IBM Information On Demand 2008 >>> Comes To You

IBM

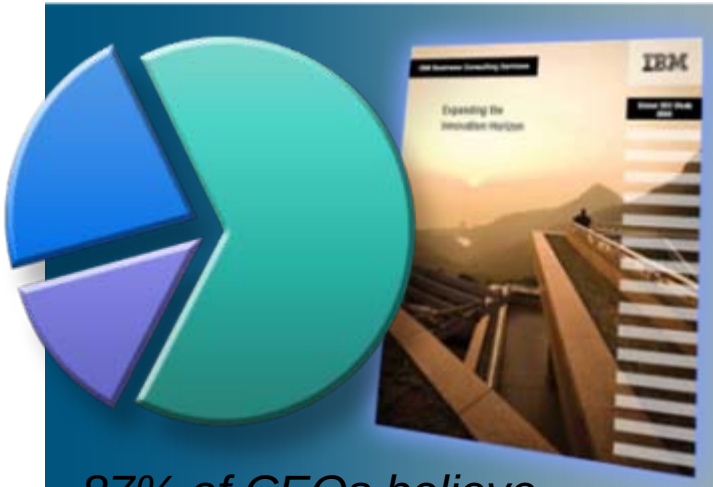
DB2 9.5's pureXML and Storage Compression Enable Rapid Data Analysis and Decision Making

Dwaine R. Snow

Senior DB2 Technical Evangelist - IBM

Corporate View of Information Architecture Is Changing

- **Information is the key to Business Innovation**
  - Organizations highly effective at driving information integration are 5 times more likely to drive business value
  - Companies that wait for their application vendors to innovate are losing the opportunity to differentiate



*87% of CEOs believe fundamental **change** is required in next two years to drive innovation*

Over 60% of CEOs believe their organizations need to do a better job leveraging information

Source: IBM Global CEO Survey

Customer Business Issues

- **Multiple versions of the truth**
  - Inability to understand and tailor customer interactions
  - Inability to collaborate effectively with supply chain
  - Difficulty in complying with information-centric regulations

- **Need to do a better job leveraging information**
  - Not using demand signals to drive supply chain
  - Not using customer analysis to tailor marketing and sales
  - Not leveraging sources of unstructured information

- **Don’t have trust in the information**
  - Asking the same question in different places produces different results

- **Don’t feel like there is control over information**
  - No way to understand or control how information is used
  - No governance of sensitive information on customers or accounts

Information is…

An Asset

A Weapon

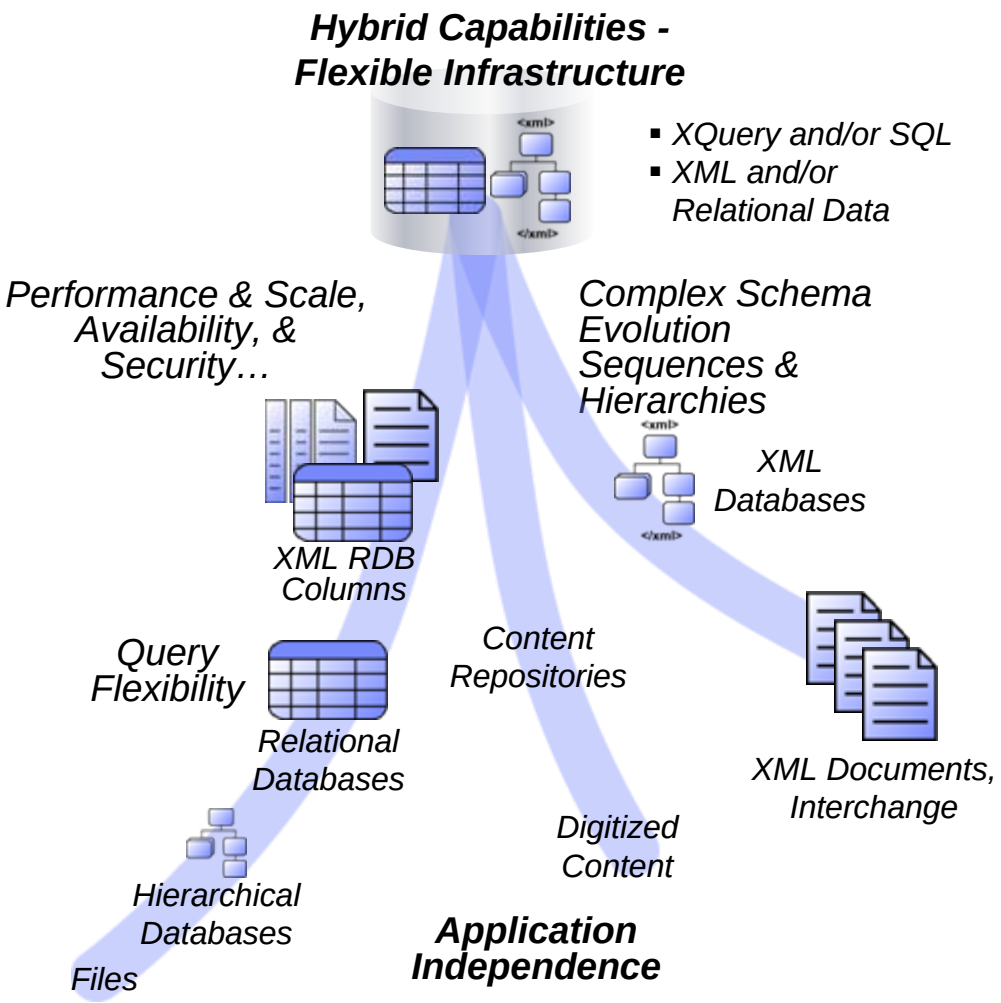
A Means to Top Line Growth

Customer Information is Critical

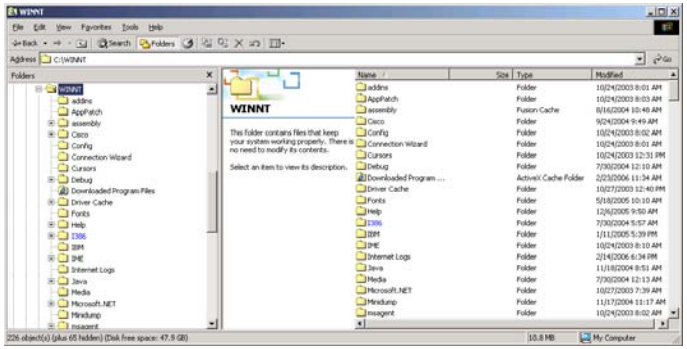
XML is changing the game

"IBM Moves the Database Goalposts…" Bloor Research

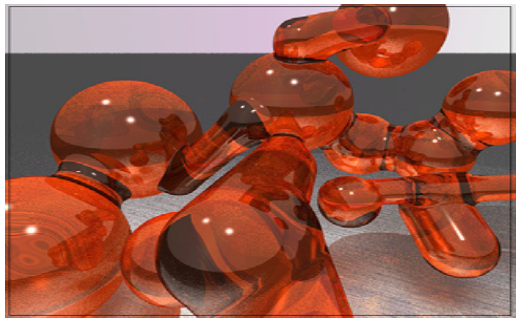
- Pure XML integration simplifies & unifies the infrastructure
  - Unifies Data and Content
  - Extends Asset Utilization
  - Leverages Existing Skills
- Pure XML document storage in DB2
  - Ensures fidelity of XML document
  - Optimizes Performance
  - Provides Flexibility
  - Leverages Mature DB2 Services



Where is your XML today?



