

TIETOVARASTOINTI TIETOJENHALLINNASSA -SEMINAARI 23.4.2008

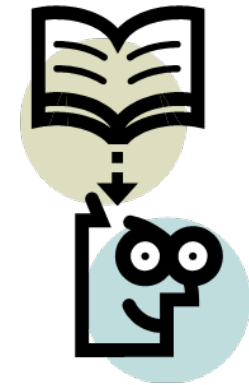


Tietovarastointi tietohallinnon uutena työvälineenä

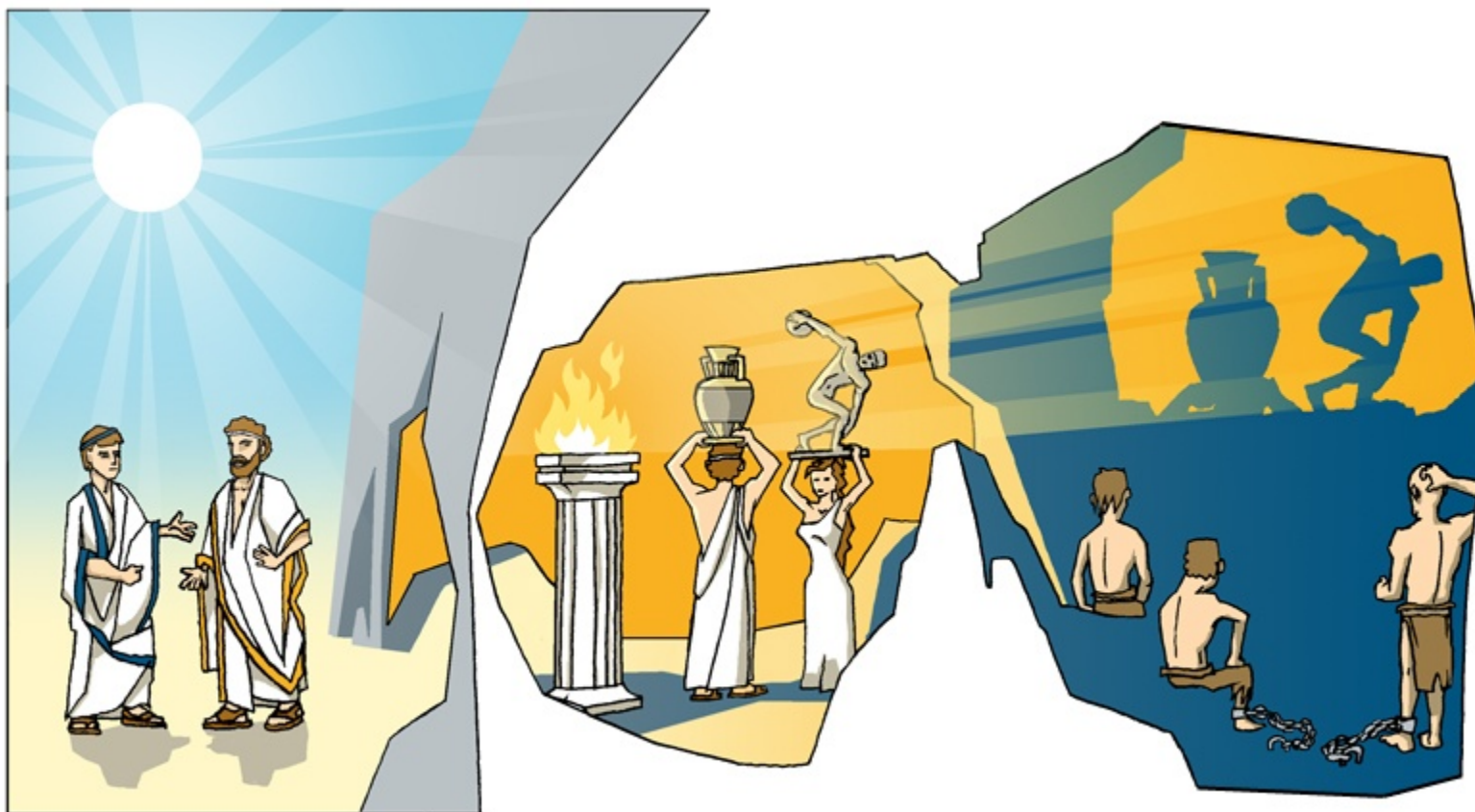
Mika Naatula



- Tiedon hallinnan merkitys kasvussa, informaatio strategisena voimavarana
- Perusteluja tietovarastointiin
- Tyypillisiä soveltamisalueita
- Tietovarastojärjestelmän erityispiirteet ja rakenne
- Tietovaraston perusprosessit
- Tietovarasto vai erillisratkaisuja



Platon 427-347 eaa



”Yhteiskunnan menestys edellyttää kansalaisten ymmärryksen lisääntymistä.”

- Perustettu 1.7.1999
- Tänään yli 160 työntekijää
- Kuusi toimipistettä:
Tanskassa, Suomessa,
Islannissa, Ruotsissa ja
Norjassa.



Toimialue
Informaation hallinta

Kattavat palvelut
Johdon konsultointi, koulutus,
toteutuksen tuki, ulkoistukset

Riippumattomuus
Vain konsultointipalveluja – ei ohjelmisto tai
laitteistomyyntiä – ei puoluellisuutta

Menetelmä
Platon Insight®
Käytännössä toimivia ratkaisuja jo 7 vuotta



Informaatio on tärkeä voimavara, josta tulisi huolehtia samalla tavoin kuin rahoitusomaisuudesta, henkilöstöstä, rakennuksista, koneista ja laitteista jne.



”Tietoon pohjautuvat perustelut vaikuttavat suoraan ihmisten tahtoon; perusteltu tahto ohjaa voimakkaasti ihmisten toimintaa.”

Platon, 427 – 347 eaa

”The way you collect, manage and use information will determine whether you become a winner or a loser . . .”

Bill Gates

”Business Intelligence and Data Warehousing - the tools for information retrieval, reporting and analysis - will be a key competitive differentiator in the 3rd millenium”

Gartner Group

Tietovaraston määritelmä

"A data warehouse is a subject-oriented, integrated, non-volatile, time-variant collection of data in support of management's decisions."

