

# Sovellustiedon elinkaari: Kehdosta hautaan

Pekka Leppänen

IBM Software Group

Information Management





## Älykkäämpi tiedon elinkaaren hallinta

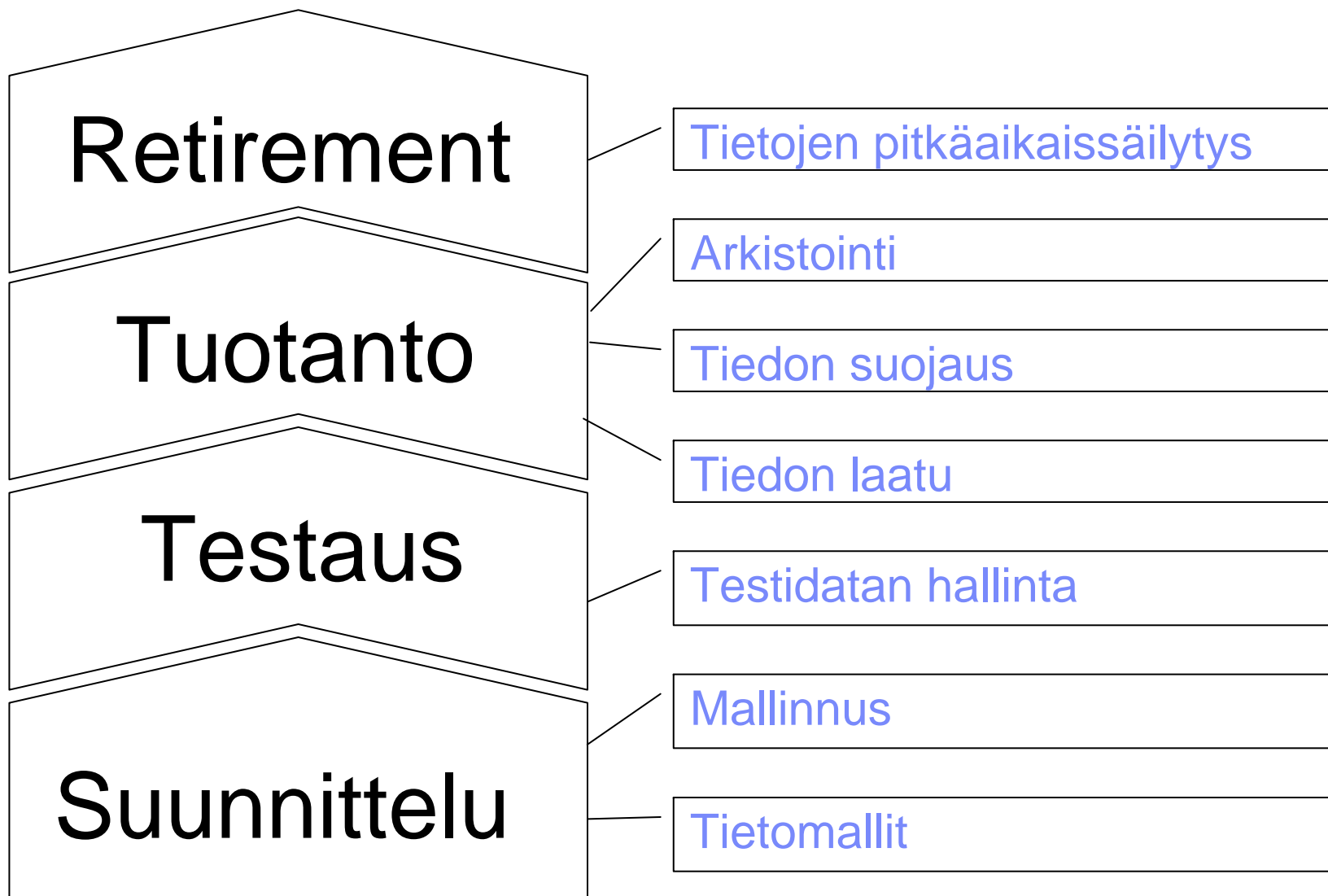
**Sovellustaso: tiedon sisältö**

**Ohjelmistoinfrastrukturi: hallinta**

**Laitteistoinfrastrukturi: tallennusjärjestelmä**

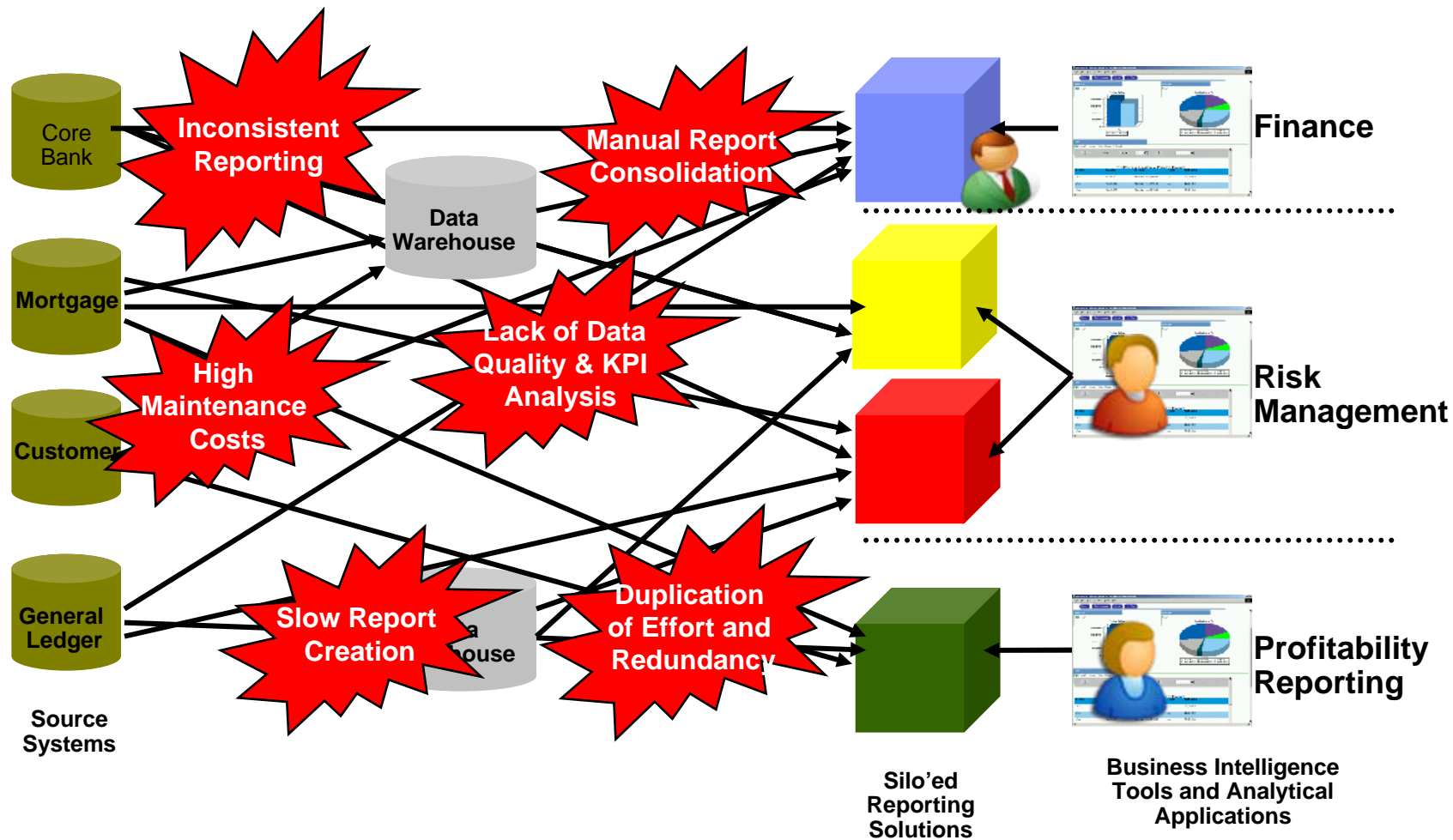


## Sovellustaso: kehdesta hautaan





# Typical Data Warehouses Deployments





## IBM Industry Data Models Address the Following Business Areas

### Banking

(Banking Data Warehouse (BDW))

- Profitability
- Relationship Marketing
- Risk Management
- Asset and Liability Mgmt
- Compliance

### Financial Markets

(Financial Markets Data Warehouse (FMDW))

- Risk Management
- Asset and Liability Mgmt
- Compliance

### Health Plan

(Health Plan Data Warehouse (HPDW))

- Claims
- Medical Management
- Provider and Network
- Sales, Marketing and Membership
- Financials

### Insurance

(Insurance Information Warehouse (IIW))

- Customer centricity
- Claims
- Operational efficiency
- Underwriting; Pricing; Profitability
- Intermediary Performance
- Compliance
- Risk Management

### Retail

(Retail Data Warehouse (RDW))

- Customer centricity