In files…
- Storage not managed and not secure

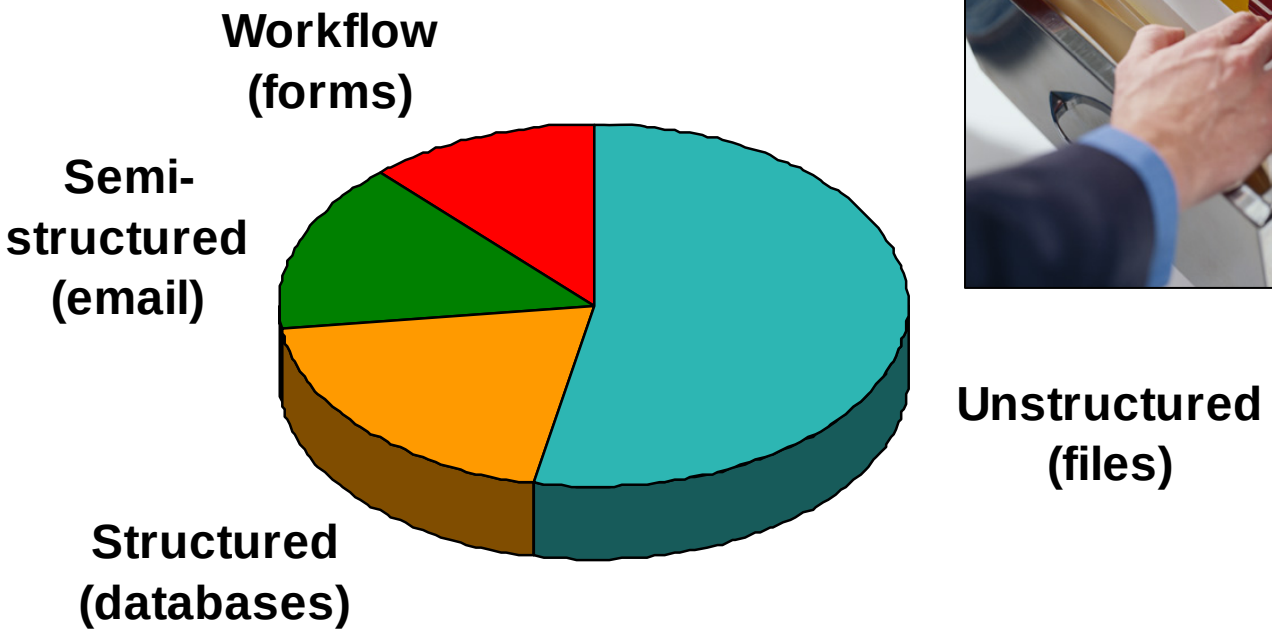


In LOBS…
- Content and business value locked up

Shred to tables…
- Complex and fragile mapping

In an XML DB…
- Scalability & integration concerns

Unstructured Data today



60 to 80% of data in today's environments is unstructured

XML – A Very Different Data Model

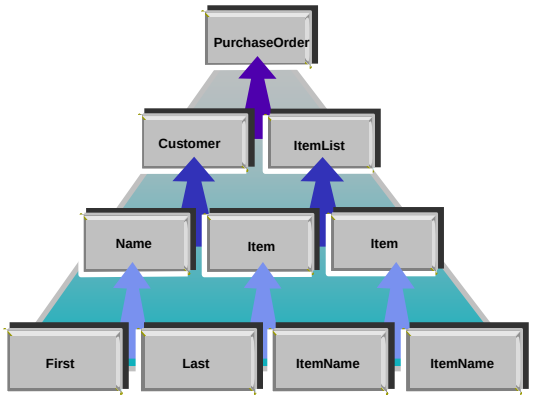
- **Relational is a data model:**
  - Relations (tables)
  - Attributes (columns)
  - Set based w/ some sequences
  - Fixed/Strict schema

POID	CustomerID	ItemID
12	1	2
162	3	4
162	3	5

Id	LastName	FirstName	Street	City	State	Zip
1	Pirahesh	Hamid	1 Harry Rd	San Jose	CA	95141
3	Selinger	Pat	555 Bailey Ave	San Jose	CA	95141

ItemID	Name
2	#6 wire nut
5	Small Walrus
4	Apollo moon rocket

- **XML is a data model:**
  - Nodes (elements, attributes, comments, etc.)
  - Relationships between nodes
  - Sequence based w/ some sets
  - **Flexible** schema

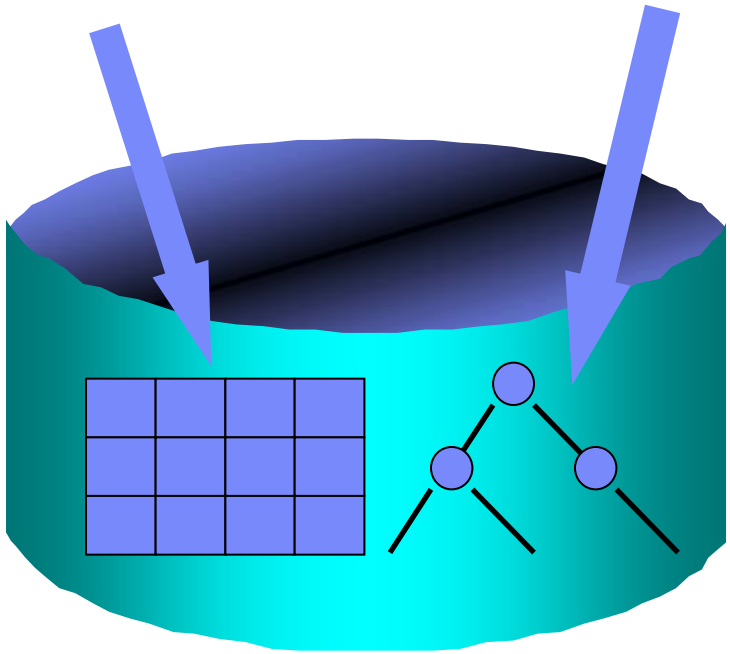


pureXML™

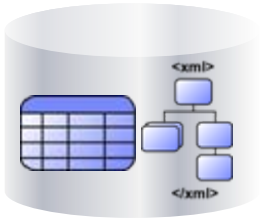
- **Must store XML in parsed hierarchical format (similar to the DOM representation of the XML infoset)**

```
create table dept (deptID char(8),…,
deptdoc xml);
```

- **Relational columns are stored in relational format (tables)**
- **XML columns are stored natively**



pureXML™ in DB2 9



SQL Person… "I see a world class RDBMS that also supports XML"

DB2 with pureXML Support

XML Person… "I see a world class XML repository that also supports SQL"

XML is integrated into all facets of DB2!

