

Are You Ready for Yottabytes?

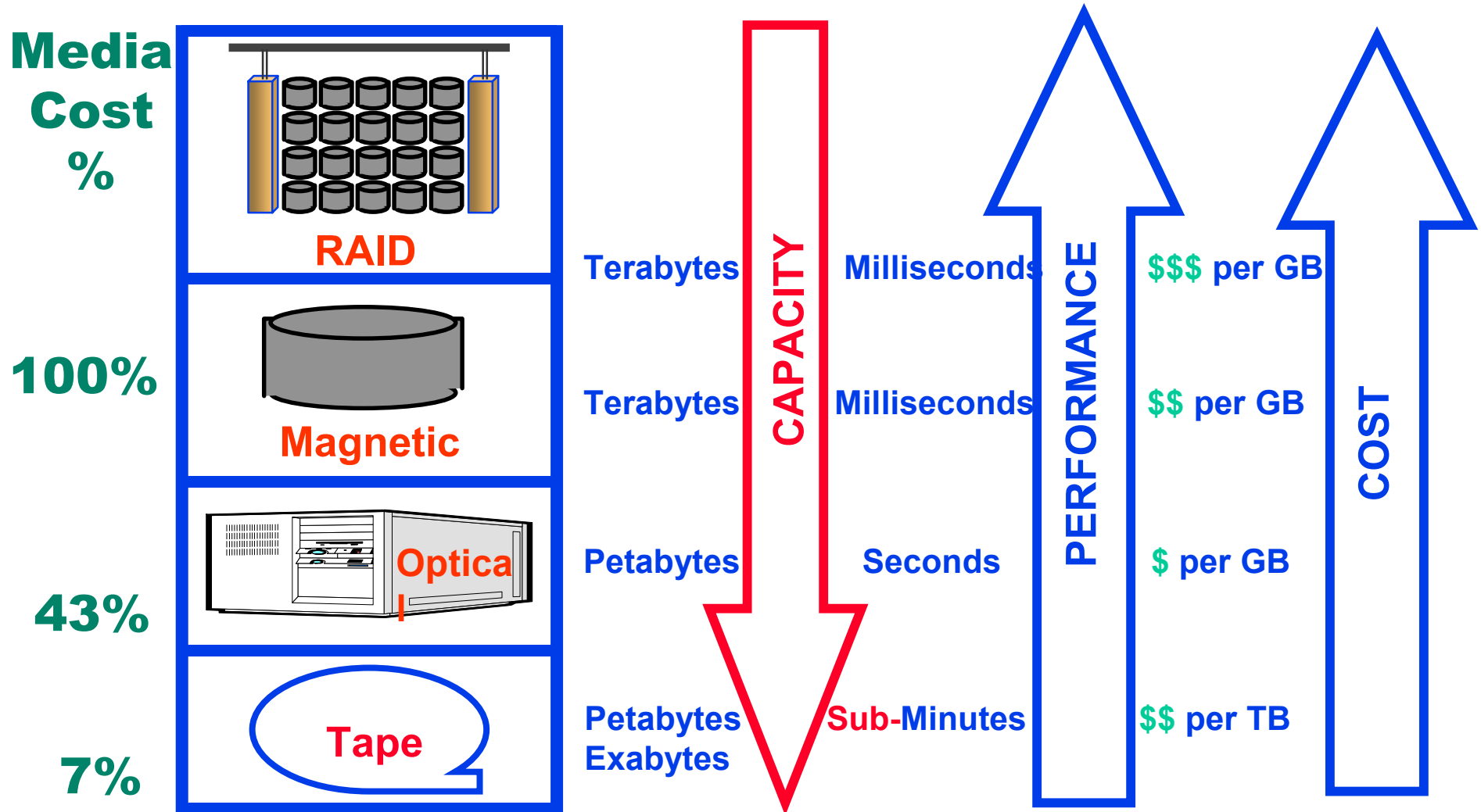


NASA/IEEE Mass Storage Symposium

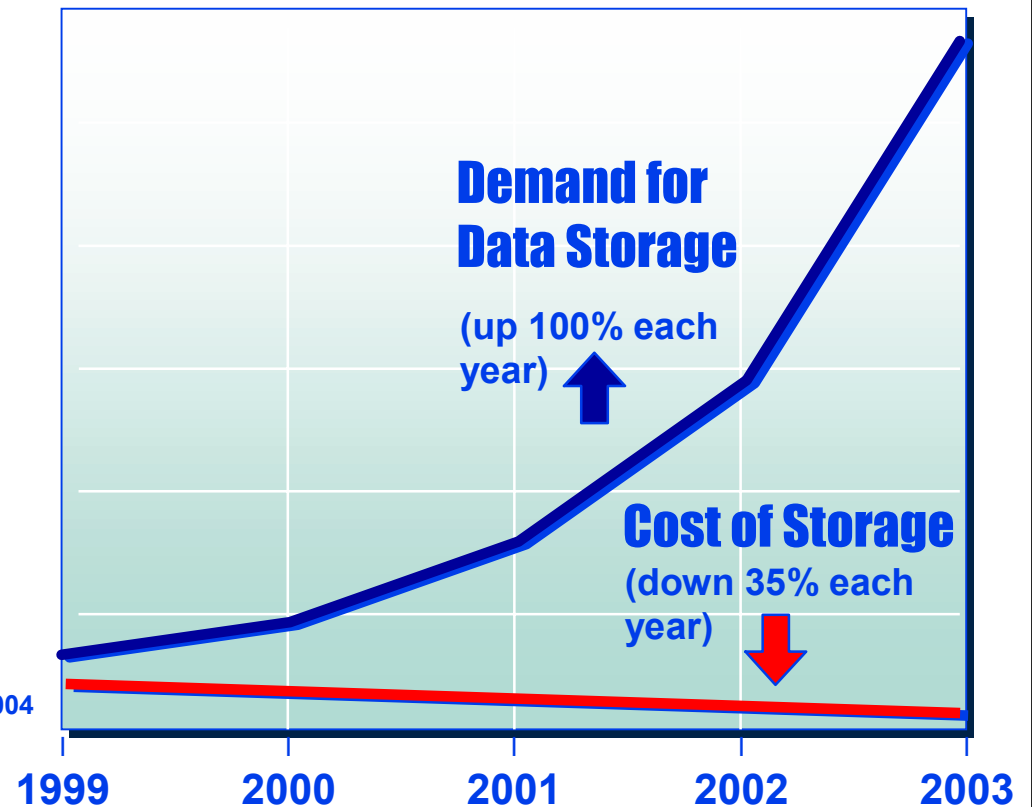
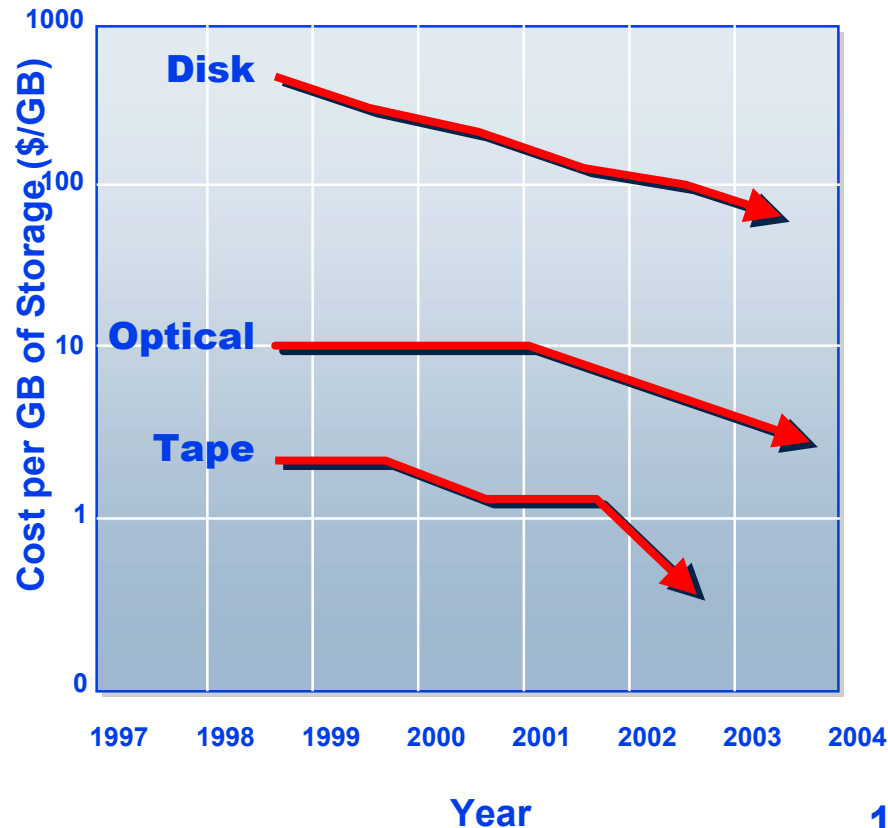
Felipe Cariño, Jr.



