



samk



Satakunnan ammattikorkeakoulu
Satakunta University of Applied Sciences

HANNES LEVÄNSUO

**Data-analytiikan hyödyntäminen
energiamarkkinoiden
tiedonvaihtosanomien käsittelyssä**

SÄHKÖ JA AUTOMAATIO KOULUTUSOHJELMA
2020

| | | |
|---|-------------------------------------|--------------------------|
| Tekijä(t) Levänsuo, Hannes | Julkaisun laji Opinnäytetyö, AMK | Päivämäärä 10/2020 |
| | Sivumäärä 27 | Julkaisun kieli Suomi |
| Julkaisun nimi Data-analytiikan hyödyntäminen energiamarkkinoiden tiedonvaihtosanomien käsittelyssä | | |
| Tutkinto-ohjelma Sähkö- ja automaatiotekniikka | | |
| <p>Tiivistelmä</p> <p>Tutkitaan energiamarkkinoiden sanomaliikenteen sanomien käsittelyä ja työjonoja, joita pitkin sanoma kulkee. Tavoitteena oli löytää pullonkaulaprosessit tai työjonot, joiden vuoksi tiettyjen sanomien käsittelyaika on pidempi, kuin tarvitsisi olla.</p> <p>Tutkimus toteutettiin Excelin Power Queryllä sekä Excelissä pyörivillä VBA-skripteillä. Näiden avulla työjonoista löydettiin ongelmaprosessi, jonka korjaus pienentäisi tiettyjen sanomien käsittelyä huomattavasti. Excelin päälle tehdyllä työkalulla on mahdollista tutkia otantoja sanomien kulusta tietyllä aikavälillä.</p> <p>Tavoitteena oli löytää myös tapoja parantaa analytiikkaa ja tässä tapauksessa paras tapa olisi luoda erillinen tietokanta analytiikkaa varten. Tuotantotietokannan tarkemmat tiedot säilytetään vain viikon, joten tutkinta suuremmalla aikavälillä ei ole mahdollista. Erillisen tietokannan avulla voitaisiin tutkia esimerkiksi eri vuosien työjonojen käsittelyaikoja toisiinsa ja löytää pitkäaikaisia vaikutuksia sanomaliikenteen käsittelyprosesseissa.</p> | | |
| energia, sanomaliikenne, tiedonvälitys | | |

| | | |
|--|--|-------------------------------------|
| Author(s) Levänsuo, Hannes | Type of Publication Bachelor's thesis | Date 10/2020 |
| | Number of pages 27 | Language of publication: Finnish |
| Title of publication Using data analytics in energy market message processing | | |
| Degree program Electrical and automation engineering | | |
| Abstract Investigation of message processing in electricity market and balance settlement. Main goal was to investigate and find slow processes in message processing queues and nodes. Work was done with Excel Power Query and VBA-scripts on Excel platform. With these tools we were able to find slowness on some messaging processing queues. We created a platform that you can manually search messages processes on a specific timeframe. One object was also to find a way to make analytics better. One of the best option in this situation was to create a new data warehouse for storing summary data. We store message queue data for a week so a long term summary data would allow yearly investigation of message queue processing. | | |
| energy, messaging, message exchange | | |

SISÄLLYS

| | |
|--|----|
| 1 JOHDANTO | 5 |
| 2 TYÖN TOIMEKSIANTAJA | 5 |
| 3 SANOMALIIKENNE | 6 |
| 4 SANOMIEN KÄSITTELY | 9 |
| 4.1 Tiedonsiirto sanomajärjestelmien välissä | 9 |
| 4.1.1 API-rajapinnat | 10 |
| 4.1.2 FTP ja SFTP-tiedonsiirto | 11 |
| 4.2 Sanomien käsittely IXS:ssä | 12 |
| 5 IXS-JÄRJESTELMÄ | 13 |
| 6 SANOMAN NYKYINEN METATIETOSISÄLTÖ | 14 |
| 7 SANOMIEN SIJAINTI JA YHDISTÄMINEN | 15 |
| 7.1 Työkalut | 15 |
| 7.2 Tiedonhaku | 18 |
| 7.3 Excelin päälle tehty työkalu | 19 |
| 8 YHDISTETYN AINEISTON TUTKIMINEN | 22 |
| 9 TULOKSET | 24 |
| 10 YHTEENVETO JA JOHTOPÄÄTÖKSET | 25 |
| LÄHTEET | |

1 JOHDANTO

Nykyisellään sanomaliikenteen käsittely hoidetaan IXS-järjestelmällä. Järjestelmä käsittelee ja muuntaa sanomat haluttuun muotoon, jonka jälkeen ne toimitetaan eteenpäin. Sanomia käsitellään yli 100 miljoonaa vuoden aikana ja ne kulkevat eri prosesseja ja työjonoja pitkin.

Sanomista jää metadata-tietoa lokiin, jota hyödynnetään yleisesti sanomavirheiden käsittelyssä ja sanomien tutkimisessa. Tavoitteenamme on hyödyntää tuota metadataa ja tehdä siihen liittyvää analytiikkaa, jonka avulla olisi mahdollista löytää sanomankäsittelyn pullonkaulat ja kehitystarpeet.

Sanomaliikennettä tutkitaan tietokantaan tallennettujen käsittelytietojen avulla ja niitä yhdistelemällä erillisiksi summa kentiksi, joiden avulla käsittelyprosessien työaikoja voidaan tutkia.

2 TYÖN TOIMEKSIANTAJA

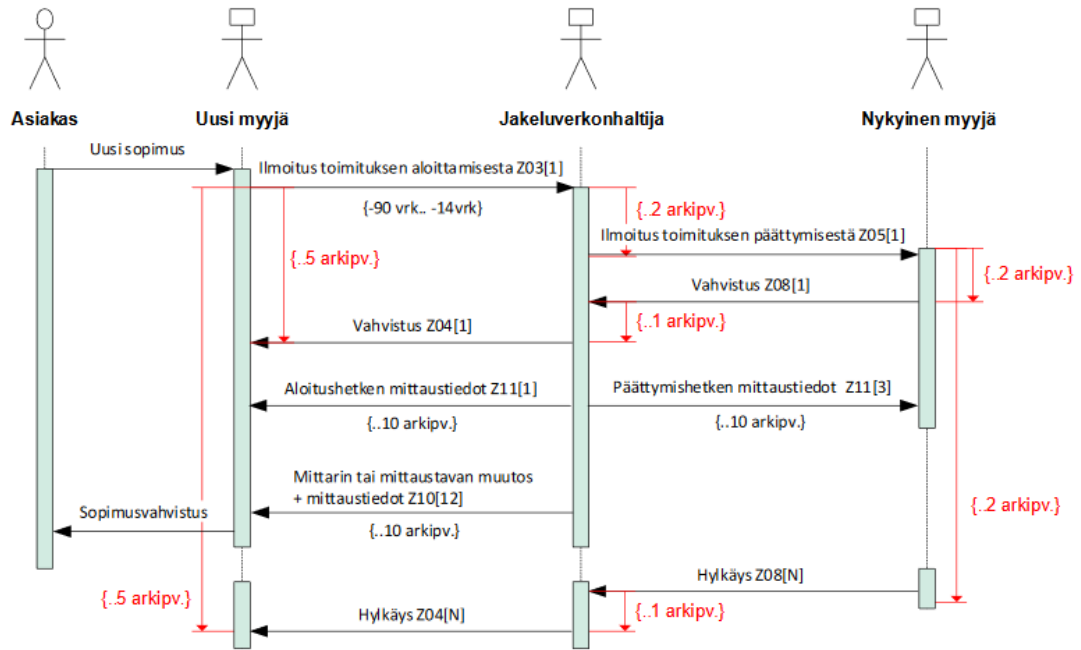
Työ tehdään Empower IM Energy Intelligence Divisionissa (EID). Yritys toimii Suomessa, mutta toimintaa on myös muissa pohjoismaissa. EID:n sisällä toimeksianto kuuluu sanomaliikenne-tiimiin, jonka alaisuudessa opinnäytetyö tehtiin.

