HAAGA-HELIA
ammattikorkeakoulu

Kansallisarkiston tiff valokuvien siirto ja metatietojen päivittäminen

Jouni Repo

Opinnäytetyö
Tietojenkäsittelyn koulutusohjelma
2014

HAAGA-HELIA
ammattikorkeakoulu

Tietojenkäsittelyn koulutusohjelma

Tiivistelmä

17.5.2014

Tekijä tai tekijät Jouni Repo	**Ryhmätunnus tai aloitusvuosi** 2012
Raportin nimi Kansallisarkiston tiff valokuvien kopiointi ja metatietojen päivittäminen	**Sivu- ja liitesivumäärä** 22 + 0
Opettajat tai ohjaajat Cristian Brade, Tiina Koskelainen	

Tämän produktin aiheena oli valmistaa migraatiotyökalu, joka kopioi tiedostoja vanhalta nauhalta uudemmalle nauhaformaatille. Sovellusta piti saada helposti muokattua myös tulevaisuudessa. Sovellus toteutettiin Linux palvelinympäristölle.

Produktissa keskityttiin PHP -ohjelmointiin siltä osin kun se oli sovelluksen valmistuksen kannalta tarpeellista. Ohjelmoinnin lisäksi opinnäytetyössä keskityttiin varmuuskopiointiin ja erilaisiin tallennusvälineisiin. Myös erilaisista kuvaformaateista kerrotaan eroavaisuudet.

Produktin lopuksi kerrotaan miten työ eteni Digitaaliarkistossa ja pohditaan miten sovellusta voitaisiin vielä kehittää.

Asiasanat
Migraatio, Ohjelmointi

HAAGA-HELIA
University of Applied Sciences

Abstract

17.5.2014

Information Technology

Authors Jouni Repo	**Group or year of entry** 2012
The title of thesis The National Archives tiff photo copying and meta data updating	**Number of pages and appendices** 22 + 0
Supervisor(s) Cristian Brade, Tiina Koskelainen	

The topic/aim of this project was to create a migration software tool, which copies files from an old tape to newer tape format. One requirement of the application was that it should be easily configured in the future as well. The application software was engineered for Linux-environment.

PHP-programming was used when needed. In addition to programming the thesis concentrates on backing up-technology and different kinds of recording devices. The thesis also introduces differences between various image formats.

The thesis ends in a discussion of how the project proceeded in Digital Archives and how the application can be further developed.

Key words
programming, migration

Sisällys

1 Johdanto

Produktin aiheena oli kopioida tif tiedostoja SDLT -nauhoilta, uudemmille LTO-4 -nauhoille. Kopioinnin yhteydessä kuviin liitettiin uudistetut metatiedot vanhojen metatietojen tilalle.

Työ toteutettiin Kansallisarkiston Digitaaliarkistossa, joka hallitsee Kansallisarkiston asiakirjojen digitoinnin pitkäaikaissäilytyksen ja samalla asiakirjojen julkaisun Internetissä.

Digitaaliarkisto perustettiin vuonna 2003 kun EU:ssa painotettiin kulttuuriperinnön digitointia ja niiden näyttämistä Internetissä. Tuohon aikaan levykapasiteetti ei ollut samanlainen kuin nykypäivänä, jonka takia kaikki tif kuvat pitivät tallentaa nauhoille. Siihen aikaan nauhoina olivat SDLT -nauhat.

Kun digitointi aloitettiin, kuviin ei ollut saatavilla kaikkia metatietoja vaan kuvat jouduttiin tallentamaan nauhoille riittämättömillä metatiedoilla. Nyt kopioinnin yhteydessä kuviin liitettiin puuttuvat metatiedot, kopioidessa tif kuvia uudelle nauhaformaatille.

Vanhoilla nauhoilla on testiaineistoa ja tietokantakopioita joita ei viedä uusille nauhoille. Vanhoja kuvia siirrettäessä niistä tarkistettiin md5-summia siirtojen välissä, että voitiin varmistaa kuvan eheys.

Produktin lopussa kerrotaan miten sovellusta voitaisiin jatko kehittää ja miten sovellus toimii tänä päivänä Digitaaliarkistossa.

2 Varmuuskopiointi

Varmuuskopioinnilla tarkoitetaan tiedon kahdennusta alkuperäisestä paikasta toiseen tallennusvälineeseen. Toinen tallennusväline voi olla palvelin, ulkoinen kiintolevy, cd/dvd-levy tai nauha.

Yleensä varmuuskopioinnilla tarkoitetaan tietokoneen sisällä olevien tietojen kopiointia turvallisempaan ympäristöön talteen. Turvallisempi ympäristö voi olla palvelin, jolla on tarpeeksi levytilaa tai palvelinympäristö on pilvessä.

Varmuuskopiointi voidaan toteuttaa ajastettuna, jolloin käyttäjä ei huomaa tiedon varmuuskopiointia. Myös käyttöjärjestelmissä on varmuuskopiointisovelluksia, joilla tieto saadaan varmuuskopioitua.

Varmuuskopion voi tallentaa myös ulkoiselle kiintolevylle, mutta sen tietojen ylläpito ja hallinnointi yrityksen kannalta eivät olisi helppoa. Ulkoiset kiintolevyt voisivat olla missä vaan ja niiden tietoturvasta ei voitaisi pitää huolta.

2.1 Yleistä

Vuonna 2008 aloitettiin KDK (Kansallinen digitaalinen kirjasto) -hanke. KDK:sta tehtiin kokonaisarkkitehtuurikuvaus, joka määrittelee KDK:n käytännön toimintamallit ja ohjaa niiden toteuttamista. Kokonaisarkkitehtuuri määrittelee yhteisiä palveluita, tietosisältöjä, sovelluksia ja teknologiaa koskevat yhteen toimivuuden vaatimukset Kansalliseen digitaaliseen kirjastoon liittyville organisaatioille (Kansallinen digitaalinen kirjasto 2013).

Digitaaliarkisto liittyi KDK:n pitkäaikaissäilytyksen hankkeeseen, jonka seurauksena vanhoihin tif tiedostoihin on liitettävä standardin mukaiset metatiedot (Opetus- ja kulttuuriministeriö 2010b).

2.2 Kuvaformaatit

Yleisimpiä kuvaformaatteja ovat jpeg, tif ja raw kuvat. Tif ja raw kuvat on niin sanottuja häviöttömiä tiedostomuotoja, joissa ei ole muokattu kuvaa kuvan oton yhteydessä.

2.3 Tiff

Tiff (Tagged Image File Format) kuva mahdollistaa tallennuksen suurilla värimäärillä ja tarkkuuksilla. kuva sisältää sekä RGB ja CMYK värimallit (FinInk 2008).

RGB (Red, Green, Blue) on additiivinen eli lisäävä värijärjestelmä. Eli punaisen, vihreän ja sinisen värin summa on valkoinen. Värit ovat valkoisen valon aallonpituuksia. Valojen aallonpituuksia sekoitetaan yhteen ja niistä syntyy eri värit. Kun värejä on niin sanotusti liian paljon yhdessä, siitä syntyy musta väri.

Melkein kaikki käytettävissä olevat värit saadaan sekoittamalla kolmea pääväriä yhteen. Värisyvyyttä käytetään jokaisen värin kohdalla 0-255. Musta väri on 0,0,0 kun taas valkoinen on 255,255,255.

CMYK -värijärjestelmä (Cyan, Magenta, Yellow, Black) on subtraktiivinen, eli vähentävä värijärjestelmä. Käytetään yleisimmin painotuotteissa ja aineistoissa. CMYK värijärjestelmän värit ovat valittu kuvien painamiseen soveltuviksi.