Capacity, Cost & Performance Tradeoffs

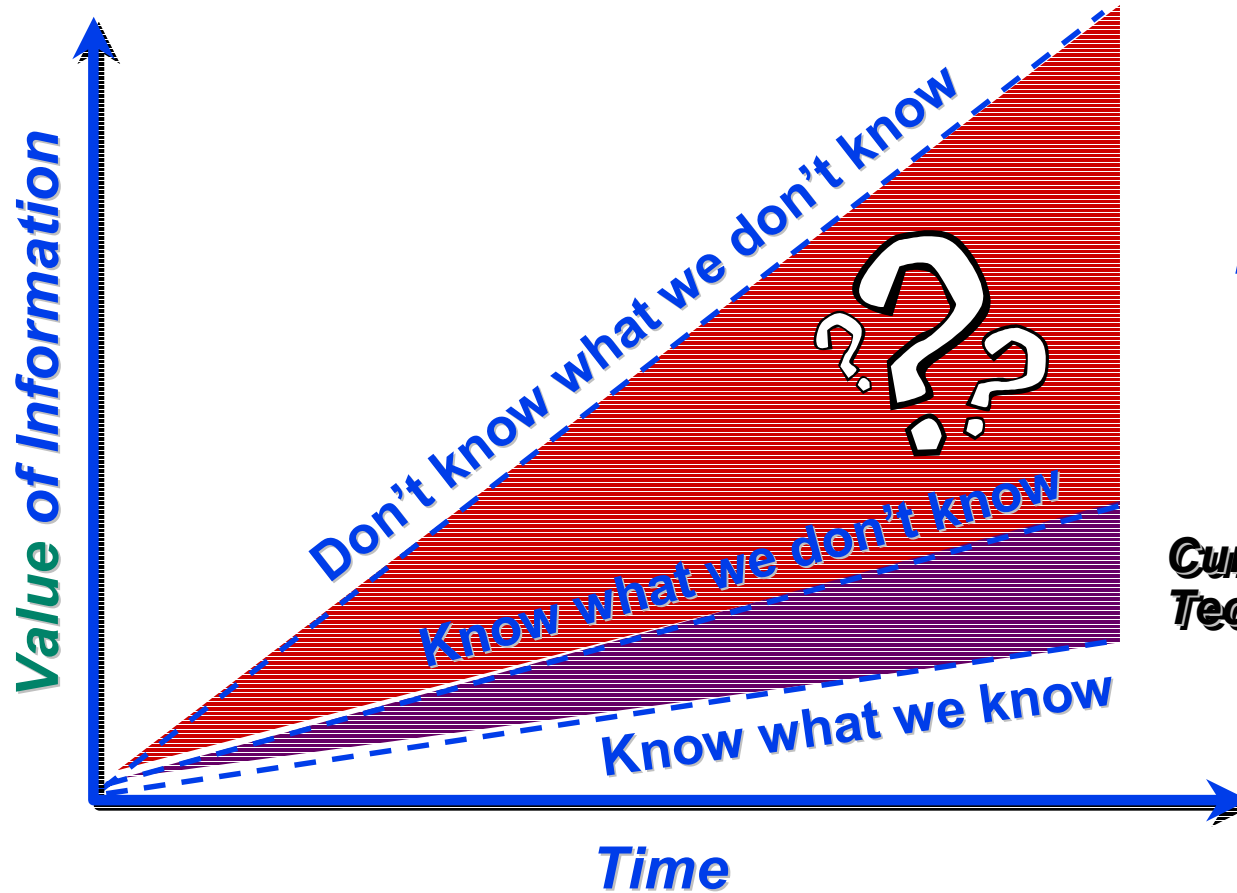


Storage Story - Cost & Usage Trends



Therefore, it doesn't matter what the price of RAID/DASD does -- the customer's data needs will continue to outpace price decreases!

