaa varten.
DSV	Data Source View, datalähdenäkymä. Käyttäjänäkymä datalouhintatyökalussa.
ER	Entity-Relationship Model. Tietovaraston tietomalli, ulotteisen mallinnuksen vaihtoehto.
ETL	Extract, Transform, Load. Standardoitu termi tiedon liikuttamiselle ja muuntamiselle.
FTP	File Transfer Protocol, Tiedonsiirtoprotokolla.
KPI	Key Performance Indicator. Suorituskykyilmaisim.
Measure	Mittapiste. SSAS:n mittamääre.
Measure Group	Mittausryhmä. SSAS:n mittapistekokoelma.
MCTS	Microsoft Certified Technology Specialist. Microsoftin sertifiointiohjelma.

MDX	Multi Dimensional Expressions. Moniulotteinen kyselykieli OLAP-tietokantoihin.
MOLAP	Multidimensional OLAP. Moniulottoinen, reaaliaikainen ja analyttinen prosessointi, datalouhinnan prosessi.
MSBI	Microsoft Business Intelligence. Microsoftin liiketoimintatiedon hallintaohjelmisto.
MSDTC	Microsoft Distributed Transaction Coordinator. Nykyaikaisten Windows-käyttöjärjestelmien palvelukomponentti, joka ohjaa useita resurssimanagereja koskevia tapahtumia.
ODBC	Open Database Connectivity. Avoin tietokantayhteyden rajapintastandardi.
OLAP	Online Analytical Processing. Reaaliaikainen analyttinen prosessointi, datalouhinnan prosessi.
PDF	Portable Document Format. Adoben standardoitu dokumenttiformaatti.
QA	Quality Assurance. Laadunvarmistus.
SSAS	SQL Server Analysis Services. Business Intelligence - ohjelmiston komponentti.
SSIS	SQL Server Integration Services. Business Intelligence -ohjelmiston komponentti.
SMTP	Simple Mail Transfer Protocol. Sähköpostin lähetysohjelma.
SSMS	SQL Server Management Studio. Microsoftin kehitysympäristö SQL Server -tuotteelle.

SSRS	SQL Server Reporting Services. Business Intelligence -ohjelmiston komponentti.
SQL	Simple Query Language. Standardoitu kyselykieli tietokantahakuihin.
VSTA	Visual Studio Tools for Applications 2.0. Työkalukokoelma modifioimaan Visual Basic -toteutuksia.
XML	Extensible Markup Language. Merkintäkieli, jonka tarkoitus on olla sekä ihmisen että tietokoneen luettavissa.
XMLA	XML for Analysis. Merkintäkieli analyyttisiä järjestelmiä, kuten OLAPia varten.



## **1 Johdanto**

Business Intelligence eli liiketoimintatiedon hallinta on yrityksen suorittamaa liiketoiminnan tietojen hankintaa, tallennusta ja analysointia. Aiemmin tietoa kerättiin lähinnä ulkopuolisista toimijoista, kuten kilpailijoista, mutta nykyisin BI on erittäin tärkeä osa myös yrityksen sisäisen toiminnan analysointia ja kehitystä. BI-sovellusten avulla saatuja tietoja käytetään toiminnan parantamiseen, tehostamiseen sekä päätöksenteon tukena.

Yrityksien tietokantoihin kertyy valtava määrä tietoa, josta voidaan nähdä historiallinen kehitys, seurata nykyisiä trendejä ja tehdä tarkkojakin ennusteita tulevista liiketoiminnallisista trendeistä. Tähän tehtävään on kehitetty laaja valikoima erilaisia ohjelmistoja, jotka helpottavat datan analysointia yrityksissä. Tunnettuja BI-tuotteiden valmistajia ovat esimerkiksi perinteiset tietokantayritykset Oracle, IBM ja myöhemmin Sybase.

Tämä työ keskittyy Microsoft Business Intelligence -ratkaisun toiminnan ja eri osa-alueiden syvempään tarkasteluun. Microsoft BI tulee osana Microsoft SQL Server -palvelinohjelmistoa ja käytön oletuksena on, että Microsoft SQL Server on entuudestaan tuttu. Työ tehtiin virtualisoidussa ympäristössä tekijän kotona ja keskittyy Microsoft Certified Technology Specialist 70-448 (MCTS) -sertifikaattiin valmentavan kurssin materiaaliin ja tehtäviin.

## 2 Ympäristö

Insinööriyön ympäristö asennettiin virtuaalipalvelimelle käyttäjän kotitietokoneeseen.

Koneen kokoonpano oli seuraava:

AMD Phenom II X3 720 (2.80GHz kolmiydinprosessori)

4Gt RAM

Windows 7 Professional 64-bit, Service Pack 1

### 2.1 Virtuaalialusta, VMware

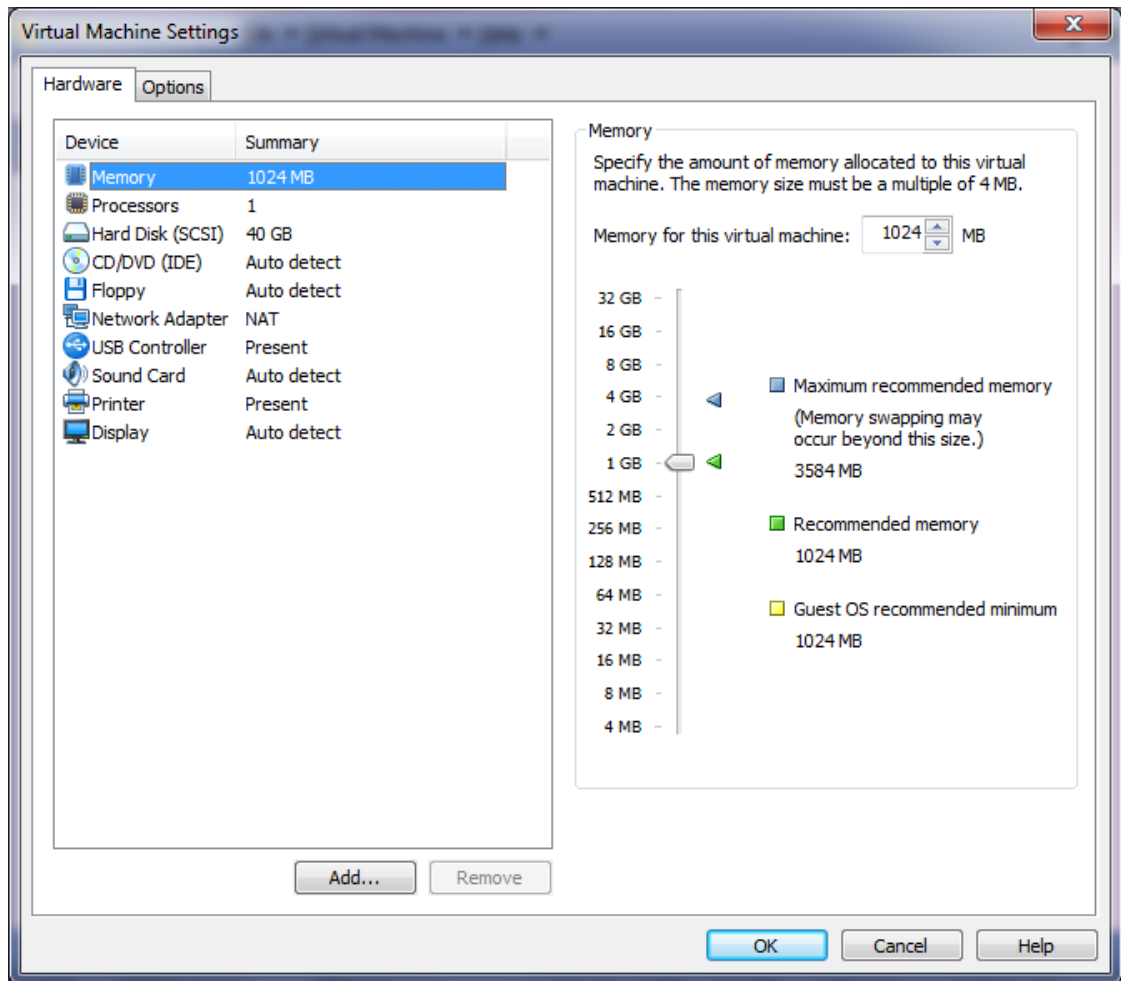
VMware on yleisesti käytetty virtualisointiohjelmisto. Ohjelmisto mahdollistaa eri käyttöjärjestelmien ajamisen virtuaalisessa tietokoneessa fyysisen pääkoneen sisällä erillisinä ohjelmina. Tämä helpottaa uusien ympäristöjen rakentamista ja vähentää kustannuksia merkittävästi.

Diane Greene, Mendel Rosenblum, Scott Devine, Edward Wang ja Edouard Bugnion perustivat VMware Incorporation -yhtiön vuonna 1998. EMC Corporation osti VMware Incorporationin vuonna 2004 ja VMware Inc toimivat tytäryhtiönä.

VMware alustojen kirjo on valtavan laaja. Esimerkkeinä mainittakoon Pilvi-ratkaisut (Cloud Infrastructure), Palvelin ja datacenter-ratkaisut (Server & Datacenter products) ja Työpöytä- ja loppukäyttäjäratkaisut (Desktop & End-User Computing Products).

Vapaasti käytettävät työpöytäversiot löytyvät kaikille suurille käyttöjärjestelmille eli Microsoft Windowsille, useille Linux-jakeluille (distroille) ja Mac OS X:lle.

Työssä käytetty virtuaalipalvelin oli Vmware Player 3.1.5 build-491717. Tämä on yksityiskäytössä maksuton. Virtuaalikoneelle oli allokoitu 1Gt muistia ja 1 prosessori. Tämä riitti varsin hyvin palvelimen käyttöä varten.



Kuvio 1. Virtuaalikoneen fyysiset asetukset

### 2.1.1 Windows 2008 Server R2

Fyysisessä tietokoneessa oli asennettuna Windows 7 ja siihen olisi voinut asentaa Microsoft SQL Serverin. Päätettiin kuitenkin tehdä puhdas Windows Server 2008 R2 -asennus, jonka päälle asennettiin Microsoft SQL Server, jolloin aiemmin asennetut ohjelmat eivät häiritsisi ohjelmistojen toimintaa. Virtualisoitu palvelin oli tähän tarkoitukseen erinomainen valinta helppoutensa takia. Itse käyttöjärjestelmän asennuksessa ei kohdattu ongelmia.

### 2.1.2 SQL Server 2008 R2 + Adventureworks2008R2 -tietokannat

Microsoft Business Intelligence on osa Microsoft SQL Server -tuoteperhettä ja asennetaan osana MS SQL -palvelinta. Asennuksen apuna käytettiin kurssin "Microsoft SQL Business Intelligence (BI) Development and Maintenance 70-433" -materiaalin lisäksi "Microsoft SQL Server 70-432" -materiaaleja. Asennusohjeet poikkesivat hieman toisistaan, eikä pieniltä lisäasennuksilta vältytty.

Tietokantoina käytettiin Microsoftin varta vasten opiskelu- ja kurssikäyttöön tehtyjä AdventureWorks 2008 R2 -kantoja. Tietokannat tulevat "Microsoft SQL Server 2008 - Business Intelligence Development and Maintenance Self-Paced Training Kit" -teoksen mukana CD:llä tai ne voi ladata Microsoft Database Product Samples -sivustolta [1].

Alun perin ympäristö asennettiin "Microsoft SQL Server 70-432" -suoritusta varten. Alkuperäinen asennus oli varsin riisuttu ja siksi myöhemmin jouduttiin asentamaan useita komponentteja erillisenä asennuksena. Mainittakoon näistä Integration Services, Analysis Services ja Reporting Services, jotka kaikki ovat Business Intelligence -käytön kannalta aivan olennaisia. Lisäosien asentaminen osoittautui kuitenkin helpoksi ja valmiit kantayhteydet voitiin ottaa käyttöön ilman erillistä konfigurointia. Lisäksi MS SQL Serverin hallinta tuli tutuksi.

### 3 ETL ja ulotteinen mallinnus (Dimensional modeling)

#### 3.1 ETL - Extract, Transform, Load

ETL eli Extract, Transform and Load on standardoitu termi esittämään datan liike- ja muutosprosesseja [2]. ETL on keskeinen komponentti, kun ladataan tietoa lähteestä, muunnetaan tieto ja tallennetaan se kohteeseen.

Extract eli poiminta - ETL:n ensimmäinen vaihe voi olla myös haastavin toteuttaa. Useissa ETL-projekteissa lähdetiedot sijaitsevat useissa eri lähteissä. Lähteinä voi toimia laaja kirjo tietokantoja, tiedostoja tai erillisiä tiedonhallintajärjestelmiä.

Ongelmaksi usein muodostuu lähdetiedon formaattien monimuotoisuus. Yleisesti extract-osan tarkoitus on hakea haluttu tieto kohteesta ja konvertoida se yhteen haluttuun formaattiin, jota seuraava vaihe, muunnos, osaa lukea. Poiminnan luontainen tehtävä on myös parsia noudettu data ja tarkastaa, onko muoto tai rakenne haluttu. Jos näin ei ole, voidaan data hylätä ja näin välttyä epävalidin datan jatkokäsittelyltä.

Transform eli muunnos - ETL:n toisen vaiheen tehtävä on muokata poimittua dataa haluttujen sääntöjen ja funktioiden mukaan ja toimittaa se eteenpäin valmiiksi kohteeseen latausta varten. Joissakin tapauksissa lähdetietoja tarvitsee muokata hyvin vähän tai ei ollenkaan. Yleensä muunnoksia on kuitenkin tarve tehdä erilaisten kaupallisten tai teknisten syiden takia. Seuraavana on yleisimpiä esimerkkejä.