New XML applications benefit from:

- Ability to seamlessly leverage relational investment
- Proven Infrastructure that provides enterprise-class capabilities

Use Case: Financial Data (FIXML)

Buying 1000 Shares of IBM Stock..

Old FIX'ed Protocol

```
8=FIX.4.2^9=251^35=D^49=AFUNDMGR^56=ABROKER^34=2
^52=20030615-01:14:49^11=12345^1=111111^63=0^64=2003
0621^21=3^110=1000^111=50000^55=IBM^48=459200101^22=
1^54=1^60=2003061501:14:4938=5000^40=1^44=15.75^15=USD
^59=0^10=127
```

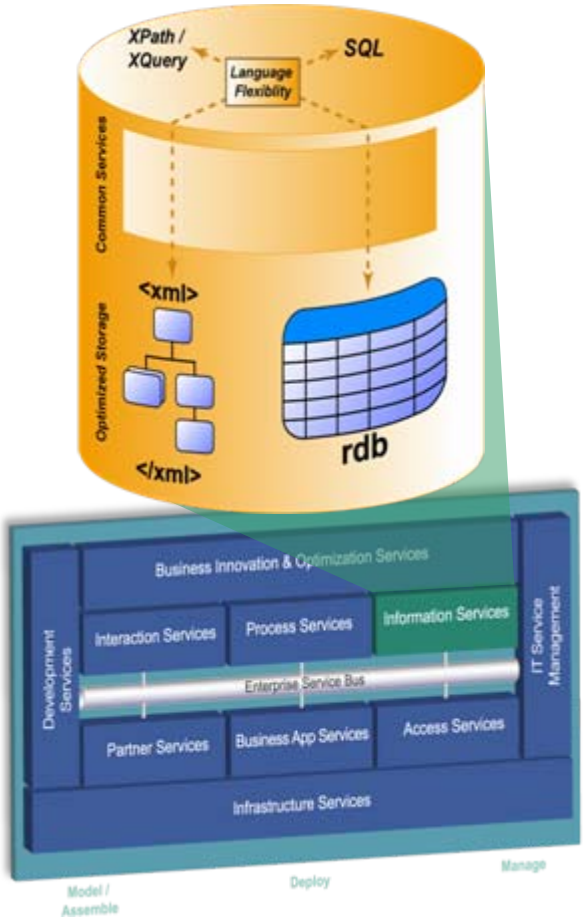
New FIXML Protocol

- Extensible
- Lower application development & maintenance costs

```
<FIXML>
  <NewOrdSingle ClOrdID="123456"
    Side="2"
    TransactTm="2003-06-15T01:14:49-05:00"
    OrderType="2"
    Price="93.25"
    Acct="26522154">
    <Header Sent="2001-06-21T01:31:28-05:00"
      PosDup="N"
      PosRsnd="N"
      SeqNum="521">
      <Sender ID="AFUNDMGR"/>
      <Target ID="ABROKER"/>
    </Header>
    <Instrument Symbol="IBM"
        ID="459200101"
        IDSrc="1"/>
    <OrderQuantity Qty="1000" Cur="USD"/>
  </NewOrdSingle>
</FIXML>
```

Customer Example 1

storebrand



Results using DB2 9 based SOA solution

- Fast, easy access to richer product & client information
- Quickly create customized products that customers want
- Expected to process five times more business

Development and app. performance re: XML data	**with relational data server**	**with DB2 9 hybrid data server**
Development of search & retrieval business processes	CLOB: 8 hrs Shred: 2 hrs	30 min.
Add field to schema	1 week	5 min.
Relative lines of I/O code (65% reduction)	100	35
Queries	24 - 36 hrs	20 sec - 10 min
Query non-shredded XML element	1 week	½ day