Kun värijärjestelmän värejä painetaan paperille tai kankaalle, puhutaan yleensä neliväripainosta. Tällöin painetulle materiaalille painetaan jokainen väri erikseen, jolloin on mahdollista saada neljästä perusväristä miljoonia eri sävyjä. Painoa hallinnoidaan rasteripisteillä, jolloin koko ja tiheys vaikuttaa lopulliseen sävyyn.

Tif kuvia käytetään pääsääntöisesti jos kuvista tarvitaan tarkka tulostus tai kuvien värejä joudutaan vielä muuttamaan. Toisaalta tif kuvia harvemmin käytetään kotona kuvien tallentamiseen, koska tiedostot ovat suuria ja niiden siirtely ja käsittely on raskaampaa (FinInk, 2008).

2.4 Raw

Raw kuva on valmistajakohtainen pakkaamaton tiedostomuoto. Kuvan yhteydessä tallentuu kuvaan kaikki kameran asetukset, jotka vaikuttavat kuvaan. Nämä asetukset saa otettua jälkikäteen pois, jolloin kuva on aivan samanlainen kuin se on kennolle tulostunut. Kennon jokainen pikseli myös tallentuu, eli 12 megapikselin kuva tallentuu 12 megabitin raw tiedostoksi (Digital photography school 2006).

Raw kuvaa ei voi muokata, vaan siitä syntyy aina uusi kuvatiedosto. Myös kuvan katsominen vaatii erilaisen ohjelma.

2.5 Jpeg

Jpeg (Joint Photographic Experts Group) kuva on pakattu kuva, joka tarkoittaa että kuvaa on muokattu kuvan oton yhteydessä tai sen jälkeen. Jpeg kuvia käytetään Internet-sivuilla kuvien näyttämiseen. Jpeg formaatti ei ole suositeltu tapa pitkäaikaissäilytykseen, koska kuva on pakattu (DigiWiki, 2011).

2.6 Metatieto

Kuvassa on kahta erityyppistä metatietoa. Kuvaan liittyvää metatietoa IPTC ja kuvan tekniikkaa liittyvää metatietoa EXIF.

IPTC (International Press Telecommunications Council) on alun perin suunniteltu uutistoimistokäyttöön kuvien arkistointiin ja julkaisemista helpottavia metatietojen säilyttämiseen (Digiwiki, 2011). Kutsutaan myös kuvaan liittyväksi metatiedoksi.

Kuvaan liittyvää metatietoa voidaan muokata solukohtaisesti ja siihen voidaan lisätä myös ylimääräisiä soluja.

Exif (Exchangeable image file format) on kuva datan ja kuvan tekniikan kuvaava tallennusmuoto. Kuvan tekniikkaa kuvaava metatieto pitää sisällänsä resoluution, kameran tyypin ja muuta kuvaamisen yhteydessä saatuja tietoja. Kuvan teknistä metatietoa ei

voida solukohtaisesti muokata, mutta kaikki EXIF solutyypit voidaan pyyhkiä (Digiwiki 2011).

3 Välineitä

Yleisempiä tallennusvälineitä ovat CD- ja DVD- levyt ja nauhat, markkinoilla on myös Blue-ray levyjä. Kiintolevyjen hinnan lasku on mahdollistanut niiden käytön pitkäaikaissäilytykseen.

Optisia levyjä ei suositella pitkäaikaissäilytykseen, koska niitä ei ole suunniteltu siihen. Vielä ei ole tehty tutkimusta kuinka kauan tieto säilyy eheänä levyllä. (Tutkimusaineistojen tiedonhallinnan käsikirja 2012) Levyt ovat myös herkkiä naarmuille ja auringonvalolle. Levyn sisältö menee nopeasti pilalle jos se alttiina suoralle valolle (Digiwiki 2011).

3.1 Nauhat

LTO (Linear Tape-Open) on kehitetty 1998 Hewlett-Packardin toimesta. Nauhoista puhutaan nykyään LT nimikkeellä ja lopussa oleva numero kertoo kehittämisnumeron. Mitä suurempi numero, sitä suurempi kapasiteetti on nauhalla (LTO 2013).

Nauhuri lukee tiedon kiintolevyltä ja pakkaa sen 2:1 nauhalle. Nauhojen kapasiteetistä puhutaan aina pakattuna, mutta pakkauksen voi ottaa pois päältä tarvittaessa, jolloin nauhan kapasiteetti puolittuu.

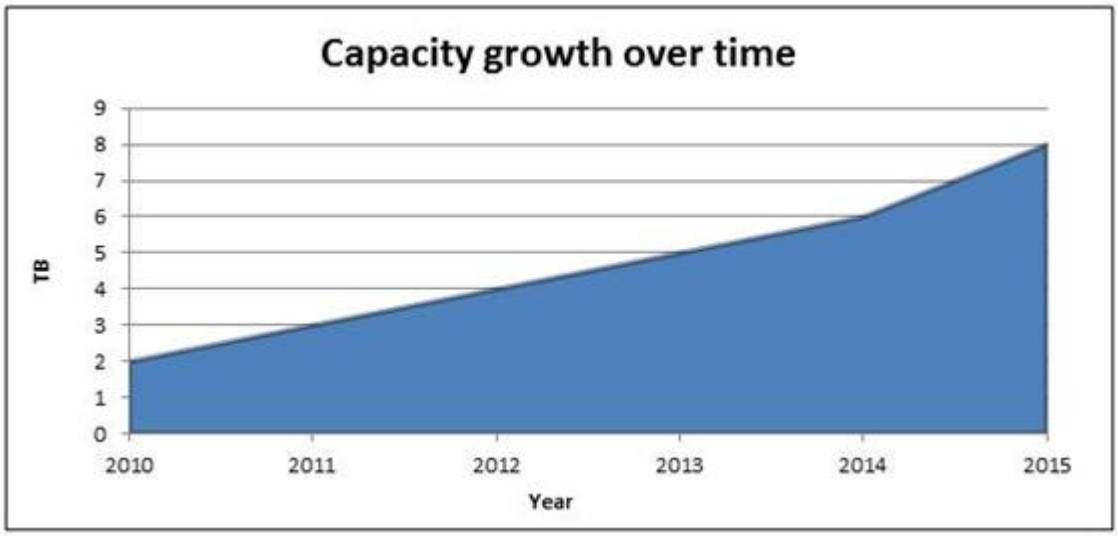
Markkinoille on tulossa LT-8 nauha ja sen kapasiteetti on pakattuna 32 teratavua. Uusien nauhojen pakkaussuhde on 2.5:1.

Nauhojen vahvuus on niiden hinta-kapasiteetti suhde verrattuna optisiin levyihin tai kiintolevyihin (Gleeson, P). Tämän hetken suurin nauhakoko on 6.25 teratavua, pakattuna (LTO 2013) ja nauhan hinta on 24 €.
Markkinoilla on myös nauhavaihtajia, joilla saadaan kasvatettua tallennuskapasiteettia beetatavuihin saakka. Nauhavaihtajia saadaan erikokoisina, mutta koon kasvaessa myös nauhansiirto lukijaan hidastuu.

Vaihtajan heikkous on sen hitaus. Nauhat vaihdetaan vaihtajan säilytyslokeroista mekaanisesti vaihtajan lukijaan ja vasta sen jälkeen kyseisen nauhan tiedot on käytettävissä. Työpaikallani mitattiin yli 10 sekunnin aika vaihtaessa nauhaa nauhalukijasta.