- Ladataan vain tietyt sarakkeet. Lähdetiedoissa voi olla kymmeniä kenttiä, esimerkiksi nimi, ikä, sukupuoli, osoite, puhelinnumero jne. Mahdollisesti halutaan poimia esimerkiksi vain ikä ja sukupuoli.
- Koodattujen arvojen muunnos (esim. lähdetiedoissa sukupuoli voi olla M/F mutta kohdejärjestelmä käyttää arvoja M/N).
- Vapaamuotoisten arvojen uudelleenkoodaus (esim. lähteen "Mies" muunnetaan kohdejärjestelmän arvoksi "1").
- Johdetaan laskettuja arvoja (esim. kokonaisynti = määrä\*kappalehinta).
- Järjestetään tiedot (sorttaus).
- Liitetään tietoja useista lähteistä ja duplikaattien poistaminen.

- Lasketaan yhteen (esim. tuoteryhmän yhteenlaskettu myynti).
- Sijaisavainten generoiminen.
- Useiden sarakkeiden muuntaminen useiksi riveiksi ja päin vastoin (pivoting ja transposing).
- Yksittäisen sarakkeen jakaminen useaksi sarakkeeksi (esim. pilkulla erotellut listat omiksi sarakkeiksi Exceeliin).
- Kaikenmuotoinen yksinkertainen tai monimutkainen datan validointi. Validoinnin epäonnistuessa joko osa tai kaikki datasta hylätään tai ei hylätä mitään riippuen säännöistä ja poikkeushallinnasta.

Load eli lataus huolehtii nimensä mukaisesti datan lautauksesta kohteeseen. Kohde on yleensä datavarasto (data warehouse, DW). Kurssin käyttämä kohde on AdventureWorksDW2008R2. Yleisesti lataus vaihtelee tarpeiden mukaan. Data voidaan esimerkiksi ylikirjoittaa jokaisella latauskerralla, lisätä dataa kohteeseen edellisten lisäksi ja samalla poistaa tiettyä aikaa vanhemmat tiedot ja niin edelleen.

Latausvaihe kommunikoi kohdetietokannan kanssa ja ottaa siten huomioon kohteeseen määritellyt rajoitukset ja laukaisee lataukseen liittyvät liipaisimet eli triggerit, joita ovat esimerkiksi pakolliset kentät, uniikkisuus ja viittausten eheydet.

ETL:n käyttökohteita on useita. Yrityksen eri osastot voivat tallentaa asiakkaiden tietoja eri tavalla ja järjestyksessä tai jopa täysin erilaiseen järjestelmään. Joku käyttää asiakkaasta nimen sijaan numeroa ja niin edelleen. ETL:n avulla tiedot voidaan muokata yhteneviksi ja tallentaa esimerkiksi tietokantaan tai tietovarastoon.

Toinen yleinen käyttökohde on siirtää tietoa sovellukselta kokonaan uuteen sovellukseen. Uusi sovelluspalvelin voi käyttää jopa aivan eri tietokantatoimittajaa ja siten mahdollisesti erilaista tietokantaa. Yleisempiä ovat migraatiot vanhentuneesta versiosta uudempaan.

Microsoft Business Intelligence on yksi mahdollinen toteutus ETL-ajolle. BI:ssä on täysi tuki yllämainituille toiminnoille, ja se onkin nopeasti kasvattamassa osuuttaan markkinoilla, joilla muita toimijoita ovat muun muassa Oracle Data Integrator, IBM InfoSphere ja SnapLogic muutamia mainitakseni.

### 3.2 Dimensional modeling (DM) - ulotteinen mallinnus

Ulotteinen mallinnus on nimitys koosteelle tekniikoita ja konsepteja, joita käytetään tietovaraston (data warehouse) suunnittelussa. SSAS käyttää ulotteista mallinnusta[3] toisen yleisen mallin eli entity-relationship -mallin (ER model) sijaan. Ulotteinen mallinnus ei välttämättä vaadi relaatiotietokantaa, vaikka työssä sellaista käytettiin. Loogisella tasolla samaa ajattelua voi hyödyntää niin moniulotteisissa tietokannoissa kuin flättitiedostoissakin. DM on tarkoitettu tukevan loppukäyttäjän tietokantakyselyitä ja keskittyy ymmärrettävyyteen ja suorituskykyyn.

DM käyttää faktoja (mitattavia) ja dimensioita (konteksti). Faktat ovat usein numeroarvoja, joita voidaan kerätä, ja dimensiot ovat joukkoja hierarkioita ja kuvauksia, jotka määrittävät faktat. Esimerkiksi myyntimäärä on fakta, jonka dimensio-osa ovat muun muassa aikaleima, tuote, kassa, myyjä, myyntiliike ja niin edelleen. Ulotteiset mallit rakennetaan liikeprosessin tarpeiden mukaan.

Ulotteisen mallinnuksen etuja ovat muun muassa seuraavat seikat.

- Ymmärrettävyys. Normalisoituun malliin verrattuna DM on intuitiivisempi ja helpompi ymmärtää. Tieto on ryhmitelty yhtenäisiin bisneskategorioihin tai dimensioihin, jolloin sitä on helpompi tulkita. Yksinkertaistus myös tehostaa ohjelmistojen toimintaa tietokantaa tutkiessa.
- Kyselyjen suorituskyky. DM on optimoitu datakyselyille, kun taas normalisoidut tietokannat yrittävät vähentää tietojen riippuvuutta toisistaan ja ovat optimoituja siirtoja ja päivityksiä ajatellen.
- Laajennettavuus. DM:ää voidaan laajentaa, ja se sisäistää helposti uutta, odottamatontakin tietoa. Olemassa oleviin tauluihin voidaan lisätä uusia rivejä tai nykyisiä rivejä voidaan helposti päivittää SQL-update-komennoilla.

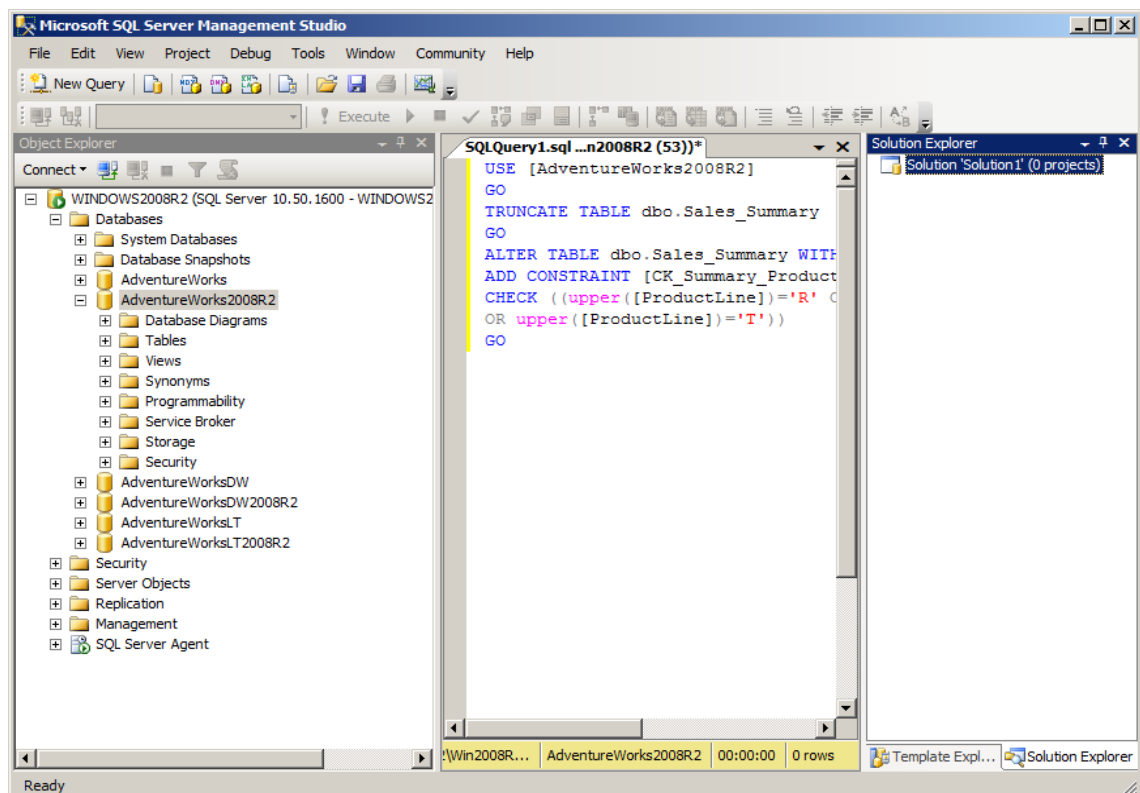
## 4 Microsoft SQL -kehitystyökalut

### 4.1 SQL Server Management Studio (SSMS)

SQL Server Management Studio on integroitu työkalu, jolla yhdistetään, konfiguroidaan, kehitetään ja hallinnoidaan lähes kaikkia Microsoft SQL Serverin komponentteja. SSMS yhdistää useita graafisia työkaluja useiden scripteditorien kanssa. Näin SSMS tarjoaa työkalut kaикentasoisille kehittäjille ja ylläpitäjille aloittelijasta ammattilaiseen.

SSMS yhdistää SQL Serverin aiemmissa erillisinä työkaluina olleet Enterprise Managerin, Query Analyzerin ja Analysis Managerin saman kehitysympäristön alle. Lisäksi SSMS toimii kaikkien SQL Server 2008:n komponenttien, kuten Reporting Servicesin, Analysis Servicesin ja Integration Servicesin, kanssa.

SSMS:ssä on kolme päätason työkaluikkunaa: Object Explorer, Solutions Explorer ja Template Explorer [4]. Nämä ovat nähtävissä oheisessa kuvakaappauksessa (kuvio 2).

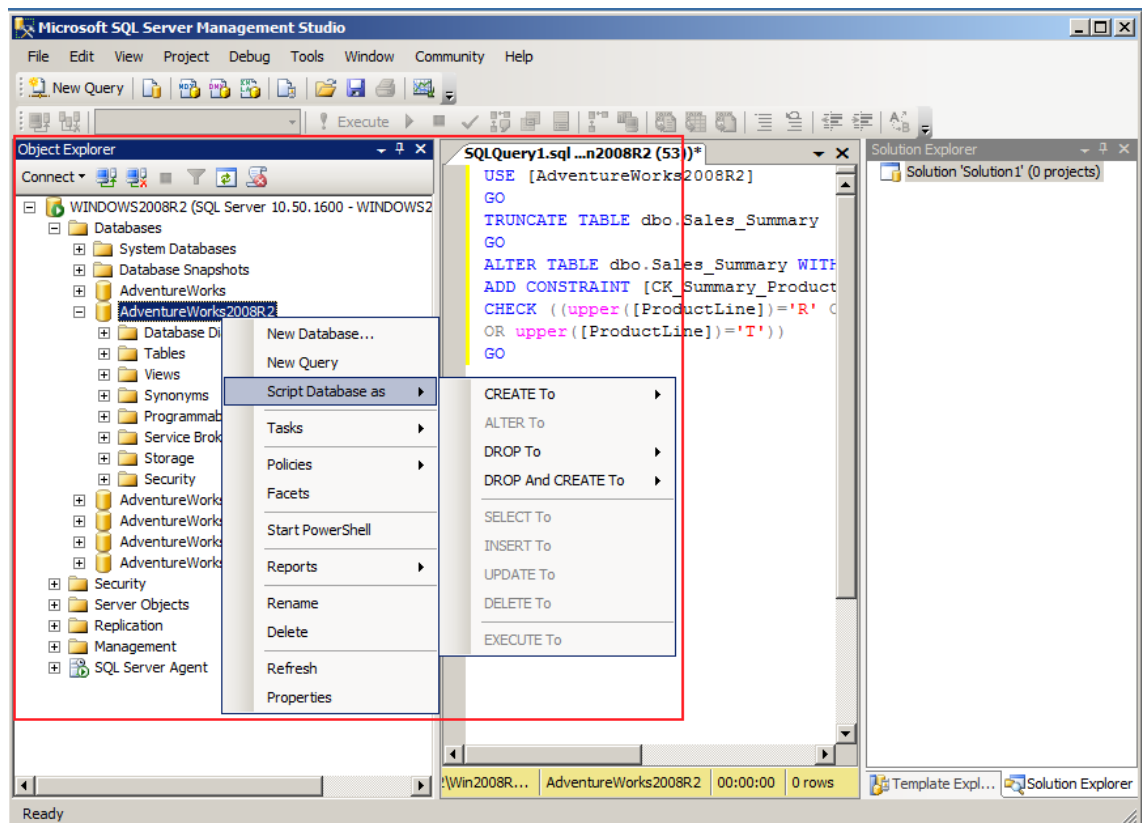


Kuvio 2. SQL Server Management Studio -käyttöliittymä



### 4.1.1 Object Explorer

Object Explorer tarjoaa hierarkkisen käyttöliittymän, jonka avulla voi nähdä ja hallinnoida kaikkia SQL Server 2008:n instansseja. Object Explorerin ominaisuudet vaihtelevat hieman kohteen tyyppin (instanssin) mukaan. Yleisesti ottaen ominaisuudet sisältävät tietokannan kehitysominaisuuksia ja ylläpito-ominaisuuksia kaikille palvelintyypeille. Esimerkkietokannan instanssit on nähtävissä Object Explorer -ikkunassa (kuvio 3).



Kuvio 3. Object Explorer -valintänäkymä.