- Merchandising Management
- Store Operations & Product Mgmt
- Supply Chain Management
- Compliance

### Telco

(Telecommunications Data Warehouse (TDW))

- Churn Management
- Relationship Mgmt & Segmentation
- Sales and Marketing
- Service Quality & Product Lifecycle
- Usage Profile

# Mallinnus ja tietomallien ylläpito



Manage Business Terms



Discover Data Relationships



Design Enterprise Models

## Requirements

- Capture business terms and classifications
- Link business terms and classifications to IT assets
- Identify data stewards and make glossary accessible

## Benefits

- Context for information is available to everyone, immediately
- IT projects are aligned with data governance
- Collaboration increases across business and IT

## Requirements

- Define business objects for archival and test data applications
- Discover data transformation rules and heterogeneous relationships
- Identify hidden sensitive data for privacy

## Benefits

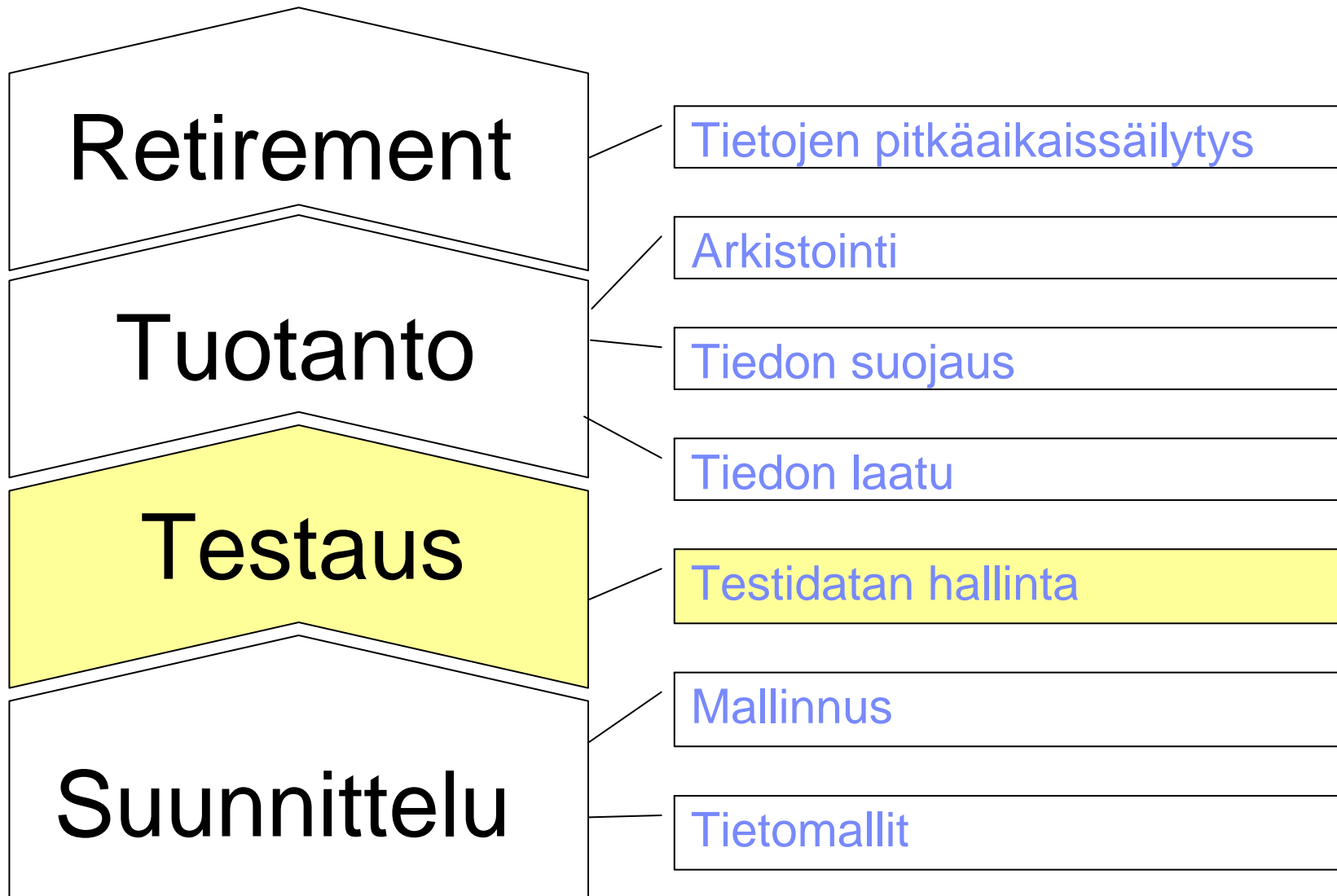
- *Automation* of manual activities accelerates time to value
- Business insight into data relationships reduces project risk
- Provides *consistency* across information agenda projects