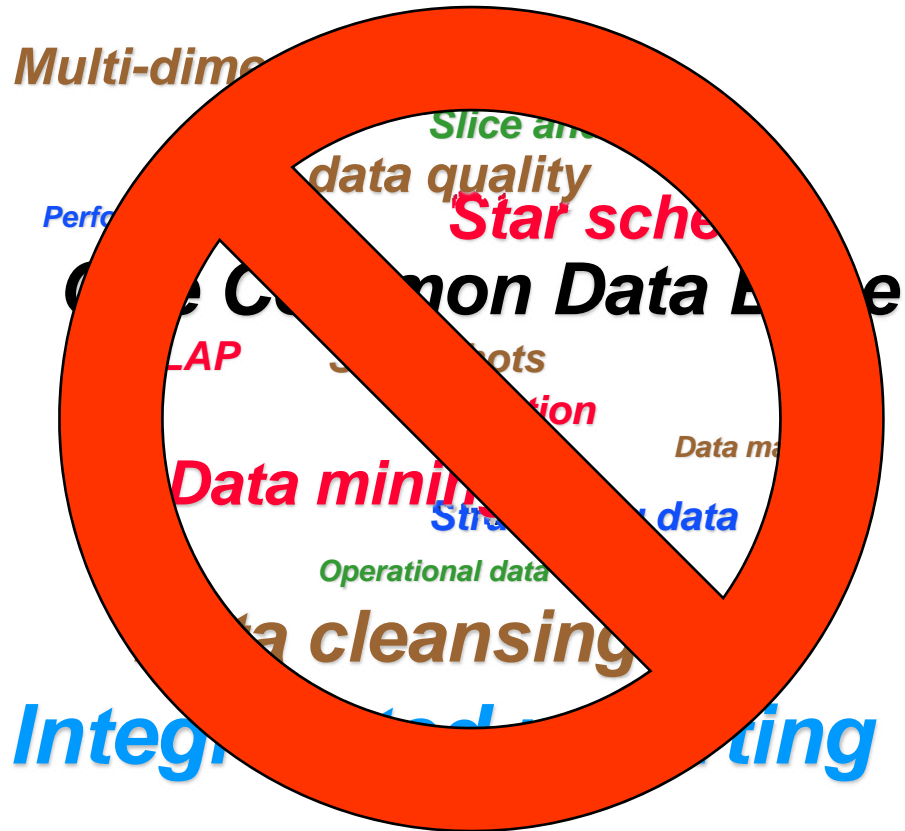
– Bill Inmon

- Subject oriented
- Integrated
- Time-variant
- Non-volatile

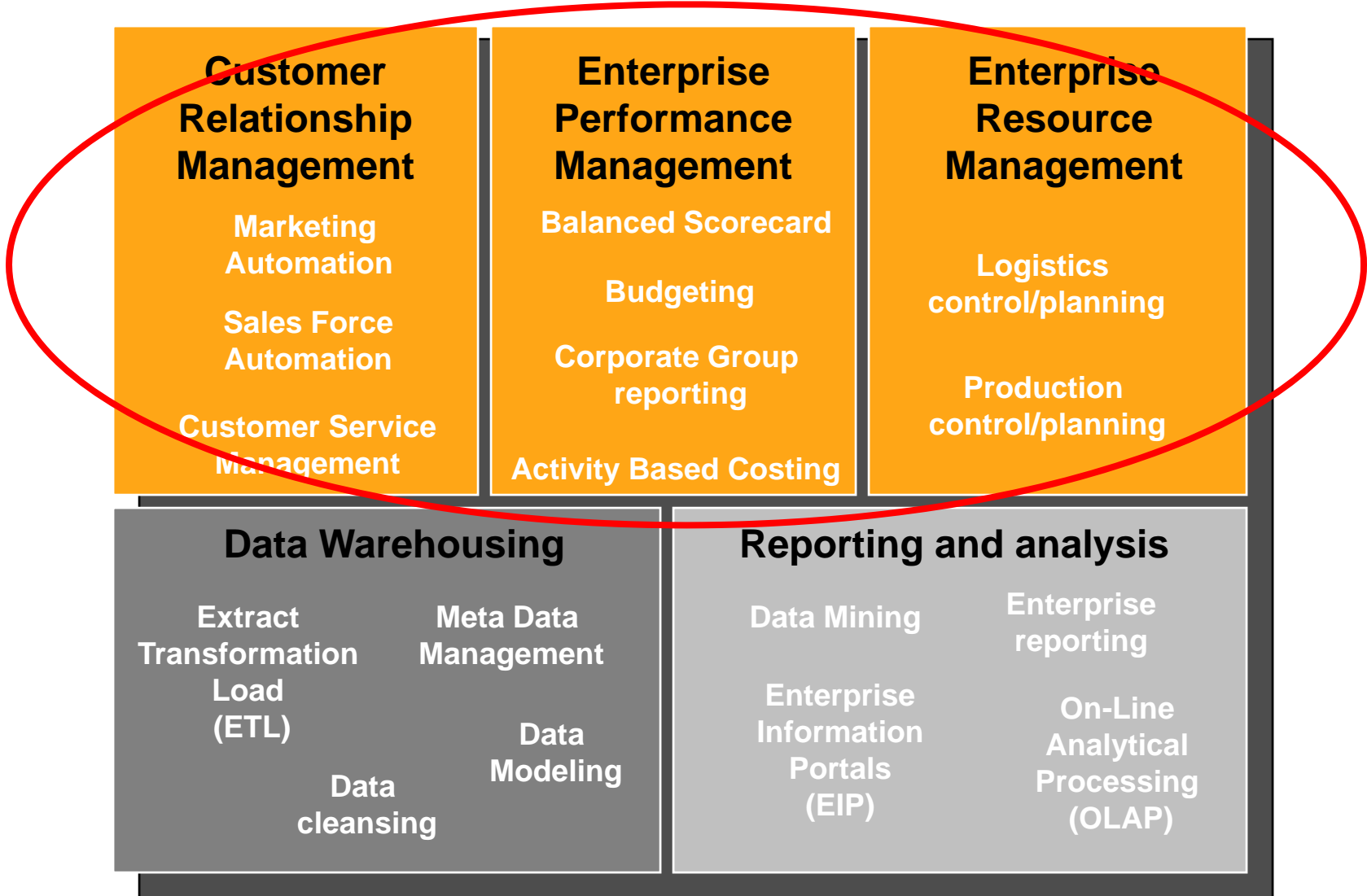




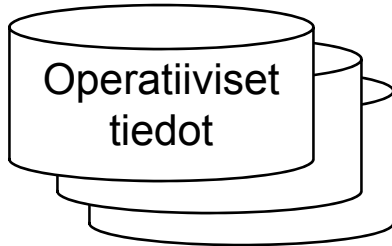
Tarve tietovarastoinnille?