3 SANOMALIIKENNE

Energiamarkkinoiden tiedonvaihto hoidetaan sanomaliikenteellä. Sanomaliikenne koostuu pääosin EDIFACT-muotoisista sanomista, mutta tässä työssä tutkitaan myös erillisiä sisäisiä sanomia, jotka liikkuvat esimerkiksi .csv- tai empxml-muodossa. Käsittelemme myös sanomat, jotka eivät suoranaisesti liity energiamarkkinoiden sanomaliikenteeseen, mutta kulkevat sanomina meidän sanomaliikenteen kautta. Näitä ovat esimerkiksi tuulitietojen haku eri rajapintoja pitkin toisiin järjestelmiin näkyville tai erillisiä laskentoja varten.

EDIFACT-sanomat koostuvat esimerkiksi PRODAT-, MSCONS-, DELFOR-, APERAK- ja CONTROL-sanomista. EMPXML-muoto ja erilaiset muunnokset tehdään asiakastarpeiden mukaan ja näitä ovat esimerkiksi tuulipuistoista tuotavien tietojen vienti eri järjestelmiin kuten EnerimEMS. EMPXML-tiedostomuotoon muunnetaan kaikki sanomat, jotka viedään EnerimEMS-järjestelmään, kuten esimerkiksi tuulipuistojen tuotantodatat, jotka tulevat meille joko API-rajapinnan kautta, JSON- tai .csv-muodossa. IXS:n päätarkoitus on hoitaa näitä muunnoksia eri sanomaformaattien välillä, sekä siirtää muutetut tai muuntamattomat sanomat eteenpäin. IXS:n avulla valvotaan sanomaliikennettä ja liian pitkät välit sanomien välillä tai sanomien virhekäsittelyt aiheuttavat hälytyksen sanomaliikenne tiimille. Sanomaliikenne tiimin yksi tarkoitus on selvittää virheen syy ja korjata asia. (Fingrid Oyj, 5)

Sähkön vähittäismarkkinoiden ohjeistukset ovat joko Energiateollisuus Ry:n tai Fingrid Oyj:n vastuulla. Erilaisia ohjeita sanomien luontiin ja käsittelyyn, kuten sanomaliikenteen ohjeistuksia ylläpidetään joko energia.fi –verkkosivulla, joka on Energiateollisuus Ry:n ylläpitämä sivusto tai ediel.fi –sivustolla, joka on Fingrid Oyj:n ylläpitämä sivusto. Näille sivustoille on kerätty vähittäismarkkinoiden sanomaliikenteen yleiset ohjeet sekä ediel sanomankäsittelijöille tarkoitettut tarkemmat ohjeet sanomien luomiseen ja käsittelyyn. (Energiateollisuus ry, 2020, 11)



Kuva 1. Esimerkki myyjänvaihdosta sanomaliikenteessä. (Energiateollisuus ry, 2020, 21)

Sanomaliikenteen käsittelyssä on paljon rajoituksia ja vaatimuksia. Yleisestä sanomaliikenteestä perehdytään enemmän EDIFACT-sanomaliikenteen säätelyyn ja sanomarakenteeseen. Järjestelmien välisissä muunnoksissa on otettava huomioon, että kaikki järjestelmät eivät automaattisesti tuota EDIFACT-muotoisia sanomia. Meidän tehtävämme on muuntaa väärämuotoiset sanomat EDIFACT-muotoiseksi, jolloin sanoma on markkinoiden mukainen ja kaikki markkinaosapuolet pystyvät tulkitsemaan lähetetyt sanomat oikein. Ilman erillistä standardia jokaisen eri osapuolen tulisi selvittää toistensa kanssa erillinen sanomaliikenteen säännöstö ja kaikki sanomaliikenne olisi erilaista toisiinsa verrattuna. Standardi auttaa sanomaliikenteen harmonisoinnissa ja siinä, että kaikkien osapuolien välinen sanomaliikenne on samanlaista. Näin ei tarvitse jokaisen osapuolen välille luoda erillistä räätälöityä sanomaliikennetarkaisua, koska standardi on olemassa ja sitä voidaan käyttää. Standardi myös vähentää työmäärää uusien osapuolien tullessa markkinoille, koska heidän tulee rakentaa vastaanotto ja lähetys vain tiettyjen sanomastandardien mukaisesti eikä montaa erilaista räätälöityä sanomaformaattia. [Energiateollisuus ry, 2020, 11]

EDI-sanomaliikenne koostuu monesta sanomatyypistä, joilla välitetään eri asioita sähkömarkkinoihin liittyen. Sanomatyypit ovat hyödyllisiä, koska vastaanottaja tietää aina, mitä tietoa vastaanotetaan ja mitä sille pitää tehdä. Sanomatyypien rakenteet ovat erilaisia eli käytännössä EDIFACT-standardi pitää sisällään monta erilaista sanomaa, joissa kaikissa on eri rakenne ja joilla kaikilla välitetään erilaista tietoa. Sanomaformaatit sisältävät pakollisia ja vapaaehtoisia kenttiä, mutta nyt tarkastellaan vain sanomatyypien niin sanottua pääsisältöä. (Fingrid Oyj, 5)

Mitattua kulutustietoa lähetetään sanomatyypillä MSCONS.96A, joka sisältää sähkökuluttajien kulutustietoja esimerkiksi normaalin kerrostaloasunnon kulutustiedot aikasarjana. Sanomassa lähetetään ainakin lähettäjä, vastaanottaja, sanoman referenssikoodi, kulutuksen aikajakso sekä kulutustiedot. Ennakkoilmoituksia ja ennusteita lähetetään sanomatyypillä DELFOR D.96A, joka pitää sisällään ennakkotietoa esimerkiksi tuotantolaitoksen seuraavan päivän arvioidusta sähköntuotannosta. Sanomassa voidaan lähettää myös ennuste arvioidusta sähkön kulutuksesta eli energiantarpeesta tietylle tunnille. DELFOR sisältää lähettäjän, vastaanottajan, referenssikoodin sanomalle, sanoman sisältämän aikasarjan rajat sekä aikasarjanarvot tunneittain. Asiakastietojen muutokset ilmoitetaan PRODAT-tyyppisellä sanomalla. Mikäli asiakkaan tiedot esimerkiksi osoite, nimi tai sähkönmyyjä vaihtuu, nämä muutokset ilmoitetaan Prodat-tyyppisellä sanomalla. Prodat sisältää lähettäjän, vastaanottajan sekä muutosta koskevan asiakkaan tiedot. Prodat-sanomaliikenne sisältää myös eri alityyppejä, kuten esimerkiksi Z05[14] (sopimuksen päättymisen ilmoitus) tai Z03[14] (uusi sopimus). Kuvasta 1 näkee miten prodat-sanomien toiminta on erilaista verrattuna esimerkiksi MSCONS- tai DELFOR-sanomiin, koska mscons- ja delfor-sanomissa ei toimiteta erilaisia alityyppisanomia, kuten Z03 tai Z05 prodateissa. (Unece.org www-sivut 2020)

Sanomaliikenne sisältää myös kuittaussanomiam, joiden tarkoitus on selvittää, onko vastaanottava osapuoli saanut sanoman. Esimerkiksi Prodat-, Delfor- tai Mscons-sanomissa kuittauksia voidaan pyytää erikseen. Mscons-sanomissa kuittausten toimitus on harvinaisempaa, kuin Delfor-sanomissa, mutta molemmissa on EDIFACT-standardin mukaan mahdollista käyttää kuittauksia. Kuittaukset toimitetaan kahdella eri sanomatyypillä joista toinen on CONTROL ja toinen on APERAK. Sanomaliikenteessä vastuu sanoman perille toimittamisesta on aina

sanoman lähettäjällä siihen asti, kunnes sanomaan liittyvä kuittaussanoma saapuu lähettäjälle. Mikäli sanomassa pyydetään kuittausta niin sanoman sisältö on lähettäjälle ja vastaanottajalle sitova vasta kun kuittaussanoma on toimitettu lähettäjälle. Control-kuittaussanoma ei ota kantaa sanoman sisältöön. Tämän kuittauksen ainoa tarkoitus on ilmoittaa lähettäjälle, että sanoma on saapunut vastaanottavaan järjestelmään. Aperak-kuittaussanoma ottaa kantaa myös sanoman sisältöön. Aperak tarkistaa onko tullut tieto hyväksytyä (positiivinen aperak) tai onko tiedoston sisällössä jotain poikkeavaa (negatiivinen aperak). Näiden kahden sanoman yhtäaikaista käyttöä ei suositella vaan yleisesti PRODAT ja MSCONS-sanomissa käytetään aperak-kuittauksia. (Fingrid Oyj, 8-9)