3.2 Kiintolevyt

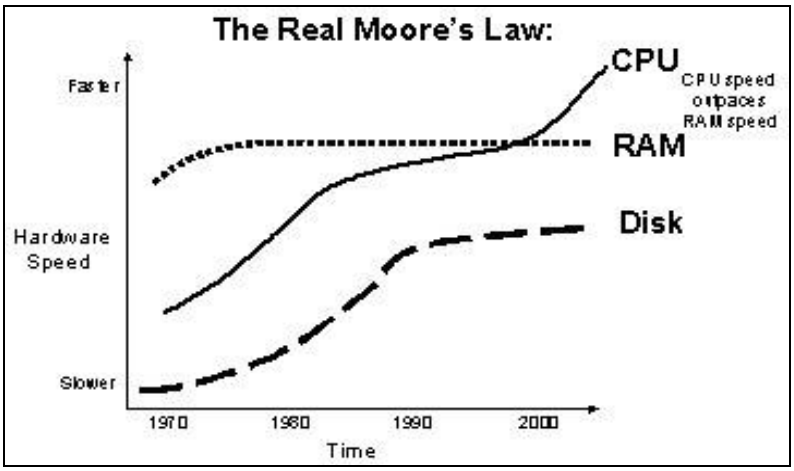
IBM kehitti kiintolevyn vuonna 1956 ja silloin siihen mahtui viisi megatavua tietoa ja painoi melkein 2 000 kiloa (Storage newsletter 2011).



Kuva 1. Kiintolevykapasiteetin kehitys (http://blogs.msdn.com/b/b8/archive/2011/11/29/enabling-large-disks-and-large-sectors-in-windows-8.aspx)

Kiintolevyjen hinta suhteessa tallennuskapasiteettiin pienenee kokoajan kehityksen mukana. Keskimäärin tallennuskoon kehitys on ollut 19 prosentin luokkaa (Kuva 1. Kiintolevykapasiteetin kehitys (http://blogs.msdn.com/b/b8/archive/2011/11/29/enabling-large-disks-and-large-sectors-in-windows-8.aspx)). Neliötuumalle vuonna 2011 mahtui 744 gigatavua. Samalla kehitysvauhdilla vuonna 2016 neliötuumalle pitäisi mahtua 1.8 teratavua, mutta tämä ei ole enää mahdollista samalla tekniikalla (Tom´s hardware 2012).

Vaikka kiintolevyn tallennuskoko on kasvussa, niin kiintolevyn käyttönopeus on pysynyt melkein samana jo useamman vuoden. Katso Kuva 2. Kiintolevynopeuden kehitys.(http://www.dba-oracle.com/oracle_tips_hardware_oracle_performance.htm)



Kuva 2. Kiintolevynopeuden kehitys.(http://www.dba-oracle.com/oracle_tips_hardware_oracle_performance.htm)

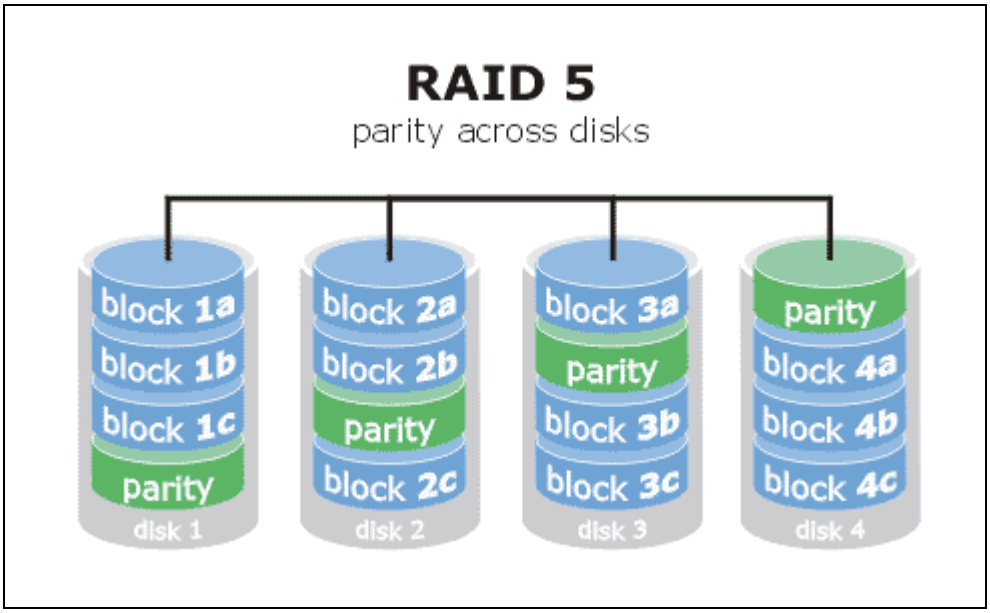
Tällä hetkellä markkinoiden isoin kiintolevy on 4 teratavua sata III väylässä ja Seagaten on tarkoitus tuoda vuonna 2014 markkinoille 5 teratavun levyn (Computerworld 2013).

Kiintolevyistä voidaan tehdä raid (Redundant Array of Independent Disks), jolloin useammasta pienemmästä kiintolevystä voidaan tehdä suurempi looginen asema ja/tai kasvattaa loogisen levyn virhesietoisuutta (Stinger web 2013).

Yleisimpiä raid tasoja on raid 0,0+1,10,1 ja 5. Raid 0 tarkoitus on vain jatkaa levyjä keskenään ja siinä ei ole vikasietoisuutta. Raid 1 on peilaava ja kaikki kiintolevyn tieto kirjoitetaan kahteen kertaan.

Raid 5 on suosituin taso ja siinä on hyvä vikasietoisuus. Tiedot jaetaan kaikille raidin levyille ja tämä antaa loogiselle levylle nopeutta. Kyseisestä raidista voi hajota yksi levy ilman että tiedot katoavat (Petri Nuutinen).

Kun looginen levy on raid levyjärjestelmän päällä, niin kyseisen kapasiteetin luku- ja kirjoitusnopeus kasvaa. Tietoa voidaan lukea ja kirjoittaa useammalle levylle yhtäaikaisesti.



Kuva 3. Raid 5

Tallennusväline	Hinta	tallennusikä	Koko	Nopeus
LT-4	23 €	30 vuotta	1.6 teratavua	n. 220 Mt/s
LT-6	Ei tietoa	30 vuotta	6.25 teratavua	n. 1.1 Gt/s
Kiintolevy (sata 3)	103 €	30 vuotta	2 teratavua	6GT/s
Kiintolevy (san)	420 € X raidin koko	Teoriassa loputon	2 teratavua	6GT/s

Taulukko 1. Hinta/kokosuhde. Lähteet: (Tape and media 2013, verkkokauppa.com 2013, Nextag 2013 ja IBM 2013)

Hintoja vertaillessa joutuu ottamaan myös kantaa virrankulutukseen. Kaikki levypakkaan liitetyt kiintolevyt ovat kokoajan päällä ja ne vievät siis virtaa vaikka niitä ei luettaisi. Nauhavaihtaja käyttää vain vähäisen määrän virtaa sen ollessa odotustilassa. Vasta kun halutaan jokin tiedosto joltain nauhalta, joutuu nauhuri käyttämään enemmän virtaa liikutellessaan nauhaa ja lukiessa siitä tietoa.

Työpaikallani seurataan laitteiden virrankulutusta ja huomasimme että nauha-aseman virran kulutus on pienempi kuin vastaavan verran kapasiteettiä olevan levypakan.

3.3 PHP

Kieli on kehitetty vuonna 1994 C -kielen pohjalta ja vastaa hyvin paljon Perl -ohjelmointikieltä. Kielen ensimmäisen ohjelman tarkoitus oli seurata kehittäjän ansioluettelon lukijoita (Gilmore 2005).