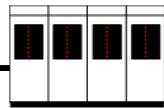
Business initiatives



Raportointi operatiivisessa maailmassa



Henkivakuutus
*150662-0315
Jens Jensen
Male



IBM MVS

Autovakuutus
*RS 29 252
J. Jensen
VW Passat



NT

Kotivakuutus
*1235677
Jens P. Jensen
East Street 66

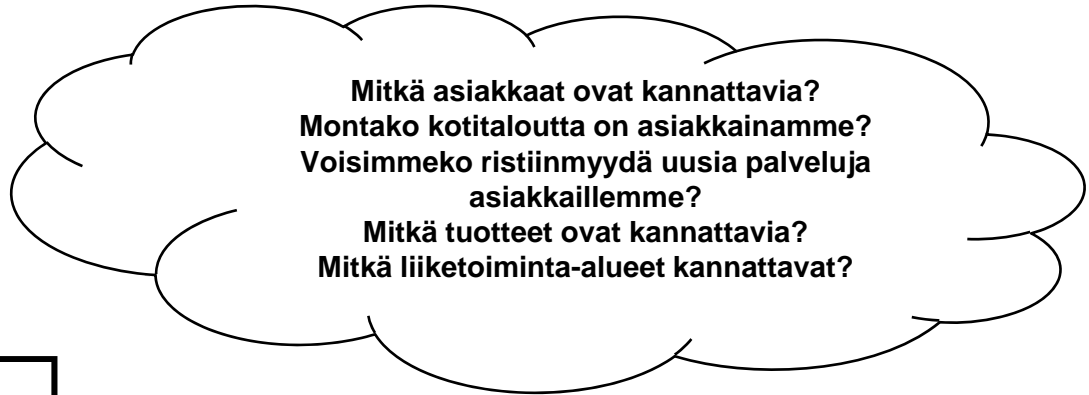


IBM AS400

Sairasvakuutus
*567-678-33-Jan-85
J Jensen
Two children



DEC VMS



Raportointi operatiivisessa maailmassa

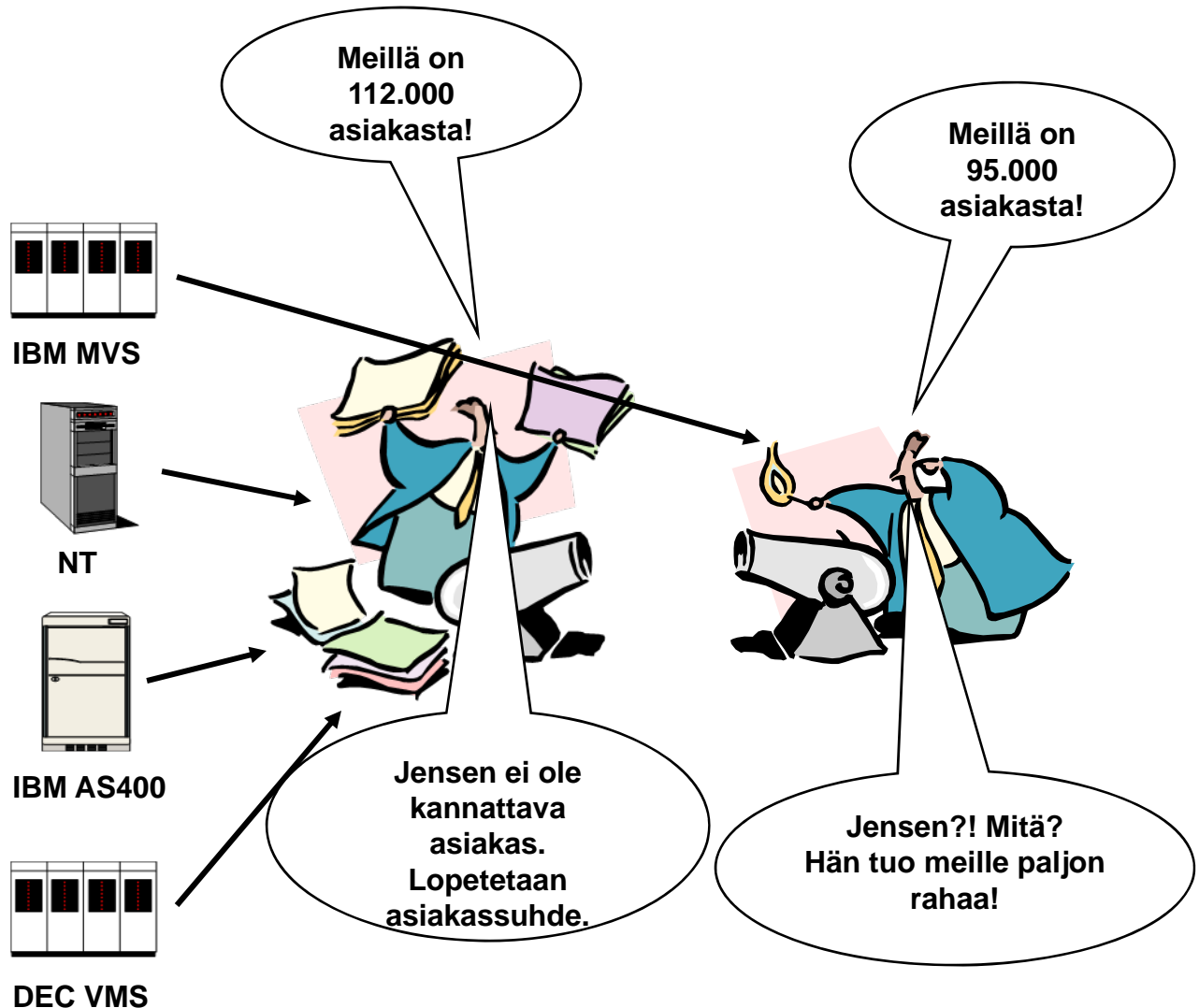
Operatiiviset tiedot

Henkivakuutus
*150662-0315
Jens Jensen
Male

Autovakuutus
*RS 29 252
J. Jensen
VW Passat

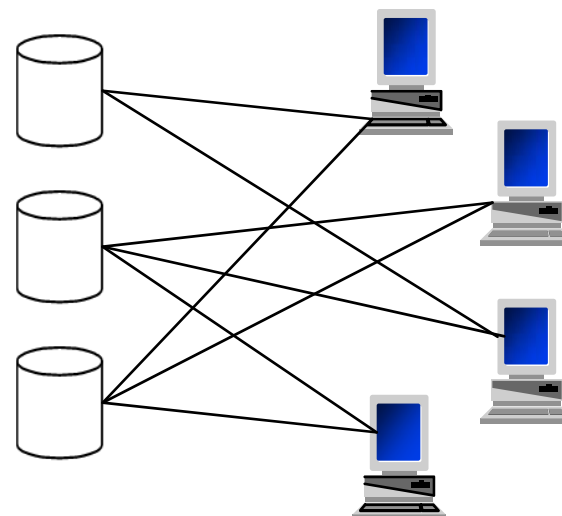
Kotivakuutus
*1235677
Jens P. Jensen
East Street 66

Sairasvakuutus
*567-678-33-Jan-85
J Jensen
Two children

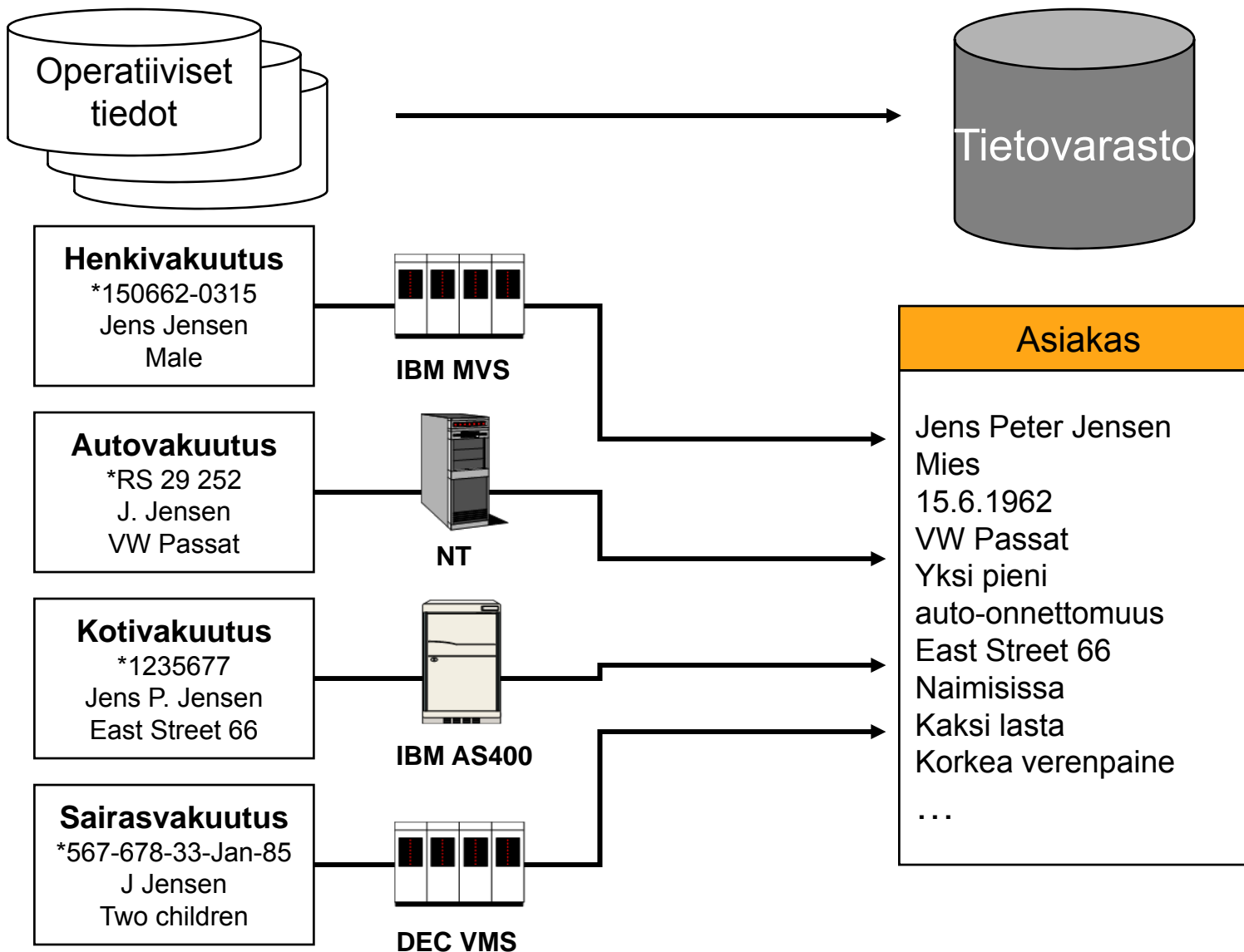


Raportointi operatiivisessa maailmassa

- Erilaisia ja erillisiä käyttöliittymiä
- Hankala tietojen saatavuus
- Monimutkaiset käsittelyt ongelmallisia
- Monta vaihtoehtoista tietolähdettä ja yhtä monta lopputulosta
- Tarvitaan paljon asiantuntemusta tietojen tulkintaan
- Historiatiedoissa puutteita
- Tietojen yhdistäminen vaikeaa tai mahdotonta



Ratkaisu: Tietovarasto



Ratkaisu: Tietovarasto

- Yksi totuus asioista
- Yhtenäistetty rakenne
- Tarkistetut ja yhteiset tiedot
- Hyvin kuvatut tietosisällöt
- Aikaleimatut tiedot, sisältää myös kattavan historian
- Helppo tietojen saatavuus, käytettävyys ja ymmärrettävyys

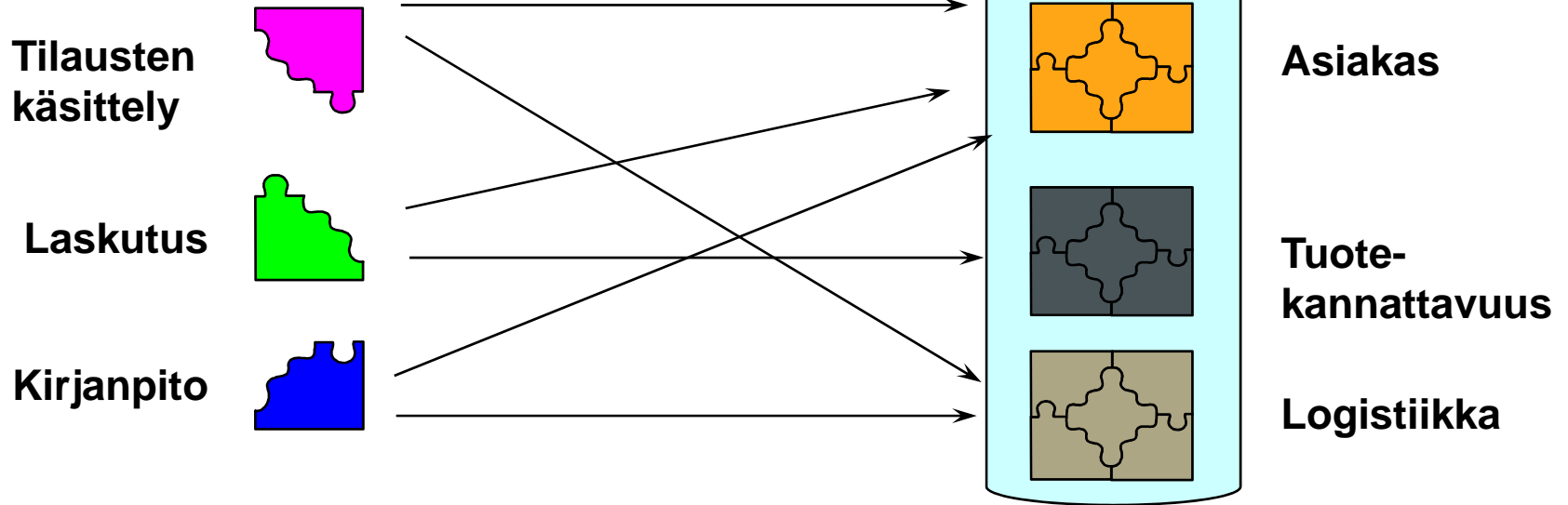


Tiedot poimitaan lähdejärjestelmästä, yhtenäistetään ja talletetaan tietovarastoon jotta saadaan laajempi käsitys asiakokonaisuudesta kuten esim. asiakkuus.