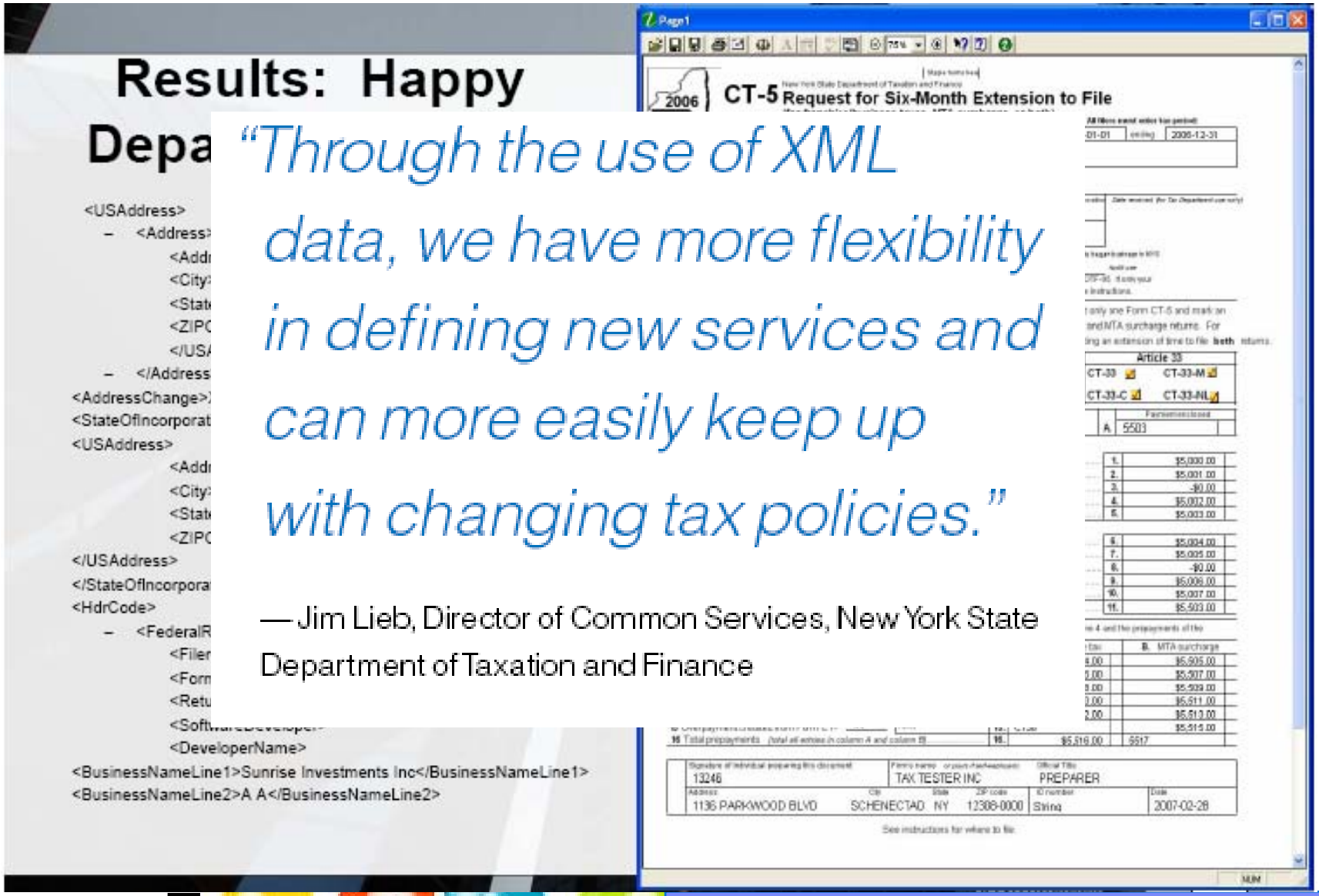
Customer Example 2

- **NY State Tax Agency**
  - Process $80B in taxes annually
  - Over 14M tax returns annually
- **Currently handling all e-filings**
  - Two more application coming online soon
- **Quick to change year to year**
  - Forms change yearly, as do Rules!
  - Thousands of different forms
    - Some can be left blank
  - No one knows how many rules ☺
- **Quickly and easily add new forms**
  - Also now able to store forms electronically

Faster, More Accurate refunds!

Up to date online tracking!

Customer Example 2



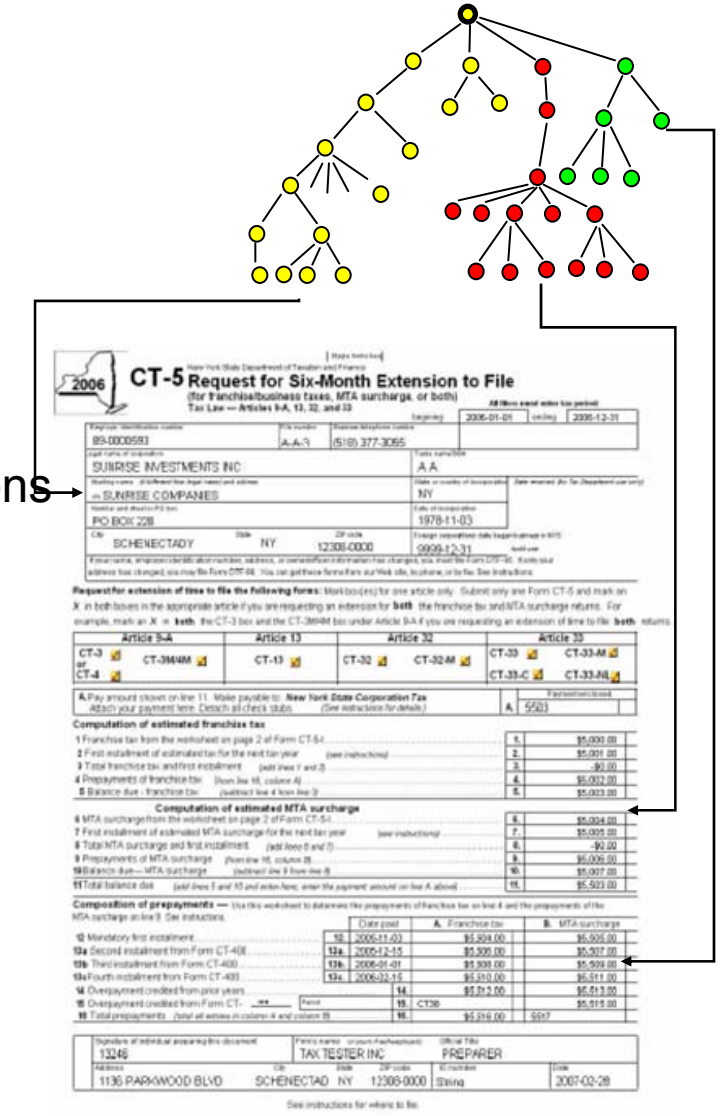
"Through the use of XML data, we have more flexibility in defining new services and can more easily keep up with changing tax policies."

— Jim Lieb, Director of Common Services, New York State Department of Taxation and Finance