Kielen alkuperäinen tarkoitus oli tehdä pelkästään Internet-sivuja (PHP 2013). ja siihen liitettiin myös funktioita tietokannan käyttöön. Alun perin PHP ja MySQL relaatiotietokanta ovat kulkeneet yhdessä, mutta vuonna 2003 PHP:hen lisättiin oma tietokantamoottori (PHP 2013).

PHP -ohjelmointikieltä käytetään pääsääntöisesti dynaamisten Internet-sivujen tekemiseen. Dynaamisen sivuston sisältö latautuu vasta kun sivun käyttäjä on avaamassa sivua. Staattinen sivusto on dynaamisen sivun vastakohta. Staattisissa sivustoissa sivut ovat valmiita palvelimella ja sivustot näyttävät aina samanlaisilta sivun käyttäjästä riippumatta (Ohjelmointi 2009).

Sillä voidaan myös tehdä komentoja erilaisiin palvelinympäristöihin (PHP 2013). Esimerkiksi Facebook ja MediaWiki sivustojen pohja on tehty PHP:llä.

Dynaaminen sivu on ainoa ratkaisu jos halutaan sivuille muuttuvia tietoja käyttäjäkohtaisesti. Joissain tapauksissa sivut ovat tallennettu relaatiotietokantaan ja sivuja selatessa sivujen sisältö haetaan tietokannasta. Hyvänä esimerkkinä on kaikki Wordpress sivustot.

Facebookin tapauksessa on tehty muutama pohja, joka vaihtuu riippuen millä laitteella sivustoa käytetään. Kirjautuessa sivulle, sivusto hakee käyttäjän tiedot ja kaverit relaatiotietokannasta ja rakentaa niiden perusteella käyttäjän oman näkymän.

MediaWiki taas toimii vähän eri tavalla. Jokainen käyttäjä näkee melkein samanlaisena kaikki sivut, mutta vain korkeamman tason käyttäjät voivat muokata yksittäisten sivustojen näkyvyyttä. Wikipedia on rakennettu MediaWiki sovelluksen päälle.

Wikipediassa on jokainen sivu tallennettu relaatiotietokantaan ja kun haetaan sivustolta jotain, tulee haetun asian tiedot tietokannasta ja normaali käyttäjä ei edes huomaa että sivu onkin dynaamisesti tehty.

3.4 Relaatiotietokanta

Kaikki relaatiotietokannat perustuvat IBM:n tutkijan E. F. Coddin vuonna 1970 julkaisemaan relaatiomalliin. Relaatiomalli määrittelee tietokantojen pohjan ja se perustuu matematiikkaan ja predikaattilogiikkaan (Ari Hovi 2004).

Tietokantojen sisällä on tauluja ja niiden sisällä on yksittäisiä sarakkeita. Jokaisessa taulussa on tunnisteena perusavain (primary key). Perusavain on oltava yksilöivä, joka estää että tauluun ei voi tulla kahta samanlaista tietoa.

Yleensä PHP kielen kanssa käytetään jotain relaatiotietokantaa, jossa säilytetään muuttuvia tietoja joita sovelluksella käytetään. Tällaisia tietoja voisi olla vaikka käyttäjien yhteystiedot tai blokin otsikot ja sisällöt.

Yleisin vapaanlähdekoodin relaatiotietokanta on MySQL (Ohjelmointiputka 2011). Sen rakensi suomalainen ja ruotsalainen yhdessä vuonna 1995 ja ensimmäinen versio tuli markkinoille vuonna 1996 (Dries Buytaert 2010). Sovellusta on ladattu jo yli 100 miljoonaa kappaletta (MySQL 2014).

Kun käytetään MySQL tietokantoja, voidaan valita mitä tietokantamoottoria käytetään. Oletuksena on ennen ollut MyISAM (Indexed Sequential Access Method), mutta version 5.5 jälkeen InnoDB tietokantamoottori on tullut oletukseksi (MySQL 2014).

MyISAM tietokantamoottorin huonouksia on kun tauluun lisätään, poistetaan tai päivitetään tietoa, niin koko taulu pitää lukita hetkellisesti. Myös jos tietokanta sammuu odottamattomasti, niin kannan palautus voi olla hankalaa. Nämä asiat on korjattu InnoDB tietokantamoottoriin (Craig Buckler 2014).

InnoDB tietokantamoottoriin on lisätty myös takaisin vieritys toiminto (rollback). Jos tauluun vietäessä tietoja tulee jotain odottamatonta ongelmia, niin tietokantamoottori osaa peruuttaa viennin ja tietokannassa ei näy edes koko tapahtumaa. InnoDB:ssä on myös pakolliset viiteavaimet joita voidaan käyttää toisissa tauluissa viiteavaimena (Craig Bucker 2014).

InnoDB moottoria pidetään parempana jos tulee samaan tauluun paljon rivien muokkaamista. (Craig Bucker 2014). Kun joutuu indeksoimaan kokonaisia tekstejä, niin silloin MyISAM tietokantamoottori on parempi vaihtoehto (Tony Stark 2013).

3.5 MD5

MD5 on tiivistealgoritmi joka lasketaan tiedostosta tai tekstistä. Kun tiedoston tai tekstin sisällöstä muuttuu yksikin merkki, niin algoritmi on erilainen. Esimerkiksi salasanaa ei tallenneta suoraan vaan siitä lasketaan md5-summa ja tämä sitten tallennetaan halutulla tavalla.

MD5 algoritmi tuottaa 128 bittisen tiivisteen, joka näytetään 32 merkkisenä heksakoodatussa muodossa. (Go hacking 2010) Esimerkiksi sanan digi MD5 tiiviste on: b445df4a839d272450bb37cfc2e440d4 ja sanan Digi MD5 tiiviste on: e5b26caca0aacee198fd5588d1ced6c6 (MD5 Generator 2014).

Tarkoituksena on että tiivisteestä ei pitäisi kyetä päättelemään mitään sille annetusta tiedosta ja tiivisteestä ei voida laskea sillä salattua sanaa tai tiedostoa, jonka takia sitä kutsutaan yhdensuuntaiseksi hajakoodausalgoritmiksi.

Myös pankit käyttävät MD5 algoritmia hyväkseen, kun tarkistavat asiakkaiden verkkomaksamisia. Kun asiakas hyväksyy maksun, lähtee maksun mukana pankille tieto mistä ollaan maksamassa ja erinäiset tiedot salattuna MD5 summalla. Kun verkkomaksamisen MD5 summa on sama kuin pankin laskema MD5 summa tietyistä asioista, niin hyväksytään maksu.

4 Riskit

Riskejä on monia, koska migraatio ympäristö on osa tuotanto ympäristöä. Tuotantopuolen verkkoyhteydet ja kiintolevyjen kirjoitusnopeudet vaikuttavat migraatioon. Migraatio käyttää tuotantopuolen levypakkaa tiedostojen välipaikkana, kun tif kuviin liitetään uusia metatietoja. Migraation yhteydessä tarkistetaan myös toisen palvelimen tietokannasta tietoja ja vertaillaan niitä nauhalla oleviin tietoihin.

Toisen palvelimen toiminta vaikuttaa myös siirtoihin ja palvelimen hitaus voi haitata migraatiota. Kyseinen palvelin on kaiken perusta Digitaaliarkistossa ja tästä syystä palvelimelle tulee paljon kyselyitä muiltakin palvelimilta ja on kovassa käytössä.