Operatiivinen järjestelmä

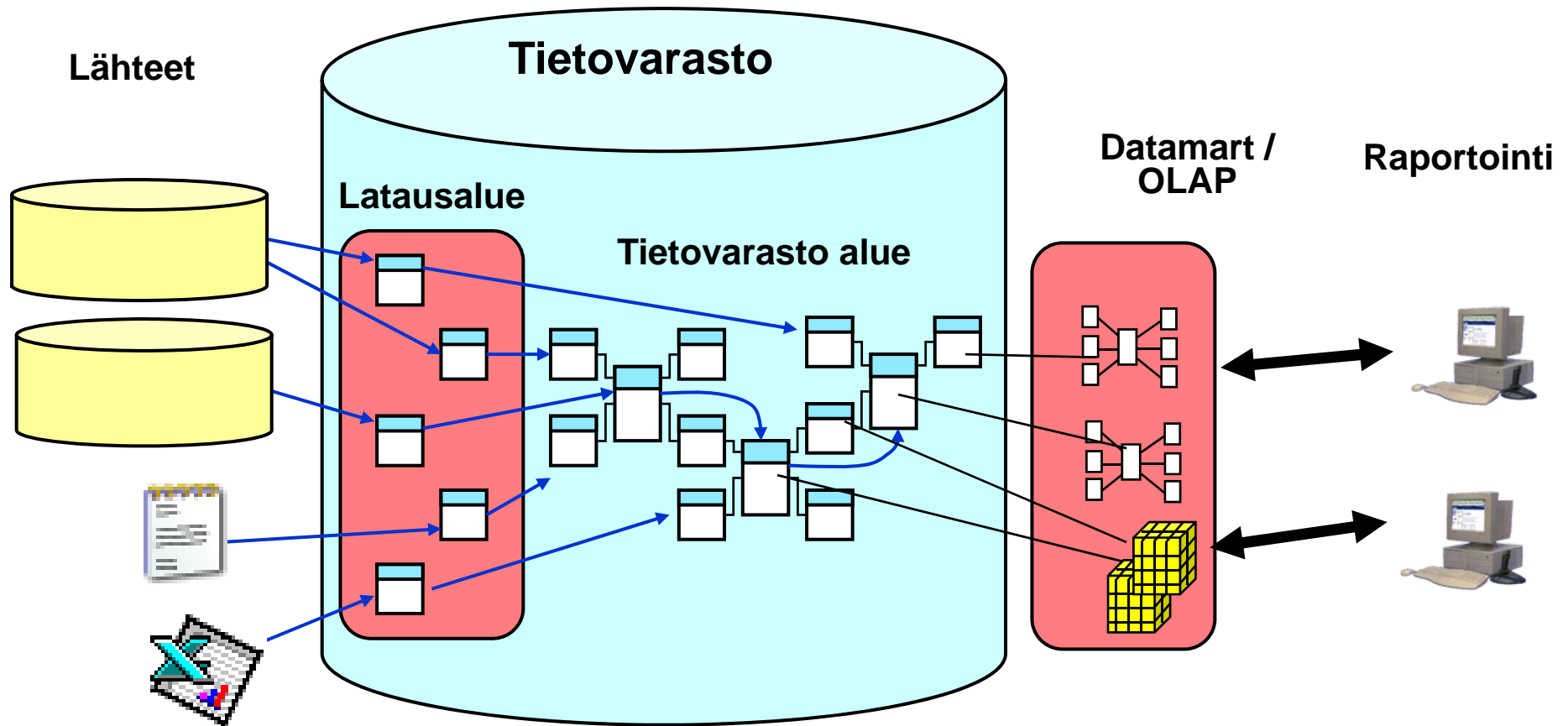
Tietovarasto



Operatiivisissa järjestelmissä tiedot on organisoitu **prosesseittain** tai tehtävittäin

Tietovarastossa tiedot on organisoitu **aihealueittain**, joihin tietoja tuodaan useasta lähteestä/prosessista.

Tietovarastojärjestelmän rakenne





- Lähdejärjestelmäriippumattomuus
- Aidon tietointegraation toteutuminen
- Hyödyntämisvälineriippumattomuus
- Laajennettavuus ja mukautettavuus uusiin tietosisältöihin
- Iteratiivinen kehitys
- Pitkäaikainen historian tallennus



Tietovaraston ominaisuuksia

- Tarkoitettu vain kyselykäyttöön, ei käyttäjien tekemään tiedonsyöttöön
- Tietomalli optimoitu kyselykäyttöön
- Sisältää usean lähdejärjestelmän tietoja
- Ladataan tyypillisesti eräajoina lähdejärjestelmistä
- Sisältää useita tietojen ”sukupolvia”, tietoihin liittyy aikaleima – vanhat tiedot jäävät talteen historian analysointia varten



Miksi tietovarasto – yleisiä teknislähtöisiä perusteluja

- Perusjärjestelmien kuormitus vähenee
- Mahdollistaa tietojen integraation useammasta lähteestä/järjestelmästä
- Mahdollistaa historiatietojen analysoinnin tehokkaammin
- Tehokkaampi ja helpompi tietojen analysointi ja raportointi (siihen optimoidut mallit)
- Parantaa tietojen laatua
- Raportoinnin irrottaminen perusjärjestelmän muutoksista



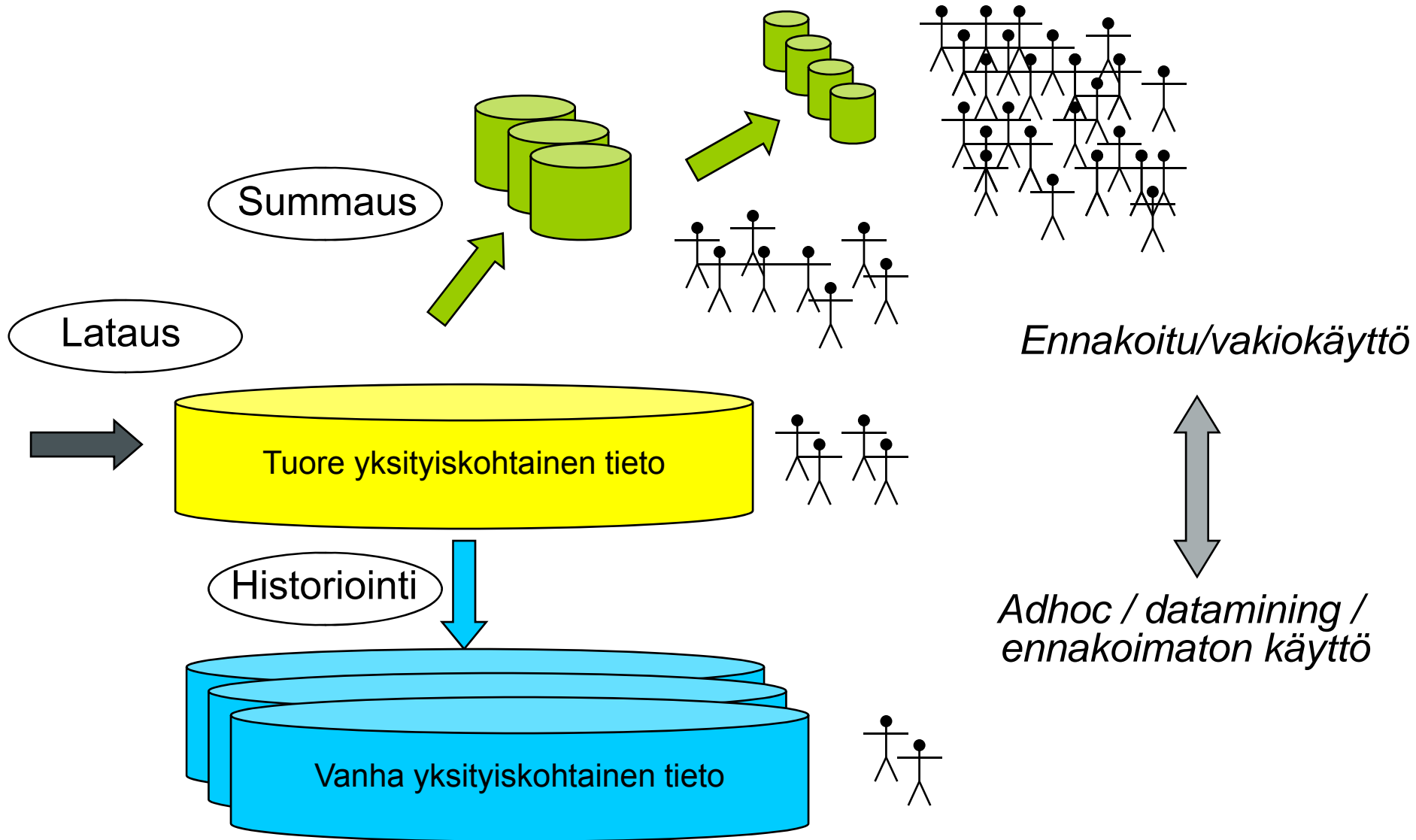
Miksi DW – yleisiä liiketoimintalähtöisiä perusteluja