The Fact Gap

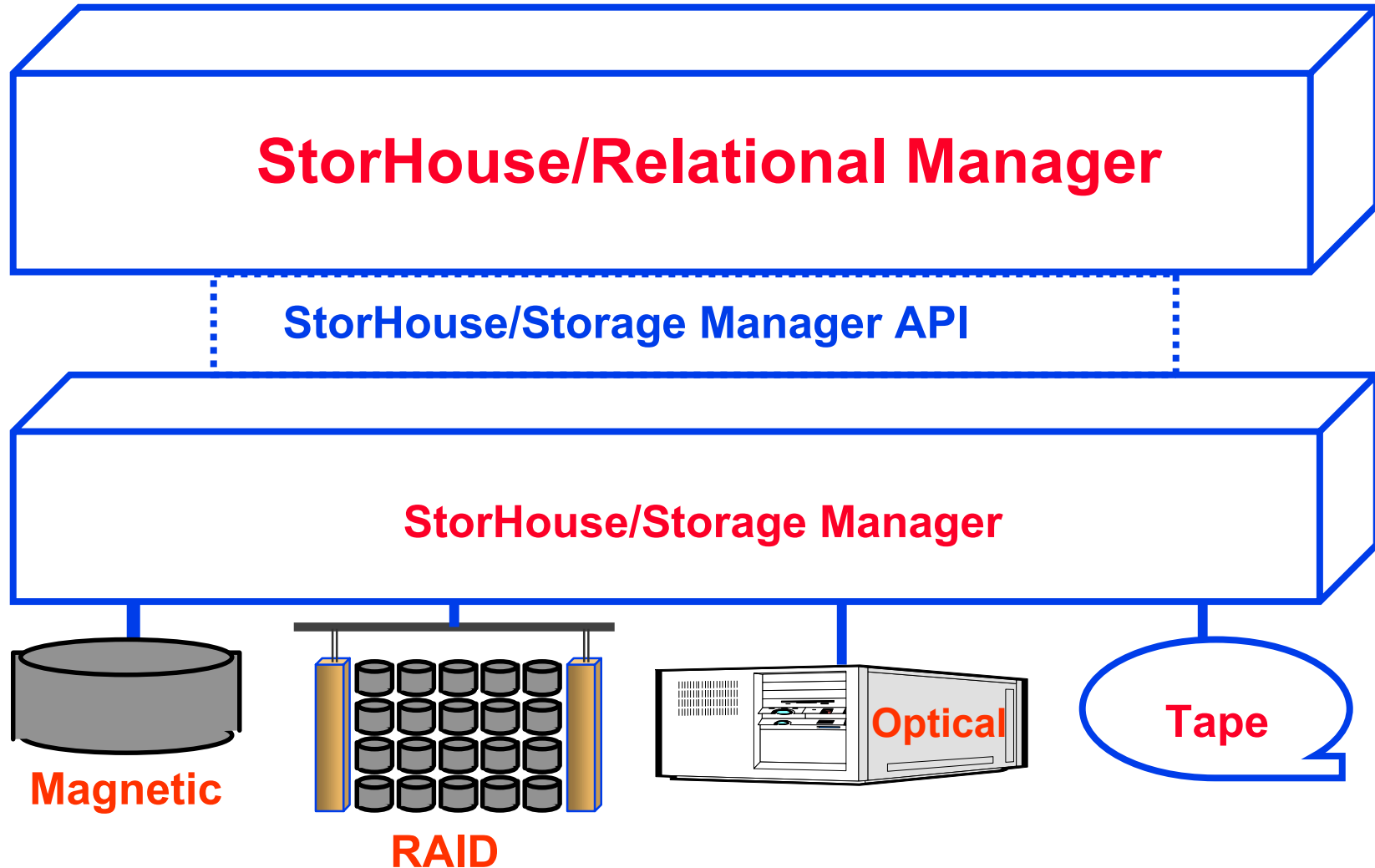


***Need for
Atomic
Data***

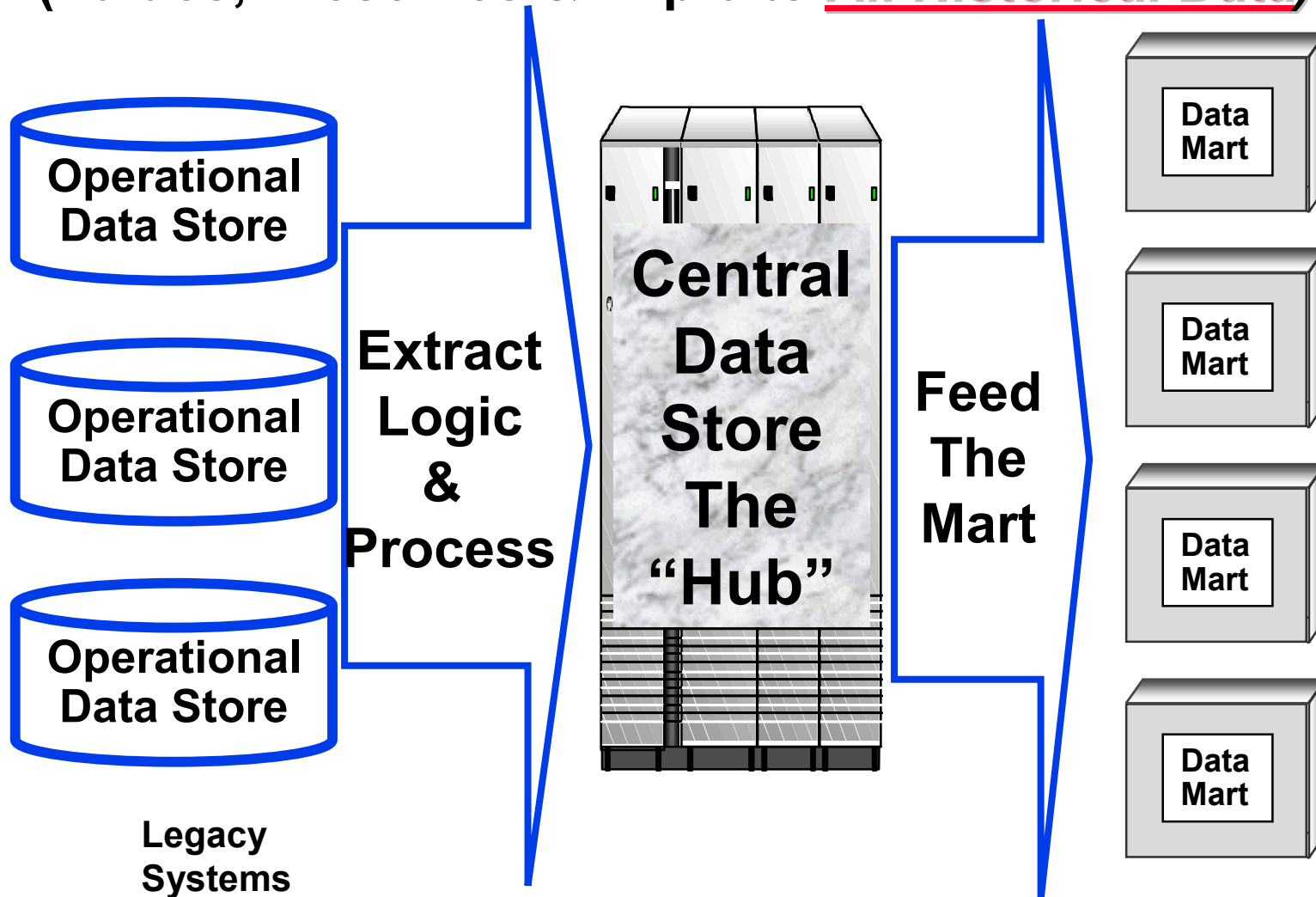
***Current Discovery
Techniques***

FileTek