## Requirements

- Design and manage enterprise models
- Enforce model conformance to enterprise standards
- Leverage industry data models for best practices

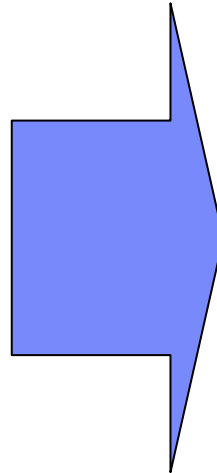
## Benefits

- Speed design activities
- Populate Business Glossary from model terms
- Validate models for enterprise conformance



## Testaustiedon hallinta: kopionnista tai manuaalisesta testidatan tuottamisesta automaattiseen prosessiin

- 'Right-sizing': kattavan, eheän ja oikeankokousen testikannan luonti tuotantotiedoista
  - Testin kattavuuden määrittäminen
    - Tarvittavien liiketoiminta-objektien määrittäminen
    - Testivolyymin asettaminen
  - Testidatan suodattaminen tuotantokannoista
- Testidatan luonti erikoistapausten tuottamiseksi
- Testikannan automaattinen virkistäminen tuotantokannoista.
- Testidatan maskaus: de-identifiointi
- Testitulosten automatisoitu tarkistus – piilevien virheiden löytäminen

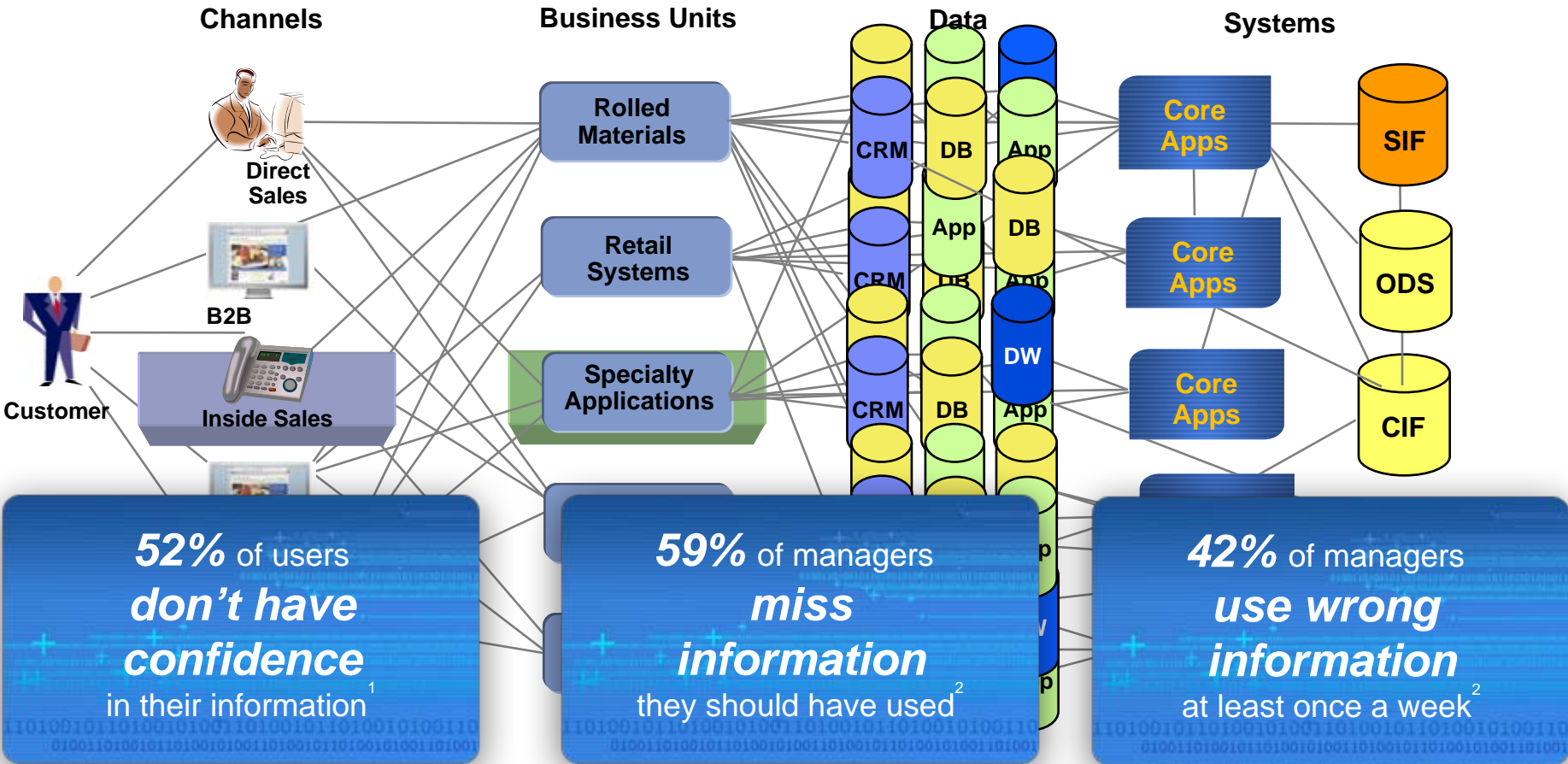


- Testauksen kattavuus ja oikeellisuus: sovellusten parempi laatu
- Testauksen nopeus ja toistettavuus: nopeammin valmista
- Testauksen tuottavuus ja tietokantojen right-sizing: säästö sovelluskehityskustannuksissa