- Auttaa datan ja siitä jalostetun informaation jakelussa sekä käsittelyssä parantaen kilpailukykyä
- Nopeampi liiketoiminnan reagointi toimintaympäristön (sisäinen ja ulkoinen) muutoksiin oikean tiedon perusteella
- Laajempi tietämys nykytilasta antaa paremmat mahdollisuudet suunnitella toimintaa tulevaisuuteen (strateginen suunnittelu ja päätöksenteko)
- Organisaation sisäisen tehokkuuden ja koordinaation parantaminen
- Parantaa operatiivisen toiminnan koordinaatiota ja kontrollia
- Yksi yhtenäinen totuus, yhtenäiset käsitteet



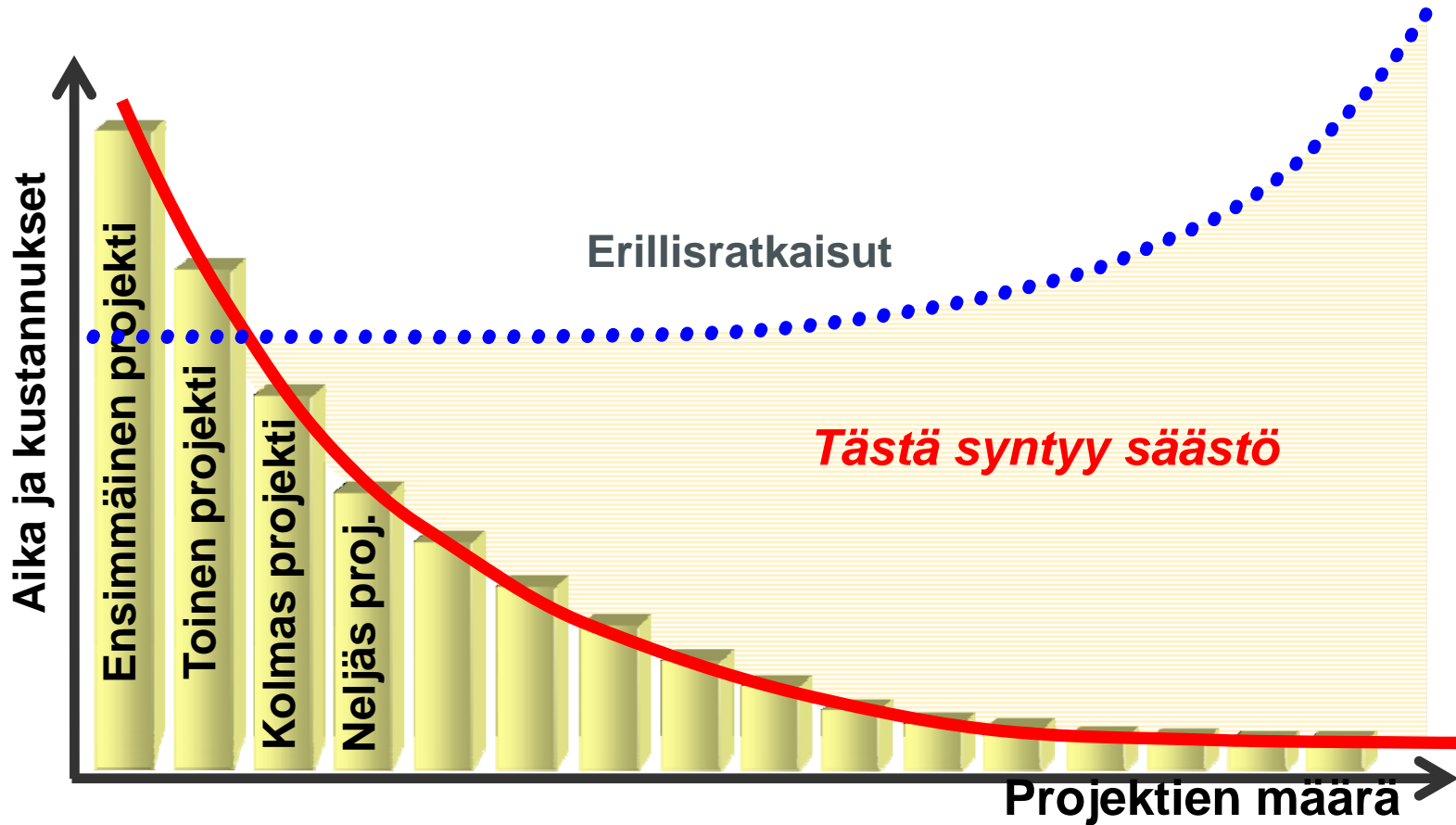


Tietovarastoinnin pääprosessit



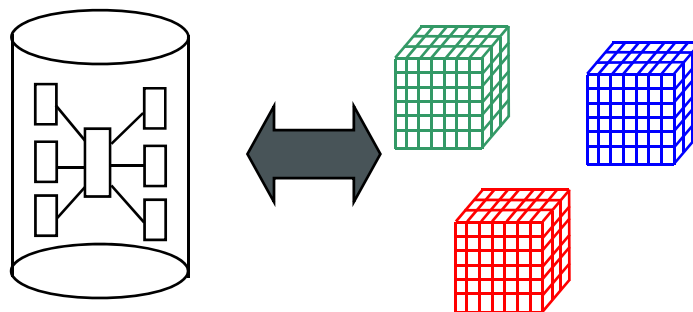


Tietovarasto vai erillISRatkaisut - kustannusnäkökulmaa



Milloin tarvitaan tietovarastointia, indikaattoreita

- Useita tietolähteitä joista tiedot on yhdistettävä raporteille
- Tarvitaan tietomuunnoksia, tiedon siistimistä ja yhdistämistä
- Tarvitaan monimutkaisia muunnoskäsittelyjä
- Manuaalityövaiheita paljon
- Raportointi henkilösidonnaista
- Tietomäärät ovat suuria ja tietorakenteet monimutkaisia
- Halutaan varautua pitemmän aikajänteen analyysiin



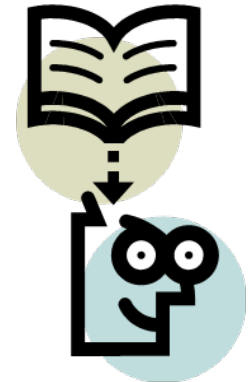


Tietovarastoinnin käytännön suunnittelu ja rakentaminen

Mika Naatula

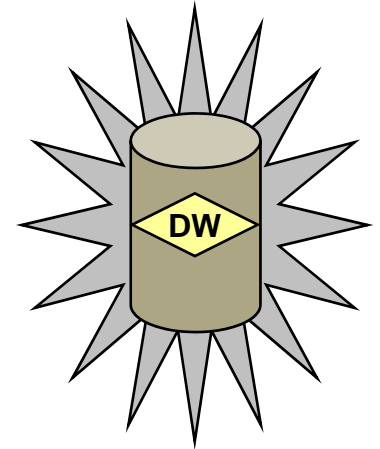


- Eri lähteistä tulevien tietojen yhtenäistäminen, semanttinen yhteensopivuus
- Laadukkaan tiedon merkitys
- Tietomallinnus ja metatietojen kuvaaminen
- Projektin läpivienti
- Tarvittavat henkilöresurssit
- Menestyksen avaimet

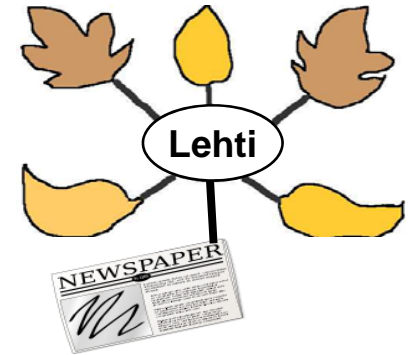


Mitä hyötyä yhtenäistämisestä?