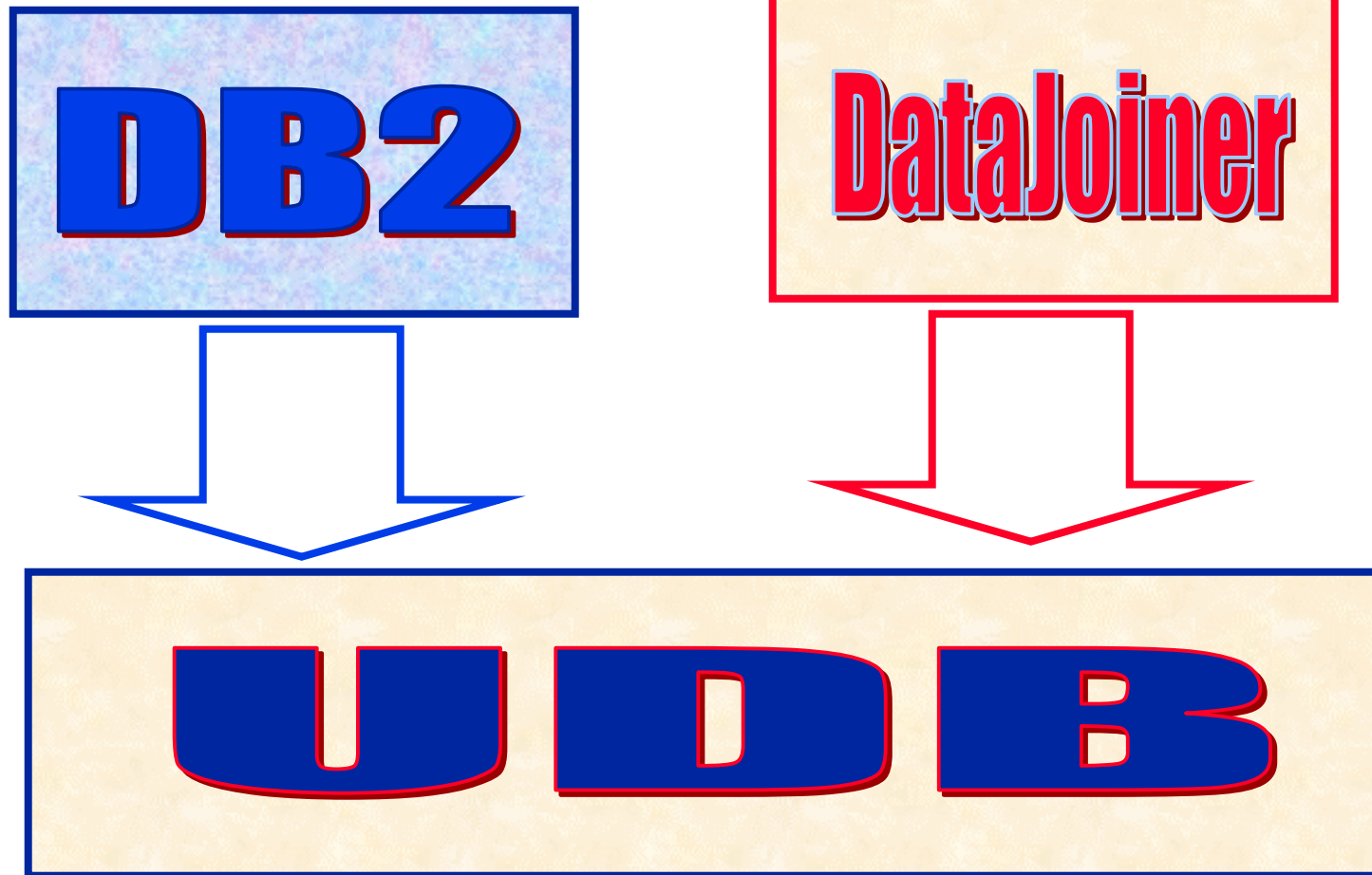
Stealth Storage Cost Barrier - Broken!



New **Atomic** (Not Summary) Data Warehouse Concept (Values, Preserves & Exploits All Historical Data)