Uusille nauhoille tallentaessa on omat riskinsä. Nauhakirjoittajan rikkoutuessa nauha voi venyä, jolloin suuri osa tai kaikki kyseisen nauhan tiedoista korruptoituu. Jos uusia nauhoja joutuu siirtämään muualle Digitaaliarkistosta, on vaarana nauhan sisällön kato. Nauha sisältää magneettinauhaa, jolloin nauhan sisältö voi kadota nauhan koskettaessa magneettikenttää.

Vanhalta nauhalta lukiessa tietoa kiintolevylle voi lukemisen aikana tulla lukuvirheitä jos nauha on kyseisestä kohdasta likaantunut. Tällöin joko kyseinen tieto on korruptoitunut tai pahimmassa tapauksessa kaikki tieto nauhalta siitä eteenpäin ovat korruptoituneet.

Vanhan nauhan kopioinnin yhteydessä voi kiintolevyllä tapahtua kirjoitusvirheitä jolloin tiedosto korruptoituu kiintolevyllä vaikka nauhan luku olisi onnistunut. Myös kiintolevyjä voi rikkoontua kopioinnin aikana, jolloin sen aikaiset siirrot voivat kadota.

Digitaaliarkistossa ei ole kuin yksi vanhoja SDLT nauhoja lukeva laite. Jos kyseinen laite menee rikki, niin tämän jälkeen migraatiota ei voida enää jatkaa.

5 Kuvien kopiointi ja metatietojen liittäminen

Digitointi aloitettiin vuoden 2003 loppupuolella kuvaamalla kirkonkirjoja. Ensimmäisenä neljänä vuonna asiakirjoja saatiin digitoitua vain noin 2 miljoonaa kappaletta. Vuonna 2009 aloitettiin digitointi hanke joka kesti kaksi vuotta ja sinä aikana digitointiin yli 6 miljoonaa asiakirjaa. Tällä hetkellä digitoituja asiakirjoja on noin 19 miljoonaa.

Aikanaan digitaaliarkisto on lähtenyt käyntiin pienellä budjetilla, jonka takia kaikissa siihen liittyvissä valinnoissa on haettu joko ilmaista tai vapaanlähdekoodin tyyppisiä ratkaisuja. Palvelinympäristönä on Linux Centos ja relaatiotietokantana on MySQL. Myös migraatioprojektissa aiotaan käyttää samankaltaisia ratkaisuja.

Digitaaliarkisto valitsi nauhan pitkäaikaissäilytykseen kiintolevyjen sijasta, kun digitoinnin aloittaessa se oli huomattavasti halvempi vaihtoehto. Ensimmäisten skannereiden tallennusmuoto oli pelkästään tiff, jonka seurauksena pitkäaikaissäilytysmuodoksi se myös valittiin.

Digitaaliarkisto vaihtoi nauhatyypin vuonna 2010 ja on käyttänyt nauhatallennusta vuodesta 2003. Työn aiheena oli saada kopioitua vuoden 2003–2010 vanhat nauhat uusille nauhoille liittäen tiedostoihin samalla KDK:n yhdenmukaiset metatiedot.

Kaikissa vanhoissa tiff tiedostoissa ei ole IPTC metatietoja kiinnitetty, jotka kertovat kuvan paikasta (SanastoWiki 2011). Migraation yhteydessä kiinnitetään myös tiff tiedostoihin uudet metatiedot, jos niitä ei vielä ollut tiff tiedostoissa.

Migraatio tarkoittaa tietojen siirtoa mediasta toiseen. Yleensä migraatiota ruvetaan tekemään ennen kuin kyseisen median teoreettinen maksimi-ikä on saavutettu, ettei tietoa katoa bittitasolla. Usein on myös järkevämpää siirtyä uudemmalle medialle, koska vanhan median saatavuus vaikeutuu median vanhetessa (Migraatio 2011).

5.1 Digitaaliarkisto ja projektisuunnitelma

Aluksi otettiin selvää kaikista riskeistä mitä voi tapahtua migraation yhteydessä ja tehtiin suunnitelma miten voitaisiin estää mahdolliset ongelmat.

Varsinaista projektisuunnitelmaa ei ollut, vaan mietittiin mahdollisia riskejä ja niiden hallintaa. Sovelluksen kehitys oli myös siinä mielessä haasteellinen, kun minulla ei ollut kunnon tuntemusta digitaaliarkiston toiminnasta.

Tarvitseeko migraation rajaamaan digitaaliarkiston tuotantopuolesta vai tarvitaanko tuotantopuolelta muita palvelimia hoitamaan tiedostojen siirtelyä tai mahdollista laskentatehoa?

Migraatio käyttää kahta nauhuria ja on yhteydessä myös toiseen palvelimeen. Voiko toisen palvelimen tietokannasta ottaa tarvittavat tiedot migraatiopalvelimelle, jotta ylimääräisiä tietokantakyselyitä ei tarvitsisi tehdä?

Koska vanhoja nauhoja on enemmän kuin nauhavaihtajaa mahtuu, niitä joutuu myös fyysisesti vaihtamaan. Kuka hoitaa vanhojen nauhojen vaihdon?

Tarkistetaan tarvitaanko vanhoilta nauhoilta palauttaa kaikki tiedot vai onko siellä tietoa, mitä emme enää tarvitse. Esimerkiksi vanhat tietokantavarmuuskopiot ja testitiedostot ovat myös nauhoilla.

Migraatio sovelluksesta pitäisi saada mahdollisimman automaattinen, koska digitaaliarkistossa henkilö vähyyden vuoksi ei ole ylimääräisiä henkilöitä katsomaan migraation perään. Vikatilanteissa lähetetään sähköpostia ylläpidolle, joka voi korjata mahdollisen ongelman. Sähköpostissa pitää selvitä mistä vika johtuu ja mahdollisesti miten vian voi korjata.

Migraatiotyökalun hallinta halutaan mahdollisimman helpoksi. Paras vaihtoehto tähän olisi Internet sivusto, jossa voi seurata missä vaiheessa migraatio menee. Tehdäänkö itse sivusto kokonaan vai haetaanko siihen jokin valmis ratkaisu?

5.2 Sovelluksen pohjatyötä

Migraatiota varten tehtiin kokonaan uusi palvelin, jonka päätyönä on migraation pyörittäminen. Palvelimelle myös laitettiin tietokanta- ja HTTP palvelu. Palvelin joutuu olemaan vähän iäkkäämpi, koska palvelimella pitää olla SCSI tuki ja sitä ei ole valmiina meidän uusimmilla palvelimilla. Palvelimen levyille haluttiin myös vikasietoisuutta, jonka takia palvelimelle laitettiin oma levypakka hoitamaan palvelimen fyysiset levyt.

SCSI (Small Computer System Interface) on vanhempi standardi tiedon välittämiseen oheislaitteen ja tietokoneen välillä. Uudemmat laitteet eivät käytä tätä standardia ja sen takia SCSI kortteja tietokoneisiin on hankala löytää.

