The Intersection of Cloud and Metadata

Some initial thoughts

Benjamin S. Woo
Program Vice President
WW Storage Systems Research, **IDC**

Agenda

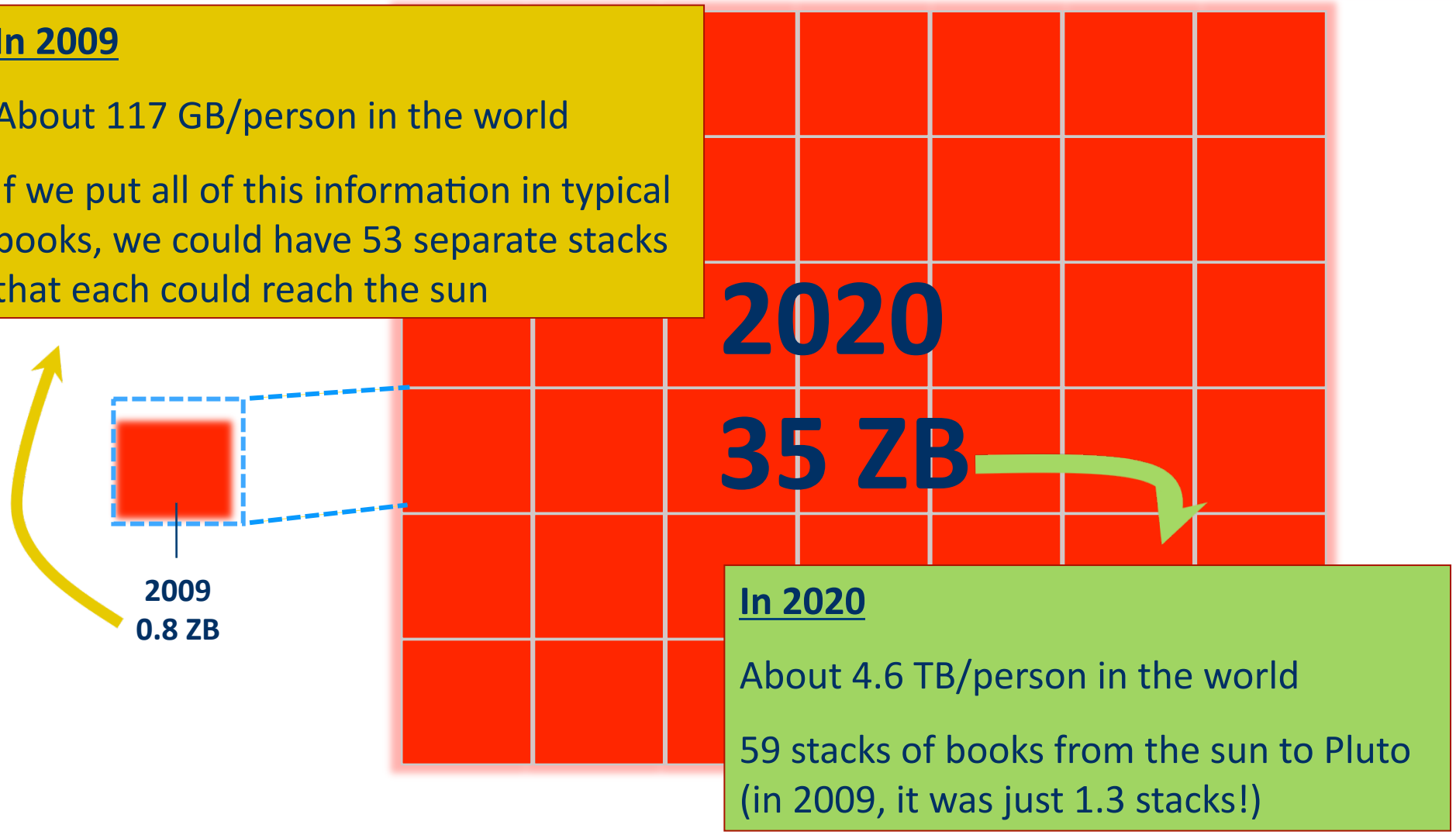
- It's *just* metadata
- Fine! It's metadata, so what?
- OK, *now* what?