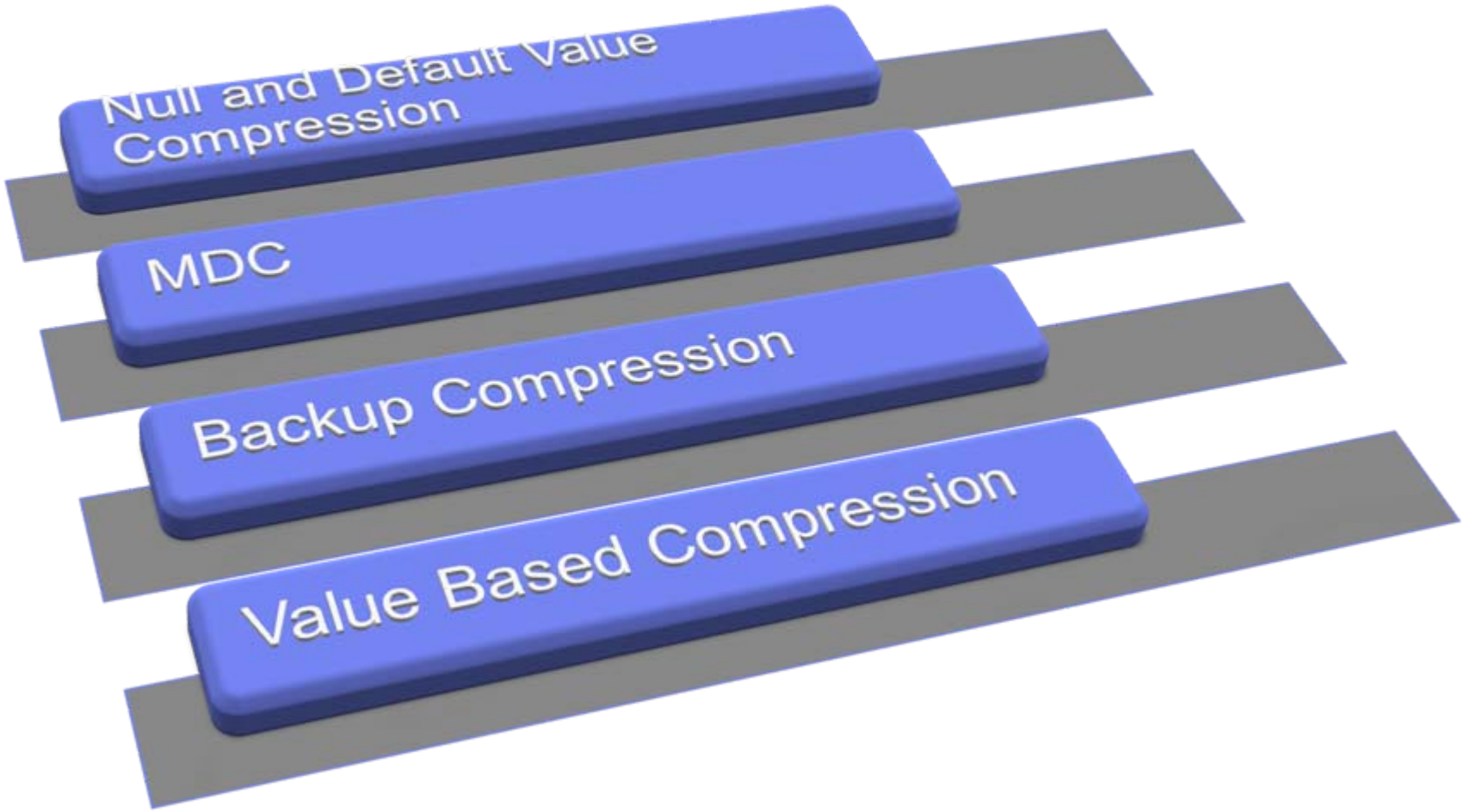
Customer Example 2

- **Handling Changing Forms Through**

- **Sub-Document Update**
  - New draft specification for Xquery Update language support
  - Provide more support for structural modifications to XML documents
  - Replace section, append section, add description, etc.
  - Use internal DB2 implementation rather than external stored procedures
- **Quickly and easily modify forms/rules**



History of DB2 Compression



Row Compression – How it Works

DB2 9 Deep Compression example – Table GLPCA

- **Table GLPCA: EC-PCA - 102 columns**

Uncompressed table

RCLNT	GL_SIRID	RLDNR	RRCTY	RVERS	RYEAR	RTCUR	RUNIT	DRCRK	POPER	DOCCT	DOCNR	DOCLN	RBUKRS	RPRCTR	RHC	RT
800	4751	8A	0	0	1995	DEM		S	1	A	21	1	1000	1402		
800	4752	8A	0	0	1995	DEM		S	1	A	21	2	1000	1402		
800	4753	8A	0	0	1995	DEM		S	1	A	21	3	1000	1402		
800	4754	8A	0	0	1995	DEM		S	1	A	22	1	1000	1402		
800	4755	8A	0	0	1995	DEM		S	1	A	22	2	1000	1402		

Compression dictionary

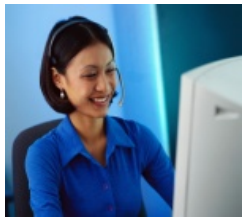
x'01C	8A,0,0,1995,DEM, ,S,1,A
F'67t	1000,1402
…	…

800	4751	x'01C	21	1	F'67t
800	4752	x'01C	21	2	F'67t
800	4753	x'01C	21	3	F'67t
800	4754	x'01C	22	1	F'67t
800	4755	x'01C	22	2	F'67t

Compressed table

Common sequences of consecutive bytes in row replaced with symbol

Compress/Decompress Process Adding Data



LOAD, Insert, Import

	Smith John 120000 Jones Jimmy 140000 Snow Dwaine 20000

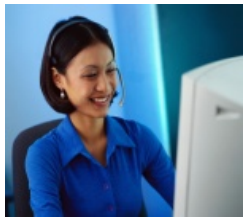
Check Dictionary

Bufferpool(s)

x01 x13 12x44
x11 x99 14x44
Snow Dwaine 2x44

x01 x13 12x44
x11 x99 14x44
Snow Dwaine 2x44

Compress/Decompress Process Selecting Data



Select .. where lname ='Smith'