# Paradoksi: tiedon määrän lisääntyessä informaation laatu heikkenee

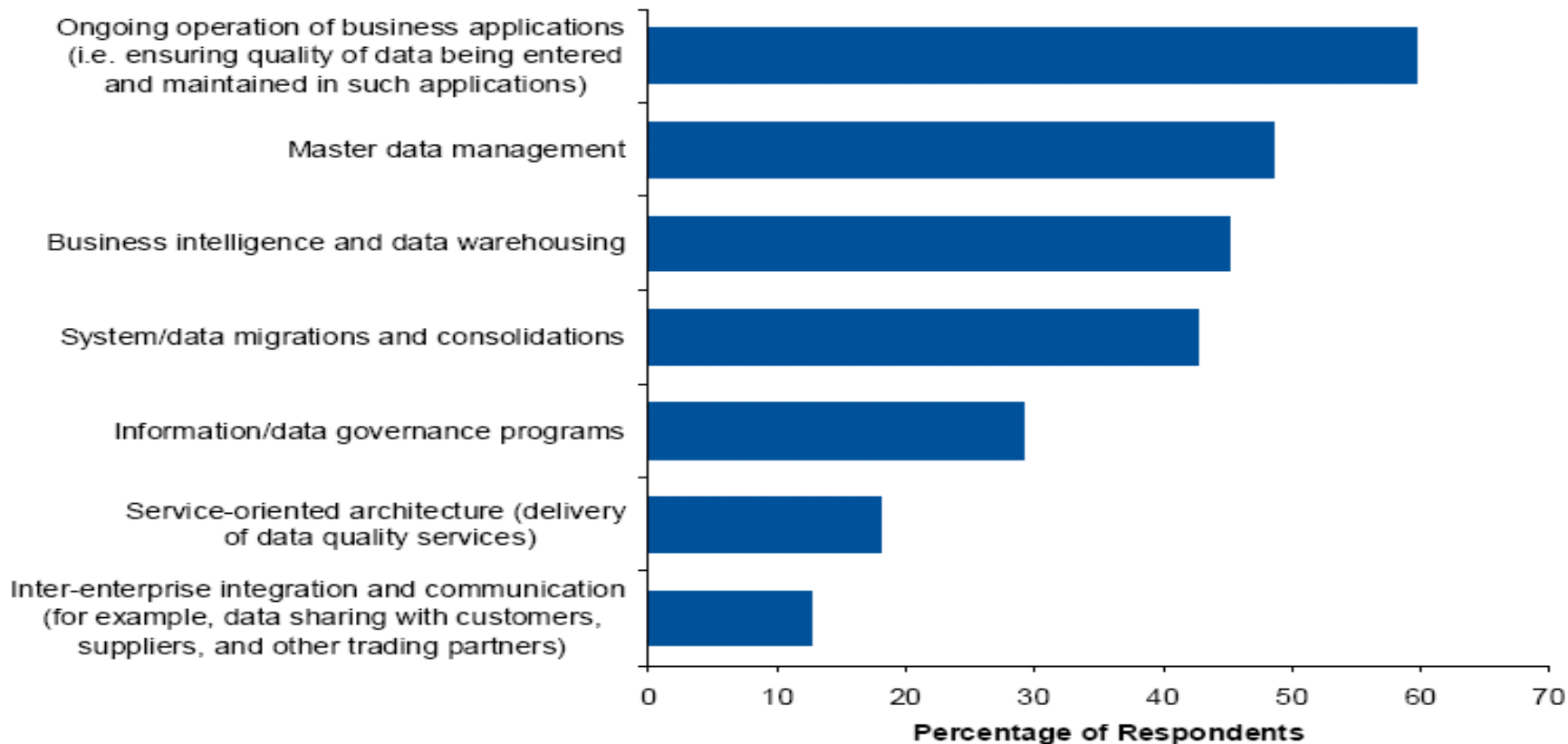


<sup>1</sup>AIIM 2008 Survey  
<sup>2</sup>Accenture 2007 Managers Survey

# Survey: Data Quality Software Is Viewed As Critical Technology



**Figure 1. Types of Applications/Initiatives in Which Data Quality Tools are Being Deployed**



***79% of survey respondents indicated they had deployed their tools of choice in more than one project or deployment, as compared to only 58% in 2008.***



## So, What Constitutes Data Quality?

**Data is standardized**

**Data is fit for purpose (conforms to rules)**

**Each record is unique**

**View of information is complete**

**Records are certified against authoritative sources**

**Lineage is understood**

**Data quality is measured over time**





## Recipe For Success: Involve the Business Early

### Recruit an executive sponsor

- Signals that the initiative is important
- Assures that funds continue to be available
- Discourages other business units from implementing conflicting projects

### Convene a data quality working group

- Assess and report on quality early in the process
- May coincide with implementation teams or data warehousing teams
- Business leads, but IT coordinates and facilitates
- Strive for consensus

### Have the business appoint a data quality steward for each business unit

- For business units with large user populations, several stewards are appropriate



## Recipe For Success: Control Scope Ruthlessly & Focus On Benefits

### Business must own scope

- Business should be owners, not renters
- IT maintains its independence by not taking sides
- Controlling scope encourages project discipline