### 4.1.2 Solutions Explorer

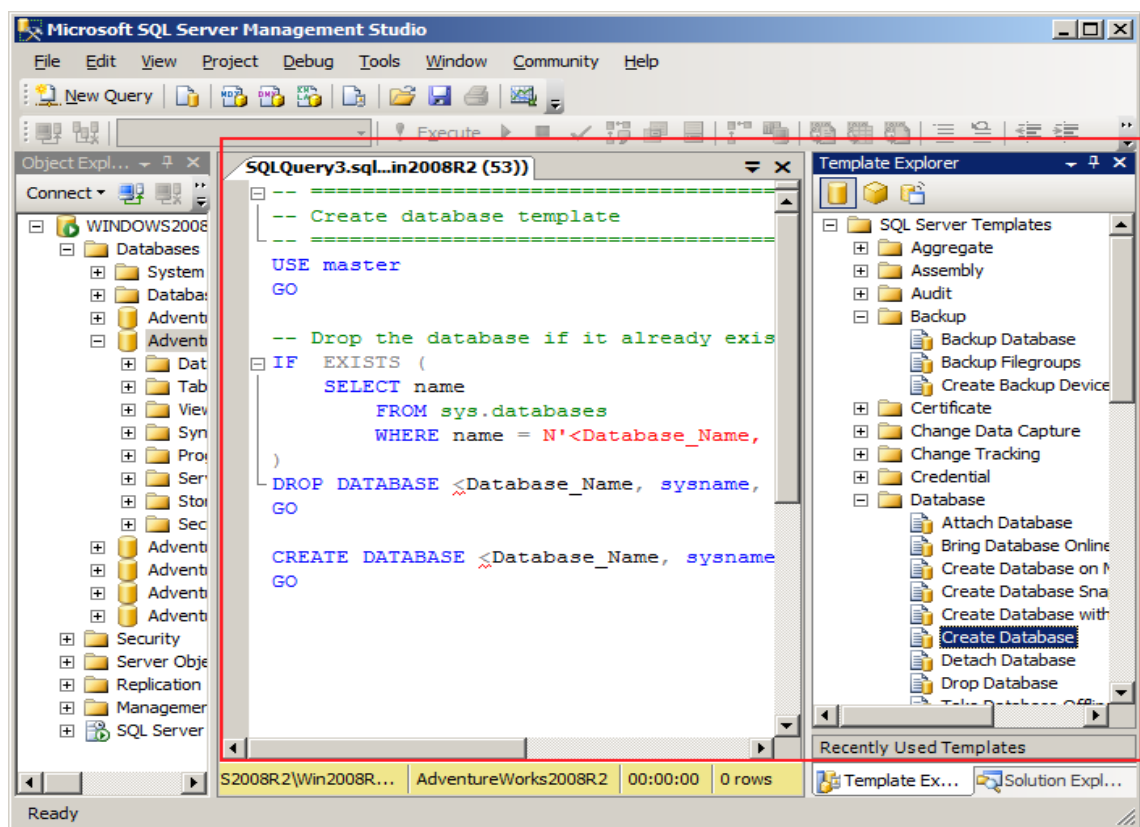
Solutions Explorer -näkymä tarjoaa projekteiksi (project) kutsuttuja kokonaisuuksia, joiden avulla voidaan hallinnoida tietokantaskriptejä, hakuja, datayhteyksiä ja tiedostoja. Yksi tai useampi toisiinsa liittyvä projekti voidaan yhdistää yhdeksi kokonaisuudeksi, ratkaisuksi (solution).

Ratkaisu sisältää yhden tai useamman projektin sisältäen tiedostot ja metadatan, mikä auttaa määrittämään ratkaisun kokonaisuutena. Projekti koostuu tiedostoista ja projektiin liittyvästä metadatasta, kuten yhteysmäärytyksistä. Ratkaisut ja projektit sisältävät kohteita, jotka edustavat skriptejä, hakuja, yhteysmäärytyksiä ja tiedostoja, joita tarvitaan tietokantaratkaisua luotaessa.

Solution Explorer tullaan poistamaan tulevista Microsoft SQL Server -versioista (Microsoft SQL Server 2012 jälkeen), eikä sitä suositella käytettäväksi uusia tietokantaratkaisuja suunniteltaessa.

#### 4.1.3 Template Explorer

Microsoft SQL Server tarjoaa runsaasti erilaisia valmiita malleja (templates). Mallit ovat skriptikokonaisuuksia sisältäviä tiedostoja, joiden avulla voidaan helposti tehdä erilaisia tietokantaoperaatioita. Esimerkkeinä voidaan mainita kannan luonti, muokaus, varmuuskopiointi jne. Kustomoitujen mallien luonti ja tallennus on helppoa.

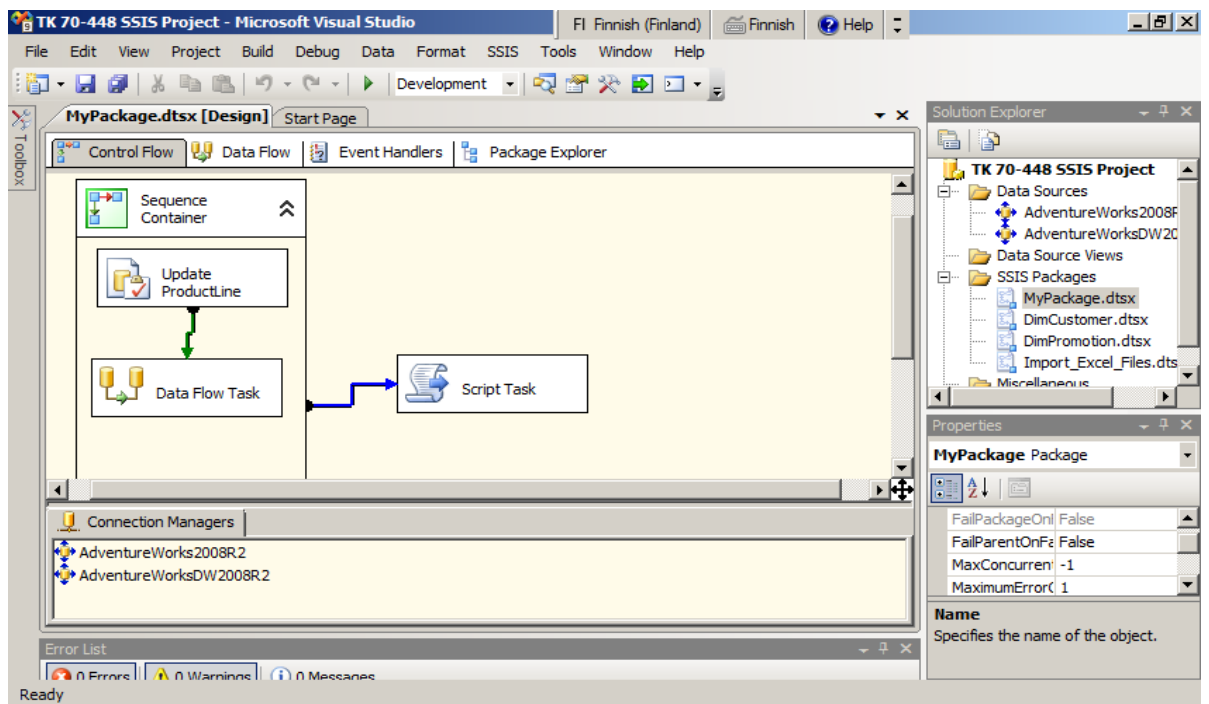


Kuvio 4. Template Explorer ja Create Database -malli.

## 4.2 Business Intelligence Development Studio (BIDS)

Business Intelligence Development Studio on pohjaltaan Microsoft Visual Studio 2008, johon on liitetty Microsoft Business Intelligencelle spesifisiä projektityyppejä. BIDS on pääasiallinen työskentely-ympäristö luotaessa bisnesratkaisuja, jotka sisältävät Analysis Services-, Integration Services- ja Reporting Services -ratkaisuja (näistä lisää luvussa 5). Jokaisen projektityypin mukana on käytettävissä mallit, joiden avulla voidaan luoda objektit BI-ratkaisuja varten. Lisätyökaluina on myös erilaisia kehittämiä, työkaluja ja velhoja (wizard), joiden avulla objekteja luodaan ja käsitellään.

BIDSin näkymä on hyvin samankaltainen kuin SSMS:ssä, toki muokattuna BIDSin tarpeita ajatellen. Esimerkiksi Object Exploreria ei ole mutta SMSS:stä poikkeavuutena löytyy Properties-ikkuna. BIDSin käyttöön tutustutaan varsin tarkasti työn harjoitusten yhteydessä (ks. luku 6).



Kuvio 5. BIDSin näkymä sisältäen designerin, Toolboxin, Solution Explorerin, Properties-ikkunan ja Connection explorerin.

## 5 BIDS-komponentit

### 5.1 SQL Server Integration Services (SSIS)

SSIS:n tärkein objekti on *paketti (package)*. Paketti sisältää logiikan, jonka mukaan haluttu tieto poimitaan, käsitellään, muunnetaan ja siirretään haluttuun kohteeseen. Paketit sisältävät myös työvuoto-elementtejä (workflow elements), joiden avulla tietoa prosessoidaan. Tällaisia työvuoto-elementtejä ovat esimerkiksi tallennettujen kantaproseduurien ajaminen, tiedoston siirto FTP-palvelimelta kohdekansioon tai sähköpostiviestin lähetys prosessivirheen tapahtuessa. Kun paketti suoritetaan, sisäinen logiikka huolehtii suunniteltujen tehtävien ajosta.

Paketti sisältää myös yhteystiedot käytettyihin lähteisiin ja kohteisiin, kuten myös muut ulkoiset yhteydet. Tällaisia ovat muun muassa tietokannat, tiedostot, FTP-palvelimet ja SMTP-palvelimet. Yhteyksiä käytetään sekä SSIS-tiedonkäsittelymoottorissa (data flow engine) että työvuoto-moottorissa (control flow engine).

SSIS-paketin voi luoda kahdella tavalla. Käyttämällä Import And Export Wizardia SSMS:ssä, jolloin paketin voi suorittaa heti tai ajastaa suoritettavaksi myöhemmin. Vaihtoehtoisesti paketin voi luoda valmiin projektin sisään tai luoda uuden projektin, jonka sisään paketti tulee. Tämä tehdään BIDSissä, joka on Microsoft Visual Studio 2008 -lisämalleilla (template) varustettuna.

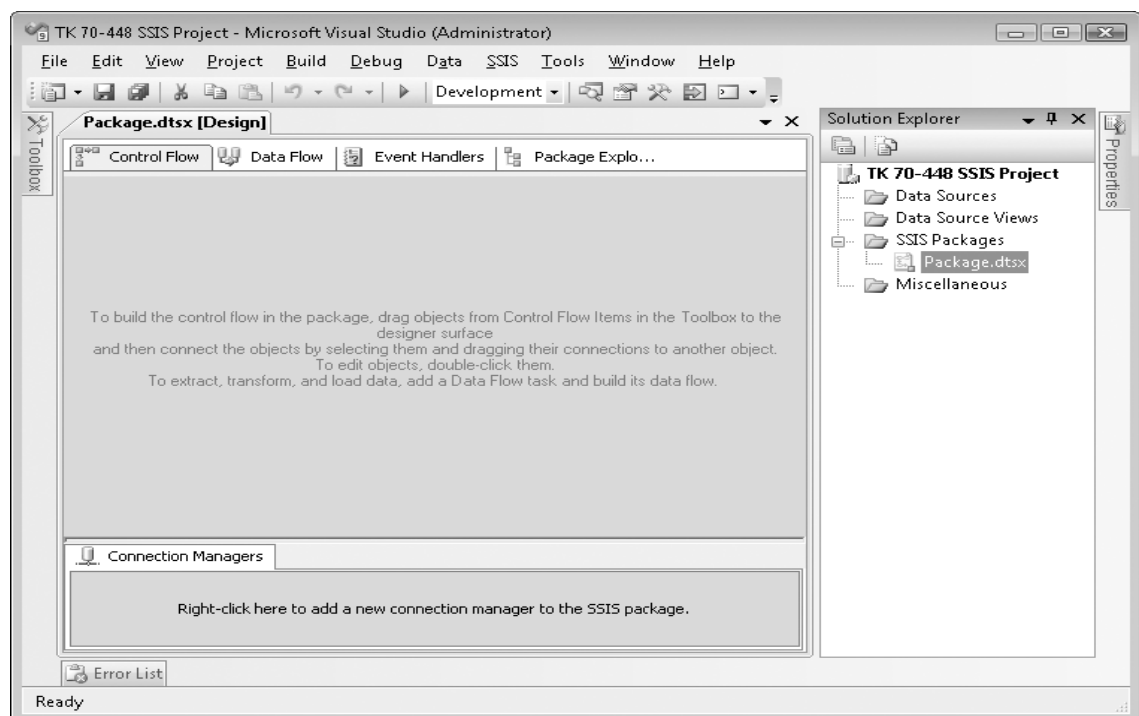
Import and Export Wizard on nimensä mukaisesti velho, joka kysyy ja opastaa määreiden kanssa. Erilaisia kysyttäviä määreitä on esimerkiksi datalähde ja kohde. Datalähde voi olla esimerkiksi tietokanta, Excel-taulukko tai muotoilematon tiedosto. Kohteeksi voi määritellä esimerkiksi tietokannan. Velholla luotu paketti voidaan joko tallentaa ja suorittaa heti tai tallentaa myöhempää käyttöä varten. Talletettua pakettia voidaan muokata myöhemmin käyttämällä BIDSiä. Velholla voidaan luoda nopeasti ja helposti yksinkertaisia paketteja, jotka ovat käyttökelpoisia esimerkiksi kertaluonteisiin ajoihin, mutta samalla velholla luodulla paketilla on eräitä rajoituksia:

- 1) Voidaan määritellä vain yksi lähde ja kohde.
- 2) Kehittyneempiä työvuoto-määrittelyjä ei voida tehdä.
- 3) Velholla tehdyt paketit eivät jaa lähdetietoja muiden pakettien kanssa.