	Smith John 120000

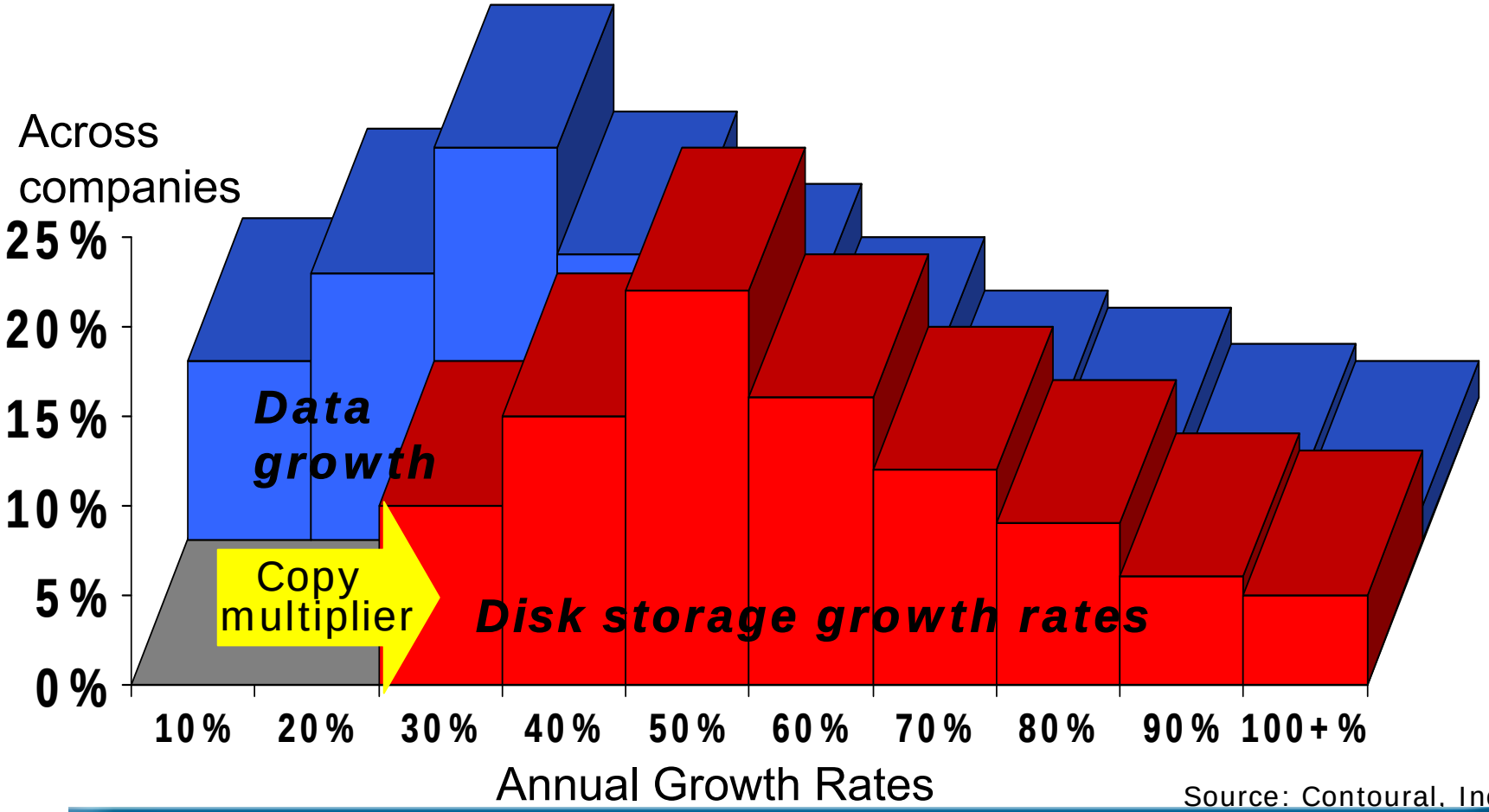
Check Dictionary

Bufferpool(s)

x01 x13 12x44
x11 x99 14x44
Snow Dwaine 2x44
…

x01 x13 12x44
x11 x99 14x44
Snow Dwaine 2x44
…

Data Growth vs. Storage Growth



Source: Contoural, Inc.

Actual storage growth much higher than data growth

Results as Presented at IOD

SUNTRUST

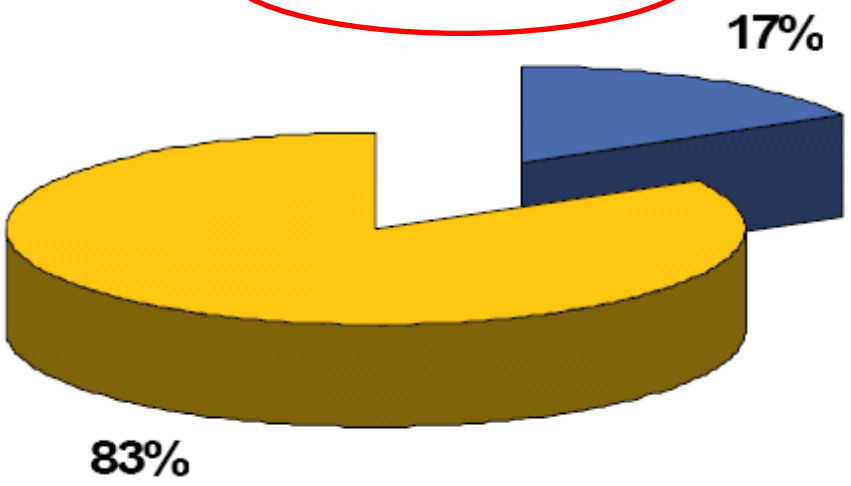
Compression Results

- Tested 3 Schema's, 96 Tables
- Total Size 4.3 TB
- Size After Compression 865 GB
- Table Compression Ranged from 65% to 88%

Schema	Tables	Un-compressed (GB)	Compressed (GB)	Compression Ratio
A	80	888	158	82%
B	12	2,736	603	78%
C	4	718	104	86%

Results as Presented at IOD

- Average compression of 83% on our Data Warehouse
- Data Warehouse went from 24 TB to 4.8 TB
- Reduced backup time and number of backup tapes
- More rows stored in bufferpools
- Reduces the number of nodes needed in the future
- Projected initial cost savings: Over $2 Million
- Ongoing saving of: $500,000 per year