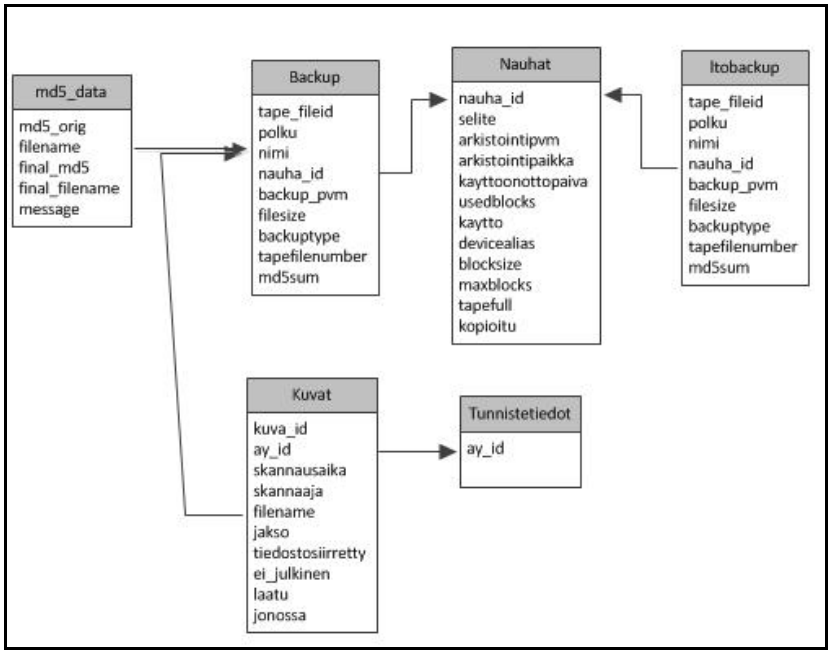
Sovelluskielenä käytetään PHP ohjelmointikieltä ja kieli toimii myös palvelinpohjaisena ohjelmointina. Sovellus toimii komentokehotteella ja sovelluksen tilaa voi seurata myös Internet sivustolla.

Koska ohjelmointikielenä käytetään PHP:tä, niin päädyimme myös käyttämään MySQL relaatiotietokantaa migraatiosovelluksen pohjana.

Migraatiotyökalulle tehtiin myös oma Internet sivusto jossa voidaan syöttää uusien nauhojen tunnukset ja seurata missä vaiheessa migraatio menee.

Hallintasivustoa käytetään usealla eri laitteella, jolloin tyylitiedoston pitää osata skaalautua käytettävän laitteen ruudun mukaan. Rajallisen ajan takia haimme tähän vapaanlähdekoodin sovelluksia ja otimme käyttöön Bootstrap tyylipohjan.

Myös oma relaatiotietokanta tehtiin palvelimelle, jotta saadaan vähennettyä liikennettä palvelimien välillä. Relaatiotietokantaan kopioitiin toisen palvelimen tietokannasta muutamia tauluja, jotka olivat migraation kannalta pakollisia.



Kaava 1. Tietokannan luokkakaavio

Toisen palvelimen tietokannasta kopioitiin kuvat, nauhat, tunnistetiedot ja ltobackup taulut. Sovellusta varten tehtiin taulut md5_data ja backup.

5.3 Menetelmän esittely

Migraatio toteutettiin tekemällä useampi pienempi sovellus jotka keskustelevat keskenään. Sovellukset ajastetaan palvelinpuolella, jolloin koko migraatio on automaattinen.

Koska tiedostoja ei voida silmämääräisesti tarkistella, pitää tiedostoista ottaa md5 summa ja laittaa se talteen relaatiotietokantaan.

Sovellus alkaa tarkistamalla migraatiopalvelimen tietokannasta onko vanhan nauhavaihtajan lukijassa oleva nauha jo kopioitu ja mihin kategoriaan se kuuluu. Kategorioita on kolme ja nauhat kopioidaan kategoria kerrallaan. Ehtojen täyttyessä tarkistetaan vielä että kiintolevyllä on tarpeeksi tilaa nauhan sisällön kopiointiin.

Jokainen kopioitu tiedosto tarkistetaan md5 summalla että tiedosto on sama kuin se oli nauhalle tallentaessa. Jos tiedoston md5 summa vastaa samaa kuin toisen palvelimen tietokannassa oleva md5 summa, niin merkitään migraatiopalvelimen omalle tietokan-

nalle että kyseinen tiedosto on onnistuneesti kopioitu. Kun nauha on kopioitu, merkitään se kopioiduksi myös tietokantaan jotta ei kopioitaisi samaa nauhaa uudestaan.

Jos tiedoston kopiointi ei onnistu nauhalta, yritetään sitä kopioida uudestaan toisen kerran. Kun on samaa tiedostoa yritetty kopioida maksimissaan viisi kertaan, siirretään tullut tiedosto ongelma hakemistoon.

Ongelmahakemistosta voi tarkistella käsin onko tiedostoissa jotain ongelmaa, täyttävätkö tiedostot KDK:n määrittelemät metatiedot. Kun tiedosto on tarkistettu käsin, se voidaan tallentaa uudelle nauhalle talteen. Ongelma hakemistoon päätyvät myös onnistuneesti kopioidut tiedostot, mutta kopioidun tiedoston md5 summa ei vastannut tietokannassa olevaan saman tiedoston md5 summaan.

Migraatio palvelin on aika iäkäs, niin joudutaan rajoittamaan sovelluksien yhtäaikaista ajoa. Kopioinnin jälkeen estetään seuraavan nauhan kopiointi ennen kuin jo kopioituihin kuviin on lisätty tarvittavat metatiedot. Metatietojen jälkeen kirjoitetaan tiedostot uusille nauhoille, jonka valmistuttua sovellus kertoo palvelimelle että voi kopioida seuraavan nauhan.

Kun nauhan sisältö on kokonaisuudessaan kopioitu kiintolevylle niin sanottuun nauhan omaan hakemistoon, niin seuraava sovellus tarkistaa tiedosto kerrallaan onko metatiedot tarvittavat vai joutuuko tiedostoon lisätä jotain. Kun tiedostoon joutuu lisäämään metatietoja, muuttuu myös kyseisen tiedoston md5 summa.

Metatietojen uudelleen kirjoittamisen jälkeen tiedostoon muuttuvat md5 summat joudutaan ottamaan talteen ja ne lisättiin myös migraatiopalvelimen tietokantaan. Koska md5 summa muuttuu metatietojen liittäessä, jouduttiin turvautumaan toiseen tapaan tarkistella että tiedosto säilyisi eheänä.

Tiedoston koko tarkistetaan ennen ja jälkeen metatiedon kirjoittamisen tiedostoon. Jos tiedoston koko muuttui merkittävästi, siirretään kuva hakemistoon josta sitä tarkisteltiin käsin. Muutoin tiedosto siirretään ”TO-TAPE” hakemistoon, josta se on helpompi

kopioida uudelle nauhalle. Kun nauhan oma hakemisto on tarkistettu kokonaan, niin siitä lähtee tieto seuraavalle sovellukselle.

Uudelle nauhalle kirjoitetaan vasta kun metakirjoitus sovellus antaa tiedon että kaikki nauha hakemiston tiedostot on tarkistettu. Sovellus joka kirjoittaa uudelle nauhalle vanhan nauhan tiedostot, käyttää kahden palvelimen tietokantoja. Vanhasta tietokannasta kysytään metatietojen liittämisen jälkeen olevaa md5 summaa ja toiseen tietokantaan kirjoitetaan tiedoston täydelliset tiedot, missä kohtaan tiedosto sijaitsee ja millä nauhalla.

Kun metatietoja sisältävät tiedostot on siirretty uudelle nauhalle ja ”TO-TAPE” hakemisto on tyhjä, niin voidaan tarkistella että tiedostot ovat eheänä uudella nauhalla. Uusi nauha siirretään fyysisesti toiselle nauha-asemalle ja uudella nauha-asemalla luetaan kaikki tiedostot kyseiseltä nauhalta. Jos luetun tiedoston md5 summa ei vastaa samaa kuin se oli toisen palvelimen ottama, niin kyseinen vanha nauha kopioidaan uudestaan.

Migraatiosovellus oli osa harjoitustyötä, johon kuului myös muidenkin sovelluksien tekemistä. Varsinainen uusien nauhan lukeminen toisella palvelimella jäi vähäiseksi ajan loputtuani ja Digitaaliarkiston henkilöiden kiireiden takia.