IBM DB2 DataJoiner Product Plans



DB2 UDB - New Options

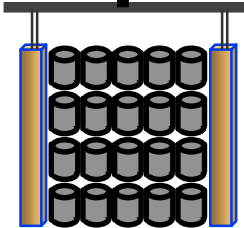
Today

New Today

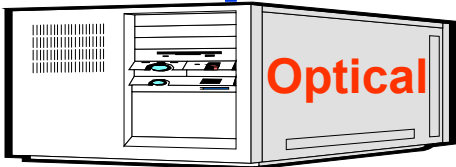
UDB

UDB

StorHouse/RM



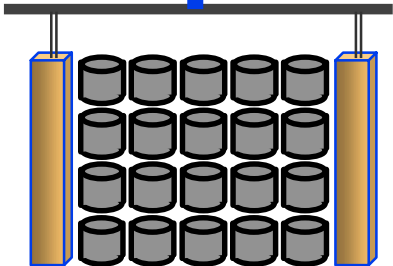
RAID



Optical



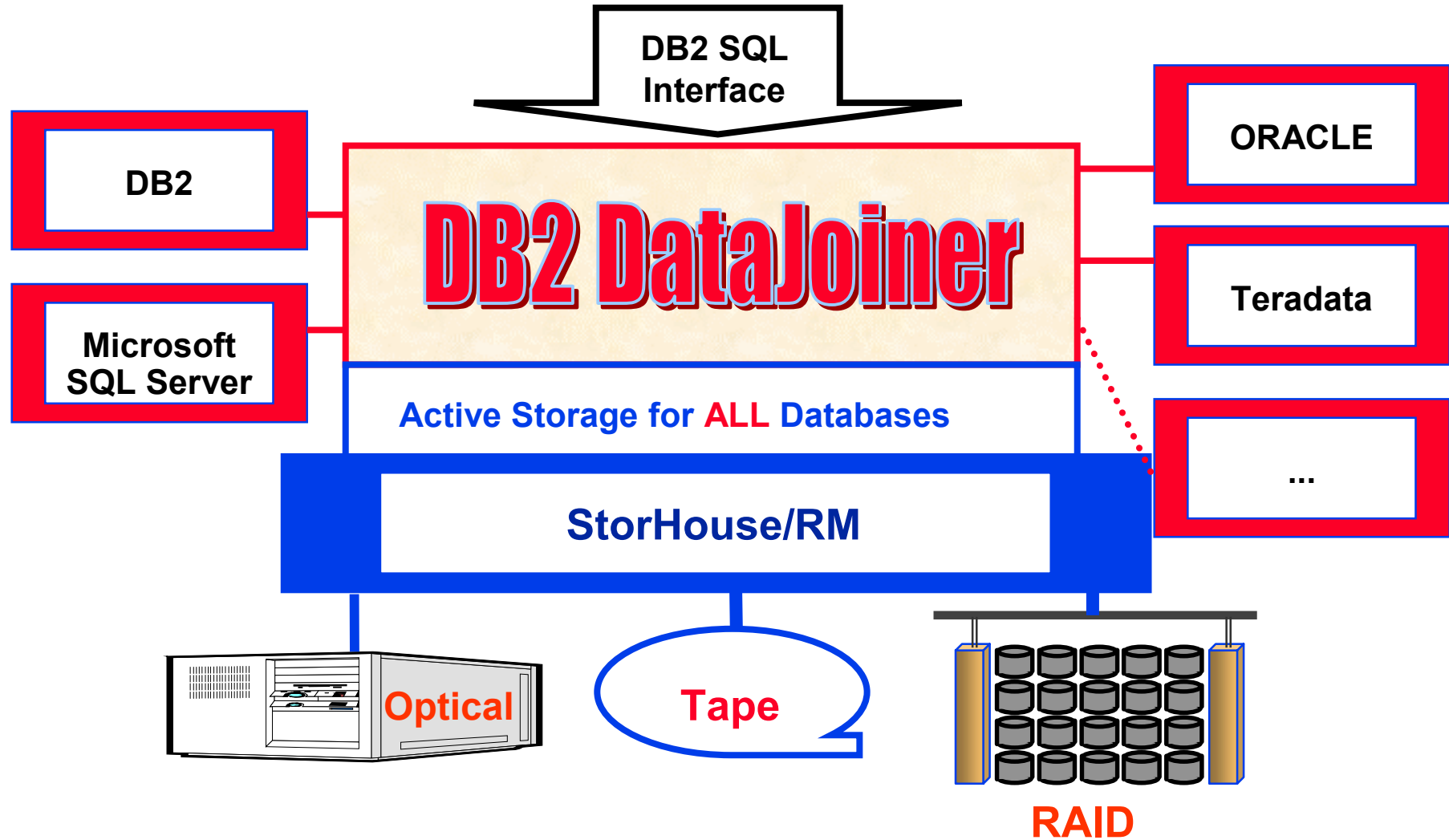
Tape



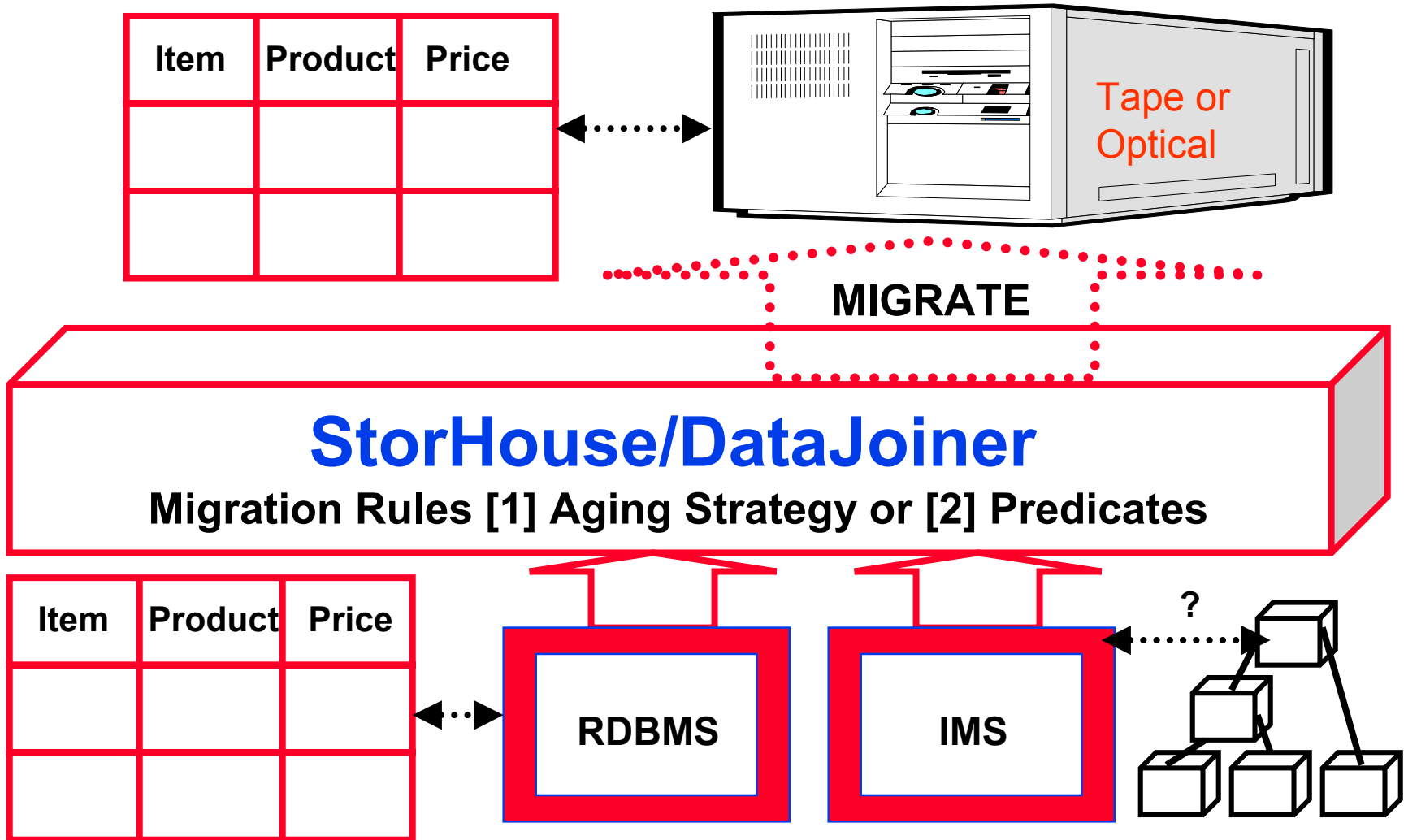
RAID

FileTek

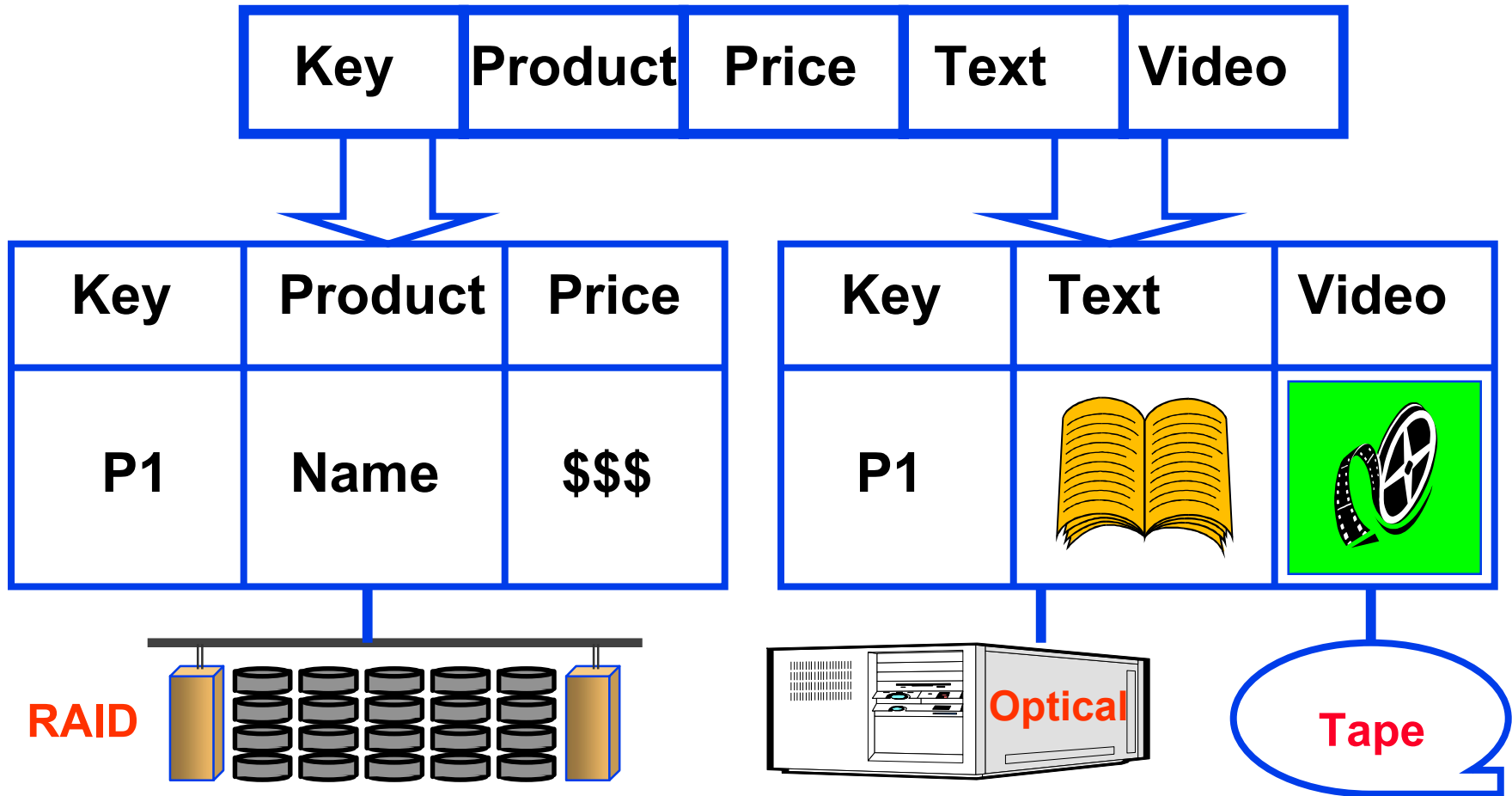
Federate Multiple Databases



Logical Horizontal Partitioning (LHP)



Logical Vertical Partitioning (LVP)



StorHouse/DataJoiner

- **Extend Current DB2 DataJoiner - with StorHouse/RM**
 - ◆ **HSM to DB2**
 - ◆ **Nth Quarter Stealth Cost Carrier - No Mas!**
- **Transparency**
 - ◆ **Application: Transparent & Plug-and-Play-StorHouse**
 - ◆ **Suggest: Plug-and-Play; Study and “Fine-Tuning”**
 - FileTek (and IBM) - Tune APPs; Change/Add DB2 DJ**