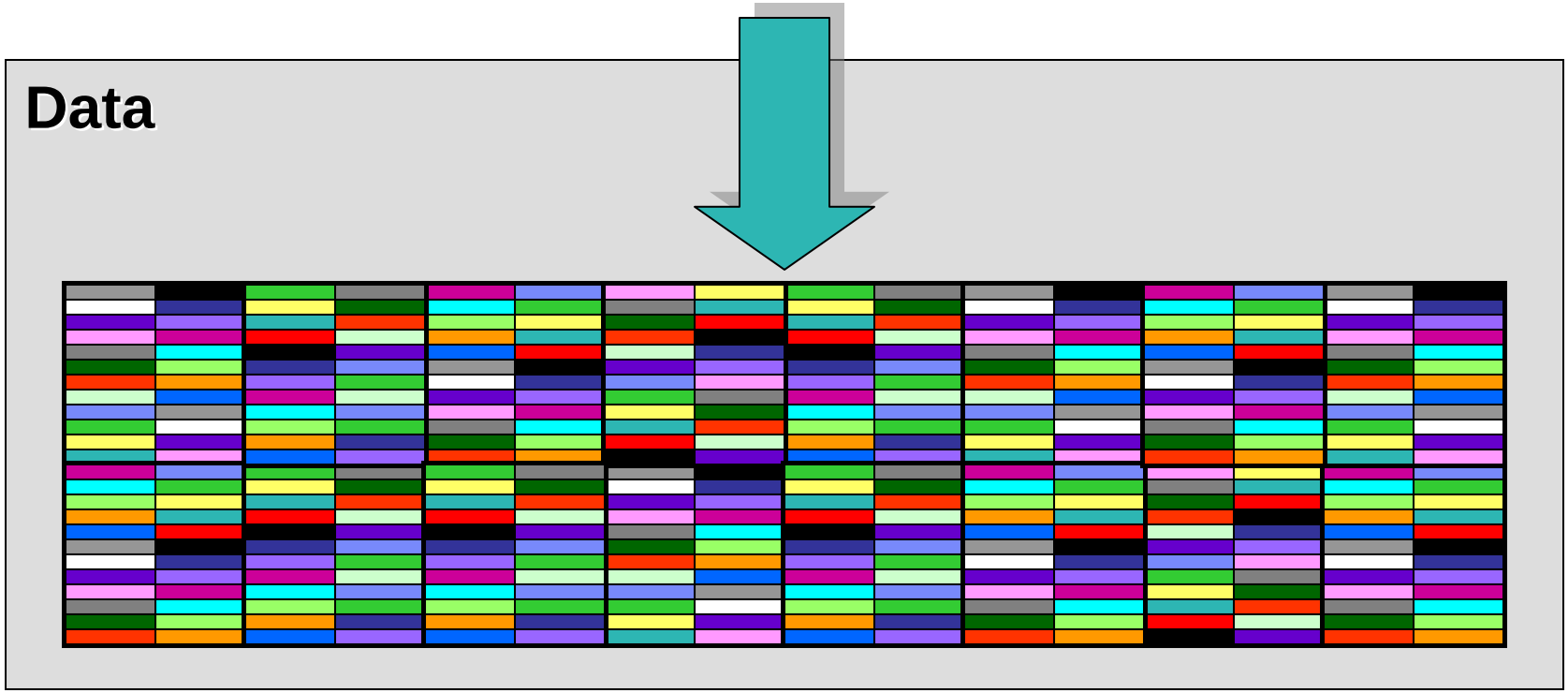


After compression Database is 83% smaller

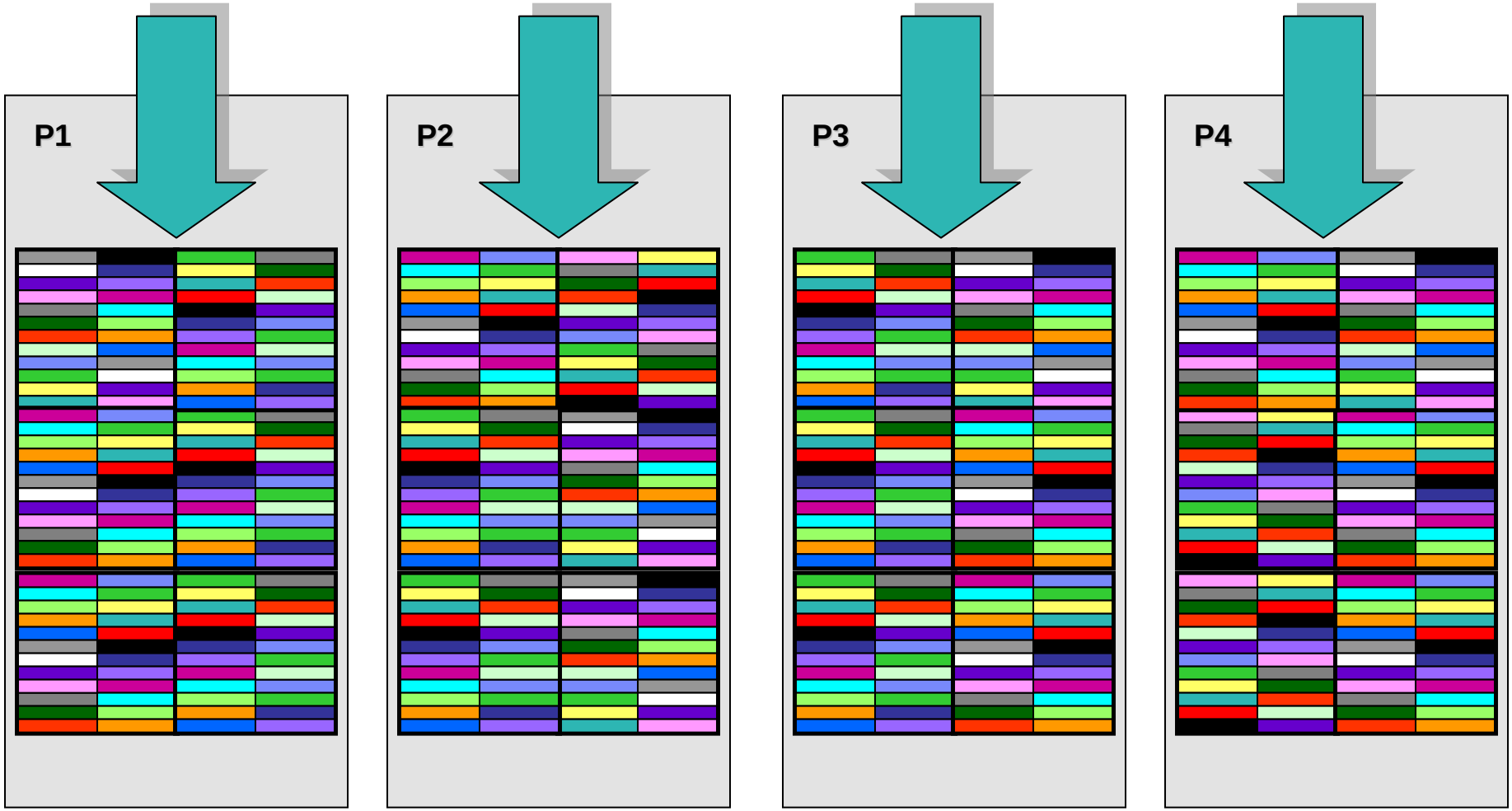
Results as Presented at IOD

"With the new compression technology in DB2 Viper, we realized an **80 Percent improvement in space savings for our most critical tables** in our data warehouse," said Donny Ledbetter, Senior Database Administrator, AutoZone. "We were even more pleased with this technology when we found that Viper's compression capability helped us **process queries to the database an average of 40 Percent faster than before**. We're looking forward to seeing the same results with our Operational Data Store and OLTP systems."

