



Education

# Using Classification to manage File Servers

Nir Ben-Zvi, Microsoft Corporation

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - ◆ Any slide or slides used must be reproduced in their entirety without modification
  - ◆ The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

**NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.**

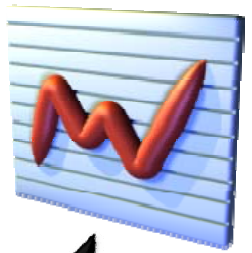
## ➤ Using classification to manage File Servers

- ◆ As data growth is exploding, companies are struggling to manage the “Risk” and “Cost” of the increasing amounts of files stored on file servers. Traditionally, data management applications use a directory based approach to manage file servers. This session introduces the concepts and opportunities for using classification to manage data based on its business value. The guiding principle is that an organization can classify files on file servers and then apply data management policies based on this classification.

- Discuss how classification can be used to solve business problems
- Walkthrough a proposal for a classification infrastructure implementation

# File server trends

Storage growth



Storage cost



Data sharing and search



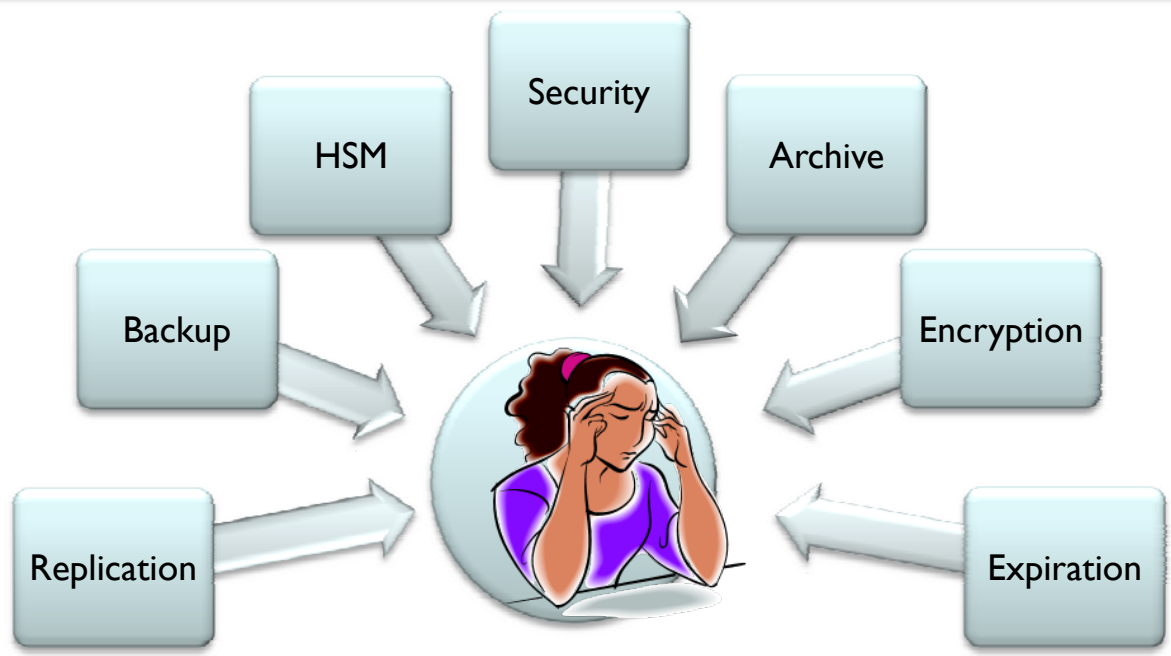
Compliance



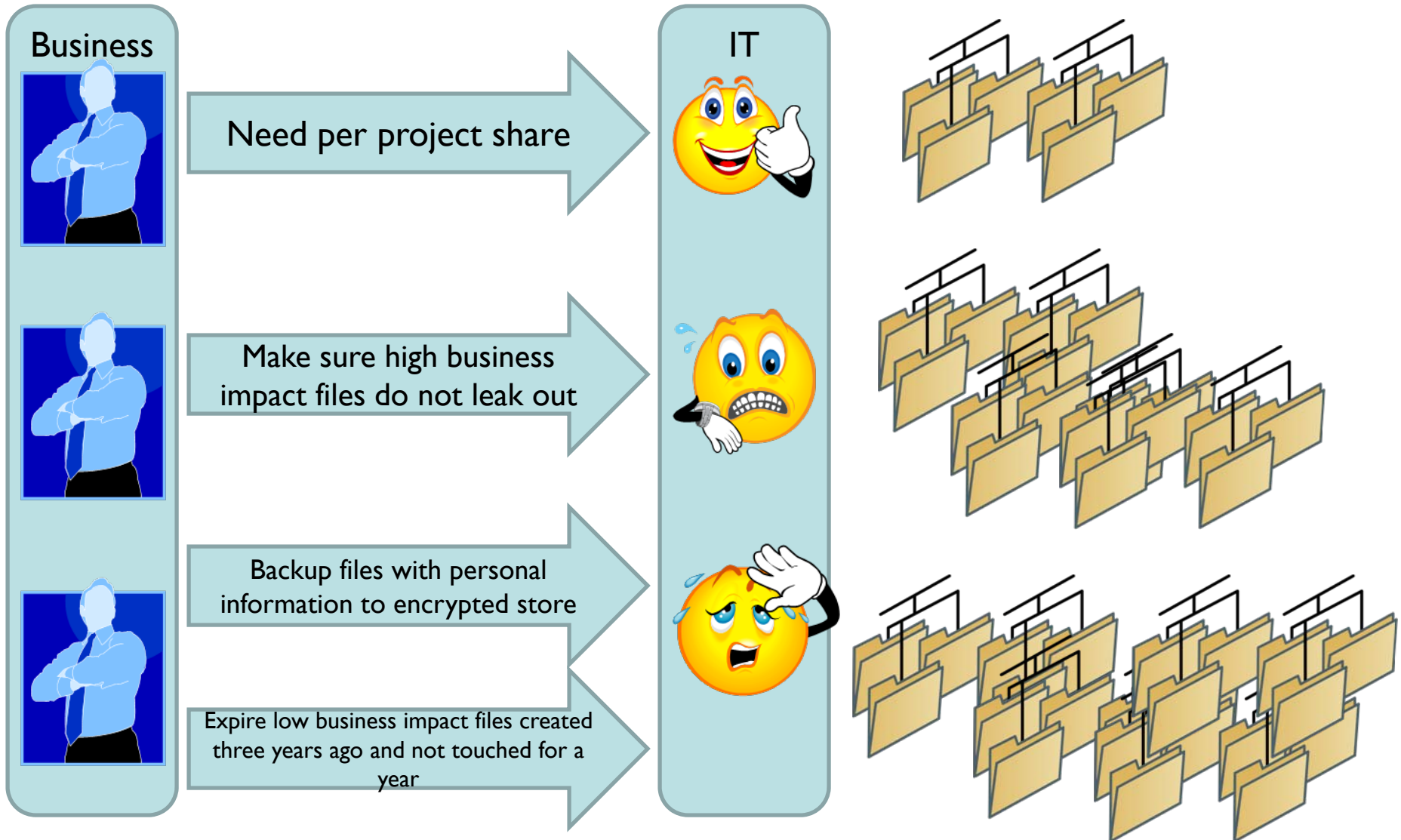
Security and Information leakage



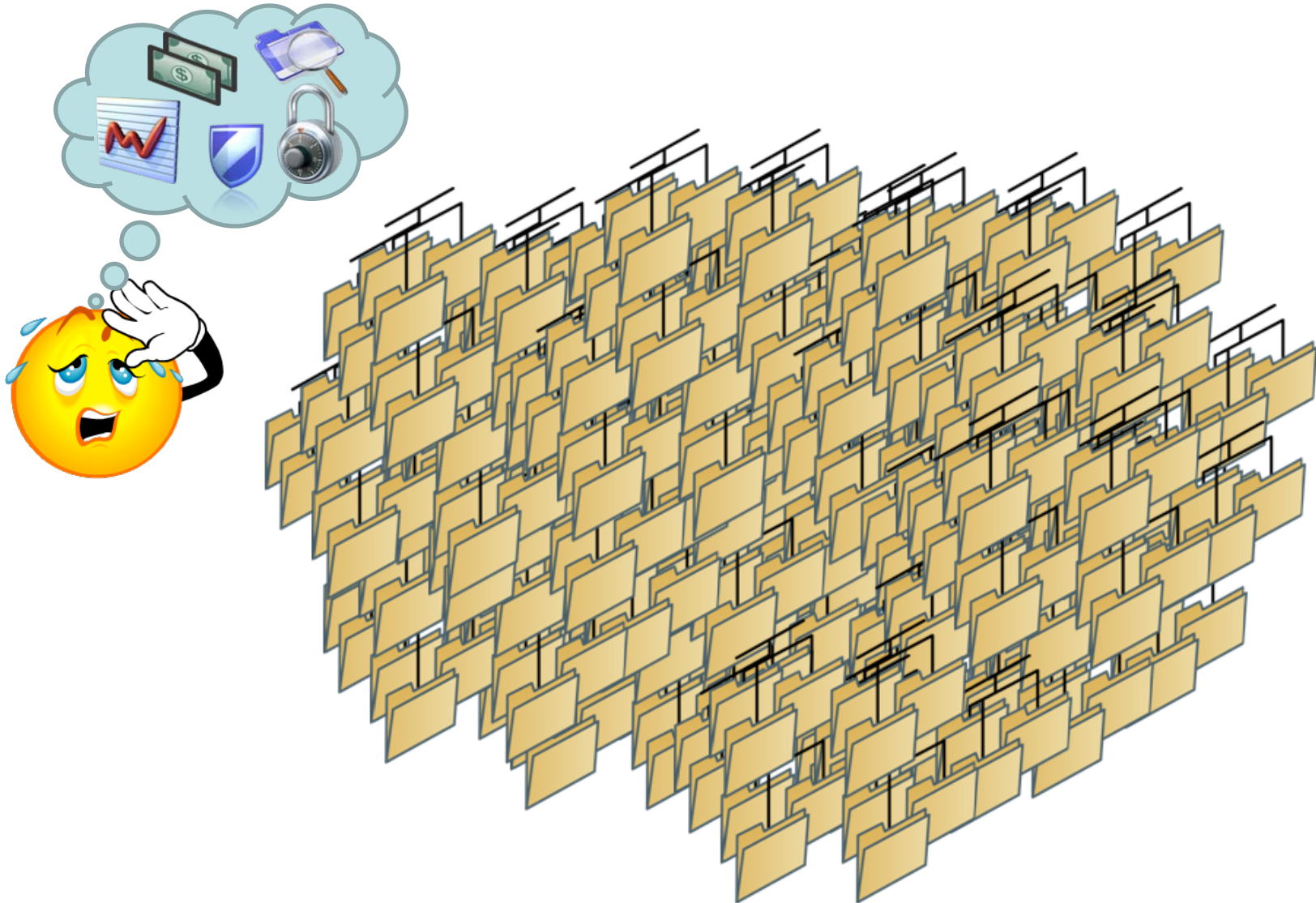
Increasing data management needs / many data management tools



# File shares and business requirements



# Some time later ...



Cost and Risk



Manage data based on business value





# Classify and apply policy

Step 1:  
Classify data

## Classification methods

IT Scripts

Manual

Line Of Business application

Automatic classification

- Location
- Content
- Owner



Step 2:  
Apply policy  
based on  
classification

## Actions based on classification

Backup

Expiration

Search

Archive

Replication

Leakage prevention

HSM

Security

Reports

Encryption

- **Enable companies and organization to:**
  - ◆ Define their classification properties (e.g.: Secrecy, Personal)
  - ◆ Control which data should be classified
  - ◆ Apply classification policies (e.g.: What is considered personal information)
  - ◆ Manage data based on classification
- **Interoperability between products:**
  - ◆ Classification products used to classify files
  - ◆ Data management products used to apply data management policies based on classification
- **Provide flexibility to adjust in continually changing business environments**