- **DB2 DataJoiner - Best Federated Product in Market (Period!)**
 - ◆ **Migrate ROWs to StorHouse/RM**
 - ◆ **Vertical Partition: split columns (some on RM)**
 - ◆ **Federate MULTIPLE Databases (I.e. Oracle, Teradata, etc.)**
- **Feed The Data Source**



SQL-99 Object-Relational Table

- SQL-92 (Built-In) Data Types:
 - ◆ Scalar: INTEGER, CHARACTER, FLOAT, ..
- SQL-3 (Objects & Functions)
 - ◆ User-Defined Types (UDTs)
 - ◆ User-Defined Functions (UDFs)

SQL-99 TABLE TYPES							
SQL-92 Scalar Columns				SQL-99 UDT Columns			
Char(n)	Integer	Date	Float	BLOB	Image	Audio	Video



SQL Financial Query (StorHouse/RM-LOB)

```
SELECT * FROM StockInfo  
WHERE Price / Earnings <= .5;
```

Store & Retrieve
Data using BLOBs:

Text

Video

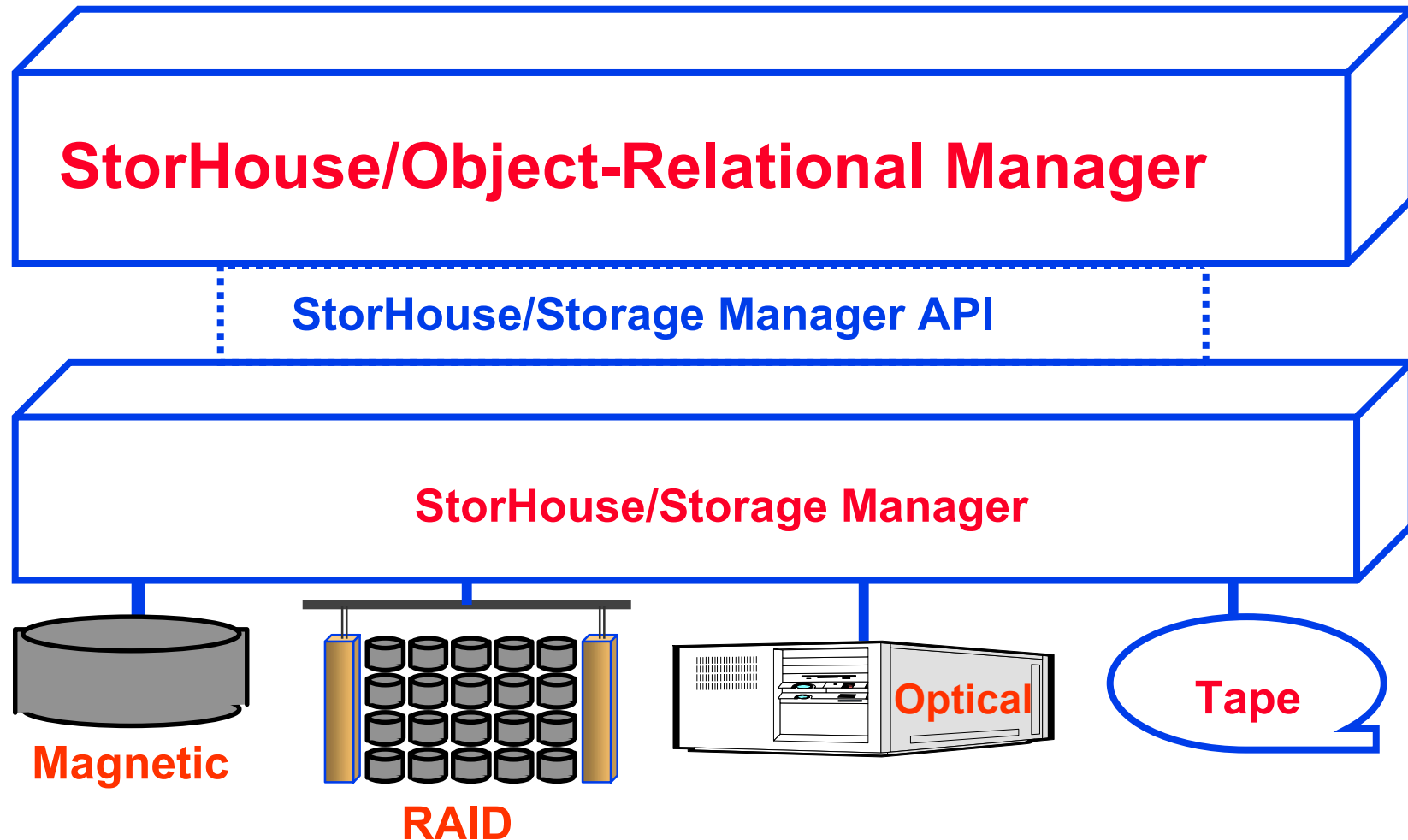


Stock	Price	Earnings	Reports	Video
InternetWare	10	100		
GreatSoftWare	25	50		
MultimediaSoft	20	45		