The Growing Digital Universe

In 2009

About 117 GB/person in the world

If we put all of this information in typical books, we could have 53 separate stacks that each could reach the sun



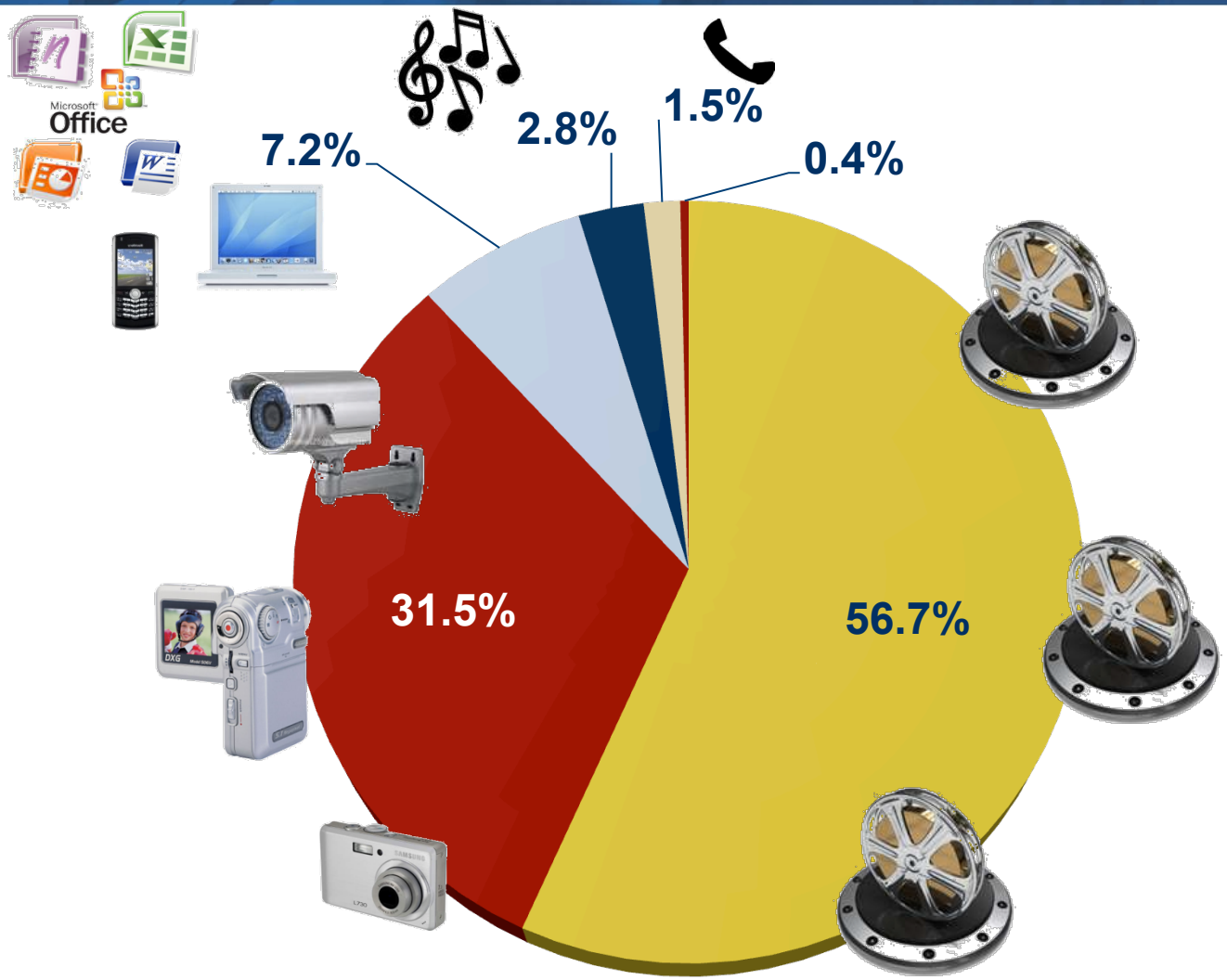
In 2020

About 4.6 TB/person in the world

59 stacks of books from the sun to Pluto (in 2009, it was just 1.3 stacks!)