Tästä syystä keksittiin nopeasti tarkistus aina uuden nauhan kopioinnin yhteydessä, jotta nauhaa ei tarvitsisi siirtää. Eli kun nauhalle kirjoitetaan uutta tietoa ja ”TO-TAPE” hakemisto on tyhjä, kopioidaan kaikki juuri kirjoitetut tiedostot takaisin ja tarkistetaan että md5 summa vastaa samaa kuin se oli nauhalle kirjoittaessa.

Jotta sovelluksen tilannetta on helppo seurata, tehtiin myös loki jokaisesta tapahtumasta mitä sovellukset tekivät. Tätä lokia pystyy seuraamaan migraatiosovellusta varten tehdyllä Internet sivulla.

Kuva 4. Migraatiotyökalun Internet sivusto

Kuva 4. Migraatiotyökalun Internet sivusto nähdään miten parhaillaan olevaa lokia käytetään hyväksi Internet sivustolla. Samaisella sivustolla on myös ohjeet miten vaihdetaan vanhat nauhat nauhavaihtajassa ja mikä sovellus on parhaillaan käynnissä migraatio palvelimella.

Internet sivulla annetaan myös uuden nauhan ID tunnus ja seurataan onko kokonainen kategoria kopioitu, jotta vaihdetaan kategoriaa. Kategorian vaihto tapahtuu komentokehotteen päällä, koska haluttiin estää mahdolliset väärät vaihdot.

Vanhoja nauhoja on 314 kappaletta ja vanhaan nauhavaihtajaan mahtuu vain 24 kappaletta. Jouduttiin tekemään sovellukseen toiminto joka lähettää sähköpostia kun kaikki nauhurissa olevat nauhat on kopioitu. Sähköposti kertoo mitä kategoriaa ollaan kopioimassa ja paljon on nauhoja kopioimatta kyseisestä kategoriasta.

Uudella nauhakirjoittajalla ei ole nauhavaihtajaa, eli nauha pitää vaihtaa kerran viikossa. Tämäkin lähettää sähköpostia kun nauha on täynnä ja sen näkee myös sovelluksen Internetsivuilla.

6 Yhteenveto ja johtopäätökset

Tätä sovellusta tehdessäni PHP ohjelmointi ei ollut minulle kovin tuttua. Olin koulussa opiskellut ohjelmointikieltä pintapuolisesti, mutta projektissa tuli vastaan haasteita joita en osannut ratkoa koulussa opitun perusteella.

Tutkimme riskejä ennen projektin tekemistä ja olimme varautuneet melkein kaikkeen. Tiesimme että levypakka ei kestä loputtomiin, eikä uuden nauhurin vaihto tuota ongelmia.

Uskoimme että vanha nauhavaihtaja kestäisi migraation tai ostamme uuden tilalle jos siihen tulee hälyttäviä ongelmia. Emme kuitenkaan ottaneet huomioon että vanhan kaltaistamme vaihtajaa ei löydy.

Työharjoittelun aikana syntynyt sovellus on osoittautunut toimivaksi ja sitä ei ole tarvinnut paljoa muuttaa alkulähtökohdasta. Sovellusta tehdessä otimme kantaa mahdollisiin laitevaihtoihin. Kun ongelmat koskivat laitteita, olivat ne hyvin helppo vaihtaa myös koodiin.

Yrityksessämme jätämme palvelimen migraation jälkeen odottamaan uutta migraatiota. Jos emme ulkoista uusia LTO-4 nauhojen säilytystä, niin jossain kohtaa niidenkin tiedot joudumme siirtämään uudemmalle nauhajärjestelmälle.

6.1 Tulokset

Keskimäärin yhdellä nauhalla on noin 3000 tiedostoa. Yhden tiedoston kopioiminen kestää noin 8 sekuntia ja md5 summan lasku sekunnin. Tässä tapauksessa puhutaan todella pitkästä projektista joka pitää toimia mahdollisimman automaattisesti.

Sovellus käynnistyi kesällä 2012 ja tällä hetkellä on kopioitu 143 nauhaa. Sovellus toimi kunnolla ja sähköpostilähetys kertoivat virhetilanteissa tarkasti mistä syystä migraatio keskeytyi. Virhe oli helppo korjata sähköpostiviestin ansiosta.

Ensimmäinen kategoria saatiin kopioitua ilman ongelmia, mutta toisen kategorian kohdalla rupesi tulemaan laiteongelmia.

Ensimmäisenä meni rikki sovellukselle tarkoitettu levypakan yksi kiintolevy. Levy toimi raid 5 järjestelmällä, eli tietoa ei tässä vaiheessa kadonnut. Kun raidista menee rikki levyjä, niin raid järjestelmä tekee loogisesta levystä kirjoitussuojatun. Eli levyn rikkouduttua, emme voineet enää jatkaa vanhojen nauhojen kirjoittamista kyseiselle levypakalle.

Vanhan levypakan kiintolevyinä toimi 300Gbit SCSI väylää käyttävä vanhempi kiintolevy. Näytä levyjä ei saatu hommattua enää kohtuulliseen hintaan, jonka takia vaihdoimme levypakkaa.

Digitaaliarkistossa on noin 68 teratavua kiintolevyä käytössä, joten otimme tuotantopuolen levypakasta osion migraatiopalvelimelle.

Tämän korjauksen jälkeen migraatio toimi moitteettomasti noin kuukauden kunnes tuli seuraava ongelma. Migraatiopalvelimen uudelle nauhalle tarkoitettu nauhakirjoitin meni rikki.

Koska käytämme jo tuotantopuolen levypakkaa migraation yhteydessä, oli helpompi ottaa käyttöön myös tuotantopuolen nauhakirjoittajat. Kirjoittimen vaihto myös helpotti migraation suoritusta, koska levykirjoittimet ovat osa uutta levypakkaa. Enää ei tarvitse vaihtaa erikseen uusia nauhoja, vaan uusi nauhavaihtaja tekee sen puolestamme.

Kun koodia oli vaihdettu toimimaan uudella tavalla, niin vanha nauhavaihtaja meni rikki. Tällä hetkellä odotamme löydämmekö jostain toista nauhavaihtajaa vanhan tilalle ja voimme jatkaa migraation tekemistä.

Migraatio on nyt keskeytetty digitaaliarkistossa ja se oli kestänyt ennen keskeytystä noin vuoden. Ennen keskeytystä oli saatu kopioitua noin 60 % kaikista vanhoista nauhoista. Keskeytyksen syy oli yksi riskeissä esitetyistä syistä. Vanha nauhalukija ei enää toimi luetettavalla tavalla ja tällä hetkellä ollaan etsimässä uutta lukijaa.

6.2 Jatkokehitys

Sovellus on tehty helposti muokattavaksi ulkoisten laitteiden osalta. Pienellä vaivalla siitä pystyisi tekemään sovelluksen, jota myös muutkin yritykset voisivat käyttää vastaavanlaiseen migraatioon. Jossain vaiheessa tallennusmuodosta tulee vanha ja sen sisältöä halutaan migratoida uudemmalle tallennuskapasiteetille.

Sovelluksen teon jälkeen tehtiin kaikille digitaaliarkiston sovelluksille samanlainen käyttöliittymä. Yhden sivuston kautta voidaan nykyään hallinnoida ja digitoida tiedostoja digitaaliarkistossa.