Database query
with Multimedia
answer




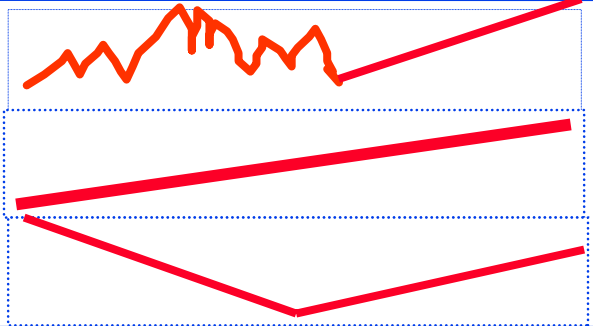
The Object/Relational Future



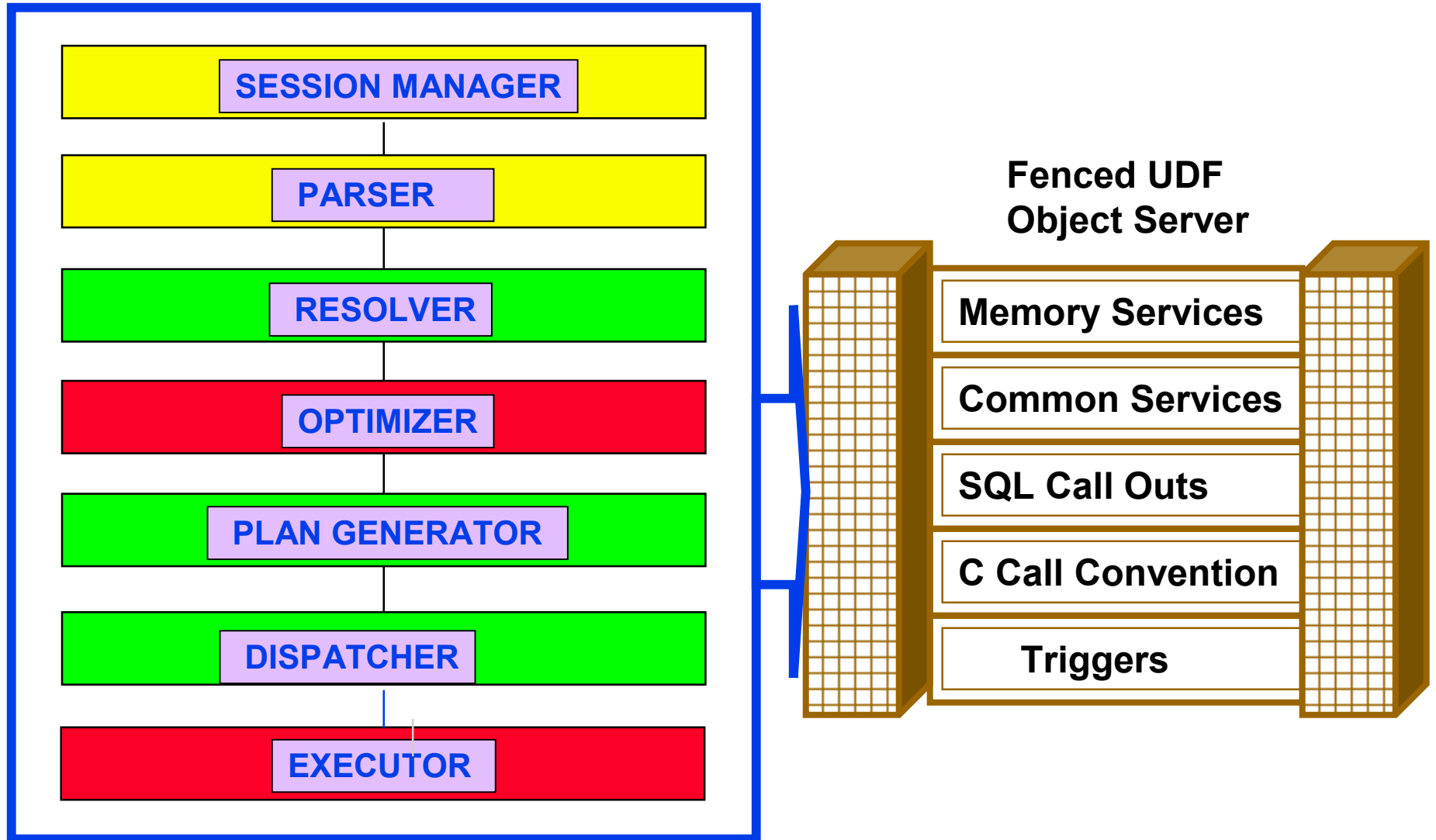
SQL Financial Time-Series

(UDFs In: Projection & Predicate)

```
SELECT Stock, Price, Video, PlotDailyValues(History)
FROM   StockHistory
WHERE  MovingAverage (History, 30) /
       LastPrice (History) < .8;
```

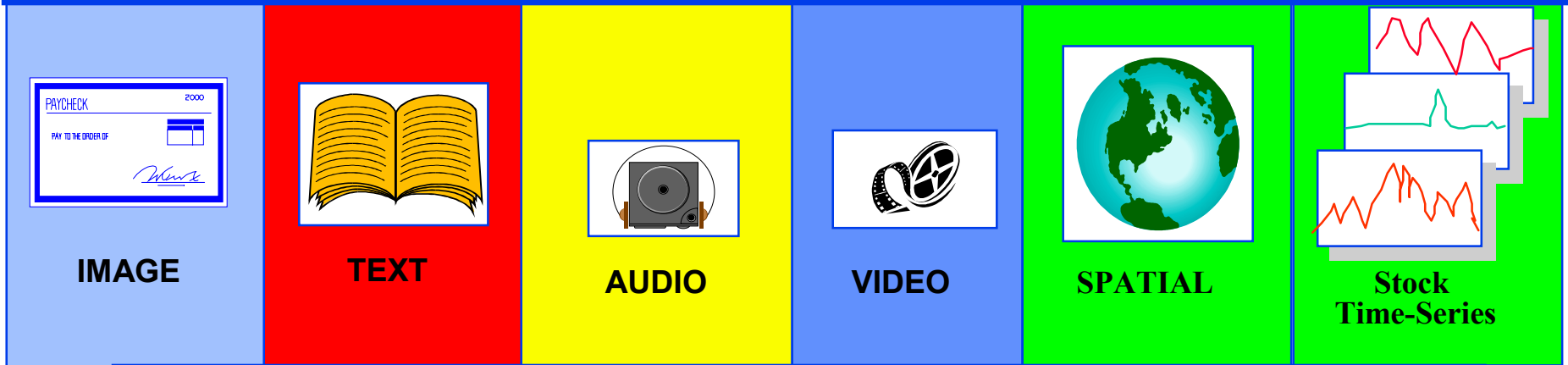
Stock	Price	Video	PlotDailyValues
Felipe.com	500		
Art.com	75		
Pekka.com	100		