### Iterate

- Projects which try to do it all in one pass generally fail

### Measure, Report, and Deliver benefits regularly

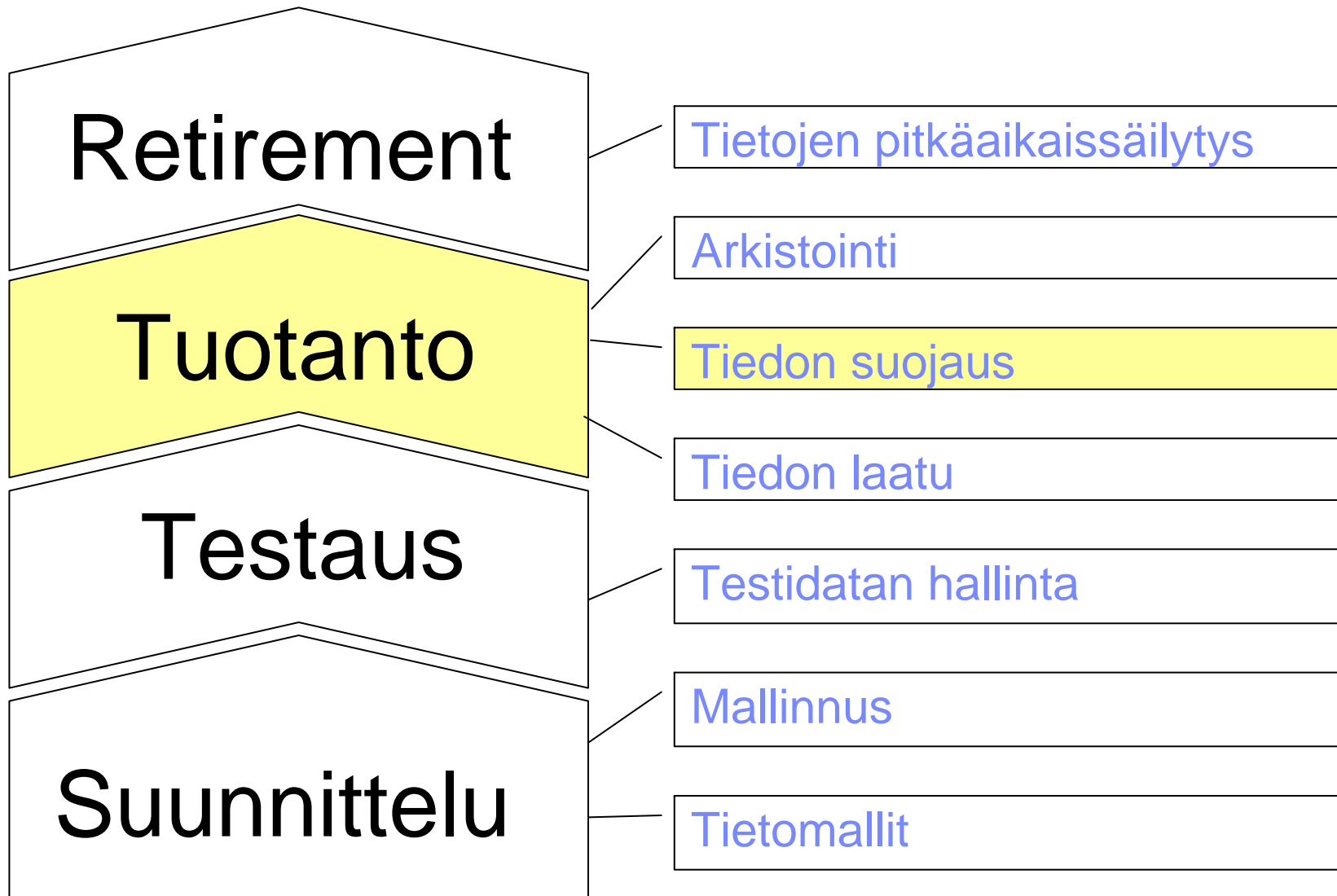
- Initial projects must provide some benefit within 6 - 9 months at the minimum (even if a small benefit)
- Subsequent phases should provide benefits every 3-6 months



## What Do You Need To Establish A Data Quality Program?

- A platform that can help analyze source data, centralizes quality rules, and provides auditable data quality
- Business-driven, data-centric design environment for data quality rules
- An ongoing process for data quality, and a way to measure quality over time
- A way to actively monitor and manage data
- The ability to remediate inconsistencies in data
- Data quality ownership and data governance
- Management sponsorship and a corporate mandate for data quality improvement







## Tarve aktiiviselle sovellustiedon suojaukselle

### Estä tietovarkaudet

- Minimoi ulkoiset ja sisäiset uhat



### Varmista tiedon eheys

- Estä auktorisoimattomat muutokset sensitiivisiin tietoihin



### Vähennä compliance-kustannuksia

- Automatisoi ja keskitä kontrollit
  - Eri tietokannoissa ja sovelluksissa
  - Eri vaatimuksia varten SOX, PCI, SAS70, ...
- Yksinkertaista prosesseja





# Database Danger from Within

“Organizations overlook the most imminent threat to their databases: authorized users.” (Dark Reading)

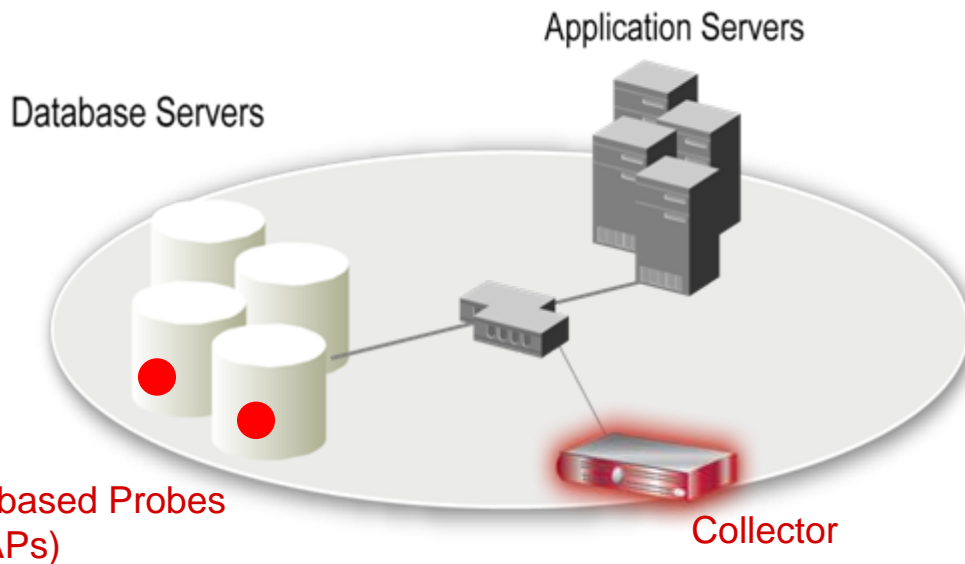
“No one group seems to own database security ... This is not a recipe for strong database security” ... 63% depend primarily on manual processes.” (ESG)

Most organizations (62%) cannot prevent super users from reading or tampering with sensitive information ... most are unable to even detect such incidents ... only 1 out of 4 believe their data assets are securely configured (Independent Oracle User Group).