Velhon rajoituksista johtuen sitä ei yleensä voida käyttää kuin yksinkertaisiin tiedonsiirtopaketteihin, jotka ovat vain projektin alkupiste. Useimmiten joudutaan kehittämään paketti, jossa on ominaisuuksia monimutkaisempien vaatimuksien mukaan tai joudutaan kehittämään joukko toisistaan riippuvaisia paketteja. Tällaisissa tapauksissa kehitys tapahtuu BIDSissä, ei SSMS:n velhon avulla.

Kaikki BI:n komponentit (SSIS, SSAS, SSRS) luodaan manuaalisesti samalla lailla. SSIS:ää luodessa BIDSissä avataan uusi projekti ja luonnin yhteydessä määritellään, mitä malleja käytetään. Kun tyhjä projekti on avattu, Solution Explorer näyttää kansiot joita SSIS käyttää. Kansioden alle määritellään lähteet, näkymät, paketit ynnä muut määritteet.

SSIS-paketin tiedostonimen pääte on dtsx. Dtsx on XML-tiedosto, joka sisältää luodun paketin logiikan ja sisällön, joka BIDSissä on pakettiin luotu. Paketin voi siirtää toiseen projektiin, manuaalisesti asentaa toiselle palvelimelle tai tehdä siitä asennuspaketin osa. Paketit eivät siis ole projektispesifisiä, vaan niitä voidaan uusiokäyttää.



Kuvio 6. Tyhjän projektin ensimmäinen paketti luotuna ja näkyvässä Solution Explorerissa.

SSIS:n päätehtävä on siirtää tietoa lähteestä kohteeseen, joten paketin luonnin jälkeen tärkein tehtävä on lähde- ja kohdepoinnterien määrittely. Näitä pointtereita kutsutaan datalähteiksi (data sources) ja datayhteyksiksi (connections). Datalähteet määritellään projektitasolla, ja ne löytyvät loogisena kansiona Solutions Explorerissa (katso kuvio 6 yllä). Datayhteydet määritellään paketin sisällä, ja ne voivat perustua projektin datalähteisiin tai voivat olla yksittäisiä yhteyksiä paketin sisällä.

Datalähde on siis SSIS-projektin objekti. Ne sisältävät yhteyssäikeitä, jotka osoittavat tiedostoihin tai tietokantoihin ja joihin voi viitata yhdestä tai useammasta paketista. Datalähteet eivät ole pakollisia mutta ovat erittäin hyödyllisiä, jos käytetään suurta määrää paketteja, jotka käyttävät samaa tietokantaa tai tiedostoyhteyttä. Useiden pakettien käyttämää datalähdettä on helppo muokata, koska sen voi tehdä missä paketissa tahansa. Muiden pakettien avaaminen synkronoi datalähteen kyseisen paketin kanssa. Kannattaa kuitenkin muistaa, että jos paketti viedään projektin ulkopuolelle, datalähdettä ei enää päivitetä.

Paketin yhteysmanagerin (Package Connection Manager) avulla SSIS:n eri komponentit voivat kommunikoida paketin ulkopuolisen kohteen kanssa. Pakettiyhteyttä (package connection) voidaan käyttää lähdeadapterina ja FTP- tai postipalvelimena. Pakettiyhteys on siis itsenäinen ja erillään projektin datalähteistä.

Tämä on toinen SSIS:n sisäisistä pääkomponenteista Data Flow´n lisäksi. Control flow sisältää control flow taskeja, containereja ja järjestysrajoitteita (precedence constraint) [5], jotka määrittävät, milloin taskit ja containerit suoritetaan. Kolme pääasiallista control flow objectia ovat:

- Control flow tasks - Suorittaa käytännön tason tehtäviä.
- Control flow containers - Sisältää ryhmittelymekanismeja taskeille ja muille containereille.
- Constraints - Yhdistää taskeja ja containereita sekä määrittelee suoritus- ja arvojärjestyksen.

## 5.2 Server Analysis Services (SSAS)

Microsoft SQL Server Analysis Services on Microsoft SQL Serverin komponentti, joka tarjoaa reaaliaikaisen analyttisen prosessoinnin (Online Analytical processing, OLAP) ja datalouhinnan (data mining) palvelut Business Intelligence -ohjelmille. OLAP mahdollistaa loppukäyttäjän käyttävän tiedon analysointiin työkaluja, kuten Microsoft Office Excel, SQL Server 2008 Reporting Service (SSRS), kustomoituja sovelluksia tai kolmannen osapuolen työkaluja. Datan louhinta auttaa käyttäjiä löytämään toistuvia tietokaavoja, joita ei ole helppo havaita, kuten millaisia tuotteita asiakkailta on tapana ostaa samanaikaisesti.

Vaikka onkin mahdollista toteuttaa SSAS-ratkaisu, millaisen tietokantakaavion päälle tahansa, se toimii parhaiten sellaisten kantakaavioiden kanssa, jotka on suunniteltu tukemaan data-analyysiä ja raportointia. Tähän päästään vähentämällä raportointitaulujen määrää ja yksinkertaistamalla kaaviota. Tällaisen kaavion suunnittelun metologiaa kutsutaan ulotteiseksi mallinnukseksi (katso luku 3).

Kun ulotteinen kaavio on valmis, voidaan sen päälle rakentaa yhdenulotteinen malli (Unified Dimensional model, UDM). UDM:n osaa kutsutaan SSAS-kuutioksi (SSAS Cube). Kuutio on looginen varastokohde (logical storage object), joka yhdistää dimensiot ja mittapisteet yhdeksi moniulotteiseksi näkymäksi dataan. Unified viittaa ideaan, että SSAS-kuutio on enemmän kuin perinteinen OLAP-kuutio, koska se yhdistää relaatiomallin ominaisuuksia (atribuutit ja joustavat taululiitokset) ulotteiseen malliin (ulottuvuudet, mittapisteet ja hierarkiat). UDM (tai yksittäinen kuutio) muodostuu useista eri osista, joita ovat seuraavat:

- Data source - Yhteys tietokantaan, jossa käsiteltävä data sijaitsee. SSAS käyttää datalähdettä hakeakseen dataa ja ladatakseen UDM-objekteja, kun niitä prosessoidaan.
- Data source view (DSV) - Tiivistää tietokantakaavion. Vaikka DSV voi vaikuttaa turhalta, se on erittäin hyödyllinen esimerkiksi kaavion laajentamiseen. Voi esimerkiksi lisätä välilaskentasarakkeita tilanteessa, jossa turvallisuusasetukset estävät muutokset tietokaavioon.
- Dimensional model - Kun DSV on valmis, rakennetaan seuraavana kuution ulotteinen malli. Tuloksena saadaan kuution määrittely sisältäen mittapisteet ja

ulottuvuudet (measures and dimensions) attribuutteineen ja/tai monitasoisine hierarkioineen.

- Calculations (valinnainen) - Ulotteinen malli ei yleensä täytä kaikkia tarpeita suoraan. UDM-kehittäjä voi lisätä kuutioon räätälöityä bisneslogiikkaa MDX-lausekkeilla kuten Quarter-To-Date- (QTD) ja Year-To-Date (YTD) - aikaperiodeilla.
- End-user model (valinnainen) - Ulotteisen mallin pääasiallinen suunnittelutarkoitus on tarjota loppukäyttäjälle mahdollisimman havainnollinen raportointi- ja datan selauskokemus. Loppukäyttäjämalli määrittelee lisäominaisuuksia, joita voi rakentaa dimensionaalisen tason päälle. Näihin ominaisuuksiin kuuluvat esimerkiksi suorituskykymittarit (KPI:t), toimenpiteet (actions), perspektiivit ja käännökset (translations).
- Management settings - Projektin administraattori voi määritellä kuutiolle erilaisia käyttövaatimuksia kuten saavutettavuus, latenssi ja turvallisuusmääritykset. Administraattori voi esimerkiksi määritellä kuution käyttöoikeudet, milloin ja miten kuutiota voidaan päivittää, kuution varastomallin ja niin edelleen.

Kuution voi luoda niin sanotusti tyhjästä BIDSin kehittämissä tai käyttää BIDSin Cube Wizardia (velho), joka rakentaa raakamallin valmiiksi muutamien päämääritysten perusteella. Velhon luomaa mallia voi jatkokehittää ja muokata BIDSin Cube designerissä, jossa on 10 kehitysvälilehteä:

- Cube Structure - Käytetään kuution perusrakenteen muokkaukseen.
- Dimension Usage - Määrittelee dimensioiden ja mittaryhmien (measure group) keskinäiset suhteet kuution sisällä.
- Calculations - Sisältää MDX-kaavat ja skriptit, joilla määritellään lasketut jäsenet ja nimetyt setit (named sets).
- KPIs - Määritellään kuution sisäiset suorituskykyilmaisimet (key performance indicator).
- Actions - Käytetään luomaan toimenpiteitä kuten raporttien luontia tai osoitteita, joita loppukäyttäjä voi käyttää selatessaan kuutiota.
- Partitions - Määrittää varastorakenteen jokaiselle mittausryhmälle kuution sisällä.



- Aggregations - Käytetään määrittelemään kustomoituja koontiryhmiä.
- Perspectives - Luodaan kuution sisällön näkymiä helpottamaan loppukäyttäjän kuution selausta ja tulkintaa.
- Translations - Määritellään kielelliset käännökset kuution metadataan kuten mittapisteiden ja dimensioiden nimet lokaalilla kielellä.
- Browser - On nimensä mukaisesti kuution sisällön selain.

### 5.3 SQL Server Reporting Services (SSRS)

SQL Server Reporting Services (raportointipalvelut) on tärkeä osa Microsoft SQL Server Business Intelligence -ratkaisua. SSRS tarjoaa kokonaisvaltaiset työkalut ja palvelut luomaan, ottamaan käyttöön ja hallinnoimaan organisaation raportteja. Se tarjoaa myös ohjelmointiominaisuuksia, joiden avulla raportointiominaisuuksia voidaan laajentaa ja muokata [6].

Reporting Services on palvelinperustainen raportointialusta, joka tarjoaa laajat raportointiominaisuudet useille erilaisille datalähteille. SSRS sisältää työkalut raporttien luontiin, hallinnointiin ja toimittamiseen sekä rajapinnan kehittäjille yhdistää tai lisätä dataa ja raporttiprosessointia kustomoiduissa työkaluissa. Kuten SSIS ja SSAS:kin, SSRS toimii Microsoft Visual Studio -ympäristössä. Se on täysin integroitu SQL Serverin työkalujen ja komponenttien kanssa. Kyseessä on siis sama BIDS-ympäristö kuin kahdessa muussakin pääkomponentissa.

SSRS:n avulla voidaan luoda interaktiivisia, graafisia tai vapaamuotoisia raportteja relaatio-, moniulotteisista tai XML-pohjaisista datalähteistä. Raportteihin voidaan lisätä havainnollistavia kaavioita ja karttoja. Raportteja voidaan julkaista, ajoittaa ja niihin pääsee käsiksi tarvittaessa. Erilaisia näkymiä on runsaasti. Raportteja voi esimerkiksi viedä Exceliin ja tilata muiden julkaisemia raportteja.

BIDS on SSRS-raporttien pääasiallinen kehitystyökalu ja siihen kuuluu kolme SSRS-spesifistä mallityyppiä (template):

- Report Server Project - Visual Studio -malli, joka sisältää SSRS-osiot mukaan lukien käyttöönottotiedot (deployment), datalähteet ja raportit. Tämä luo

automaattisesti tyhjän projektin, jossa on kansiot "Jaetut datalähteet" ja "Raportit". Projektiin voi luoda uusia raportteja raporttivelhon avulla ja niitä voi myöhemmin muokata raporttikehittimen (Report designer) avulla. Jokaisella raportilla on datalähde, joka voi olla sen oma tai usean raportin kesken jaettu lähde.

- Report Server Project Wizard - Hyvin samankaltainen edellisen kanssa mutta tämä käynnistää automaattisesti velhon, kun luodaan uusi projekti (sama velho kuin edellisessä kohdassa). Raporttivelho on automaattinen työkalu, jonka avulla luodaan taulukko- ja matriisimallin (tabular and matrix report). Taulukkoraportti on raportti, jossa määritellään raportin sarakemäärä. Matriisitaulukko käyttää dataa lähteenä sarakemäärän määrittämiseen.
- Report Model Project - Report Model (raporttimalli) on liiketoimintalähtöinen metadatakuvaus datalähteistä ja sen liitoksista.

Raportit muodostetaan dataseteistä, ja ne ovat SSRS:n tärkeitä lähdetietoja. Dasetti (dataset) on kokoelma tietoa ja koostuu, kuten taulukin, riveistä ja sarakkeista. SSRS käyttää yhtä tai useampaa datasettiä raporttien lähteenä. Dasetti on kaksiulotteinen taulu, joten jos lähteenä käytetään moniulotteista lähdetä, kuten OLAP-kuutio, tai louhintamalleja tai hierarkkisia lähteitä, kuten XML, täytyy kysely kirjoittaa siten, että tulos on kaksiulotteinen. Dasetissä on kolme pääelementtiä:

- Data source - Datalähde joka määrittelee ajurin ja yhteysmäärytykset
- Command type - Komentotyyppi joka voi olla tekstiä, tietokantataulun sisältö tai tallennettu proseduuri. Komentotyyppien luomiseen voi käyttää graafisia työkaluja.
- Query string - Tekstikäskyt käyttävät datalähteen käyttämää kyselykieltä. Esimerkkeinä T-SQL, MDX tai DMX. Tallennettuja proseduureja kutsutaan nimen perusteella.

Täydellinen SSRS-asennus sisältää seuraavat komponentit:

- SQL Server catalog - Kaksi tietokantaa SQL Serverillä, ReportServer ja ReportServerTempDB joita käytetään tiedon tallentamiseen ja välimuistina.