4 SANOMIEN KÄSITTELY

4.1 Tiedonsiirto sanomajärjestelmien välissä

Kun tietoa liikutetaan kahden eri järjestelmän välillä, pitää olla selkeät määrittelyt minkälaista tietoa ja missä muodossa tiedot liikkuvat. Sovelluksen välisiä yhteyksiä kutsutaan rajapinnoiksi eli ”pinta” kahden eri sovelluksen välissä, jonka kautta tieto liikkuu sovelluksesta toiseen. Rajapintoja on myös erilaisia esimerkiksi pelkästään API-rajapintoja voi olla useita erilaisia. Palveluntarjoajat tai sovelluskehittäjät voivat julkaista rajapintaluettelon sovelluksen julkaisun yhteydessä eli luettelon rajapinnoista, joiden avulla sovelluksen kanssa voidaan keskustella. Kahden eri sovelluksen keskustelua rajapinnan yli kutsutaan yleensä integraatioksi eli järjestelmät ovat yhdistyneet toisiinsa. (Visma-blogi [www-sivut](http://www.sivut) 2019)

Tietoa liikutettaessa internetin yli tulee huomioida tietoturva-asiat. Tietoa ei voida suoraan lähettää verkon yli vaan yhteys asiakkaan ja palvelimen välissä tai kahden eri järjestelmän välissä tulee suojata. Ilman yhteyden suojaamista kaikki kahden järjestelmän välissä kulkeva tieto on suojaamatonta ja salaamatonta eli osaava henkilö voisi lukea tiedot selväkielisenä järjestelmien välillä tai mahdollisesti muuttaa tietoja.

Tiedonsiirron turvaamiseksi on kehitetty erilaisia tapoja. Mikäli pelkän internetin yli liikutetaan tiedostoja, tiedostot voidaan salata esimerkiksi AES-256 kryptausmetodilla. Tällöin itse tiedosto on salattu ja vaatii salausavaimen, jotta tiedoston voi avata, mutta yhteys ei ole salattu. Pelkällä AES-256 salauksella voidaan hyvin toimittaa yksittäisiä tiedostoja tarvittaessa esimerkiksi henkilötietoja sisältäviä dokumentteja organisaatioiden välillä. Mikäli siirrot ovat yksittäisiä esimerkiksi sähköpostilla tehtäviä ei kannata erikseen lähteä luomaan VPN-tunnelia eli salattua yhteyttä kahden pään välille vaan kannattaa toimittaa tiedosto salattuna esimerkiksi 7-zip -sovelluksella. (Helsinki YO www-sivut 2020)

Mikäli tehdään integraatio kahden järjestelmän välille, se tehdään yleensä salaamalla koko yhteys ei vain tiedostot tai tiedot, jotka toimitetaan. Tämän hyöty on, että tiedonsiirrossa ei tarvitse salata tiedostoja tai tietoja, jotka liikkuvat yhteyttä pitkin. Kun yhteys on salattu, kaikki sen sisällä liikkuva tieto menee salattuna, joten ulkopuoliset eivät voi kuunnella tai lukea tietoja, joita toimitetaan yhteyden kautta. Salattuja yhteyksiä kutsutaan yleensä VPN-tunneleiksi eli Virtual Private Network – yhteyksiksi. Hyvä esimerkki jatkuvasta tiedonsiirrosta, joka pitää salata on etätyö. Etätyötä tekevä henkilö on jatkuvasti yhteydessä yrityksen verkkoon internetin yli ja tässä tapauksessa vain tiedostojen salaus ei onnistu vaan koko yhteys tulee salata. Yhteyden salaamiseksi luodaan erillinen VPN-yhteys työntekijän koneen ja yrityksen sisäisen verkon välille, jolloin kaikki työntekijän tekemät toimet internetin yli menevät salattuna yritysverkolle. (Cloudflare www-sivut 2020)

4.1.1 API-rajapinnat

API-rajapinnat (Application Programming Interface) eroavat FTP- ja SFTP-tiedonsiirroista, koska API-rajapinnat ovat nimensä mukaisesti ohjelmointirajapintoja eli datanhaku yleensä suoritetaan ohjelman sisäisesti sisäisillä prosesseilla. API-rajapinta voi olla sovelluksen sisäinen tai se voi olla sovelluksen ulkopuolinen, kuten esimerkiksi julkaistu SOAP- tai REST-palveluilla. FTP- ja SFTP ovat tiedoston siirtoprotokolla ja seuraavassa kappaleessa syvennyttään niihin. (CGI – blogi www-sivut 2018)

JSON/REST (JavaScript Object Notation/Representational State Transfer) palvelu on yleisesti käytössä mobiilisovelluksissa ja sen suosio on nousussa SOAP-palveluidenkin ylitse. REST voi käyttää myös muita esitysmuotoja JSON:in lisäksi, kuten esimerkiksi XML. IXS-järjestelmässä suuri osa API-rajapintahauista ovat juuri JSON/REST-tyyppisiä. JSON esitysmuodossa hyödynnetään tiedon nimi/arvo pareja sekä taulukoita tiedon esittämiseen. (CGI – blogi www-sivut 2018)

```
{
  "glossary": {
    "title": "example glossary",
    "GlossDiv": {
      "title": "S",
      "GlossList": {
        "GlossEntry": {
          "ID": "SGML",
          "SortAs": "SGML",
          "GlossTerm": "Standard Generalized Markup Language",
          "Acronym": "SGML",
          "Abbrev": "ISO 8879:1986",
          "GlossDef": {
            "para": "A meta-markup language, used to create markup languages such as DocBook.",
            "GlossSeeAlso": ["GML", "XML"]
          },
          "GlossSee": "markup"
        }
      }
    }
  }
}
```

Kuva 2. Tiedon esitystapa JSON-formaatissa. (Json.org www-sivut 2020)

4.1.2 FTP ja SFTP-tiedonsiirto

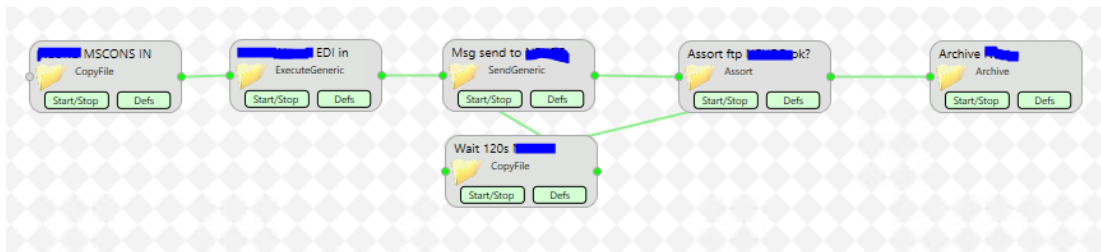
FTP eli File Transfer Protocol on kehitetty standardi tiedoston välitykseen verkon ylitse. FTP on asiakas - palvelin protokolla, jossa asiakas pyytää palvelimelta tiedostoon liittyviä toimintoja, kuten tiedoston lataus, nimen muuttaminen, poistaminen, kopiointi jne. Yleisesti asiakas ottaa yhteyden palvelimeen ja palvelin vaatii käyttäjätunnusta ja salasanaa yhteyden muodostamiseksi. FTP-yhteys sellaisenaan on täysin salaamaton ja kaikki liikenne, jopa salasanat ja käyttäjätunnukset, liikkuvat FTP-yhteyden yli täysin salaamattomana. Käytännössä siis kuka tahansa voisi lukea liikennettä, joka kulkee asiakkaan ja palvelimen välissä. (Techtarget www-sivut 2019)

Liikenteen suojausta varten on kehitetty tapa suojata FTP-yhteys ja silloin puhutaan SFTP-yhteydestä eli Secure File Transfer Protocollasta. SFTP:ssä FTP yhteys suojataan päästä päähän SSH:lla. Suojattua FTP-yhteyttä ulkopuoliset eivät pysty

lukemaan. SFTP-liikenteessä tiedostojen sisältö ja tunnistautumistiedot ovat salattuja päästä asiakkaalta palvelimelle. (Techtarget www-sivut 2019)

4.2 Sanomien käsittely IXS:ssä

Sanomien käsittely IXS:ssä hoidetaan työjonoilla ja työjonoihin kuuluvilla prosesseilla. Työjono sisältää yleensä useita eri prosesseja, joilla on eri tehtävä työjonossa. Esimerkiksi yksi prosessi hakee ja lukee sanoman esimerkiksi SFTP:llä tai verkkosivuilta erillisen rajapinnan kautta. Toinen prosessi muuntaa lähdesanoman kohdetta vastaavaan muotoon esimerkiksi .csv-tiedoston EMPXML-tiedostomuotoon. Kolmannessa prosessissa tutkitaan muunnoksen onnistumista ja epäonnistunut sanoma viedään virheenkäsittely-työjonossa eteenpäin. Onnistunut sanoma viedään eteenpäin esimerkiksi lähetysprosessiin, joka toimittaa sanoman eteenpäin kohdeasiakkaalle tai kohdejärjestelmään, josta järjestelmän sanomankäsittelijä lukee sanoman sisälle. Samaan aikaan viimeinen työjonon prosessi käsittelee sanoman ja vie sen arkistoon, johon tallennetaan sanoman alkuperäinen muoto sekä sen muunnettu muoto.