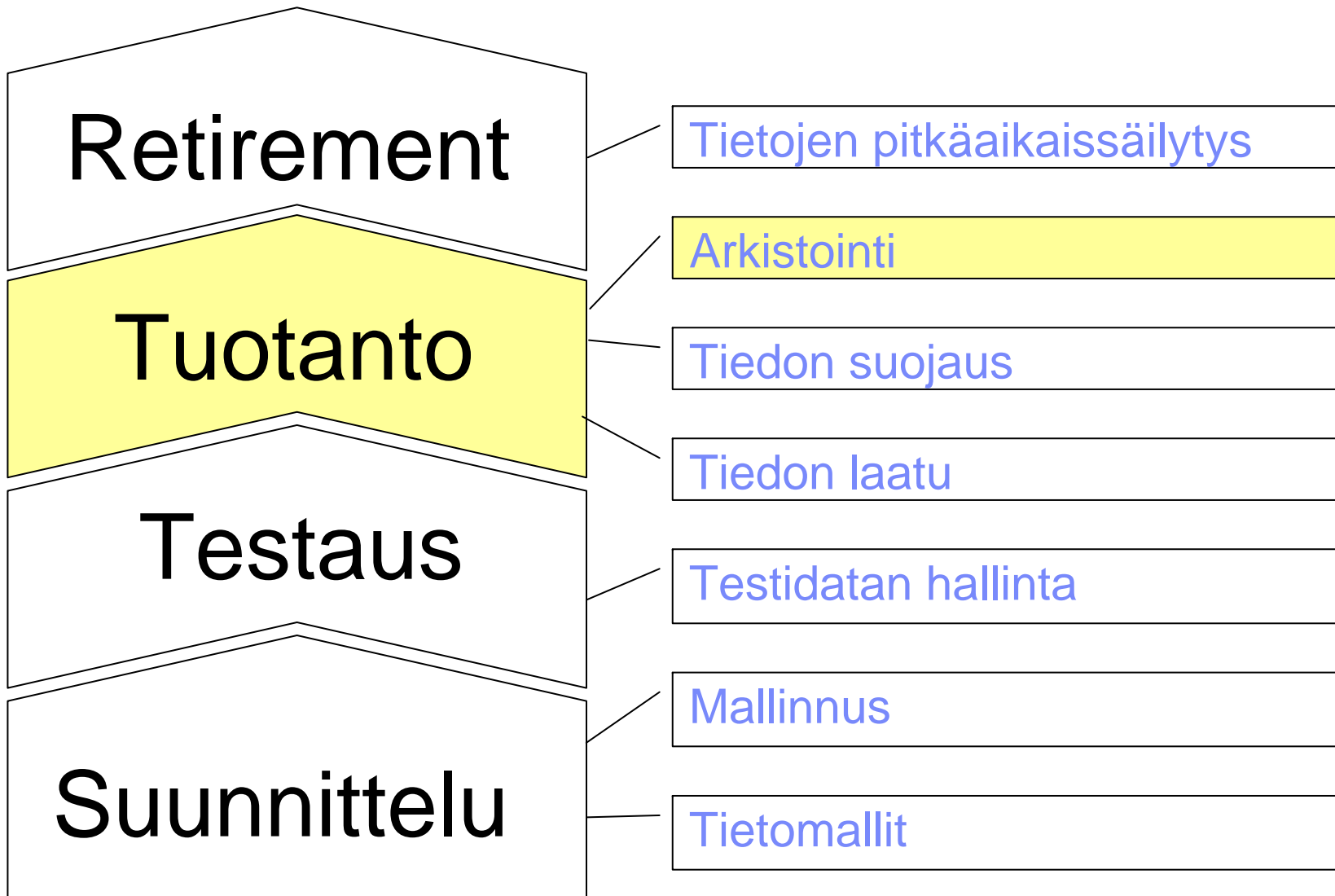


[http://www.darkreading.com/database\\_security/security/app-security/showArticle.jhtml?articleID=220300753](http://www.darkreading.com/database_security/security/app-security/showArticle.jhtml?articleID=220300753)  
<http://www.guardium.com/index.php/landing/866/>