- Vastaa kunnolla liiketoiminnan kaikkiin kysymyksiin (etenkin ylin johto)
- Uudelleenkäyttö ja hallittavuus
- Varmistaa tietojen
 - oikeellisuus
 - yhteismitallisuus
 - yhdistämiskelpoisuus
- Muutosvalmius
 - Vaatimuksia kohti kasvavaa synergiaa
 - Organisaatiomuutokset
 - Uudet tarpeet, vanhat tiedot
 - ”Agile organisation”



- Semanttinen yhteensopivuus tarkoittaa, että tietojärjestelmä pystyy yhdistelemään eri lähteistä vastaanottamaansa tietoa niin, että tietojen merkitys säilyy.
- Ongelmia aiheuttavat tietojärjestelmien ja organisaatioiden erilaiset tietosisällöt ja käsitteistöt, esimerkiksi:
 - Käsitelmärittelyä tai dokumentointia ei ole tehty, edes peruskäsitteistä ei ole sovittu
 - Yhdellä käsitteellä on useita käyttötarkoituksia tai yhteen käyttötarkoitukseen käytetään useita eri käsitteitä
 - Luokitukset ja kooditukset (arvojoukot) vaihtelevat
 - Sama tietokenttä, eri kooditus
 - Sama tietokenttä, samat kooditukset, mutta eri tulkinta



=> Yhtenäistämistä ei voi jättää toteuttajien vastuulle



Mitä laatu maksaa?

“Quality is free...What costs money are the unquality things — all the actions that involve not doing jobs right the first time.”

- Philip Crosby, Philip Crosby Associates

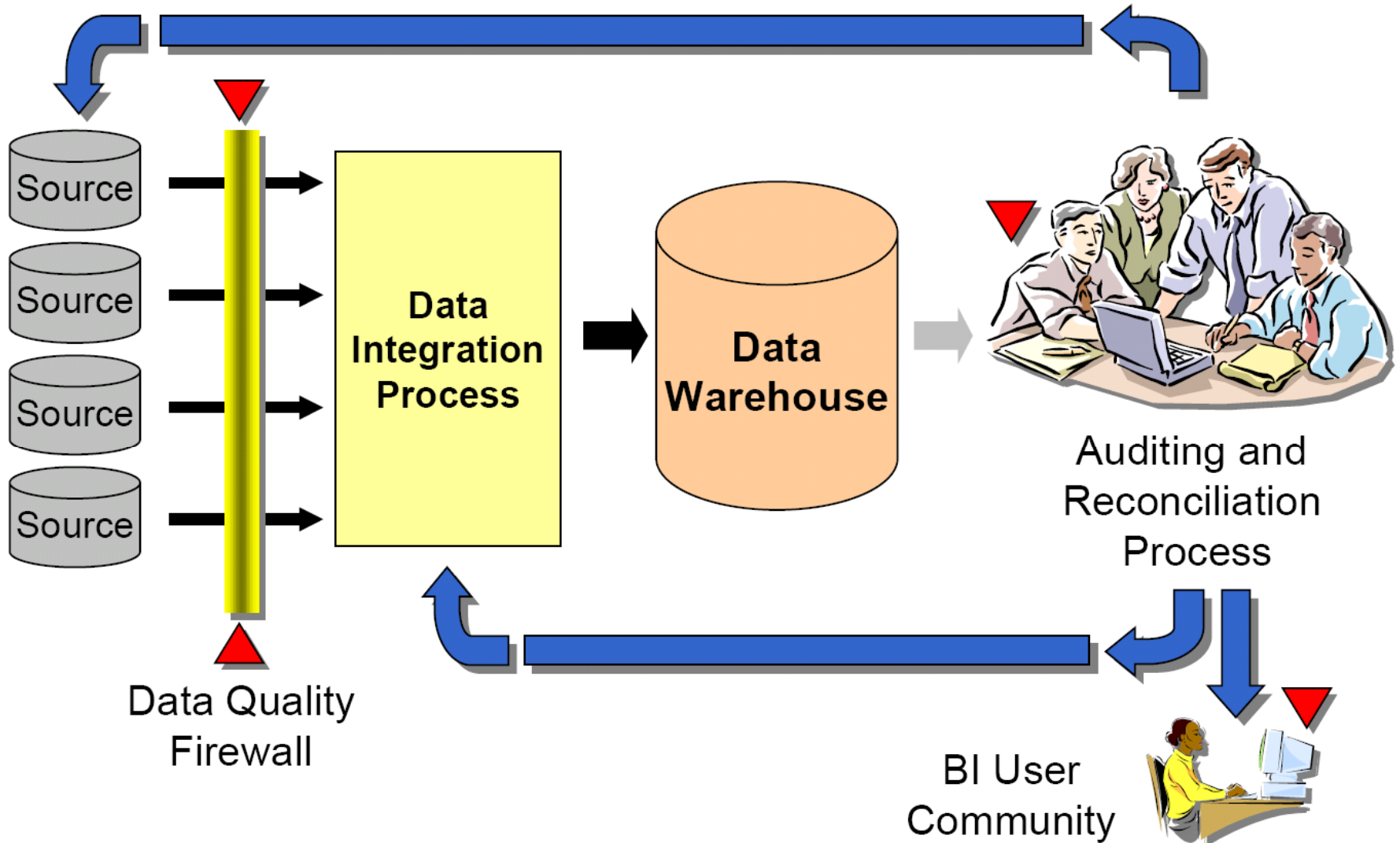


“To maximize acceptance of the data warehouse, construct a "firewall" to mitigate data quality issues. Coupled with ongoing data quality analysis, this is a powerful tool for reducing data warehouse project risk.”

- Ted Friedman, Gartner Group

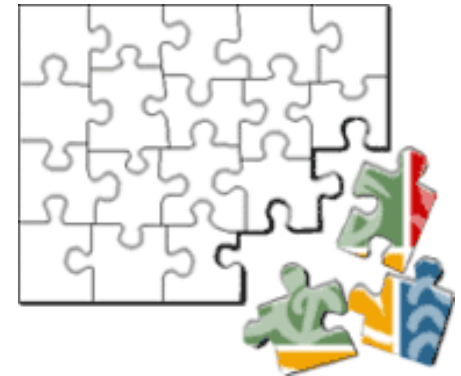


Tiedon laadun palomuuuri



Laadukkaan tiedon ominaisuuksia

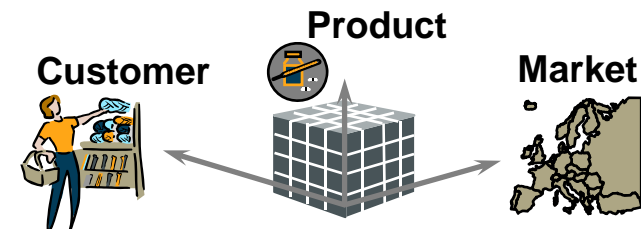
- Metatiedot kunnossa
 - Tiedon merkitys on selvillä ja siitä on kuvaustieto talletettuna
 - Tiedetään mistä lähteestä tieto on tullut ja milloin se tuotu
 - Tiedetään kuka tiedosta vastaa
 - Tunnetaan tiedon laatuominaisuudet (esim. paljonko virhemarginaali on)
 - Tiedot on yhdenmukaistettu eli ”standardoitu”
 - Samaa asiaa tarkoittava informaatio on samassa muodossa ja samassa kentässä (vaikka olisikin peräisin eri lähteistä)
 - Tietojen koodaukset on yhtenäistetty ja selvät kirjoitusvirheet on korjattu
 - Kaksoiskappaleet on yhdistetty
 - On yhdistetty eri järjestelmistä peräisin olevat tiedot niin että alkuperäisten tietueiden avaintiedot säilytetään
 - Jos mikään lähdejärjestelmä ei ole varsinainen master, luodaan rekisteritiedoille uusi keinotekoinen avain
- ⇒ Yhtenäinen tietomalli ja avaimisto, tietoihin voi luottaa!