Kuva 3. Työjonon osa ja siinä olevat nodet eli sanomankäsittelyprosessit.

IXS:ssä on tarkat lokitiedot tallennettuna msg.log –tauluun ja lokitiedot säilytetään siellä n. viikon. Näitä lokitietoja on tarkoitus hyödyntää työjonojen tehokkuuden tarkkailussa sekä siinä paljonko sanomia tietyn työjonon läpi kulkee tietynä ajanjaksona.

Tämän työn tavoite on tutkia sanomankäsittelyä sanomankäsittelyjonoissa eli kuinka pitkään sanoma kulkee jonossa ja kuinka kauan sitä käsitellään prosesseissa. Olisiko mahdollista kehittää sanomankäsittelyä tehokkaammaksi tai parantaa käsittelyprosessia jotenkin, jotta sanoman käsittely nopeutuisi. Tavoite on myös löytää

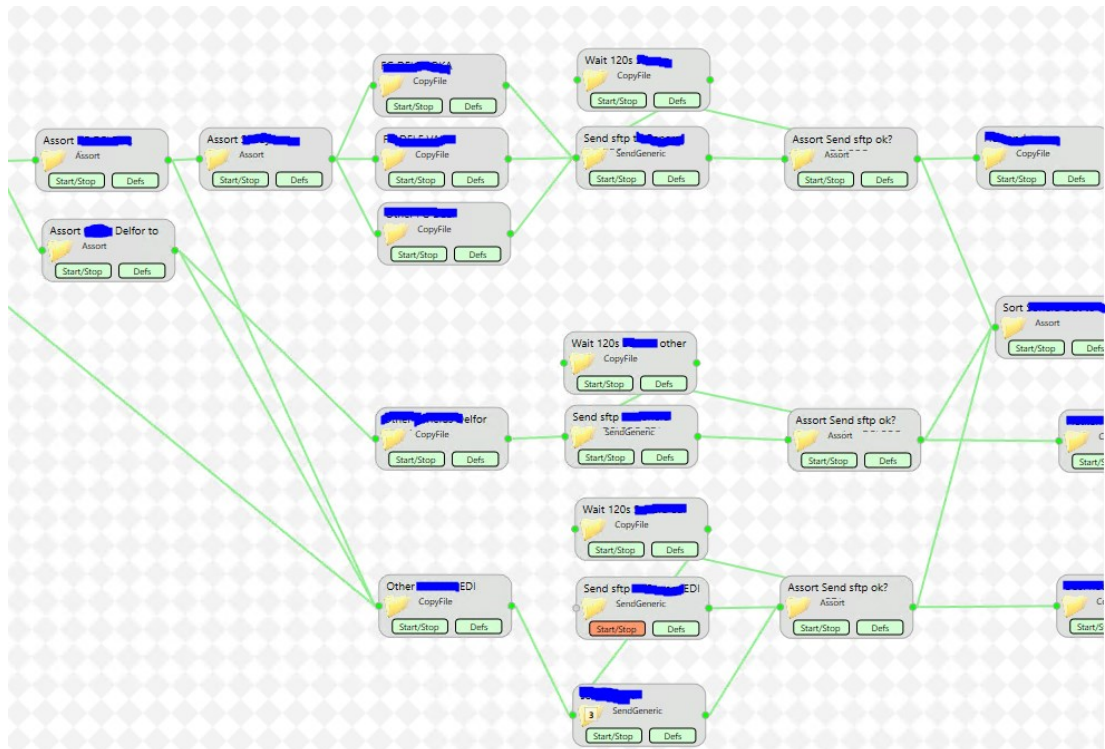
erilaisia pullonkauloja eli prosesseja tai työjonoja, joissa sanomankäsittely kestää pitkään.

Sanomankäsittelyprosessit tuottavat lokitietoa ja metatietoa sanomasta sekä sanomatiedostosta. Metatietoja kuten käsittelyaika, onnistuminen, odotusaika jne. hyödynnetään työjonojen tehokkuuden tutkimisessa ja siinä voitaisiinko tuota käsittelyä parantaa.

5 IXS-JÄRJESTELMÄ

IXS on Empowerin oma sovellus, jonka toteutus on aloitettu vuonna 2013. Sanomatiimin päätyökalu on IXS, jonka avulla ratkaistaan mahdollisia sanomankäsittelyongelmia, kuten virheellinen muunnosprosessi, yhteysvirhe joko toimittavassa tai hakevassa päässä jne. IXS:ää ajetaan erillisellä palvelimella ja erillistä asiakaskäyttölittymää ei ole olemassa, joten IXS on Empowerin sisäinen työkalu. IXS:n kautta on mahdollista tutkia sanomaliikenteen ongelmia ja virheilmoituksia sekä luoda erilaisia työjonoja tai muokata työjonoja. Työjono tarkoittaa niiden prosessien sarjaa, jonka sanoma käy läpi ennen, kuin sanoma on kokonaan käsitelty. Uudet muunnosprosessit tehdään Visual Studiolla, joko Visual Basicillä tai C#:lla. Käännetyt .dll-tiedostot viedään tuotantopalvelimelle ja otetaan tuotantokäyttöön IXS:llä, mutta niiden koodaus tehdään IXS:n ulkopuolella.

IXS koostuu työjonoista ja prosesseista, jotka toimivat kansiotasolla eli jokaisella prosessilla on omat käsittelykansiot, joiden avulla sanomaliikennettä hoidetaan. Prosessi käsittelee oman käsittelykansionsa sisällä olevat sanomat ja siirtää ne eteenpäin seuraavalle prosessille. Prosessien lokitiedot tallennetaan tietokantaan ja metadata-tiedostoon, jossa viedään seuraavalle prosessille kyseisen sanoman käsittelytiedot ja käsittelyn historiatiedot. Tämän työn tarkoitus on tutkia tietokantaan erillisiin lokitauluihin vietyjä tietoja ja selvittää miten niistä saataisiin parhaat hyödyt irti.



Kuva 4. IXS työjonojen haarautuvuus.

IXS:n sisällä olevat prosesseja kutsutaan nodeiksi ja ne näkyvät IXS:n käyttöliittymällä ”laatikoina” ja näitä laatikoita yhdistävinä viivoina. Työjonoissa kulkevat sanomat kirjoittavat omat tietonsa tietokantaan ja prosessien välillä liikkuvat tiedot kulkevat erillisessä metadata-tiedostossa, joka sellaisenaan ei ole ihmisille luettavassa muodossa. Jokainen node käsittelee tiedoston ja lukee sekä muokkaa kyseisen tiedoston metadata-tiedostoa seuraavaa prosessia varten, jolloin koko työjonon ajan metadata-tiedosto kasvaa ja muuttuu tarpeen mukaan. Metadata-tiedosto voidaan myös poistaa prosessien välissä, mikäli halutaan ettei vanhan metadata-tiedon arvot vaikuta käsittelyyn enää siitä pisteestä eteenpäin.