# Real-Time Database Monitoring with InfoSphere Guardium

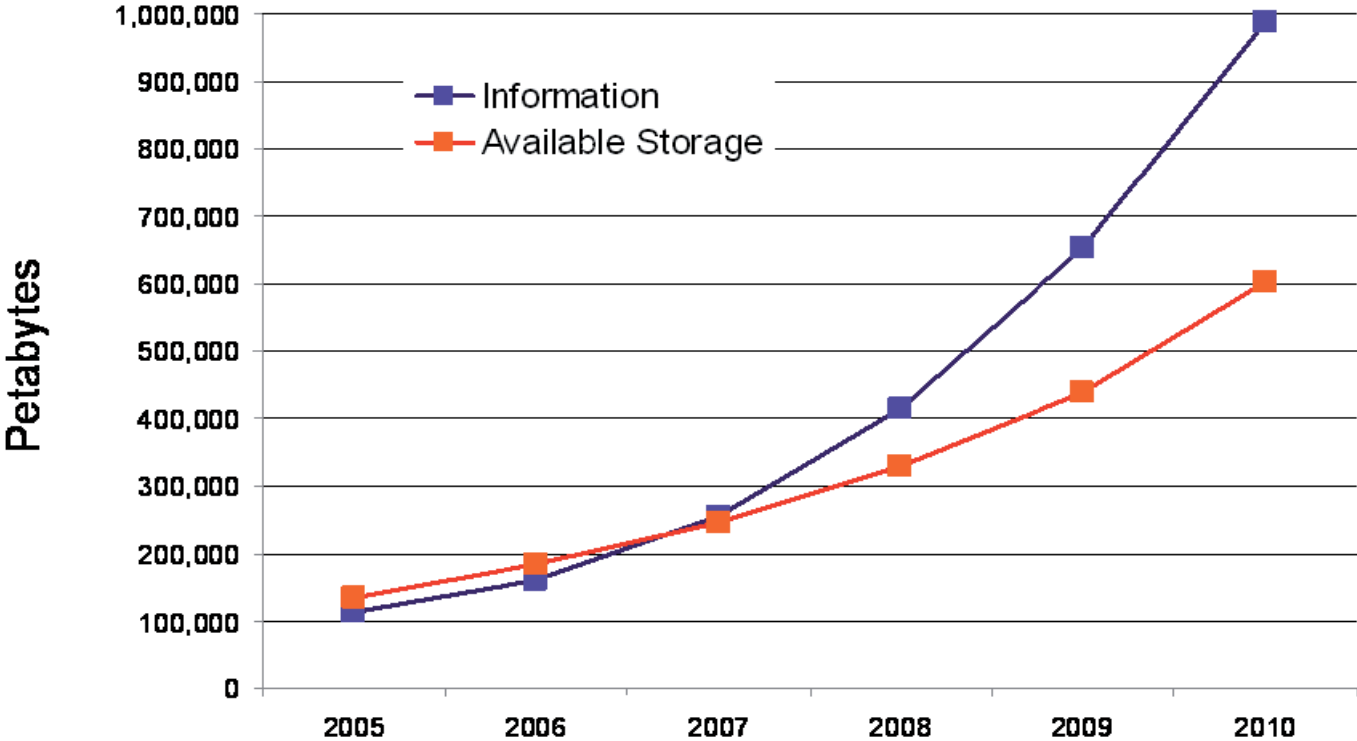


- Non-invasive architecture
  - Outside database
  - Minimal performance impact (2-3%)
  - No DBMS or application changes
- Cross-DBMS solution
- 100% visibility including local DBA access
- Enforces separation of duties
- Does not rely on DBMS-resident logs that can easily be erased by attackers, rogue insiders
- Granular, real-time policies & auditing
  - *Who, what, when, how*
- Automated compliance reporting, sign-offs & escalations (SOX, PCI, NIST, etc.)





# The Problem: Information Exceeds Storage Capacity!



\* Source: The Expanding Digital Universe , John F. Gantz, Research Director, IDC , March 2007

**Forrester estimates that, on average, data repositories for large applications grow by 50% annually (structured data)**

\* Source: Noel Yuhanna, Forrester Research, Database Archiving Remains An Important Part Of Enterprise DBMS Strategy, 8/13/07



## Challenges associated with Data Growth



### Business

- Performance degradation
  - Key business processes - payroll, shipping, financial period close, information reporting - not being completed on time
  - Applications not available when customers want to do business with you
- Regulatory compliance
  - ▶ Data retention (Sox, Euro-Sox, J-Sox)
- Budget Pressures
  - ▶ Increased hardware storage costs due to growing data

### IT



- Storage cost spiral
- Deteriorating service levels
  - Customer & employee complaints
  - Increased fire fighting
- Increased maintenance burdens
  - More time spent on tuning and partitioning
  - Longer backup windows
- Batch window creep



## Solution Requirements

Discover existing relationships across systems to ensure complete business object archiving

Manage data across the enterprise including multiple applications, databases, and platforms

Segment and manage data at the complete business object level

Archive to selected target format

- Compressed, indexed file
- XML file
- Archive database

Implement tiered storage strategies to minimize costs

- Nearline Storage
- CAS devices (EMC Centera, IBM DR550)
- Existing tape libraries

Multiple access methods to archived business records

- Native Application access (Siebel, PeopleSoft, JD Edwards, E-Business Suite)
- Application Independent access (Original app/version is not needed)



## How Does Archiving Improve Performance?

### Improved Availability

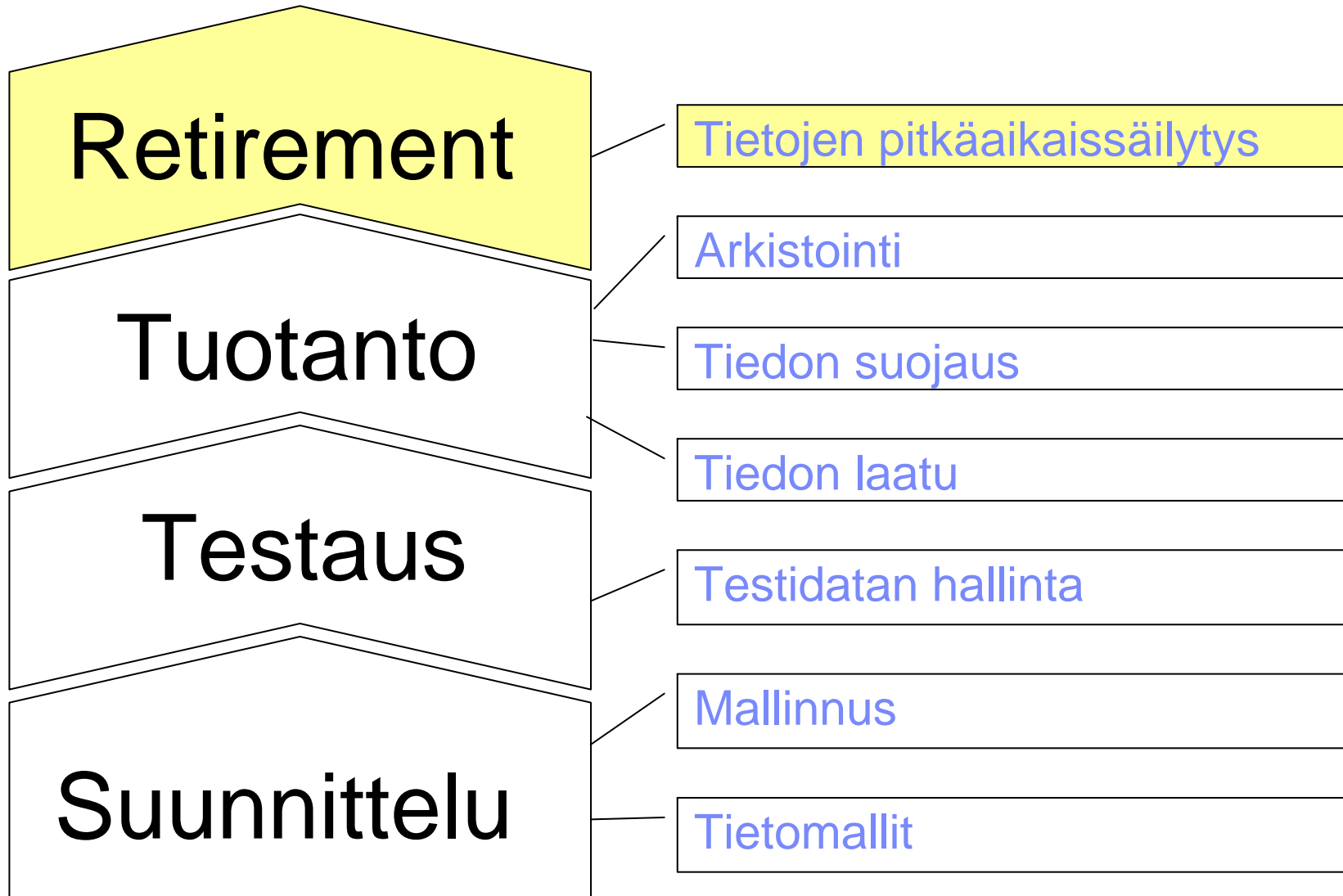
- No downtime caused by batch process overruns
- Uptime during crunch time
- Meet SLAs