- Tiedon laadun varmistaminen on edellytys onnistumiselle
 - Aihealueperusteinen malli pakottaa tiedon läpikäyntiin
 - Tietojen jäljitettävyys
- Käytettävyys
 - Tietomalli ymmärrettävä myös ei IT ihmisille, tiedon haku helppoa
 - Uusien henkilöiden perehdytys nopeaa sekä edw kehitykseen että etenkin raportointiin
 - Porautuminen tiedoissa alimmalle tapahtumatasolle mahdollista
- Ylläpidettävyys, mukautettavuus erilaisiin tilanteisiin
 - Toteutetaan lisää aihealueita samaan kantaan
 - Samaa tietoa alkaa tulla useammasta lähteestä
 - Tietolähteenä oleva perusjärjestelmä vaihdetaan
 - Samoista perustiedoista jalostetaan lisätietoja
 - Tarvitaan lisää määreitä tai kytkentöjä samaan perustietoon
- Suorituskyky skaalautuu tarvittaviin tietomääriin
 - Vasteajat analyysikäytössä
 - Latausajat
 - Poiminnat datamart latauksia varten



- Tiedot jaotellaan tietoa luokitteleviin referenssitietoihin eli perustietoihin (dimensiot) ja tapahtumia kuvaaviin transaktiotietoihin (faktat)
- Dimensiot
 - ajan suhteen pysyviä tai hitaasti muuttuvia
 - käytetään luokittelemaan faktatietoja
 - voivat muodostaa sisäisiä hierarkioita, jaetaan mahdollisesti ali-dimensioihin (lumihiihtale malli)
 - normaalit päivitysoperaatiot lisäyksiä tai muutoksia (ei poistoa)
- Faktat
 - sisältävät aika-ulottuvuuden
 - sisältävät numeerisia kenttiä
 - normaali päivitysoperaatio tietojen lisäys

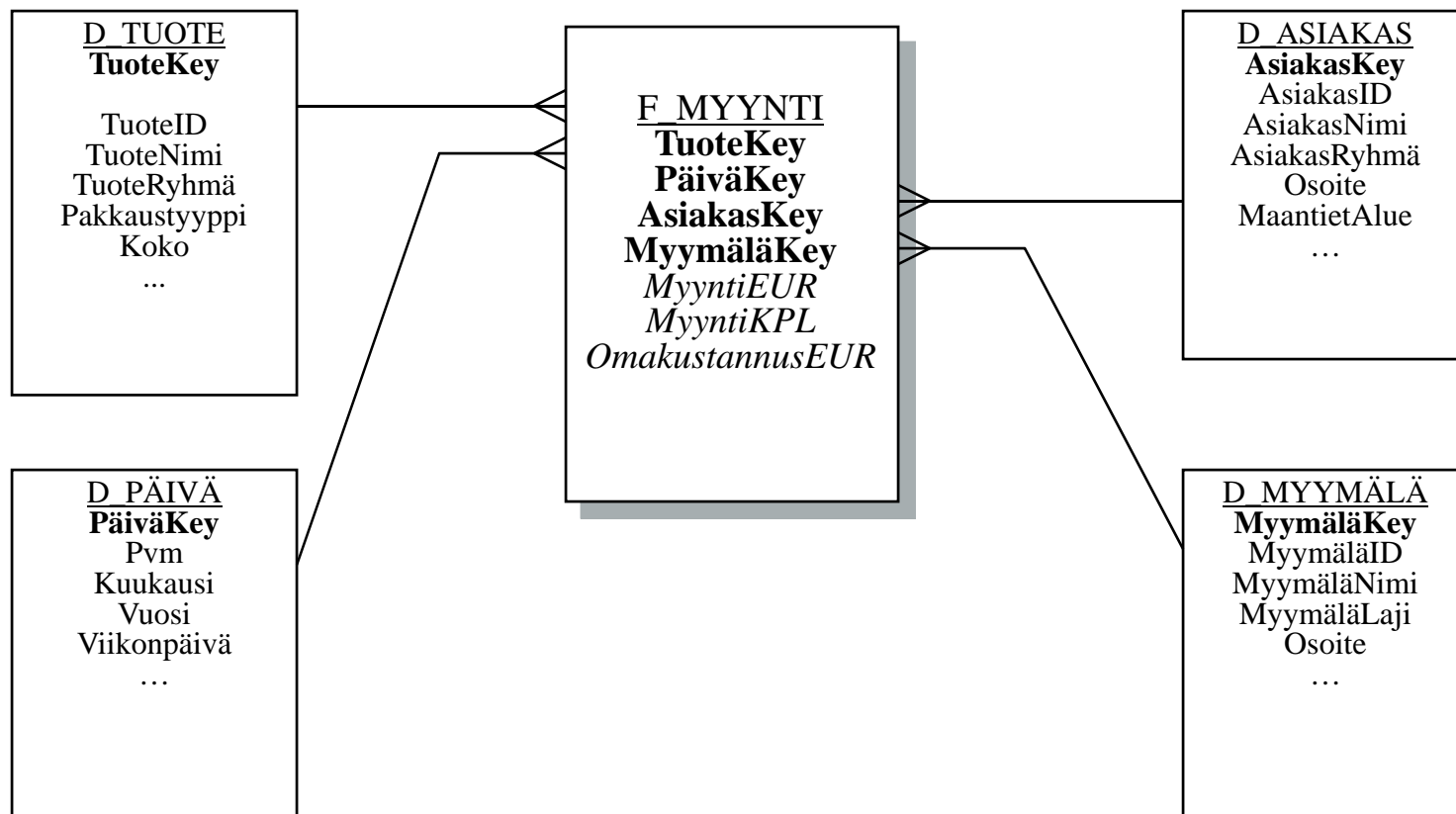


Tietojen aikariippuvuus

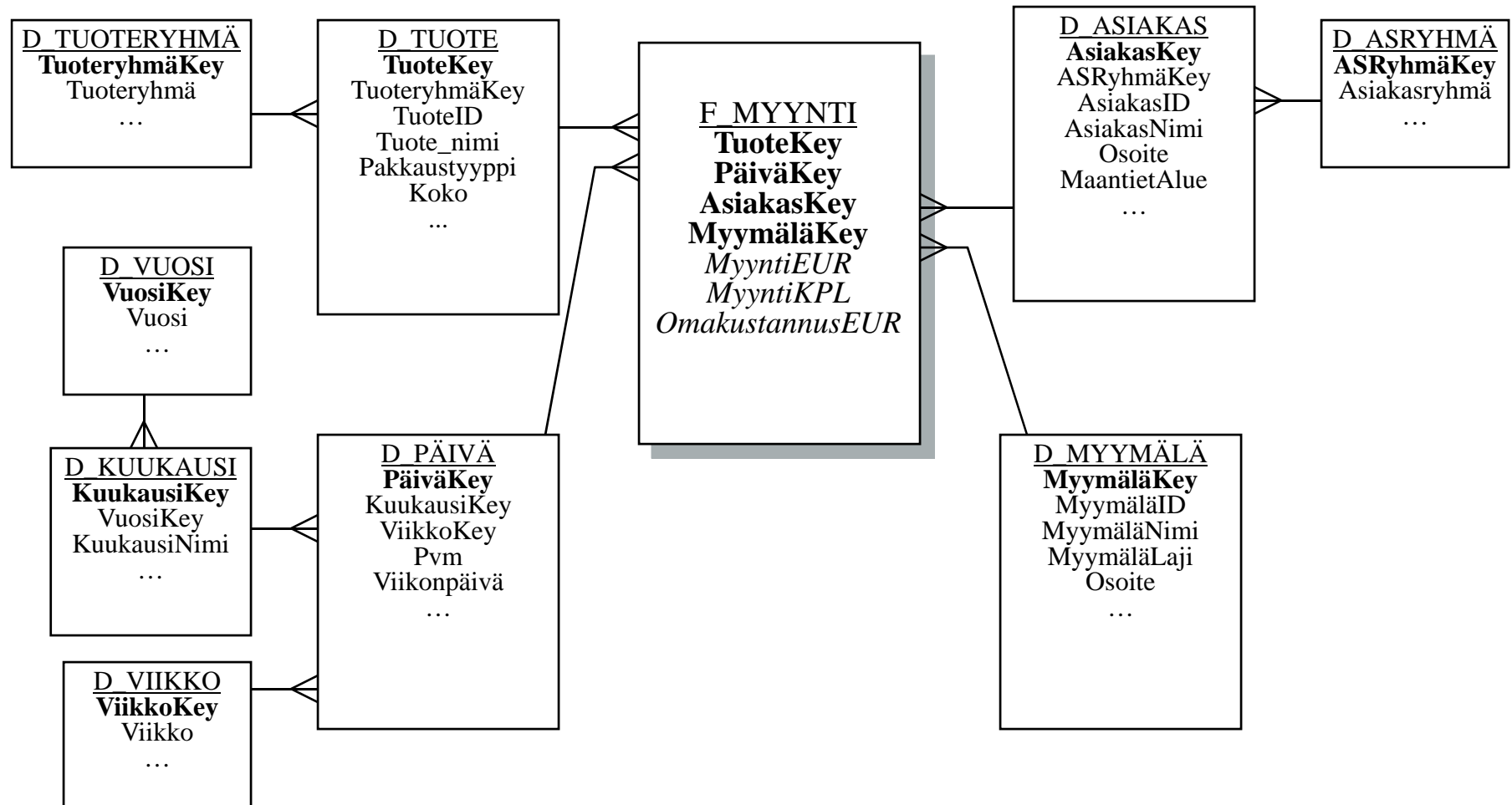
- Kaikki alimman tason faktataulut edustavat jompaa kumpaa kahdesta aikariippuvuuden lajista
- Tapahtuma (transaction)
 - Tietue faktataulussa kuvaa yhtä ainoaa tapahtumaa, joka on tapahtunut tietyssä aikana.
 - Numeeriset faktakentät ovat tapahtumaan liittyviä määriä.
 - Aikaleima perustuu tapahtuma-aikaan.
- Tilannekuva (snapshot)
 - Faktakentät kuvaavat otoshetkellä olevaa tilaa (saldoa), mukana voi hyvin olla myös otosajankohdan tapahtumien summa.
 - Tyypillinen ”otoskanta” sisältää tilannekuva-faktatauluja, joista pidetään tallessa vain viimeisin versio.
 - Aikaleima perustuu otoshetkeen.
 - Kumulatiivinen tilannekuva voidaan mieltää tilannekuvan johdannaiseksi, joskin se usein esitetään käytännön syistä erillisenä kolmantena lajina.