- Report Server service - Raporttimoottori (report engine), joka huolehtii raporttien muodostamisesta, datan käsittelystä, turvamäärityksistä, ajastuksista ja toimituksesta.
- Web Service and URL Access - Raporttipalvelimen virtuaaliset hakemistot sallivat web-pohjaisen pääsyn raportteihin.

## 6 Kurssin harjoitukset

Tässä luvussa käydään läpi itseopiskelukurssiin "Exam 70-448: TS: Microsoft SQL Server 2008, Business Intelligence Development and Maintenance" liittyviä harjoituksia. Kurssin harjoituksia ei ole tarkoitus kuvata tässä työssä tarkasti vaan lähinnä käydä läpi, mitä missäkin harjoituksessa käytiin läpi ja mitä oli tarkoitus oppia.

Kurssin päätarkoituksena on opettaa Business Intelligence käyttäjiä kehittämään ja ylläpitämään ratkaisuja, jotka on rakennettu Microsoft SQL Server 2008 BI -alustalle. Komponentteja yllämainitulla alustalla ovat SQL Server Integration Services (SSIS), SQL Server Analysis Services (SSAS) ja SQL Server Reporting Services (SSRS).

### 6.1 Luku 1: SSIS-paketin kehitys

Harjoituksissa käydään läpi ensimmäisen SSIS-paketin luontiin liittyvät vaiheet. Vaikka kurssi olettaa, että BIDS on entuudestaan tuttu, pääsee kurssin ohjeilla helposti alkuun. Luontiin opastetaan kahdella eri tapaa: joko käyttämällä velhoa SMSS:n puolella tai luomaan kokonaan uusi projekti BIDSillä.

Paketit luodaan AdventureWorks2008-tietokantoja silmälläpitäen ja ohjeita onkin syytä noudattaa kirjaimellisesti. Ensimmäistä pakettia luodessa ohjeet ovat tarkkoja ja huolellisesti lukemalla epäonnistumisen mahdollisuus on pieni.

Ensimmäinen opinto-osio opastaa projektin ja pakettien luomisessa ja tallentamisessa. Projektimallina käytetään Integration Services Project -mallia. Harjoituksessa paketteja

luodaan kolme. Tästä saadaan kurssin pohjana toimivan projektin tyhjä malli, jossa tietokantayhteydet ovat olemassa ja jota kehitetään seuraavissa harjoituksissa. Lisäksi valitaan datalähteet (data sources). Projektin datalähteiksi valitaan AdventureWorks2008R2 ja AdventureWorksDW2008R2. Nämä tulevat olemaan projektin datalähteet ja ovat siis Microsoftin kehittämät opetus- ja demonstrointitietokannat, jotka voi vapaasti ladata Microsoftin sivuilta [1].

Toisessa opinto-osiossa tutustutaan Control Flow´hon. Ensin avataan ensimmäisessä opinto-osiossa luotu projekti ja jatketaan sen pakettien kehittämistä. Työkalulaatikosta (Toolbox) tuodaan ensimmäinen tehtävä (task) control flow -välilehdelle, nimetään tehtävä ja määritellään tälle yksikertainen päivityskomento. Sequence container liitetään control flow´hon ja siirretään juuri luotu tehtävä sen sisään. Tämän jälkeen testataan ensimmäisen kerran paketin suoritusta. Lisäksi muokataan ETL-pakettia (DimCustomer) ja luodaan uusi tietokantataulu päivityksiä varten SSMS:ssä.

Kolmannen opinto-osion aiheena on Data flow ja transformaatio. Aiemmin luotuun control flow´hon lisätään data flow task ja tätä muokataan data flow -välilehdellä. Tällöin lisätään OLE DB Source -objekti ja lähde. Kantaan tehdään SQL-kysely, jonka tulos muunnetaan ja ryhmitellään Aggregate transformationin avulla. Tämän jälkeen määritellään OLE DB Destination eli kohde, johon muokattu tieto viedään. Kohdeobjektin sisään luodaan taulu (create table), johon haku tallennetaan. Kun objekti tallennetaan, tietokantaan syntyy uusi taulu.

Lisäksi luodaan Data flow destination eli kohde, joka lataa uusia tietoja ja päivittää olemassa olevia tietoja DW-tietokantaan (DataWarehouse). Tähän tehdään kaksi erillistä taskia eri lähteistä, joiden tulokset liitetään lopuksi Merge Join taskilla yhteen. Tässä vaiheessa tehdään myös attribuuttien mappays, joka liittää arvot toisiinsa column-nimen ja datatyyppin mukaan. Mappays vaatii myös manuaalista yhdistämistä tapauksissa, joissa esimerkiksi column-nimet poikkeavat, eikä BIDS osaa niitä yhdistää.

Opinto-osio 1	Creating SSIS Packages and Data Sources - SSIS-paketin ja datalähteiden luonti
Opinto-osio 2	Creating and Editing a Control Flow Objects - Control Flow Objectien luonti ja muokkaus

Opinto-osio 3	Using Data Flow Adapters and Transformations - Datavuosoovittimien ja muunnosten käyttö
---------------	---

Taulukko 1. Luvun 1 sisältö.

## Luku 2: Virheetutkinta ja korjaus SSIS:ssa

Useat relaatiotietokannat kuten Microsoft SQL Server suorittavat operaatioita hyvin tarkalla tasolla. Tässä sillä tarkoitetaan, että yksittäinenkin käsky onnistuu ja vaikuttaa dataan tai käsky epäonnistuu ja data palautetaan alkuperäiseksi. Tällaista käskyä sanotaan transaktioksi (transaction).

SSIS:ssä transaktiot voivat olla paketin eri tasoilla ja paketin voi suorittaa uudelleen siitä, mihin edellinen suoritus jäi tai tätä edeltävään vaiheeseen. Tätä kutsutaan tarkistuspisteiden (checkpoints) lisäämiseksi. Transaktioiden lisäämiseksi Microsoft Distributed Transaction Coordinator (MSDTC) -palvelun täytyy olla käynnissä.

Ensimmäisessä opinto-osiossa käynnistetään MSDTC-service ja lisäksi transaktio-ominaisuus määritellään arvoksi required. Tämän jälkeen luodaan uusi SQL Task, joka yrittää muuntaa tyyppin teksti tyyppiksi kokonaisluku. Tämä epäonnistuu ja siten transaktio epäonnistuu. Paketti suoritetaan loppuun ja tuloksista näkyy, että päivitykset olisi tehty. Tästä huolimatta kantatarkastus osoittaa, että muutoksia ei tehty. Tämä johtuu required-arvosta. Koska transaktio epäonnistui, koko SQL-kysely peruttiin ja alkuperäiset arvot palautettiin.

Harjoituksessa määritellään FailPackageOnFailure-määre taskeille arvoksi True ja pakotetaan taskin tulos epäonnistuneeksi. Tämä tehdään, jos halutaan demonstroida virheenkäsittelyä tai checkpointin ominaisuuksia. Checkpoint määritellään erilliseen xml-tiedostoon, johon tallennetaan epäonnistuneen paketin suorituksen tila.

Opinto-osio 2 käsittelee paketin tilaa, logituksen asettamista ja task-virheiden käsittelyä. Harjoituksessa luodaan precedence constraint, jolla määritellään, milloin control flow -taskit suoritetaan. Suorituksen aikana seurataan constraintin määrittelyjen seurauksia. Taskina suoritetaan popup-ikkuna, joka kertoo taskin suorittuneen loppuun. Määrityksistä riippuen popup-ikkuna näkyy tai ei näy. Molemmat vaihtoehdot

käydään läpi. Harjoituksissa määritellään login kirjoitus tietokantatauluun ja tarkastetaan login sisältö SSMS:stä tietokantakyselyllä.

Opinto-osio kolmessa tutkitaan data flow -virheiden käsittelyä ja debuggausta eli virheenetsintää. Polut (Paths) ovat kuten precedence constraintsit, mutta taskien suorittamisen sijaan käsittelevät tietorivejä. Polkuja on kahta lajia: datapolut ovat vihreitä ja virhepolut ovat väriltään punaisia.

Tehtävässä luodaan error flow -polku ja tämän jälkeen tunnistetaan virhe tarkastelemalla ulostulorivejä data viewerin avulla. Tehtävässä tutustutaan muuttujan käyttöön BIDSissä. Muuttujan voi määrittellä task-, container- tai projektitasolla. Tässä tehtävässä muuttujan avulla seurataan taskin virhemäärää.

Taulukko 2. Luvun 2 sisältö.

Opinto-osio 1	Configuring Package Transactions and Checkpoints - Pakettitransaktioiden ja tarkistuspisteiden konfigurointi
Opinto-osio 2	Identifying Package Status, Enabling Logging, and Handling Task Errors - Paketin tilan tarkastus, logien määrittely ja virhetehtävien käsittely
Opinto-osio 3	Handling Data Flow Errors and Debugging - Datapolkujen käsittely ja tutkinta

## 6.2 Luku 3: SSIS-pakettien konfigurointi ja asentaminen

Opinto-osio 1 käsittelee pakettikonfiguraatioita ja expressioneita. Pakettikonfiguraatiot (package configurations) ovat oletuksena paketissa pois päältä. Jos pakettikonfiguraatiota käytetään, konfiguraatiotyyppin voi valita seuraavista: XML-konfiguraatiotiedosto, ympäristömuuttuja (käyttöjärjestelmän), rekisterimerkintä, isäntäpaketin muuttuja tai SQL Serverin taulu. SQL Serverin taulu voi sijaita toisessa tietokannassa, muut tyypit pitää olla lokaaleja, eikä esimerkiksi XML-tiedostoa voida lukea verkon yli toiselta koneelta. Pakettikonfiguraatiot luetaan, kun paketti käynnistetään.

Konfiguraation ominaisuudeksi (configuration property) voidaan määrittää muuttuja, yhteysmanageri, dynaaminen logitus, pakettitason ominaisuus tai suoritettava taski tai container. Harjoituksissa luodaan ensin XML-konfiguraatio. Tämä lisätään myös toiseen luotuun pakettiin.

Luvussa myös määritellään SQL-konfigurointityyppi (Configuration type) ja määritellään se kolmannen luodun paketin konfiguraatioksi. Luonnin yhteydessä kantaan muodostuu taulu nimeltä [dbo].[SSIS Configurations] ja tämä tarkastetaan SSMS:stä. Lisäksi tarkastellaan expressionin käyttöä yhteyden päivittämiseksi. Expressionia ei suoriteta, kun paketti käynnistetään vaan se suoritetaan määriteltyyn aikaan ja se voi muuttaa esimerkiksi paketin yhteyden arvoa kesken suorituksen.

Opinto-osio 2 käsittelee paketin käyttöönottoa. Kun paketin konfiguraatiot ja expressionit on määritelty, paketti voidaan viedä uuteen ympäristöön. Paketti voidaan ottaa käyttöön tiedostojärjestelmässä (file system) tai SQL Serverillä. Paketti muodostuu XML-tiedostoista. Ne voidaan manuaalisesti kopioida uuteen sijaintiin ja käyttää source-control-toolia, joka tekee käyttöönoton (deployment).

Jos paketti on SQL Serverillä, käyttöönotto voidaan suorittaa DTUtil-komentorivityökalua käyttäen tai tuoda paketit manuaalisesti SSMS:n import-toolia käyttäen. Myös SSIS:n sisäänrakennettua Deployment Utilitya voidaan käyttää. Sen käytön puolesta puhuu tieto, että kokonainen projekti voidaan tuoda samaan aikaan joko tiedostojärjestelmään tai SQL Serverille. Utility luo tiedostojoukon, josta asennus voidaan myöhemmin helposti tehdä.

Harjoituksissa käydään läpi DTUtil-työkalun käyttö eri optioilla sekä Deployment Utilityn käyttö esimerkein.

Taulukko 3. Luvun 3 sisältö.

Opinto-osio 1	Using Package Configurations and Expressions - Pakettikonfiguraatioiden ja lausekkeiden käyttö
Opinto-osio 2	Deploying SSIS Packages - SSIS-paketin asennus

### 6.3 Luku 4: SSIS-pakettien hallinnointi, suojaus ja suorittaminen

Viimeisessä SSIS-komponenttia käsittelevässä luvussa tutustutaan valmiiden pakettien käyttöhallinnointiin, suojaukseen ja suoritukseen. SSIS-service on Windows service nimeltä "SQL Server Integration Services 10.0", ja se ajaa tiedostoa MsDtsSrve.exe. Tämä toimii kuten muutkin Windowsin servicet, kunhan huolehtii, että käynnistysparametrit on oikein määritelty SSMS:n puolelta.

Kun palvelu on käynnissä, kaikki konfigurointi tapahtuu SMSS:n puolelta. Palvelu näkyy Object explorer -ikkunassa ja sen kautta voidaan hallinnoida rajoitetusti, määritellä turvallisuusasetuksia ja suorittaa paketteja, jotka on asennettu palvelimelle. Object explorerissa on kahden tason kansioita: käynnissä olevat paketit (Running Packages) ja tallennetut paketit (Stored Packages).

SSIS:n suojaaminen on tärkeitä kahdesta syystä: yhteystiedot ja skeematiedot. Vaikka paketin määrykset eivät sisällä dataa lähteestä tai kohteesta, se sisältää molempien yhteystiedot. Nämä on syytä suojata. Paketti sisältää dataa, joten suojaamatonta siirtodataa voi lukea, vaikkei lähde ja kohde ole luettavissa.

Suojaamiseen on kolme tapaa. Ensiksi on SQL Serverin turvallisuusroolit (security roles), jolloin pääsy palvelimelle ja pakettien suoritus on rajattu. Toiseksi paketti voidaan salata BIDSissä kehityksen yhteydessä kokonaan tai osittain, jolloin sensitiivinen data on suojattu. Kolmanneksi voidaan käyttää käyttöjärjestelmätason tiedostojen salausta, jolloin pakettien lähdetiedostot ja -kansiot ovat salattuina tiedostojärjestelmässä.