Source: IDC Digital Universe Study, sponsored by EMC, May 2010

Fueling the Digital Universe

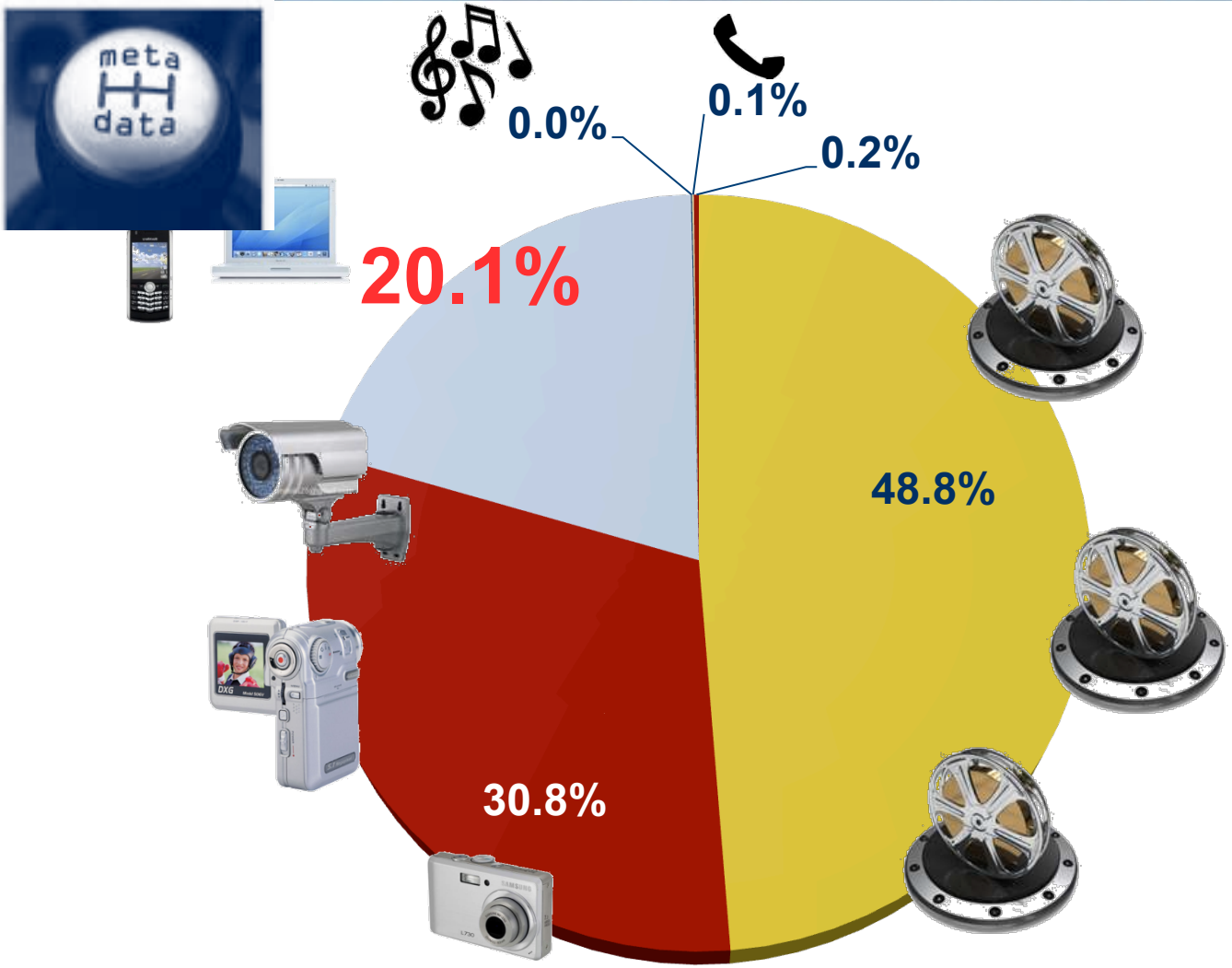


2009

0.8 ZB

- Media
- Cameras
- Data
- Music
- Voice Capture
- Other

Fueling the Digital Universe

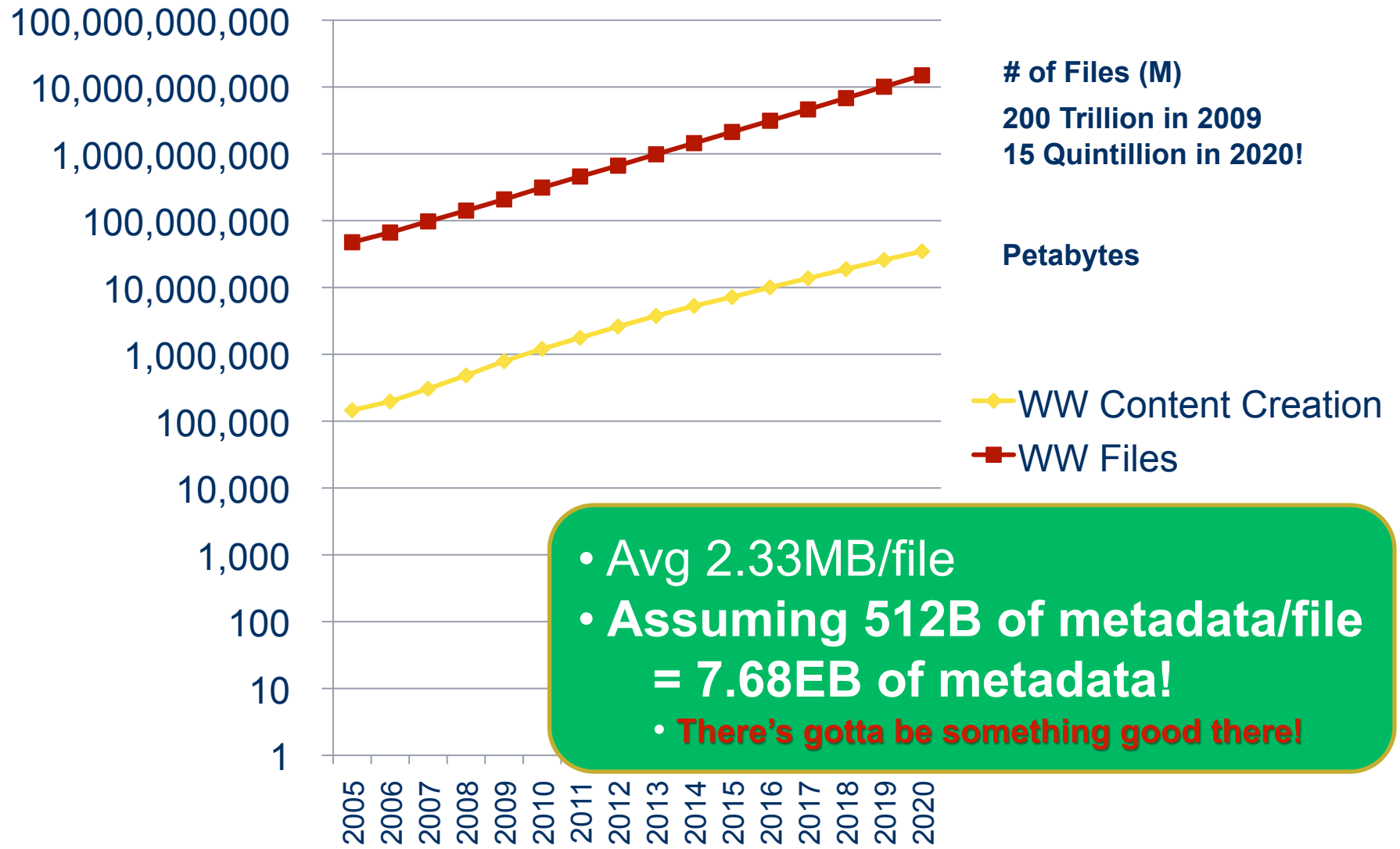


2020

35 ZB

- Media
- Cameras
- Data
- Music
- Voice Capture
- Other

Digital Universe Growth – Content vs # of Files



- Avg 2.33MB/file
- **Assuming 512B of metadata/file = 7.68EB of metadata!**
  - **There's gotta be something good there!**