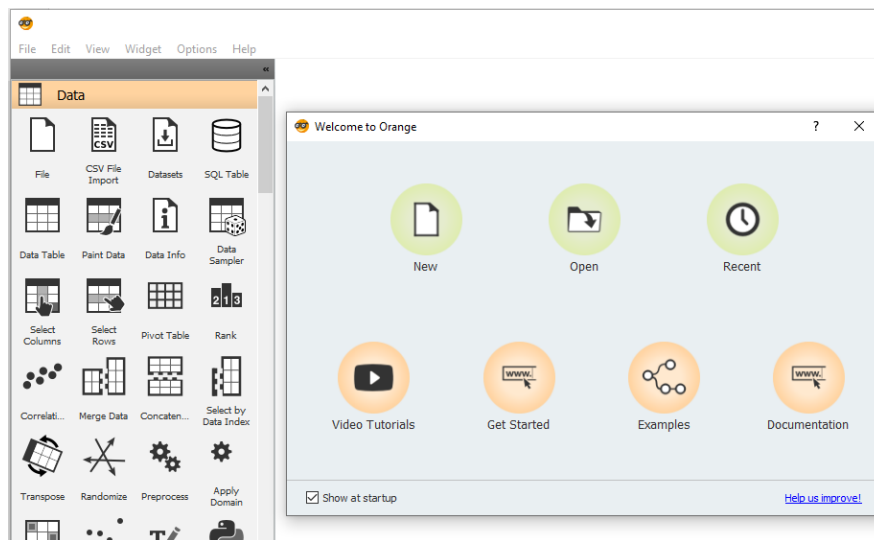
6 SANOMAN NYKYINEN METATIIETOSISÄLTÖ

Metadatan sisältämät tiedot vaihtuvat riippuen työjonoista ja prosesseista, joissa sanoma kulkee. Yleisesti metadataan pyritään lisäämään tiedot niin, että tiedoston kulkua on helppo seurata. IXS toimii tiedostopohjaisesti eli metadata vie sanomasta

Tableau ja Microsoftin PowerBI ovat molemmat hyviä työkaluja ETL-prosessien hallintaan ja analytiikkaan. Molemmissa työkaluissa on myös erilliset dashboard-näkymät, joilla tietoja voidaan esittää automaattisesti interaktiivisessa muodossa. Eli raportin selaaja voi itse suodattaa helposti näytettävää tietoa. Molemmat näistä työkaluista ovat myös lisenssien takana olevia maksullisia sovelluksia, joten niitä ei hyödynnetty tässä työssä.

Tableau on perustettu 2003 tietokonetieteiden projektissa Stanfordissa. Projektin tarkoituksena oli kehittää tiedonkeruun ja tiedonkulun analysointia sekä tehdä siitä helpommin lähestyttävää ihmisille visualisointien avulla. Tableau-sovelluksen perustana toimii VizQL, joka on suunniteltu siihen, että tiedon haku ja muokkaus onnistuisi raahaamalla komponentteja visuaalisessa näkymässä. Tableautta on kehitetty koneoppimisen, statistiikan ja puhekielen osalta, jotta käyttäjän luovuutta voitaisiin paremmin hyödyntää analytiikassa. Tableau on tuoteperhe, joka koostuu monesta eri analytiikkasovelluksesta. Tässä työssä olisi voitu hyödyntää Tableau server- ja Tableau desktop-sovelluksia. (Tableau [www-sivut](http://www.tableau.com) 2020)

Microsoft PowerBI on Microsoftin vastaus visuaaliseen data-analytiikkaan. Microsoftilla on myös ilmaiset SQL Server-lisenssin mukana tulevat SSIS (Sql Server Integration Services), SSRS (SQL Server Reporting Services) ja SSAS (SQL Server Analytics Services) työkalut. Näiden SQL Server-lisenssin mukana tulevien sovellusten käyttäminen ei ole yhtä visuaalista ja ne eivät hyödynnä Dashboard-näkymiä, kuten Tableau ja PowerBI. PowerBI mahdollistaa käyttäjille interaktiivisen Dashboardin tekemisen, kuten Tableaukin. PowerBI hyödyntää myös tekoälyä laskennoissa ja suuri ero Tableau:n ja PowerBI:n välillä on PowerBI:n Office-ympäristöön ja Azureen liittämisen helppous. (Microsoft Power BI [www-sivut](http://www.microsoft.com) 2020)



Kuva 7. Orange-työkalun käyttöliittymä.

Ensimmäisessä datan haussa pyrittiin käyttämään Orange-sovellusta, joka on tarkoitettu tietojen analysointiin. Orangen suurin ongelma kuitenkin oli, kun kyse on avoimesta ja ilmaisesta sovelluksesta niin sen vakaus ei ole markkinoiden huipputuotteiden luokkaa. Orangen avoimuuden hyödyt tulevat siinä, että se on ilmainen sovellus ja siihen on mahdollista asentaa ja käyttöönottaa erilaisia käyttäjien tekemiä moduuleita. Käytännössä Orangeen voidaan luoda toiminnallisuuksia, mikäli ne puuttuvat perusversiosta tai käyttäjien luomista moduuleista. Orangea käytetään myös ilmaisuutensa vuoksi kouluissa ja yliopistoissa data-analytiikan opettamisessa. Sovellukseen on suunniteltu myös erilaisia moduuleita opetuksen helpottamiseksi. Orangella ETL-prosessi ja tietojen haku kaatui siihen, ettei Orange kyennyt käsittelemään enempää, kuin 20000 riviä kerrallaan. Tämä aiheutti aina, että kyselyt kestivät todella pitkään ja tietojen käsittely ei onnistunut. Ensimmäisten muutaman yrityksen jälkeen siirryttiin käyttämään toista vaihtoehtoa, joka tässä tapauksessa osoittautui Excelin Power Query -työkaluksi. (Orange www-sivut 2020)

Microsoft Power Query on tietojen yhdistämiseen ja analysointiin suunniteltu työkalu. Power Queryn ominaisuuksia voidaan hyödyntää joko PowerBI:n päällä tai Excelissä ja se on yksi Excelin kolmesta eri Power analyysityökalusta Power Pivot, Power View ja Power Query. (Microsoft PQ yleiskatsaus www-sivut 2020)

Excelin Power Query osoittautui sopivaksi työkaluksi tämän tietomassan käsittelyyn. Tietoa ei kuitenkaan voitu analysoida suuria määriä kerrallaan vaan tässäkin

tapauksessa piti analysoida pientä osaa tiedosta kerrallaan. Excelin käyttö työkaluna mahdollisti kuitenkin erillisten VBA-skriptien tekemisen kyselyiden ja laskentojen kustomointiin.

7.2 Tiedonhaku

Yhden päivän työjonojen haku onnistuu Excelin Power Queryn avulla niin, ettei odotusajat kyselyiden suorittamiselle kasva liikaa. Työjonojen haku on kaikkein raskain kysely, koska aineiston prosessi- ja työjonotietoihin yhdistävät tiedot ovat metadata-kentässä sekaisin. Metadata-kenttä pitää rikkoa useampaan tietokenttään, jotta sieltä saadaan poimittua msgref-tieto, jonka avulla saadaan sanoman msgid haettua. Sanoman identifioivan numeron avulla työjonojen seuranta ja tarkkailu onnistuu helposti, koska sanoman id voidaan yhdistää prosessien lokiriveihin. Nämä haut ovat kaikkein raskaimpia tämän aineiston tutkimisessa, koska lokirivien määrä on kaikkein suurin ja kyselyssä pitää rikkoa metadata-kenttä useammaksi kentäksi. Tätä kyselyhakua voitaisiin tehostaa rikkomalla metadatasta msgref kenttä erilliseen sarakkeeseen omaksi alkiokseen, joka voitaisiin indeksoida. Indeksointi nopeuttaa tiedonhakua todella paljon, mutta vie myös enemmän tilaa.

Tietojen yhdistäminen tehdään pattern-hakua hyödyntämällä eli haetaan stringin sisältä tietyn mallista tekstiä. Tekstimallin haku on yleisesti hidaskuusi hakumetodi, joten tuota ei hyödynnetä, kuin yhden rivin hakemisessa, josta saadaan selville sanoman ID. ID:n kanssa haettiin loput rivit, jolloin sanoman käsittely nopeutuu ja yhdistäminen erillisiin jonotietoihin helpottuu.