Harjoituksissa käydään läpi pakettitason suojaus salasanalla (EncryptAllWithPassword) ja katsotaan, miten suojaus vaikuttaa paketin suoritukseen. Lisäksi tuodaan paketti SSIS-serviceen ja sitten määritellään turvallisuusroolit paketille. Tämän lisäksi käytetään salanasuojausta.

Opinto-osio 2:ssa käsitellään pakettien suoritusta ja ajastusta kehitysympäristön (tässä BIDS) ulkopuolella. Tähän on kaksi tapaa: ohjelmallisesti (programmatic execution) käyttäen SSIS:n objektimallia tai käyttämällä komentorivityökaluja. Objektimallia käytettäessä paketit voidaan ladata ja suorittaa sovelluksen sisällä.



Yleisempi käyttötapana on komentorivityökalu. SSIS:n mukana tulevat työkalut nimeltään DTEXEC sekä DTEXECUI, jotka auttavat komentorivikäskyn muotoilussa. Lisäksi SQL Server 2008:n SQL Server Agent tukee suoraan komentorivisuoritusta.

Harjoituksissa käydään läpi komentorivityökalun käyttö ja ladataan SSIS-paketti DTEXECUI:lla SQL Serverille. Tämän jälkeen paketti suoritetaan komentoriviltä. Lisäksi paketin suoritus ajastetaan SQL Server Agentin avulla.

Taulukko 4. Luvun 4 sisältö.

Opinto-osio 1	Managing the SSIS Service and Configuring Package Security - SSIS-palvelun hallinta ja pakettisuojausten konfigurointi
Opinto-osio 2	Executing and Scheduling Packages - Pakettien suorittaminen ja ajastaminen

#### 6.4 Luku 5: SSAS-kuution kehitys

Opinto-osio 1:ssä käsitellään datalähteen ja datalähdenäkymän (Data source view, DSV) luontia. DSV:tä määriteltäessä huomataan usein, että lähteenä käytetyn tietokannan suunnittelu ei täytä SSAS-ratkaisun vaatimuksia. Esimerkiksi jokin tietty taulu saattaa sisältää sarakkeita, jotka täytyy yhdistää (etunimi ja sukunimi yhdeksi sarakkeeksi "nimi") tai tarvittavat dimensiot löytyvät useammasta kuin yhdestä tietokantataulusta. Tällaisissa tapauksissa voidaan DSV:tä laajentaa käyttämällä nimettyjä laskelmia (calculations) ja kyselyitä (queries).

Nimetty laskelma on sarake, joka perustuu lausekkeeseen. Lausekkeen syntaksi on lähdespesifinen ja sen täytyy noudattaa sääntöjä, jotka koskevat kohdetietokantaa. Esimerkiksi sana-tyyppisiä kenttiä ei voi laskea yhteen ja niin edelleen. Nimetty laskelma on kuitenkin osa DSV:tä, ei relaatiotietokantaa. Myös taulun primaariavaimet ja taulujen sisäiset riippuvuudet pitää määritellä lähdekannan mukaisiksi.

Harjoituksessa BIDSissä luodaan SSAS-projekti samaan tapaan kuin SSIS-projekti kohdassa 6.1. Nyt mallina käytetään SSAS-projektin templaatteja. Lisäksi määritellään datalähde ja DSV. DSV:hen määritellään nimetty laskelma (named calculation).

Kun DSV on valmis, on syytä käsitellä yhdenmätyn ulotteisen mallin (Unified Dimensional model, UDM) luonti ja muokkaus. Tämä käydään läpi opinto-osio kahdessa. UDM:ään suunnitellaan dimensiot, kuutiot, mittaryhmät (measure groups) ja mittapisteet. Nämä voi luoda manuaalisesti alusta lähtien, mutta BIDS tarjoaa kätevän velhon, jolla voi luoda UDM:n. Kun UDM on valmis, mahdollisen jatkokehityksen voi hoitaa BIDSin kehittimissä.

Opinto-osio 2 käsittelee loogisesti seuraavaa osaa SSAS:n käytössä eli kuution luontia ja muokkausta. Kun DSV on valmis, voidaan ulotteisen mallin raakamalli luoda automaattisesti käyttäen BIDS Cube -velhoa. Kun kuutio on valmis, se voidaan asentaa ja testata. Vaikka velho onnistuu luomaan melko hyvän raakamallin DSV:n tarjoaman metadatan pohjalta, tarvitsee mallia usein muokata käyttäjien tarpeiden mukaan. Tämä onnistuu BIDSin kehittimissä. Näistä lisää myös harjoitusten luvuissa 6, 7 ja 8.

Harjoituksissa käydään läpi kuution luonti velhon avulla. Velho tunnistaa jo useita asetuksia kuten mittausryhmät ja dimensiot, mutta sitä pitää myös opastaa pitkin luontia, esimerkiksi poistaa mittauskelvoton sarake mitattavien joukosta. Velhon käytön jälkeen käydään läpi luodun kuution muokkaaminen Cube Designerillä.

Opinto-osio 3 käsittelee dimensioiden luomista ja käsittelyä. Luonnin voi suorittaa jälleen joko BIDSissä alusta loppuun tai käyttämällä velhoa. Harjoituksissa käydään läpi dimension luonti velholla ja sen muokkaus myöhemmin BIDS Dimension Designerillä.

Luvun viimeinen opinto-osio käsittelee mittaryhmien ja mittapisteiden luontia. Kuutionluontivelho auttaa luomaan kuution hyvin pitkälti valmiiksi. Kannattaa kuitenkin muistaa, että kun liiketoimintavaatimukset kasvavat, on todennäköistä, että joudutaan luomaan uusia mittaryhmiä ja pisteitä. Uusi mittaryhmä joudutaan luomaan esimerkiksi, jos uusi faktataulu luodaan lähdedataan ja sen tiedot pitää tuoda kuutioon. Kun mittaryhmä on luotu, täytyy tälle määritellä myös mittapisteet ja näiden määritykset.

Harjoituksissa käydään mittaryhmän ja mittapisteen luominen ja konfiguroiminen sisäisten avainten ja riippuvuuksien kera.

Taulukko 5. Luvun 5 sisältö.

Opinto-osio 1	Creating Data Sources and Data Source Views - Datalähteiden ja datalähdenäkymien luonti
Opinto-osio 2	Creating and Modifying SSAS Cubes - SSAS-kuutioiden luonti ja muokkaus
Opinto-osio 3	Creating and Modifying Dimensions - Ulottuvuuksien luonti ja muokkaus
Opinto-osio 4	Creating Measure Groups and Measures - Mittapisteiden ja -ryhmien luonti

## 6.5 Luku 6: SSAS-kuutioiden laajentaminen

SSAS-kuutioita voidaan laajentaa hyvin joustavasti vastaamaan liiketoimintatarpeita. Attribuuttiriippuvuuksia ja hierarkioita voidaan määrittellä lisää kuution optimoimiseksi ja datan analysoinnin tehostamiseksi. Lisäksi voidaan parantaa käyttäjien raportointikokemuksia rakentamalla loppukäyttäjätaso (end-user layer), joka sisältää suorituskykyilmaisimia, toimenpiteitä, käännöksiä ja perspektiivejä.

Opinto-osio 1 käsittelee käyttäjähierarkioiden ja dimensioiden välisiä suhteita. Relaatiotietokannoista poiketen UDM ei käsittele tietoa kaksikulotteisesti riveissä ja sarakkeissa. UDM käyttää useita ulottuvuuksia ja tukee datahierarkioita. Attribuuttien väliset suhteet on syytä määrittää tarkasti. Näiden suhteiden etuna on tehokas datan tallentaminen sekä datan nopea nouto ja kerääminen. Harjoitukset koskevat yllä olevia aiheita, eli luodaan ja muokataan dimensiohierarkioita, määritellään attribuuttirelaatioita ja luodaan suhteita dimensioiden ja mittaryhmien kanssa.

Opinto-osio 2 on laaja ja käsittelee useita SSAS:n osia. Monet yritykset käyttävät suorituskykyilmaisimia (KPI) mittaamaan tehokkuuttaan. KPI:t ovat määreellisiä mittareita, jotka edustavat kriittisiä onnistumistekijöitä. Analytikot käyttävät niitä mittaamaan yrityksen suorituskykyä aiemmin määriteltyä tavoitetta vasten pidemmällä

aikavälillä. Esimerkkeinä KPI-kohteista mainittakoon liikevoitto, liikevaihdon kasvu ja asiakaskannan kasvu.

Toimenpiteet (Actions) voivat laajentaa kuutiota vaihtelevin tavoin. Käyttäjä voi olla dimension alimmalla tasolla ja haluaa nähdä yksittäisen myyntitilauksen tuotteelle. Jos tilaustieto ei ole tallennettu kuutioon, voidaan luoda raportointitoimenpide, joka sallii käyttäjän pyytää SSRS-raportti. Tämä näyttää tilaustiedon toisesta järjestelmästä.

Lokalisaatio onnistuu kuution sisäisten käännösten (translations) avulla. Käännöksen avulla data saadaan käännettyä kansainvälisten käyttäjien helposti omaksumaan muotoon. Käännösten avulla metadata (mittapisteet, dimensiot, perspektiivit ja KPI:t) ja itse data saadaan käännettyä loppukäyttäjälle.

SSAS-kuutio voi kasvaa suureksikin ja levitä useisiin mittaryhmiin, jolloin suuri ulotteinen malli voi olla loppukäyttäjälle ylivoimaisen vaikea käyttää. Tässä tilanteessa avuksi voidaan ottaa perspektiivit. Kuution perspektiivi on kuution osan näkymä. Sen päätarkoitus on vähentää suuren kuution kompleksisuutta ja näyttää vain tietty osa kuution objekteista. Perspektiivejä ei voida käyttää turvallisuusmekanismeina vaan objektien turvallisuusmääritykset ovat voimassa perspektiiveissäkin. Harjoituksessa luodaan KPI, toimenpide, käännös espanjaksi ja perspektiivi AdventureWorks-työntekijälle töidensä suorittamiseksi.

Opinto-osio 3 käy läpi MDX-laskelmia ja -kyselyitä. SSAS:n avulla voidaan tehdä nopeasti ulotteisen malli, joka tarjoaa olennaiset OLAP- ja datalouhintaominaisuudet. Yleensä kuitenkin liiketoiminnan tarpeet vaativat parempia ominaisuuksia kuin mitä ulotteinen malli voi tarjota. Moniulotteiset lausekkeet (Multidimensional Expressions, MDX) tarjoavat ohjelmointimahdollisuuden implementoida kustomoituja laskelmia. SSAS:ssa MDX:ää käytetään kahdella tapaa: kyselyihin ja moniulotteisten kuutioiden laajentamiseen. Harjoituksissa luodaan MDX-kyselyjä, laskettuja jäseniä ja nimettyjä settejä.

Taulukko 6. Luvun 6 sisältö.

Opinto-osio 1	Defining User Hierarchies and Dimension Relationships - Käyttäjähierarkioiden ja ulottuvuussuhteiden määrittely
---------------	--

Opinto-osio 2	Creating KPIs, Actions, Translations, and Perspectives - Suorituskykyilmaisimien, tehtävien, käännösten ja perspektiivien luonti
Opinto-osio 3	Creating Calculations and Queries by Using MDX - Laskelmien ja kyselyiden teko moniulotteisella kyselykielellä

## 6.6 Luku 7: SSAS-tietovaraston hallinta, prosessointi ja asennus

SSAS:n tallennusarkkitehtuuri perustuu järkevään ja joustavaan suunnittelumalliin. Se antaa administraattoreille mahdollisuuden valita tallennusmallin, joka tarjoaa parhaan tasapainon kuution suorituskyvyn ja tallennuskoon välillä. Suunnitelma on näkymätön käyttäjäohjelmistoille. Administraattori voi vaihtaa mallia milloin vain kuution asennuksen jälkeen.

SSAS tallentaa mittaryhmädatan fyysiseen yksikköön, jota kutsutaan partitioksi (partition). Partition määrittelee palan faktataulua, joka on ladattu mittaryhmään (measure group). Partitiointiin on kaksi pääasiallista syytä; parempi suorituskyky ja helpompi hallittavuus. Jos kaikki data on yhdessä partitiossa, hakutoiminnot ovat raskaita ja käytettävyys kärsii. Jos data on jaettu esimerkiksi partitioihin vuoden mukaan, suorituskyky nousee huomattavasti. Partitiolla voi olla erilaiset tallennus- ja kokoamissäännöt. On esimerkiksi mahdollista tuhota partitio ilman, että kuutiota tarvitsee prosessoida uudelleen tai käsitellä uuden datan partitiota useammin kuin arkistointipartitiota. Lisäksi partitiointi lyhentää prosessointiaikaa, koska palvelin voi käsitellä partitioita moniajona yhtä aikaa.

Välillä tuntuu, ettei ole väliä, miten data on paloiteltu. Kuutio tuntuu sisältävän kaiken mahdollisen kootun (aggregated) datan. Todellisuudessa arvot eivät ehkä ole tallennettu kuutioon. Palvelin voi johtaa kootut arvot kolmella tapaa: sisäisestä välimuistista, olettaen, että dataa on pyydetty aiemmin käyttämällä esilaskettuja datakoonteja, jotka on tallennettu kuutioon, tai jos ei muuta mahdollisuutta ole, hakea data partitiolta ja koota se reaaliaikaisesti. Tästä johtuen etukäteen laaditut hyödylliset koonnit nopeuttavat kyselyitä todella paljon. Harjoituksissa luodaan partitio vuoden mukaan, testataan erilaisia tallennusmoodeja ja niiden vaikutusta kyselyihin, luodaan proaktiivinen välimuisti ja harjoitellaan velhon käyttöä.