- Define classification properties
  - ◆ Taxonomy defined by the business owners and implemented by the IT organization
- For example:
  - ◆ Business impact = high/medium/low
  - ◆ Personal information = true/false
  - ◆ Project = data scanning
- Universal properties vs. local properties
  - ◆ Universal makes it easier when moving files between organizations
  - ◆ Local allows flexibility and agility for changes



## ➤ Identify scope of files to be classified

- ◆ Discover files to be classified
- ◆ Scan the file servers on a schedule basis

## ➤ Identify changes

- ◆ Full scan for every classification process
- ◆ Use file system change log to discover files that need to be classified
- ◆ Real time discovery of files that changed



## ➤ Label files with classification properties

- ◆ Manually by users (information worker)
- ◆ Line of business applications and IT scripts
- ◆ Automatically

## ➤ Automatic classification

- ◆ Evaluate the value of property(s) for a given file
- ◆ Examples are: Based on Location, Content, Owner ...

## ➤ Aggregation policy for property values

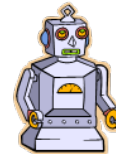
- ◆ Multiple classification mechanisms might return different results for the same property value

## ➤ Classification is best effort

- ◆ Need to deal with classification errors
- ◆ Example: White listing for personal information classification



- Classification properties can be stored in multiple places
  - ◆ In the file
  - ◆ Adjacent to file content
  - ◆ Database
  - ◆ Cloud
- Need a model for determining the authoritative value of the property for a file when it is stored in multiple places
- Maintaining classification properties is a challenge
  - ◆ When the file moves (or sent via email ...)
  - ◆ When the file is modified



- Enable actions based on classification condition
  - ◆ Example: Expire files where Business Impact=Low and Last access > a year ago
- Query file classification to match condition
  - ◆ Example: What is the value of Business Impact for a specific file
- Apply actions
  - ◆ Immediately when files are classified
    - Example: Encrypt files that are classified as having personal information
  - ◆ On a schedule/manual basis
    - Example: Backup all files where Business Impact=High every day

## Based on business value ...

### Reduce Cost

- Expire files to reduce storage purchasing needs
- Move files to less expensive storage
- Optimize backup SLAs
- Replicate only business related files

### Manage risk

- Find sensitive files on public servers
- Watermark documents
- Keep files containing personal information encrypted in backup
- Apply rights management to high secrecy files
- Comply with retention policies

- **Using classification to determine policy vs. applying policy based on classification**
  - ◆ For example: Set a property on a file to specify 3 years retention policy vs. Set a property on a file to specify SOX and then apply 3 years retention policy based on SOX classification
- **File movement classification implications**
  - ◆ Do files need to be reclassified when they are moved
  - ◆ Striping classification when files are moved through the organization boundaries
- **Aggregation of multiple potential values**
  - ◆ When classifying files
  - ◆ When retrieving property values stored for the file
- **Inaccuracies and failures in classification**

Classify data



Apply policy  
based on  
classification



Plan  
classification  
properties  
(taxonomy)



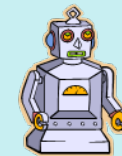
Identify  
files to be  
classified



Classify  
files according to  
organization  
policy



Store  
Classification  
properties  
assigned to files



Manage  
Files based on  
classification

# Refer to Other Tutorials

- Please use this icon to refer to other SNIA Tutorials where appropriate.



**Check out SNIA Tutorial:  
Enter Tutorial Title Here**

- Please send any questions or comments on this presentation to SNIA: [trackfilemgmt@snia.org](mailto:trackfilemgmt@snia.org)

**Many thanks to the following individuals  
for their contributions to this tutorial.**

**- SNIA Education Committee**

**Calvin Keaton  
Matthias Wollnik  
Mathew Dickson  
Adi Oltean  
Ran Kalach  
Calvin Keaton**

**Paul Massiglia**