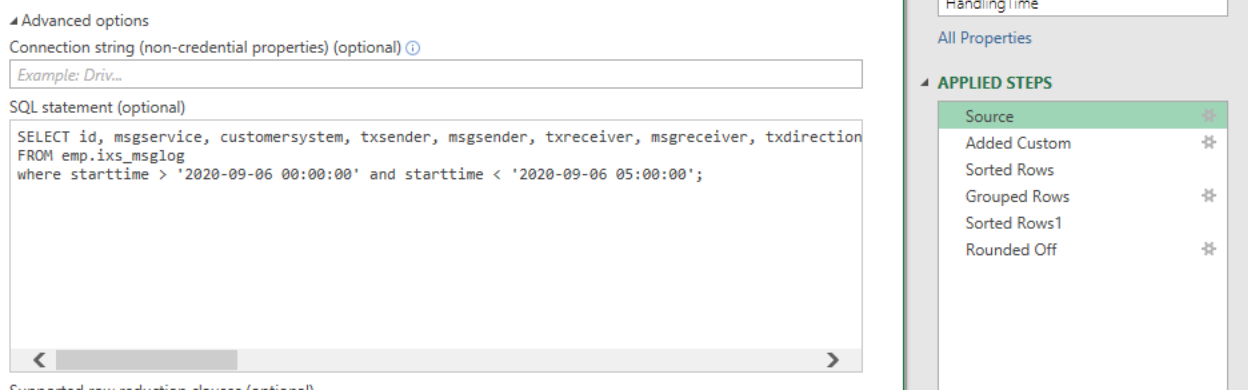
| txsender | msgsender | txreceiver | msgreceiver | txdirection | msgtype | Count | HandlingAvg | SizeAvg | HandlingMin | HandlingMax | HandlingMedian | SizeMedian |
|----------|-----------|------------|-------------|-------------|---------|-------|-------------|---------|-------------|-------------|----------------|------------|
| | | | | OUT | MSCONS | 15 | 0.58.53 | 0 | 0.58.22 | 0.59.24 | 0.58.53 | 0 |
| | | | | OUT | MSCONS | 329 | 0.46.37 | 0 | 0.34.38 | 0.58.23 | 0.46.59 | 0 |
| | | | | IN | MSCONS | 20 | 0.35.22 | 0 | 0.34.41 | 0.35.47 | 0.35.14 | 0 |
| | | | | OUT | MSCONS | 3 | 0.34.52 | 54361 | 0.33.47 | 0.35.24 | 0.35.24 | 0 |
| | | | | OUT | MSCONS | 20 | 0.34.48 | 0 | 0.34.21 | 0.35.24 | 0.34.53 | 0 |
| | | | | IN | MSCONS | 3 | 0.20.26 | 33051 | 0.13.03 | 0.35.11 | 0.13.04 | 10944 |
| | | | | IN | MSCONS | 2 | 0.23.51 | 2994 | 0.12.31 | 0.35.10 | 0.23.51 | 2994 |
| | | | | IN | MSCONS | 3 | 0.35.08 | 7858 | 0.35.08 | 0.35.08 | 0.35.08 | 5809 |
| | | | | IN | MSCONS | 2 | 0.35.07 | 2937 | 0.35.07 | 0.35.07 | 0.35.07 | 2937 |

Kuva 8. Jonotiedot

7.3 Excelin päälle tehty työkalu

Tässä työssä ainoa käytettävissä oleva työkalu, joka on monipuolinen ja muunneltavissa on Excel. Seuraavaksi piti selvittää, kuinka Power Querylla saadaan tiedot näkyviin ja VBA:lla käsiteltyä datat ihmisystävälliseen muotoon. Ensin aloitettiin perustietojen tuonnista tietokannasta Power Querylla näkyviin. Perustietojen haku tehdään kolmella eri kyselyllä.

Kysely 1: Haetaan jonotiedot, johon lasketaan sanomien käsittelyajat, sanomamäärät ja sanoma koot. Tämän kyselyn tiedot ovat työjonojen summa- ja keskiarvotietoja kaikista jonojen läpi kulkeneista sanomista. Tämä kysely on tutkimisen perusta, johon haetaan lisätiedot toisissa kyselyissä.



Kuva 9. Kysely 1 muodostus ja vaiheet

Kysely 2: Tarkoituksena on yhdistää sanomien msgref- sekä msgid-tiedot, joiden avulla voidaan seurata sanoman kulkema jono prosessikohtaisesti sekä tutkia prosessin käyttämää käsittelyaikaa.

Advanced options

Connection string (non-credential properties) (optional) ⓘ

Example: Driver={SQL Server}; Server={lo...}

SQL statement (optional)

```
SELECT distinct(msgid), case position('msgref=' in metadata) when '0' then '0' else
replace(split_part(substring(metadata,position('msgref=' in metadata)), '#@@#',1), 'msgref=', '')
end "MessageRef",
replace(split_part(substring(metadata,position('msgtype=' in metadata)), '#@@#',1), 'msgtype=', '')
as msgtype
from emp.log
where position('msgref=' in metadata) not in (0) and starttime > '2020-09-06 00:00:00' and start
```

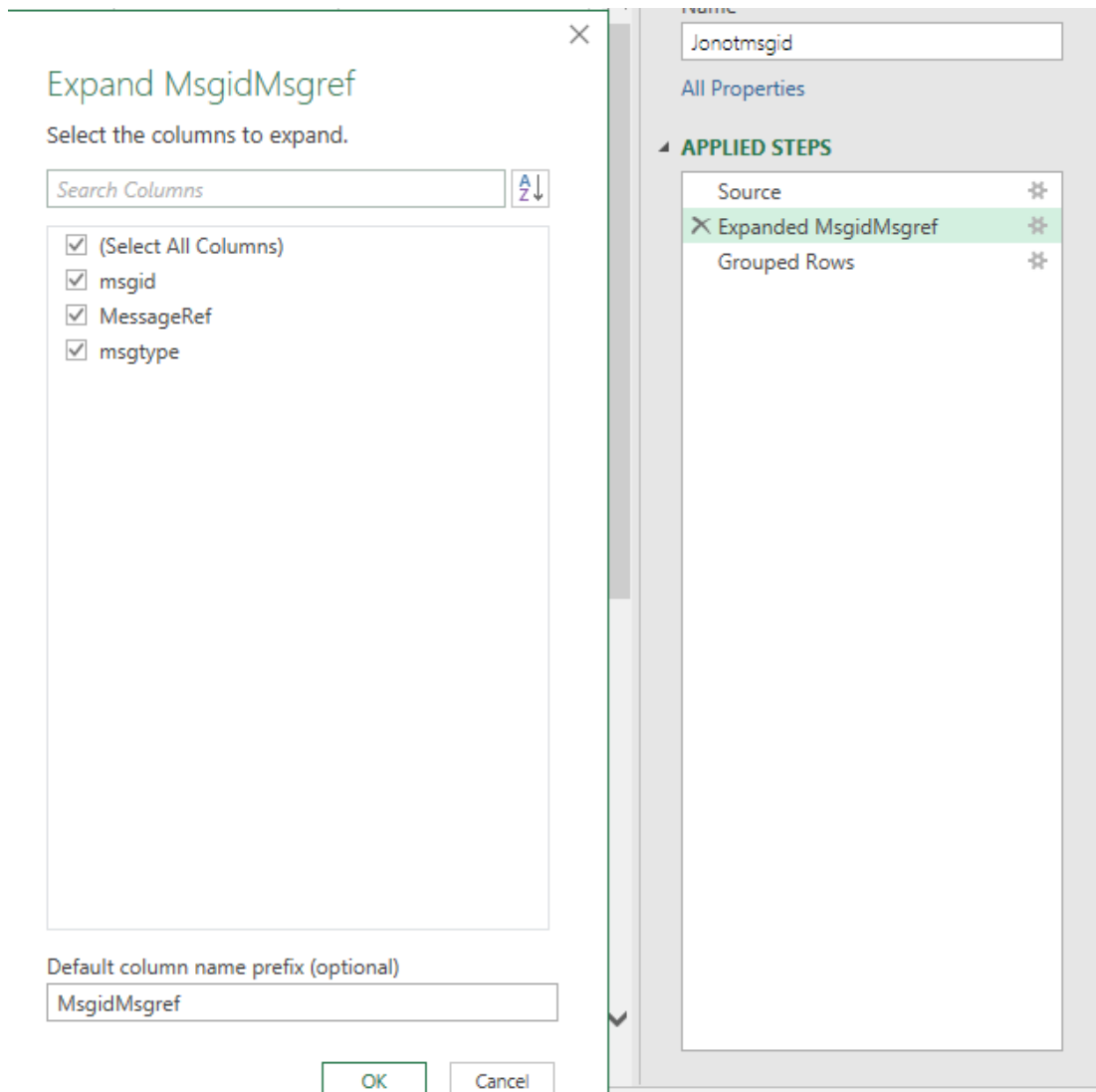
Supported row reduction clauses (optional)

(None) Detect

OK Cancel

Kuva 10. Tärkeimpänä metadatan pilkkominen ja msgid:n yhdistäminen msgref-arvoon.