Dimensionaalinen malli "tähti"

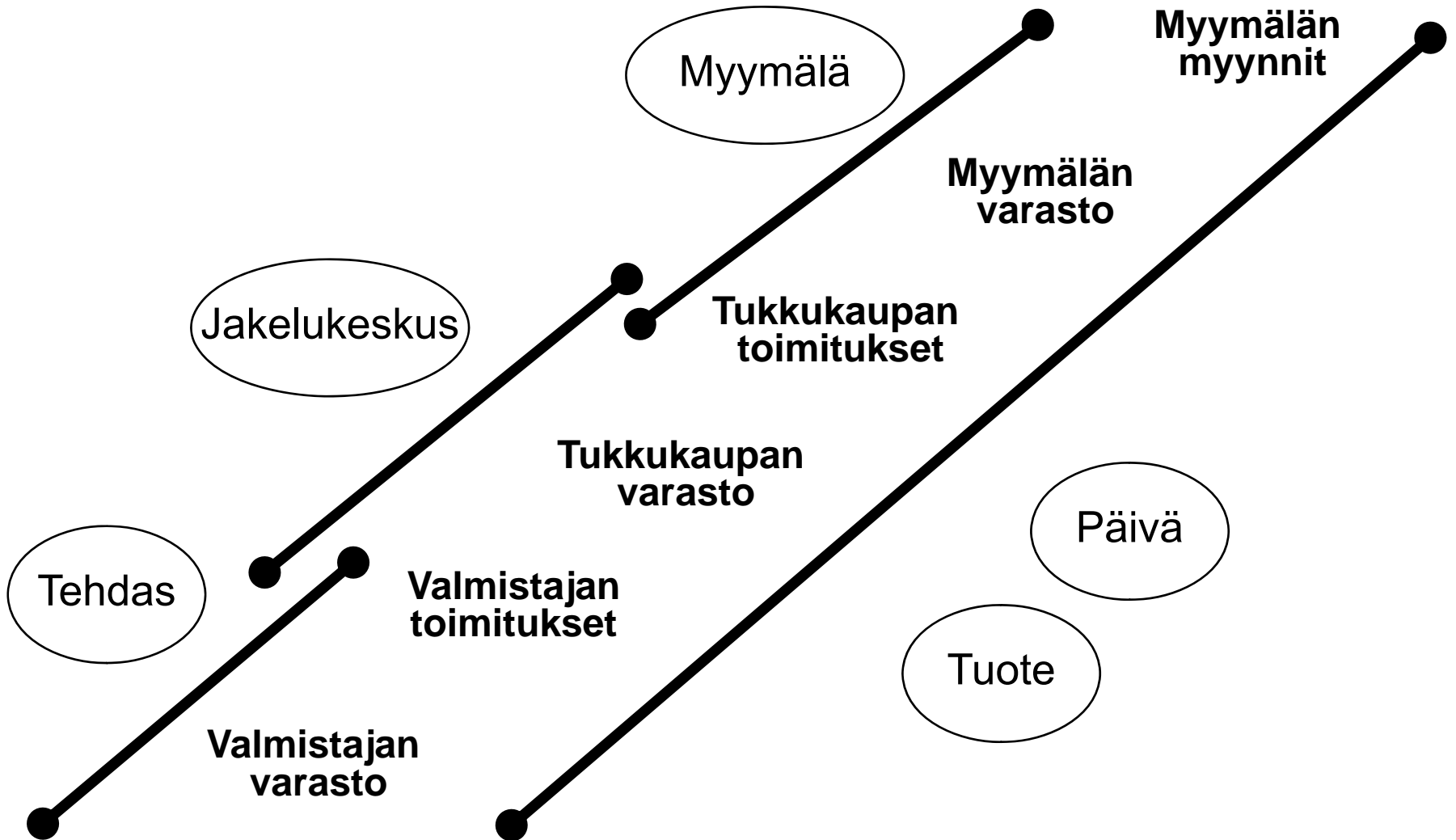


Dimensionaalinen malli "lumihutale"





Yhtenevät dimensiot, esimerkkinnä vähittäiskauppa





- Tietosisällön tarkentaminen ja määrittely
 - mukaan tulevien tietojen valinta, tiedon taso ja tarkkuus
 - perustelut tietojen tarpeellisuudesta ja hyväksikäytöstä
- Lähdejärjestelmäanalyysi
 - liittymäraajapintojen määrittely, kuvaukset lähdejärjestelmästä, rajapintojen toteutustekniikka
 - siirtotiedostojen kuvaus: kentät, tietotyypit, sisältö, muoto
 - päivitystiheydet
- Lataustaulujen ja tietovarastotaulujen tietomallin määrittely
 - taulukuvaukset, suhteet, avaimet, perusindeksit
 - dimensioiden muutoshallinta
 - raporttien määrittely karkealla tasolla



- Tietovaraston tallennussuunnitelman määrittely
 - tallennusfrekvenssit, historiointi, tietojen poistot
 - dimensioiden muutoshallinta
 - karkea tietokannan kasvuennuste
- Käsittelysääntöjen määrittely
 - tietovirran kuvaus, arkkitehtuuri
 - laskentasäännöt, tarkistusten toteutustapa
- Metadatan hallintasuunnitelman laadinta
 - mitä metadataa kerätään
 - tekninen toteutus, taulurakenteet
- Tietoturvasuunnitelman laadinta
 - suunnitelma käyttöoikeuksien hallintaan

Tietovarasto projektina: toteutus

- Toteutusvaiheessa perustetaan tietovaraston tietokanta, tehdään latausajot ja raportit määritelmien mukaisesti
- Määrittely ja mallinnus pitää tehdä huolella, mutta osa asioista kannattaa jättää selvitettäväksi vasta toteutuksen yhteydessä, joten toteutukselle on syytä varata riittävästi aikaa ja resursseja.
- Vaikka tiedon laatuun olisi kiinnitetty huomiota koko projektin ajan, lopullinen ”totuus” paljastuu vasta toteutusvaiheessa
- Ensi vaiheen toteutukseen valitaan riittävän laaja tietomäärä, jotta sillä olisi merkitystä - ja riittävän suppea jotta se olisi nopeasti tehtävissä.
- Dokumentoi toteutus riittävällä tasolla.



Tietovarasto projektina: testaus ja käyttöönotto

- Suunnittele alkulataukset
- Suunnittele testaus
- Suunnittele käyttöönotto
- Varaa työmäärissä riittävästi aikaa ja resursseja myös testaukselle.
- Puutteellinen testaus saattaa vaarantaa koko tietovaraston maineen.
- Vaikka käyttöönotto on hyvin suunniteltu, se on aina ”kertaluontoinen” toimenpide, jossa saattaa tulla yllätyksiä!



Keskeisiä henkilörooleja

- Sponsori liiketoiminnan puolelta
- Projektin johto
 - asiakkaan ja toimittajan projektipäälliköt
- Liiketoimintaintegraattori
 - vastuu liiketoiminnan tuntemuksesta, tavoitteet, sisältö
- Tietovarastoarkkitehti
 - tietovaraston suunnittelu, tietomallinnus
- Loppukäyttäjä / pääkäyttäjä / analyytikko
 - tietotarpeiden määrittely, tavoitteiden asetus, ylläpidettävyyksvaatimukset
- Lähdejärjestelmäasiantuntija
 - jokaiselle lähdejärjestelmälle ”yhteyshenkilö”
- ETL suunnittelija ja -kehittäjä
- Raportoinnin suunnittelija ja -kehittäjä
- Tietokannan ylläpitäjä (DBA)



Menestyksen avaimet

- Johdon tuki
- Rakennetaan tarvelähtöisesti
- Resurssit sekä tietohallinnon että liiketoiminnan puolelta
- Tiedon laatuun kiinnitetään huomiota
- Käytetään tarkoitukseen soveltuvia ohjelmistoja
- Tietokannan mallinnus oikein
- Nopea syklinen toteutus (3-6 kuukautta) ja palaute käyttäjiltä
- Käytön ja hyötyjen mittaaminen
- Odotusten hallinta
- Organisaation sitouttaminen kaikilla tasoilla
- Vastuun jako oikein eri osapuolten kesken
- Käytetään osaavia konsultteja 😊