Kun UDM on valmis, se täytyy asentaa (deploy) testiserverille laadunvarmistusta varten (Quality Assurance, QA) tai tuotantopalvelimelle käyttäjien saataville. Oletettavasti asennus täytyy tehdä tietyin aikaväleihin, sillä korjatut virheet ja parannukset on saatava synkronoitua testiin tai tuotantoon.

SSAS tukee useita eri asennusvaihtoehtoja: BIDSin kautta paikalliseen testiin, asennusvelhon kautta, ajastamalla XMLA-skriptin avulla, tietokantasynkronointi, jolla saadaan synkronoitua kuutiot, varmuuskopio toiselta serveriltä ja palautus halutulle serverille ja ohjelmallisesti (ei kuulu kurssin sisältöön). Harjoituksissa käydään läpi asennusvelhon ja synkronointivelhon käyttö.

Objektit, joita SSAS:ssa voidaan prosessoida, ovat tietokanta, kuutio, mittaryhmä, partitiio, dimensio, louhintarakenne (mining structure) ja louhintamalli (mining model). Vain dimensiot, partitiot ja louhintarakenteet tallentavat dataa. Kun objektia prosessoidaan, palvelin luo prosessointisuunnitelman. Esimerkiksi dimension attribuutteja prosessoidessaan palvelin analysoi riippuvuudet ja prosessoi ensin attribuutit, joilla ei ole riippuvuuksia muihin attribuutteihin. SSAS-palvelin tukee useita prosessointivaihtoehtoja palvellakseen erilaisia vaatimuksia. Prosesseja voidaan muokata sekä BIDS:ssä että SSMS:ssä. Harjoituksissa prosessoidaan objekteja sekä BIDS:ssä että SSMS:ssä ja työskennellään proaktiivisen välimuistin kanssa.

Taulukko 7. Luvun 7 sisältö.

Opinto-osio 1	Defining Measure Group Partitions and Aggregations - Mittaryhmäosioiden ja kokoomien määrittely
Opinto-osio 2	Deploying SSAS Objects - SSAS-objektien asennus
Opinto-osio 3	Processing SSAS Objects - SSAS-objektien prosessointi

## 6.7 Luku 8: SSAS:n hallinta ja turvaaminen

Yritystason laitteisto- ja ohjelmistoalustan, kuten SSAS, täytyy olla luotettava ja turvallinen. Luotettava järjestelmä turvaa tietopääomaa rajaamalla pääsyä

arkaluontoiseen materiaaliin. Järjestelmän pitää myös olla helposti hallittavissa ja monitoroitavissa.

Administraattorina suoritetaan päivittäisiä tehtäviä, kuten SSAS-palvelininstanssien hallinnointi, joihin kuuluu asennus, varmuuskopiointi, palautus ja tietokantojen synkronointi, automatisoitujen toistuvien tehtävien, kuten tietojärjestelmän hallinta, UDM:ien turvaaminen ja palvelinten suorituskyvyn tarkkailu. Muun muassa näihin aiheisiin tutustutaan tässä luvussa.

SSAS:n turvallisuusarkkitehtuuri pohjaa Windowsin turvallisuusmäärittämiin. Käyttäjät autentikoidaan Windows-tilien kautta ja oikeuksia jaetaan roolitusten mukaan. SSAS:ssa Windowsin käyttäjät ja ryhmät voidaan ryhmitellä rooleiksi. Määritellyt turvamääritykset rajoittavat kuution käyttäjänäkymiä. Harjoituksissa luodaan uusi hallintarooli, jonka jäsenet voivat muokata hakuja mutta eivät pääse käsiksi salaisiin tietoihin, supistetaan solutason oikeuksia tiettyihin soluihin ja lisätään prosessointioikeuksia tiettyihin soluihin.

Opinto-osio 2:ssa käsitellään disaster recoverya (katastrofipalautuminen) klusteroinnin, varmuuskopioiden ja objekti-skriptien kautta. Windows serverin klusterointi tarjoaa skaalautuvan ja erittäin hyvin saavutettavan (high availability) SSAS-ratkaisun, mutta tämä vaatii luonnollisesti useamman palvelimen. Tietokantojen varmuuskopioiden ottaminen säännöllisin väliajoin ja ennen suuria muutoksia kuuluu kaiken laiteylläpidon peruserätyksiin. Kuution määrittämisen tallentaminen skripteihin on järkevää varsinkin ennen kuution jatkokehitystyötä, jolloin palaaminen on helppoa, eikä tietokannan sisällöllä ole merkitystä. Harjoituksissa käydään läpi tietokannan varmuuskopion ottaminen ja ajastetun varmuuskopiointiskriptin teko SQL Server agentin avulla.

SSAS tukee useita palvelintason ominaisuuksia, jotka ohjaavat SSAS-asennuksen käyttäytymistä. Tällaisia ovat esimerkiksi datatiedostojen sijainti ja muistinkäyttö. Oletusasetukset ovat yleensä riittävät, mutta tarvittaessa esimerkiksi logiasetuksia voi muokata SSMS:n kautta. Harjoituksissa käydään läpi kyselystatistiikan keräys ja optimointivelhon käyttö.

Kun tietokanta on asennettu ja käytössä, jäljelle jää monitorointi ja virheenjäljitys. Parhaissakin järjestelmissä löytyy poikkeavaa käytöstä virheiden ja alentuneen suorituskyvyn muodossa. SQL Server Profiler ja Windows Performance -konsoli auttavat tarkkailemaan palvelimen tilaa ja huolehtimaan mahdollisista virhetilanteista. Harjoituksissa käytetään Profilerin jäljitysominaisuuksia tarkkailemaan palvelimen toimintoja ja monitoroidaan kyselyn löytymistä koontitaulusta.

Taulukko 8. Luvun 8 sisältö.

Opinto-osio 1	Setting Up SSAS Server Security - SSAS-palvelinsuojausten määrittely
Opinto-osio 2	Managing SSAS High Availability, Backups, and Object Scripting - SSAS vikasietoisuuden, varmuuskopioiden ja objektiskriptauksen hallinta
Opinto-osio 3	Managing SSAS Tuning and Logging - SSAS hienosäädön ja logituksen hallinta
Opinto-osio 4	Tracing and Monitoring SSAS Instances - SSAS-instanssien seuraaminen ja monitorointi

## 6.8 Luku 9: SSAS-datalouhintatyöskentely

BI:n yksinkertaisin työkalu, jota organisaatio voi käyttää, on varmaankin SQL Server Reporting Services (SSRS). Loppukäyttäjällä on kuitenkin rajalliset kyvyt katsellessaan raporttia. Raporttimallien käyttö on edistyneemmän käyttäjän taito, eikä sitä voida edellyttää kaikilta käyttäjiltä. Raportin analysoinnin suorituskyky on rajallinen. Sekä suorituskyvyn että taitojen puutteen voi ohittaa käyttämällä SSAS:n datavarastoja ja OLAP-järjestelmiä.

Käyttäjä saa tiedon analyysia varten nopeasti ja voi vaihtaa näkymiä reaaliaikaisesti porautuakseen syvemmälle dataan. Käyttäjillä on kuitenkin rajoitteita näitä järjestelmiä käyttäessään. Yleensä käyttäjällä ei ole aikaa tutkia kaikkia graafeja ja taulukoiden näkymiä käyttäen kaikkia arvoja ja hierarkioita. Lisäksi käyttäjät osaavat etsiä vain kaavoja, joita he itse odottavat ja analyysi perustuu usein matemaattisiin operaatioihin, kuten summien vertailuun eri ryhmien välillä. SSAS-datalouhinta (data mining) keskittyy lähinnä näihin rajoitteisiin.



Datalouhinnassa ei määritellä etsittäviä kaavoja etukäteen. Ensin haetaan dataa edistyneiden matemaattisten algoritmien avulla ja tämän jälkeen tutkitaan toistuvia kaavoja ja sääntöjä, joita algoritmit ovat löytäneet. Koska louhinta-algoritmit toimivat automaattisesti, ne voivat tutkia miljoonia erilaisia sarakkeyhdistelmiä hyvin lyhyessä ajassa. Datalouhinnan lisätehtävä on tehdä ennusteita löydettyjen toistuvien kaavojen pohjalta.

Datalouhintatekniikat jaetaan kahteen pääluokkaan: kohdistettuun lähestymiseen (directed approach) ja kohdistamattomaan lähestymiseen (undirected approach). Kohdistettu lähestyminen käyttää tunnettuja esimerkkejä ja lisää kerättyjä tietoja tuntemattomiin esimerkkeihin ennakoitakseen valitun kohdemuuttujan. Kohdistamatonta lähestymistä käytettäessä yritetään löytää uusia kaavoja koko datasetin sisältä. Kohdistettua lähestymistä käytetään esimerkiksi etsimään syytä tuotteen ostolle ja kohdistamatonta käytetään, kun halutaan tietää mitä tuotteita ostetaan samanaikaisesti. Kun ongelma on tiedossa ja etsintä määritelty, datan muoto pitää ymmärtää ja data pitää valmistella. Tämän on louhinnan tärkein tekijä.

Kun data on tutkittu, pitää luoda louhintamalli (Data Mining Model) ja rakenne. Mallit luodaan SSAS-mallien avulla BIDS:ssä. Datalähde ja DSV luodaan samoin kuin UDM:n dimensiot ja kuutiot. Yksi louhintarakenne voi sisältää useita louhintamalleja. Kannattaa luoda useita malleja, vertailla niitä ja käyttää luotettavimman oloista mallia. Jos useat mallit antavat samat tulokset, voidaan tuloksien olettaa olevan hyödyllisiä.

Datalouhinnan onnistumisen avain on tarkoitusta vastaavien algoritmien löytäminen ja käyttäminen. SSAS sisältää yhdeksän algoritmia, ja SSIS sisältää kaksi tekstinlouhintamuunnosta. Jos analysoidaan tekstiä, voidaan käyttää SSIS:n muunnoksia Term Extraction ja Term Lookup. BIDSissä on kolme päätyökalua louhintamallien luomista varten: Datan louhintavelho (Data mining wizard), Datan louhintakehitin (Data mining designer) ja Datalouhintanäkymä (Data mining viewers). Harjoituksessa tutustutaan dataan louhintavelhoa ja DSV:tä käyttäen, luodaan kuution näkymä tarkempaa tutkimista varten ja luodaan rakenne. Lisäksi liitetään rakenteen sarakkeet lähdesarakkeisiin, pilkotaan data testisetteihin ja luodaan ensimmäinen louhintamalli. Data mining designerillä liitetään lisää malleja rakenteeseen ja lopuksi tutkitaan malleja Data mining viewerillä.

SSAS noutaa tiedot datalähteestä kahdesta syystä: prosessoidakseen SSAS-tietokantaobjekteja ja suorittaakseen ennuste- ja porausoperaatioita DMX select - lausekkeella. SSAS pitää konfiguroida asianmukaisesti jokaista taskia varten, jotta se pystyy imitoimaan datalähteen tietoja. Datalouhintaa varten SSAS:lle pitää antaa erilaisia käyttöoikeuksia datalähteisiin, louhintarakenteisiin ja louhintamalleihin. Harjoituksissa disabloidaan algoritmi ja testataan testikäyttäjän louhintaoikeuksia. Testien jälkeen prosessoidaan rakenteita ja malleja SSMS:ssa.

Taulukko 9. Luvun 9 sisältö.

Opinto-osio 1	Preparing and Creating Data Mining Structures - Datalouhintamallien valmistelu ja luonti
Opinto-osio 2	Creating Models and Applying Algorithms - Mallien luonti ja algoritmien lisäys
Opinto-osio 3	Validating Models, Using DMX Queries, and Using Prediction Queries in Reports - Mallien validointi, DMX-kyselyiden käyttö ja ennusteiden käyttö raporteissa
Opinto-osio 4	Securing and Processing Data Mining Models - Datalouhintamallien suojaus ja prosessointi

## 6.9 Luku 10: SSRS-raporttien kehitys

SQL Server Reporting Services tukee useita tapoja luoda raportteja. Raportit voi suunnitella alusta lähtien Raporttikehittimellä (Report Designer), Raporttirakentimella (Report Builder) tai kolmannen osapuolen työkaluilla. Raportteja voi myös tuoda Microsoft Office Access 2002:sta tai uudemmasta. BIDS sisältää raporttivelhon kuten myös raporttikehittimen, joka on raporttien luonnin pääasiallinen kehitystyökalu. Opinto-osio 1:n harjoituksissa käytetään raporttivelhoa BIDSissä projektin ja raportin luontiin, sekä raportin sovittamista käyttöön PDF-tiedostoksi ja sen muokkaamista helpommin ymmärrettävään muotoon.

Datasetti sisältää kolme pääelementtiä: datalähteen, komentotyyppin ja kyselyjonon. Lähde määrittelee, miten SSRS yhdistyy datalähteeseen. Se tukee useita muotoja

kuten SQL Server, SSAS, XML, raporttipalvelinmalli, SSIS, .NET, mitä tahansa OLE DB tai ODBC-ajuria. Harjoituksissa luodaan kaksi eri lähdettä käyttävää raporttia käyttämättä velhoa. Toinen käyttää relaatiotietokantaa ja toinen SSAS OLAP - kuutiotalähdettä.

Tehokkaiden raporttien kehittäminen vaatii muutakin kuin uusien elementtien lisäämistä tai olemassa olevien parantamista. On tärkeää, että käyttäjä saa datasta kokonaiskuvan ja pääsee porautumaan läpi tietomäärän löytääkseen tärkeitä yksityiskohtia. Interaktiiviset raportit sisältävät ryhmiä, sarakkeita, rivejä, tauluja tai matriiseita, jotka eivät ole näkyvissä, kun raporttia katsotaan ensimmäisen kerran. Harjoituksissa harjoitellaan aiemmin luodun raportin tietojen piilottamista alkunäytöstä. Tiedot saa halutessaan näkyviin laajentamalla näkymää.