Kysely 3: Kolmannella kyselyllä jonotietoihin yhdistetään msgid, jota hyödynnetään, kun tutkitaan kyseisen jonon käsittelytietoja. Msgid joka tässä lisätään jonojen perään on tärkein komponentti VBA-skriptien lisätietojen haussa, koska msgid on indeksoitu, joten sen avulla tehdyt kyselyt ovat nopeita.



Kuva 11. Kyselyn laajennettuna haetut arvot. Tämän kyselyn tärkein msgid.

Kysely 1 hakee jonotiedot yleisellä tasolla ja yleisellä tasolla msglog-tilusta. Tämä kysely ei yksinään riitä, koska tietoa tarvitaan myös log-tilusta, jotta jonojen yhdistäminen prosesseihin ja niissä käytettävään aikaan onnistuu. Kysely 2 toimii apukyselyinä, jolla voidaan yhdistää msglog ja log-tilujen tiedot toisiinsa. Kysely 2:n ansioista voidaan yhdistää kolmannen kyselyn tiedot ensimmäisen kyselyn jonotietoihin. Kolmannessa kyselyssä haetaan toisen kyselyn tulosten perusteella ensimmäisen kyselyn tulosjoukolle msgid-arvot. Msgid on yhden ko. työjonon läpi kulkeneen sanoman identifioiva tieto, jonka avulla voidaan hakea jonon sisältämät prosessit ja niiden arvot kyselyn 1 tulosjoukolle.

Näiden kolmen Power Query:n peruskyselyn lisäksi on tehty VBA-skripti, jota hyödynnetään jonojen ja prosessien tietojen yhdistämisessä. VBA-skripti luo

lopullisen tulosjoukon, jossa voidaan vertailla sanoman käsittelyyn kuluva aikaa sanoman prosessien käsittelyaikoihin sekä prosesseihin, joita pitkin sanoma kulkee. VBA-skripti ajaa kolme eri kyselyä tietokantaan alkuperäisten kyselyiden tulosjoukkojen ja Excel-filtteröinnin perusteella. VBA-skripti on toteutettu niin, että vain tulosjoukon kymmeneen ylimpään riviin haetaan tiedot. Tämän tarkoituksena on estää käyttäjää hakemasta liian suurelle tulosjoukolle tietoja, joka voisi aiheuttaa ongelmia tietokannalle.

```
'Luodaan COUNT-kysely hakuehtojen perusteella
StrQuery = "SELECT COUNT(*) FROM [ ] where msgid = '" & msgId & "'"
'MsgBox (StrQuery)
rst.Open StrQuery, cnn
Sheets("JonotNodet").Range("R" & j).CopyFromRecordset rst
rst.Close

'Haetaan käsittelyaika nodeissa
StrQuery = "SELECT SUM(endtime-starttime) FROM [ ] where msgid = '" & msgId & "'"
'MsgBox (StrQuery)
rst.Open StrQuery, cnn

Sheets("JonotNodet").Range("S" & j).CopyFromRecordset rst
rst.Close
|
'Haetaan nodenimet
StrQuery = "SELECT processname, endtime-starttime as handlingtime, starttime FROM [ ] where msgid = '" & msgId & "' order by id asc"
'MsgBox (StrQuery)
rst.Open StrQuery, cnn

Do Until rst.EOF
  For Each fld In rst.Fields
    queryResult = queryResult & " - " & fld.Value
  Next fld
  rst.MoveNext
Loop

'queryResult = rst.GetString()
Sheets("JonotNodet").Range("T" & j).Value = queryResult
queryResult = ""
rst.Close
```

Kuva 12. Skriptin kyselyosuus, joka hakee prosessien lukumäärän, käsittelyajan ja työjonot.

8 YHDISTETYN AINEISTON TUTKIMINEN

Yhdistetyllä aineistolla tarkoitetaan, kun tarkkaillaan sanomaloki msglog- ja log- taulujen yhdistettyjä tietoja. Msglog- taulusta haetaan yksittäisen sanoman yleistiedot ja log- taulusta kerätään IXS-järjestelmän prosessien sanomankäsittelystä syntyneet tiedot. Näiden taulujen yhdistävä tekijä msgref on pitkän string alkion sisällä. Msgref- tietoa hyödynnetään vain siihen, että msglog ja log taulut saadaan yhdistettyä ja poimittua msgid-tieto. Msgid on tietojen raportoinnin kannalta tärkein tieto, koska se on indeksoitu ja sanomakohtainen yksilöivä tieto.

Yhdistämällä aineistot näistä tauluista saadaan tietoa siitä, kuinka pitkään mitään sanomaa on käsitelty eri prosesseissa ja kuinka pitkään sanoma on ollut ns. odotustilassa eli tilassa, jolloin sitä ei ole käsitelty.

Tarkastelemme sanomajonoista seuraavia arvoja: sanomien lukumäärä, käsittelyaika, käsittelyaika prosesseissa, käsittelyjonot ja sanoman koko. Tutkimalla esimerkiksi sanoman käsittelyaikaa prosessissa saadaan tarkka tieto, kuinka pitkään sanoma on ollut käsiteltävänä. Sanoman normaali käsittelyaika näyttää kauanko sanoma on ollut järjestelmässä eli siihen on laskettu mukaan odotusajat. Näiden kahden ajan ero syntyy, koska ennen jokaista käsittelyprosessia on ns. pollausaika. Pollausaika on se aika, kuinka usein prosessi käy tutkimassa oman käsittelykansionsa sisällön. Pollausaika on yleensä 5–15 sekuntia ja sanomissa, jotka eivät ole aikakriittisiä tuo aika voi olla pidempikin esimerkiksi minuutin tai kaksi. Sanoman käsittelyaika on mahdollisesti riippuvainen sanoman koosta, siksi sanoman koko on otettu mukaan myös tutkittaviin suureisiin. Käsittelyjonojen tutkiminen per työjono näyttää minkä prosessien läpi sanoma kulkee ja sen avulla saadaan jäljitettyä se työjonon prosessi, jossa mahdollinen käsittelyongelma tai pullonkaula on.

| Count | HandlingAvg | SizeAvg | HandlingMin | HandlingMax | HandlingMedian | Nodet | Pros.aika | Jono | | |
|-------|-------------|---------|-------------|-------------|----------------|-------|--------------|------|------------------------------------|--------------------------------------|
| 15 | 0.58.53 | 0 | 0.58.22 | 0.59.24 | 0.58.53 | 17 | 00:00:01.746 | | 00:00:00.005 - 6.9.2020 3.06.14 - | 00:00:01.065 - 6.9.2020 4.04.03 - |
| 329 | 0.46.37 | 0 | 0.34.38 | 0.58.23 | 0.46.59 | 15 | 00:00:01.768 | | 00:00:00.006 - 6.9.2020 3.06.13 - | 00:00:01.178 - 6.9.2020 4.03.12 - |
| 20 | 0.35.22 | 0 | 0.34.41 | 0.35.47 | 0.35.14 | 15 | 00:00:00.403 | | 00:00:00.002 - 6.9.2020 3.05.06 - | 00:00:00.15 - 6.9.2020 3.39.23 - A |
| 3 | 0.34.52 | 54361 | 0.33.47 | 0.35.24 | 0.35.24 | 12 | 00:00:00.05 | | 00:00:00.003 - 6.9.2020 3.05.06 - | 00:00:04.285 - 6.9.2020 3.39.27 - |
| 20 | 0.34.48 | 0 | 0.34.21 | 0.35.24 | 0.34.53 | 15 | 00:00:00.403 | | 00:00:00.002 - 6.9.2020 3.05.06 - | 00:00:00.15 - 6.9.2020 3.39.23 - A |
| 3 | 0.20.26 | 33051 | 0.13.03 | 0.35.11 | 0.13.04 | 20 | 00:00:00.635 | | ? - 00:00:00.018 - 6.9.2020 3.04.2 | 00:00:00.003 - 6.9.2020 3.04.30 - As |
| 2 | 0.23.51 | 2994 | 0.12.31 | 0.35.10 | 0.23.51 | 20 | 00:00:00.464 | | ? - 00:00:00.025 - 6.9.2020 3.04.2 | 00:00:00.002 - 6.9.2020 3.04.30 - As |
| 3 | 0.35.08 | 7858 | 0.35.08 | 0.35.08 | 0.35.08 | 20 | 00:00:00.441 | | ? - 00:00:00.018 - 6.9.2020 3.04.3 | 00:00:00.002 - 6.9.2020 3.04.46 - As |
| 2 | 0.35.07 | 2937 | 0.35.07 | 0.35.07 | 0.35.07 | 20 | 00:00:00.411 | | ? - 00:00:00.025 - 6.9.2020 3.04.3 | 00:00:00.002 - 6.9.2020 3.04.46 - As |
| 2 | 0.35.07 | 9260 | 0.35.07 | 0.35.07 | 0.35.07 | 20 | 00:00:00.561 | | ? - 00:00:00.017 - 6.9.2020 3.04.3 | 00:00:00.002 - 6.9.2020 3.04.46 - As |

Kuva 13. Työkalun luoma tulosjoukko.