Source: IDC Digital Universe Study, sponsored by EMC, May 2010

Distinguishing between consumption and creation using the cloud

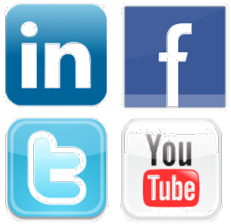
Consumption

- Many device types
  - DT / LT / NB
  - Tablets
  - Smartphones



- Loads of data types
  - Emails
  - Rich Content
  - Social Networking





Creation

- Fewer device types that can create all data types
  - DT/LT/NB → Office Docs
  - Specialized devices

- Office Docs
- Sensors


Ways to leverage metadata

Methodologies for leveraging metadata

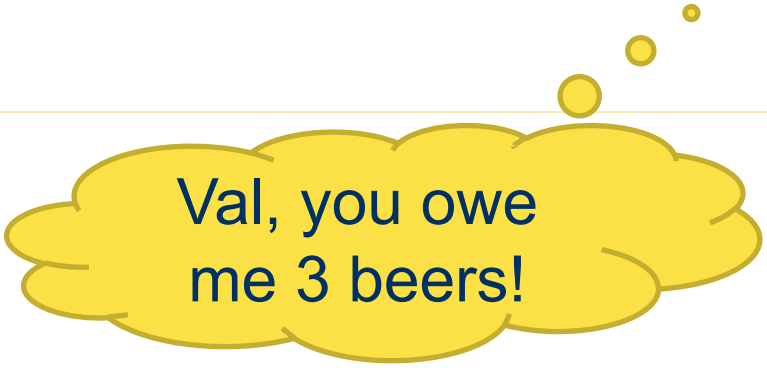
- On-Premise to Cloud → Content Creation
- Cloud to On-Premise → Content Consumption
- Cloud to Cloud → Content manipulation

MATURE

Emerging

Different Protocols

	Proprietary	Industry Pervasive	Open
On Premise → Cloud	Xdrive Backup to Cloud	Oracle Fusion	SNIA CDMI
Cloud → On Premise	Pogoplug	Oracle Fusion	SNIA CDMI
Cloud → Cloud	Google Apps Facebook	Oracle Fusion	SNIA CDMI



Use cases for metadata

Use Case	Sample Use Case
Search & Retrieval	Salesforce.com
Referential	Google search Microsoft Bing
Relational	Facebook LinkedIn
Comparative	eDiscovery

OK, that's nothing new … Tell me something I don't know

Emerging use cases for metadata will come through cloud-to-cloud implementations

- Examples: Oracle Applications and vApps built on SpringSource

Very likely to be a mash up of multiple sources of data

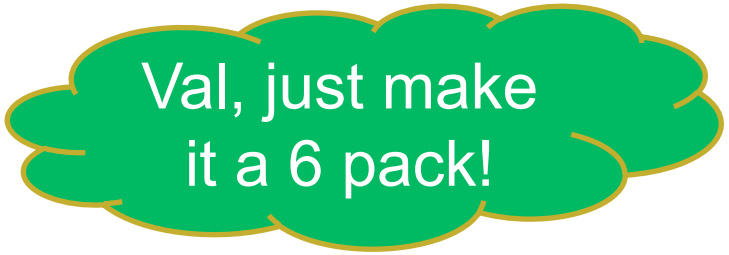
Metadata will serve to enable a "*pre*-query" to understand whether a full query of the dataset is required or desired

Also very likely to leverage a combination of industry pervasive protocols with industry standard protocols

Essential Guidance

For SNIA members …

- Storage systems need to evolve beyond on-premise protocols (CIFS/NAS, FC, iSCSI)
- Storage systems will need a WAN interface that talks open industry standard "cloud" protocols (eg. SNIA CDMI)
- … But the data packet has to be interpretable by industry pervasive platforms (eg. Oracle Fusion, VMware/SpringSource)
- Only then, will a storage system be able to be truly a "private cloud" content depot



Q&A

Shoot me an email at
bwoo@idc.com