Object/Relational Manager (ORM)

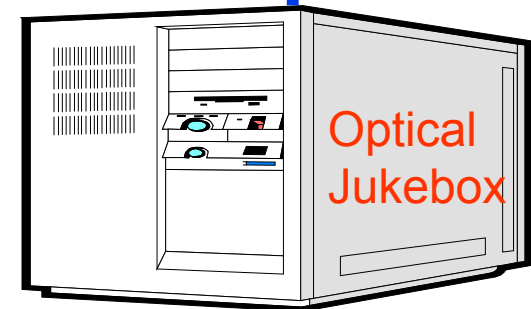
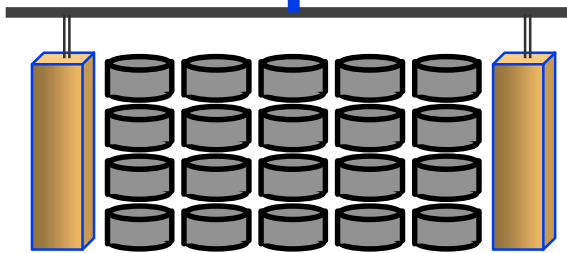
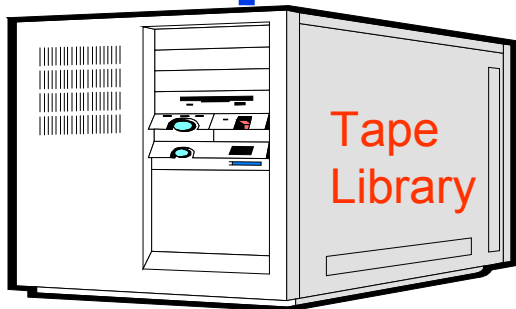


StorHouse/ORM & UDF Libraries

LOGICAL DATABASE INFORMATION BACKPLANE

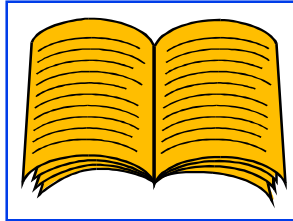


OBJECT/RELATIONAL MANAGER



FileTek

TEXT - Data Type



TEXT

TEXT: Includes Many Data Types

ASCII & EBCDIC Text

Word & Framemaker Documents

PDF Files, etc.

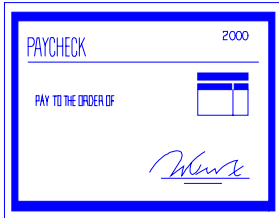
```
SELECT * FROM ProductTable
```

```
WHERE FindText (INSTRUMENTS, "IPO");
```



INSTRUMENT	Company	Prospects
IPO	FileTek	

IMAGE - Data Type



**IMAGE &
PICTURES**



IMAGE Includes *Diverse* Types

JPEG

Pictures

Emerging Standards Scanned
Images, Etc.

Excellent First Step: Use BLOB

UDF Functions

Retrieve Based on Content

Likely Use:

Retrieve Based on Scalar

Display Use UDF Transform

Image Content UDF Query

```
SELECT * FROM ResortTable  
WHERE FindImage (Photo, Blue, 50%);
```



RESORT	PHOTOGRAPH
El Conquistador	

AUDIO & VIDEO - Types

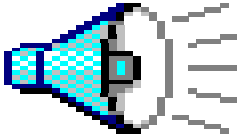
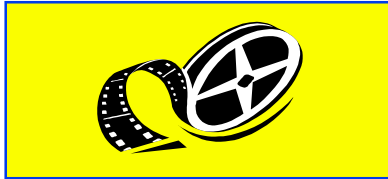


AUDIO



VIDEO

Product Information
Advertisement (for Web Apps)
Customer Interactions
Financial Transactions
Insurance Claims, etc.

SCALAR DATA	AUDIO	VIDEO
...		

VLDB Sizes: Cost-Effective System Needs

FileTek

Road to **Yottabytes**

- Terabyte (10**12) - 1,000,000,000,000 (4 commas)
 - ◆ National Retail Point-Of-Sale Data (~100TB)
 - ◆ Telecommunications Call Detail Warehouse (~200TB)
- Petabyte (10**15): 1,000,000,000,000,000 (5 Commas)
 - ◆ Text, Image Data, Audio & Video
- Exabyte (10**18): 1,000,000,000,000,000,000 (6 commas)
 - ◆ US National Medical Insurance Records
 - ◆ Yes: Image, Audio & Video (Again?)
- Zettabyte (10**21): 1,000,000,000,000,000,000,000 (7 commas)
 - ◆ Spatial & Terrestrial Data; Video & Audio Archives

The Big One - Yottabyte or Bust!

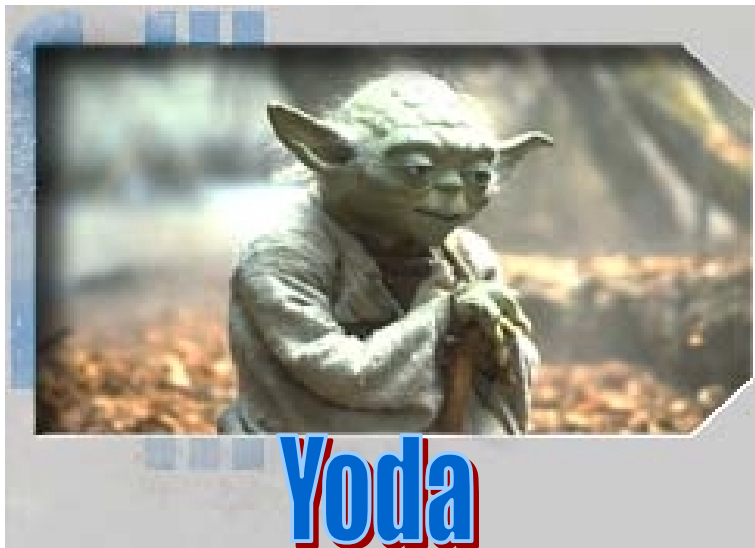
■ **Yottabyte (10^{24}): (8 commas)**

1,000,000,000,000,000,000,000,000

◆ **By Moore's Law:**

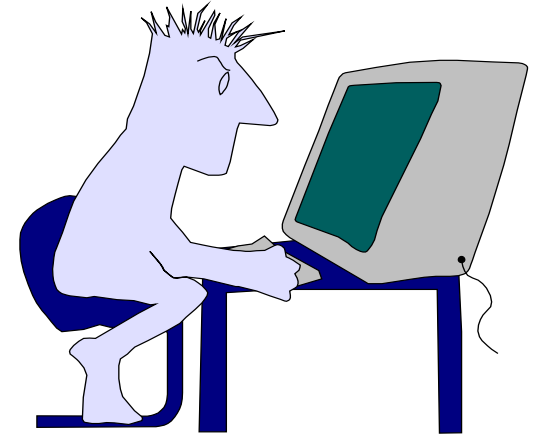
Year 2050 Database Size

Math Sizing Source: Pekka



¿ Questions or Comments?

■ FCARINO@FILETEK.COM



PHONE: (310) 607-9423
(626) 574-8285