Kuten kuvasta 8 voidaan huomata, työkalulla löydettiin muutama työjono, joiden käsittelyaika HandlingMin, -Max, -Med ovat huomattavan korkeita verrattuna prosessin tekemään käsittelyaikaan Pros.aika. Ylimmän rivin esimerkissä sanoma kulkee työjonoa pitkin keskimäärin noin 59 minuuttia ja tuosta ajasta sanoma on prosessien käsittelyssä alle 2 sekuntia. Käytännössä näiden sanomien suuri käsittelyaika ei johdu siitä, että sanoma olisi raskas käsitellä vaan sanomalla olevat odotusajat ovat pitkiä. Sarakkeessa Jono käsittelyprosessit on piilotettu, mutta

käsittelyaika ja käsittelyn aloitusaika on jätetty näkyville. Työkalun avulla löydettiin syy siihen, miksi näiden käsittelyaikojen ero on suuri. Jonon ensimmäinen prosessi alkaa käsitellä sanomaa klo 03:06:14 UTC+0 ja käsittely kestää 5 tuhannesosasekuntia. Seuraava prosessi ottaa sanoman käsittelyyn vasta 04:04:03 UTC+0 eli melkein tunnin sen jälkeen, kun sanoma on käsitelty ensimmäisessä prosessissa.

Tätä tulosjoukkoa tutkimalla huomattiin, että järjestelmässä on yksi prosessi, jonka kohdalla tämä tapahtui ja esimerkin työjonoissa kaikki sanomat kulkevat tämän yhden prosessin läpi. Esimerkin tapaus on yksi suurimmista odotusajoista ja pullonkauloista, joka työkalulla löytyi. Työkalun avulla voidaan siis tutkia työjonojen sanomankäsittelyä ja löytää olemassa olevat pullonkaulat. Yksi ongelma työkalulla on se, että sitä on käytettävä manuaalisesti ja tutkittava ajanjakso ei voi olla kovin pitkä esimerkiksi kahden päivän tutkintajakso on liian suuri johtuen sanomien lukumäärästä.

9 TULOKSET

Tutkimalla sanomankäsittelyä saatiin hyviä tuloksia ja löydettiin jo eräs prosessi, jonka sanomankäsittely on huomattavasti muista poikkeavaa. Tämä prosessi saatiin kartoitettua ja korjaustoimiin voidaan ryhtyä. Esimerkiksi tämän prosessin käsittelyaika ei ole ollut tiedossa ennen, kuin opinnäytetyön avulla tutkittiin jonojen käsittelyaikoja ja pullonkaulaprosesseja.

Opinnäytetyötä varten tehtyä Exceliä ja sen Power Queryn avulla on mahdollista tutkia ja löytää pullonkaulat tietyille aikaväleille. Yksi ongelma Excelin Power Queryä käytettäessä ja kantakyselyiden tekemisessä tuotantokantaan on se, että kyselyt tehdään niin suureen massaan tietoa, että kyselyiden tekeminen on hidasta. Työkalun avulla voidaan tutkia lyhyitä esimerkiksi 5-15 tunnin mittaisia jaksoja halutulta aikaväliltä. Suuremmat kyselyt voivat kaataa Excelin, koska Excelin prosessi menee odotustilaan ja suorittaa taustalla kyselyitä. Tätä sovellusta varten ei myöskään vielä

luotu erillistä ihmisystävällistä käyttöliittymää vaan kyselyt tehdään muokkaamalla kantanäkymien parametrejä ja hakemalla sieltä tiedot näkyville.

10 YHTEENVETO JA JOHTOPÄÄTÖKSET

Kyselyiden nopeuttamiseksi ja tiedonhaun parantamiseksi. Emp.log-taulussa olevasta metadata-kentästä pitäisi irrottaa msgref-kenttä omaksi alkiokeksi. Msgrefin indeksoiminen toisi paljon nopeutta sanomien ja työjonojen yhdistämiseen sekä niiden tutkimiseen. Mikäli msgref olisi eroteltu erilliseen alkioon niin jokaisella työjonoilla ei tarvitsisi erikseen aina rikkoa metadata-kentästä msgref-tietoa saadakseen msgid:n.

Parempaan tietanalyysiin tekemiseksi olisi myös tarpeen luoda erillinen tietokanta, johon voitaisiin tehdä suurempia kyselyitä vaarantamatta tuotantotietokannan eheyttä tai suoritustehoa. Erillisen data-analyysi -tietokannan tekemisessä voitaisiin hyödyntää myös pidempää käsittelyväliä, kuin yksi viikko esimerkiksi aggregoimalla tietosisältöjä ja seuraamalla pidemmältä väliltä sanomakäsittelyn suorituskykyä tai sanomamääriä.

Yksi data-analyysin ongelma on myös se, että tarkemmat prosessien käsittelytiedot säilytetään vain yhden viikon. Tämä on operatiivisessa toiminnassa täysin riittävä, mutta analyttisesti olisi hyvä, jos viikon datasta saataisiin summatiedot erilliseen tietokantaan, jolloin tietoa sanoman käsittelyjen pituudesta olisi saatavilla paljon pidempään, mutta ei niin tarkalla tasolla. Tallessa voisi olla esimerkiksi päiväkohtaiset keskiarvo- ja mediaanitiedot käsittelyjonojen osalta.

LÄHTEET

CGI – blogi 2018, <https://www.cgi.fi/fi/blogi/pieni-api-sanakirja>, viitattu 2.10.2020

Cloudflare www-sivut 2020, <https://www.cloudflare.com/learning/access-management/what-is-a-business-vpn/>, viitattu 5.10.2020

Energiateollisuus ry, Vähittäismarkkinoiden menettelytapa ja sanomaliikenneohje, http://energia.fi/files/1350/Vahittaismarkkinoiden_menettelytapa_ja_sanomaliikenneohje.pdf, viitattu 20.8.2020

Fingrid Oyj, Ediel sanomavälityksen yleiset sovellusohjeet, https://www.ediel.fi/sites/default/files/Ediel%20sanomav%C3%A4lityksen%20yleiset%20sovellusohjeet%203.3_0.pdf, viitattu 1.10.2020

Helsinki YO www-sivut 2020, <https://helpdesk.it.helsinki.fi/ohjeet/tietoturva-ja-pilvipalvelut/tietoturva/luottamuksellisten-liitetiedostojen-salaus>, viitattu 5.10.2020

Json.org www-sivut 2020, <https://json.org/example.html>, viitattu 2.10.2020

Microsoft Power BI www-sivut 2020, <https://powerbi.microsoft.com/en-us/>, viitattu 2.10.2020

Microsoft PQ yleiskatsaus www-sivut 2020, Power Query – yleiskatsaus ja käytön opettelu, <https://support.microsoft.com/fi-fi/office/power-query-%E2%80%93-yleiskatsaus-ja-k%C3%A4yt%C3%B6n-opettelu-ed614c81-4b00-4291-bd3a-55d80767f81d>, viitattu 2.10.2020

Tableau www-sivut 2020, <https://www.tableau.com/>, viitattu 2.10.2020

Techtarget www-sivut 2019, <https://searchnetworking.techtarget.com/definition/File-Transfer-Protocol-FTP>, viitattu 2.10.2020

Unece.org www-sivut 2020, <http://www.unece.org>, viitattu 2.10.2020

Visma-blogi www-sivut 2019, <https://www.visma.fi/blog/mita-ohjelmiston-avoimet-rajapinnat-mahdollistavat/>, viitattu 5.10.2020

Orange www-sivut 2020, orange.biolab.si, viitattu 2.10.2020