Migraatiotyökalusta tehtiin vielä erilainen versio, jolla voidaan siirtää muiden digitoituja tiedostoja digitaaliarkiston ympäristöön. Sovellus kopioi tiedoston ulkoiselta levyltä digitaaliarkistoon tiedoston, tarkistaa onko tiedostolla tarvittavat metatiedot ja lisää ne tarvittaessa. Kopioinnin jälkeen tiedoston MD5 summa ja julkisuustiedot tarkistetaan. Jos kaikki on kunnossa, tiedosto päätyy digitaaliarkistoon asiakkaiden nähtäväksi.

Lähteet

Ari Hovi 2004. SQL Opas. Jyväskylä

Bootstrap 2014. Ota Bootstrap (Get Bootstrap). Luettavissa: http://getbootstrap.com/. Luettu 24.4.2014

Craig Buckler 2014. Hyvät ja huonot puolet MyISAM moottorista (MySQL: the Pros and Cons of MyISAM Tables). Luettavissa: http://www.sitepoint.com/mysql-myisam-table-pros-con/ Luettu: 17.5.2014

Craig Buckler 2014. Hyvät ja huonot puolet InnoDB moottorista (MySQL: the Pros and Cons of InnoDB Tables). Luettavissa: http://www.sitepoint.com/mysql-innodb-table-pros-cons/ Luettu: 17.5.2014

Computerworld 2013. Seagate tuo ensivuonna markkinoille 5teratavun levyn (Seagate to produce 5TB hard drive next year, 20TB by 2020). Luettavissa: http://www.computerworld.com/s/article/9242268/Seagate_to_produce_5TB_hard_drive_next_year_20TB_by_2020. Luettu 9.11.2013

DigiWiki 2011. IPTC. luettavissa: http://www.digiwiki.fi/fi/index.php?title=IPTC. Luettu 20.10.2013

DigiWiki 2011. JPEG. Luettavissa: http://www.digiwiki.fi/fi/index.php?title=JPEG. Luettu 8.10.2013

Digiwiki 2011. Kuvien metadata. Luettavissa: http://www.digiwiki.fi/fi/index.php?title=Metadata. Luettu 8.10.2013

Digiwiki 2012. Optinen media. Luettavissa: http://www.digiwiki.fi/fi/index.php?title=Pitk%C3%A4aikaiss%C3%A4ilytys#Optinen_media. Luettu 9.11.2013

Digital photography school 2006. Raw vastaan jpeg (RAW vs. Jpeg). Luettavissa: http://digital-photography-school.com/raw-vs-jpeg. Luettu 20.10.2013

Dries Buytaert 2010. MySQL:n historia (The history of MySQL). Luettavissa: http://buytaert.net/the-history-of-mysql-ab
Luettu: 24.4.2014

FinInk 2008. Tietoa kuvatyypeistä. Luettavissa: https://www.finink.com/doc/Tietoa_kuvatiedostoista.ashx. Luettu 20.10.2013

Gilmore, J. 2005. PHP & MySQL 5. Gummerus Kirjapaino Oy. Jyväskylä

Gleeson, P. eHow tech. Paljon DLT-Nauhan käyttöikä? (What is the life of dlt tapes?) Luettavissa:
http://www.ehow.com/about_6590493_life-dlt-tapes_.html. Luettu 6.2 2013.

Go Hacking 2010. Mikä MD5 summa on (What is MD5 hash) Luettavissa: http://www.gohacking.com/what-is-md5-hash/ Luettu: 24.4.2014

IBM 2013. TS2340 tiedot. (TS2340 Tape Drive Express Model) Luettavissa: http://www-03.ibm.com/systems/storage/tape/ts2340/index.html. Luettu 17.3.2013

Kansallinen digitaalinen kirjasto 2013. Tietoa KDK -hankkeesta. Luettavissa: http://www.kdk.fi/fi/tietoa-hankkeesta. Luettu 8.10.2013.

LTO, 2013. LTO kehittäminen (LTO Ultrium Generations) Luettavissa: http://www.lto.org/technology/generations.html. Luettu 13.10.2013.

MD5 2014.MD5 Generaattori (MD5 Generator) Luettavissa: http://www.adamek.biz/md5-generator.php Luettu: 24.4.2014

Migraatio 2011. Migraatio Luettavissa: http://www.digiwiki.fi/fi/index.php?title=Migraatio. Luettu: 10.2.2014

MySQL 2014. About. Luettavissa: http://www.mysql.com/about/ Luettu: 24.4.2014

MySQL 2014. InnoDB. Luettavissa: http://dev.mysql.com/doc/refman/5.5/en/innodb-introduction.html Luettu: 17.5.2014

Nextag 2013. 2TB SAS kiintolevyt. (2TB sas hard drive) Luettavissa: http://www.nextag.com/2tb-sas-hard-drive/products-html Luettu: 17.3.2013

Ohjelmointi 2009. PHP & MySQL perusteet Luettavissa: http://www.php-perusteet.com/. Luettu: 24.4.2014

Ohjelmointiputka 2011. Osa12 - Tietokannat Luettavissa: http://www.ohjelmointiputka.net/oppaat/opas.php?tunnus=php_12 Luettu: 24.4.2014

Opetus- ja kulttuuriministeriö. 2010b. Kansallisen digitaalisen kirjaston kokonaisarkkitehtuuri: Liite B Standardisalkku. Luettavissa: http://www.kdk2011.fi/images/stories/tiedostot/kdk%20standardisalkku%202011-08-29.pdf. Luettu 26.11.2012.

Petri Nuutinen. Palvelinympäristö. Luettavissa: http://web.samk.fi/staff/petri.nuutinen/Palvelinymparisto/07RAID.pdf. Luettu 9.11.2013

PHP 2013. Mikä on PHP (What is PHP) Luettavissa: http://www.php.net/manual/en/intro-whatis.php. Luettu 8.10.2013

PHP 2013. Mitä PHP voi tehdä (What can PHP do) Luettavissa: http://www.php.net/manual/en/intro-whatcando.php. Luettu 8.10.2013.

PHP 2013. SQLite Luettavissa: http://fi1.php.net/manual/en/book.sqlite.php. Luettu: 12.2.2014

SanastoWiki 2011. Metatieto. Luettavissa: http://wiki.narc.fi/sanasto/index.php/Metatieto. Luettu 8.10.2013.

Stinger web 2013. RAID. Luettavissa: http://koti.mbnet.fi/~stinger/raid.php. Luettu 9.11.2013

Storage newsletter 2011. Ensimmäinen kiintolevy (First HDD at 55 From IBM at 100). Luettavissa: http://www.storagenewsletter.com/rubriques/hard-disk-drives/history-first-hdd-ibm-ramac-350/. Luettu 9.11.2013

Tape and media 2013. LTO-4 nauhojen hinta (Lto-4-tapes price). Luettavissa: http://www.tapeandmedia.com/lto_ultrium_4_tape.asp. Luettu: 17 3 2013

Tony Stark 2013. Koska käyttää MyISAM (When to use MyISAM and InnoDB?) Luettavissa: http://stackoverflow.com/questions/15678406/when-to-use-myisam-and-innodb Luettu: 17.5.2014

Tom´s hardware 2012. Kiintolevyjen tallennustiheys kasvaa. Luettavissa: http://www.hardware.fi/uutiset/artikkeli.cfm/2012/05/22/ennuste_kiintolevyjen_tallennustiheys_tuplaantuu_vuoteen_2016_mennessa. Luettu: 9.11.2013

Tutkimusaineistojen tiedonhallinnan käsikirja 2012. Fyysinen säilytys. Luettavissa: http://www.fsd.uta.fi/tiedonhallinta/osa9.html. Luettu 8.10.2013.

Verkkokauppa.com 2013. sata II. Luettavissa: http://www.verkkokauppa.com/fi/catalog/1272c/Kovalevyt-SATA-II. Luettu: 17.2.2013