Data-alueet (Data region) ovat raporttikohteita, jotka näyttävät datarivejä dataseiteistä. SSRS:ssä on kolme erilaista data-aluetta: Tablix [7], kaavio (chart) ja mittari (gauge). Tablix on kuten tauludata-alue sillä poikkeuksella, että tauludata-alueella on määritelty määrä sarakkeita, kun taas Tablix on dynaaminen ja sarakemäärä voi kasvaa pivot- taulun käytön aikana.

Data voidaan filteröidä, ryhmitellä ja järjestellä käyttäjän haluamaan muotoon tai järjestykseen. Harjoituksissa käytetään BIDSiä raportin rakentamiseen kahdesta eri datalähteestä. Raporttiin tulee kaksi osaa, matriisi SSAS UDM-kuutiosta graafisella esityksellä ja yksityiskohtainen SQL Server -lähteestä haettu raportti.

Taulukko 10. Luvun 10 sisältö.

Opinto-osio 1	Creating SSRS Projects and Reports in BIDS - SSRS-projektien ja raporttien luonti BIDSissä
Opinto-osio 2	Creating a Dataset from a Data Source - Datasetin luonti datalähdettä käyttäen
Opinto-osio 3	Working with Advanced Report Object Properties - Edistyneiden raporttiobjektien kanssa työskentely
Opinto-osio 4	Applying Dataset Filters and Groups - Datasettiffilttereiden ja ryhmien lisäys

## 6.10 Luku 11: Raporttien laajentaminen ja käyttöönotto

SSRS käyttää kahdenlaisia parametreja: kyselyparametreja ja raporttiparametreja. Kyselyparametri, jota kutsutaan myös datasettipermetriksi, on muuttuja, joka määrittellään datasettitasolla määrittämään tapaa, jolla SSRS tekee datalähdekyselyitä. Raporttiparametri on muuttuja, joka määrittellään raporttitasolla ja joka sallii raportin muotoilun raportin muodostamisen aikana. Harjoituksissa luodaan raportti parametrien mukaan ja määrittellään parametrit raporttiosoitteen kautta.

Raporttitoiteutuksen kriittinen osa on expressioneiden käyttö raporttien käytettävyyden parantamisessa ja dynaamisesti luotujen, upotettujen funktioiden ja oman koodin perusteella muuttuvien raporttien toteutuksessa. Harjoituksessa luodaan raportteja, jotka käyttävät expressioineita, Sen jälkeen käytetään funktioita muuttamaan raporttien ominaisuuksia dynaamisesti.

SSRS:ssä on useita mahdollisuuksia julkaista raportti. Raportit voidaan julkaista raporttipalvelimella, josta käyttäjät voivat lukea raportit raporttimanagerilla (Report Manager). Raportit voidaan myös julkaista portaalissa, esimerkiksi Microsoft SharePointissa. Helpoin tapa julkaista raportti on käyttämällä BIDSiä. Harjoituksissa julkaistaan raportti BIDSillä, Report Managerilla ja Report Builder 1.0:lla.

Taulukko 11. Luvun 11 sisältö.

Opinto-osio 1	Assigning Parameters Within Reports - Raporttiparametrien määrittäminen
Opinto-osio 2	Using Expressions to Perform Advanced Report Item Formatting - Lausekkeiden käyttö kehittyneiden raporttikohteiden muotoilun suorittamiseen
Opinto-osio 3	Deploying New Reports and Changes - Uusien raporttien ja muutosten käyttöönotto
Opinto-osio 4	Using Reports in Your Code - Raporttien käyttö ohjelmointikoodissa

## 6.11 Luku 12: Asennettujen raporttien ja datalähteiden ajastus ja turvaaminen

SSRS on yksi työkaluista, joiden avulla käyttäjät pääsevät käsiksi eri säilytyspaikoissa oleviin tietoihin. Tästä johtuen tiedon pitää olla turvattu niin lähteessä kuin työkalussa, jolla tietoa käsitellään. SSRS:ssä voidaan käyttää Windowsin tietoturvaominaisuuksia (kuten muissakin BI:n komponenteissa) määrittelemään pääsyoikeudet tietoihin. SSRS:ssä on myös turvallisuusmalli.

SSRS:n turvallisuusmalli perustuu objekteille, joita kutsutaan nimellä "securables" ja rooleilla. Objektit voidaan suojata lupamäärittelyillä tai roolille määritellyllä toiminnallisuudella. Windows-käyttäjät ja ryhmät määritellään rooliin kuuluviksi. SSRS:ssä on kahdenlaisia roolimäärittelyksiä: kohdetason roolit ja järjestelmätason roolit. Kohdetason rooleja käytetään määrittämään pääsy raporttiobjekteihin ja järjestelmätason rooleja käytetään määrittämään pääsyä muokkaamaan järjestelmää, palvelinturvallisuutta, käyttöoikeuksia, palvelintason tehtäviä ja niin edelleen. Harjoituksissa käydään läpi kohdetason roolin luonti oikeuksineen ja myöhemmin lisätään järjestelmätason rooli.

Käyttäjät voivat selata raportteja aina halutessaan. Joskus on kuitenkin tarve generoida raportti aikataulun mukaisesti. Käyttäjät voivat halutessaan tilata ajastetun raportin toimitettavaksi esimerkiksi sähköpostiin. Ajastukset tehdään jälleen SQL Server Agentin avustuksella ja lisäksi raporttipalvelin pitää konfiguroida tukemaan ajastuksia ja lähetysoperaatioita. Myös ajastetulle ajolle pitää luoda uudet käyttäjätunnukset luotettavuuden parantamiseksi. Nämä eivät ole kenelläkään käyttäjällä käytössä vaan pelkästään ohjelmalla raporttien lataamiseksi. Harjoituksissa luodaan kantakäyttäjä ja sen jälkeen jaettu aikataulu ja tilaus tiedostonjakoa varten.

Turvallisuus ja tilaajatiedot ovat vain osa komponenteista, joita tulee miettiä, kun määritellään SSRS:n ominaisuuksia. Tärkeä osa on myös raporttien suoritus, johon kuuluu välimuistiin lataaminen ja tilannevedokset (snapshots). Näiden konfiguroiminen voi auttaa palvelimen suorituskykyä ratkaisevasti.

Taulukko 12. Luvun 12 sisältö.

Opinto-osio 1	Administering SSRS Item-Level Permissions and Site Securite Roles -
---------------	---

	SSRS:n kohdetasoisten sääntöjen ja alueiden turvaroolien hallinnointi
Opinto-osio 2	Creating Report Schedules and Subscriptions - Raporttiajastusten ja tilausten luonti
Opinto-osio 3	Managing Report Caching and Execution Properties in Report Manager - Raporttivälimuistin ja suoritusominaisuuksien hallinta raporttimanagerissa

## 6.12 Luku 13: SSRS-palvelimen konfigurointi ja hallinnointi

Kun SSRS asennetaan oletusasetuksin, se on heti valmis raporttien luontia ja loppukäyttäjien raportinlukua varten. Jos kuitenkin tarvitaan erillinen SSRS-instanssi tai halutaan tehdä SSRS-farmi, tarvitaan lisäkonfigurointia. Täydellinen SSRS-asennus sisältää seuraavat komponentit: SQL Server catalog, Report Server service ja Web Service and URL Access. Kustomoidussa asennuksessa joudutaan huolehtimaan muun muassa salausavainten asettamisesta.

Report Server Configuration Managerin avulla voidaan konfiguroida lähes kaikki SSRS:n asetuksista. Monet asetuksista, kuten yhteysasetukset ja salausavainten määrittäminen, voidaan tehdä myös komentorivityökaluilla. Lisäksi raportointipalvelimella on useita palvelintason ominaisuuksia, jotka määritellään SSMS:n ja RSReportServer.config -XML-tiedoston kautta. Harjoituksissa käytetään komentorivityökalua salausavainten varmuuskopiointiin ja palauttamiseen.

Taulukko 13. Luvun 13 sisältö.

Opinto-osio 1	Installing and Configuring Initial Instances of SSRS - SSRS-instanssien asennus ja konfigurointi
Opinto-osio 2	Configuring Advanced SSRS Settings and Scale-out Deployment - Kehittyneiden SSRS-asetusten määrittäminen ja laajentuva asennus

## 7 Yhteenveto

Insinöörityöni perustui aluksi lähes kokonaan itseopiskelukurssin 70-448: TS: Microsoft SQL Server 2008, Business Intelligence Development and Maintenance -teoksen materiaaleihin. Kurssiin tutustuessani huomasin, etteivät tietoni tietokannoista olleet vielä lähellekään riittävät, ja tietoja oli täydennettävä verkosta. Tieto ei ole vielä kukaan riittävä, mutta se on lisääntynyt huomattavasti viimeisten kuukausien aikana, ainakin mitä tulee Microsoft SQL Serveriin ja sen lisäosiin.

Microsoft Business Intelligence tarjoaa varsin laajat ja monikäyttöiset työkalut datan keräämiseen, analysointiin ja raportointiin. Microsoft Officen ollessa lähes monopoliasemassa toimistotyökaluna, on useissa yrityksissä valmiina loistavat edellytykset jatkokäyttöä löydettyä tietoa. Raporttidatan jakelu ja lukeminen on helppoa edellä mainitusta syystä, sillä käyttäjillä on valmiiksi käyttöliittymä raportteihin. Työkalujen yhtenäisyys madaltaa kynnystä ottaa ohjelmisto käyttöön mutta laajuutensa takia on varmasti silti melkoinen kustannuskysymys monessa yrityksessä.

Olin asentanut Windows 2008R2 -palvelimen sekä Microsoft SQL Server 2008R2 -palvelimen virtuaalialustalle kotonani Microsoft SQL -peruskurssia varten. Koska tietokannatkin olivat jo valmiina, näytti siltä, että ympäristö oli valmiiksi pystytetty ennen kuin aloitin työn tekemistä.

Tämä ei luonnollisestikaan pitänyt aivan paikkaansa, SQL-serverin asennuksessa ei kehoitettu asentamaan BI:n käyttämiä lisäosia eli SSIS, SSAS ja SSRS -komponentteja. Tätä kautta pääsin tutustumaan MS SQL Serverin hallintaan lisäkomponenttien kautta. Asennuksen hoidin erehdys-erehdys-metodin kautta, eli ensin määrittelin komponenteille väärät käyttöoikeudet. Korjasin tämän määrittelemällä oikeudet jälleen väärin. Kolmannella kerralla, ohjeet luettuani, sain komponentit toimimaan ja pääsin tekemään harjoitustöitä.

Itse kirjoitustyö osoittautui varsin työlääksi operaatioksi. Lähdemanuaali on erittäin laaja ja olennaisten osien kerääminen vei aikaa huomattavasti odotettua enemmän. Lisäksi epäselvän termistön selvittäminen itselleni internet-hakujen kautta vei oman aikansa. Ilman aikaisempaa kokemusta SQL-palvelimista olisi työ auttamattomasti jäänyt tekemättä.

Itse ohjelmisto osoittautui hyvin monipuoliseksi kokonaisuudeksi ja on joustavuutensa myötä pätevä valinta niin pienelle kuin suurellekin yritykselle. BI tukee useita eri valmistajien tietokantoja lähdetietoina joten käyttöönoton voi tehdä vaiheittain. Microsoft Officen ollessa vallitsevassa markkina-asemassa työasemien toimistosovelluspakettina, raportointityökalut ovat monissa yrityksissä jo valmiina loppukäyttäjien saatavilla. Tämä madaltaa kynnystä ottaa käyttöön Windowsin raportointisovelluskokonaisuus.



## Lähteet

- 1 Microsoft Database Product Samples -sivusto. Microsoft, Inc.  
<<http://msftdbprodsamples.codeplex.com/releases/view/55926>>. Päivitetty 24.4.2012. Luettu 8.5.2012.
- 2 Extract Transform Load. Verkkodokumentti.  
<[http://en.wikipedia.org/wiki/Extract,\\_transform,\\_load](http://en.wikipedia.org/wiki/Extract,_transform,_load)>. Päivitetty 24.4.2012. Luettu 8.5.2012.
- 3 Dimensional modeling. Verkkodokumentti.  
<[http://en.wikipedia.org/wiki/Dimensional\\_modeling](http://en.wikipedia.org/wiki/Dimensional_modeling)>. Päivitetty 17.2.2012. Luettu 8.5.2012.
- 4 SQL Server Management Studio Tools. Verkkodokumentti. Microsoft, Inc.  
<[http://msdn.microsoft.com/en-us/library/ms174168\(v=sql.105\).aspx](http://msdn.microsoft.com/en-us/library/ms174168(v=sql.105).aspx)> Luettu 8.5.2012.
- 5 Precedence Constraints. Verkkodokumentti. Microsoft, Inc.  
<<http://msdn.microsoft.com/en-us/library/ms141261.aspx>>. Luettu 8.5.2012.
- 6 SQL Server Reporting Services. Verkkodokumentti. Microsoft, Inc.  
<[http://msdn.microsoft.com/en-us/library/ms159106\(v=sql.105\).aspx](http://msdn.microsoft.com/en-us/library/ms159106(v=sql.105).aspx)>. Luettu 8.5.2012.
- 7 Understanding the Tablix Data Region. Verkkodokumentti. Microsoft, Inc.  
<[http://technet.microsoft.com/en-us/library/dd220587\(v=sql.105\).aspx](http://technet.microsoft.com/en-us/library/dd220587(v=sql.105).aspx)>. Luettu 8.5.2012.