### Speeding Backup and Recovery

- Bring up important/recent data first
- Bring up older/reference data as conditions permit

### Improved Application Performance

- One of the most understated benefits to archiving
- Longest and most lasting benefit



# Universal Access

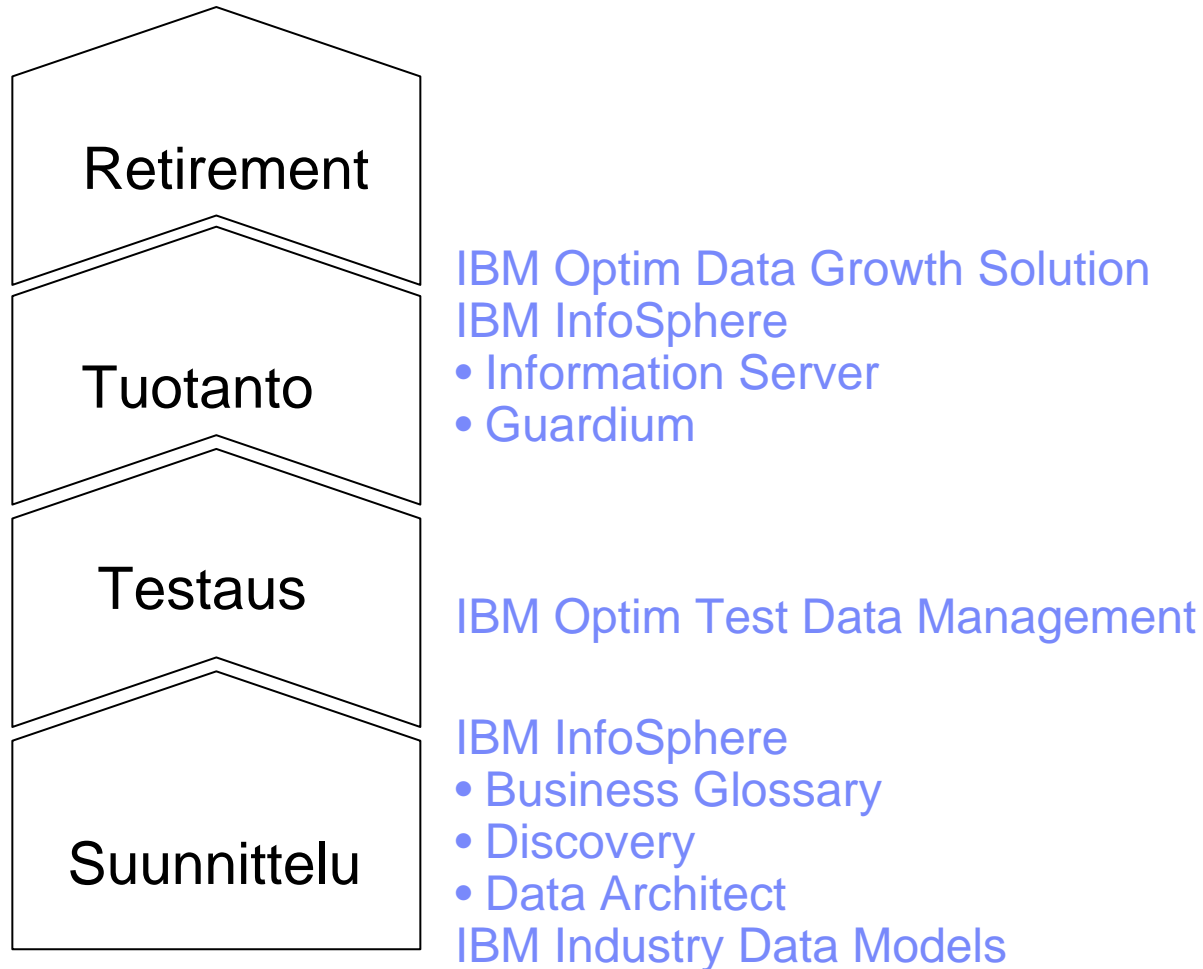


## Application independent access

- Industry standard methods: SQL, ODBC/JDBC, XML
- IBM Mashups
- Portals
- Report writers: Crystal Reports, Cognos, Business Objects, Discoverer, Actuate
- Desktop formats: Excel, CSV, MS Access
- Database formats



## Sovellustaso: kehdosta hautaan



# Lopuksi

Kyky hyödyntää tietoa on keskeinen kilpailutekijä kaikille organisaatioille

Keskittyminen operatiiviseen tietoon ei riitä. Tarvitaan

- Kyky hyödyntää tietoa konsernitasolla eri operatiivisten sovellusten yli
- Ymmärrys tiedosta uudelleenkäyttöä ja jatkojalostusta varten
- Mahdollisuus tiedon hyödyntämiseen ilman sen tuottanutta sovellusta

Tarvitaan strategia koko tiedon elinkaarelle – kehdestä hautaan