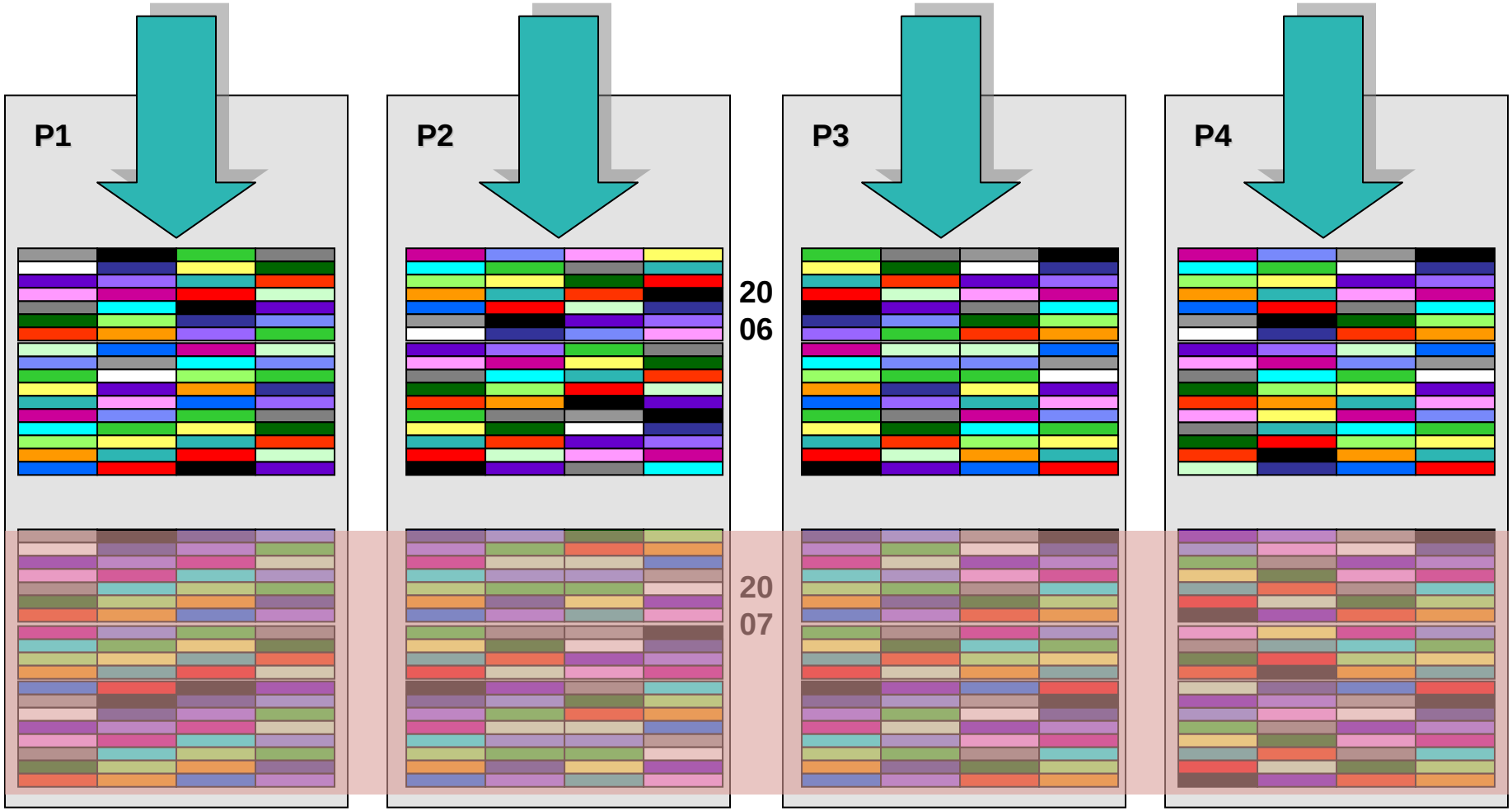
DB2 – Delivering Data Faster



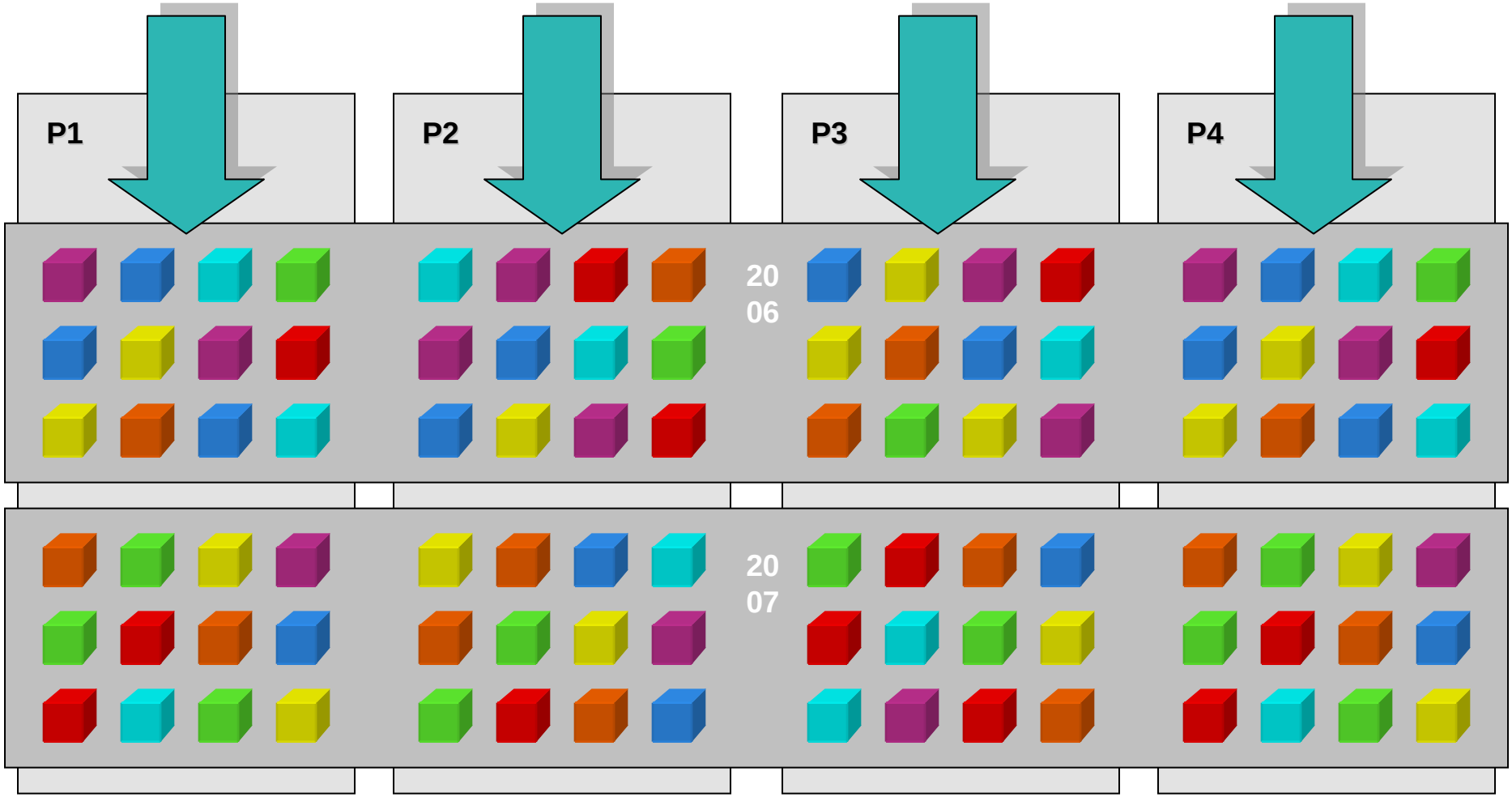
DB2 – Delivering Data Faster



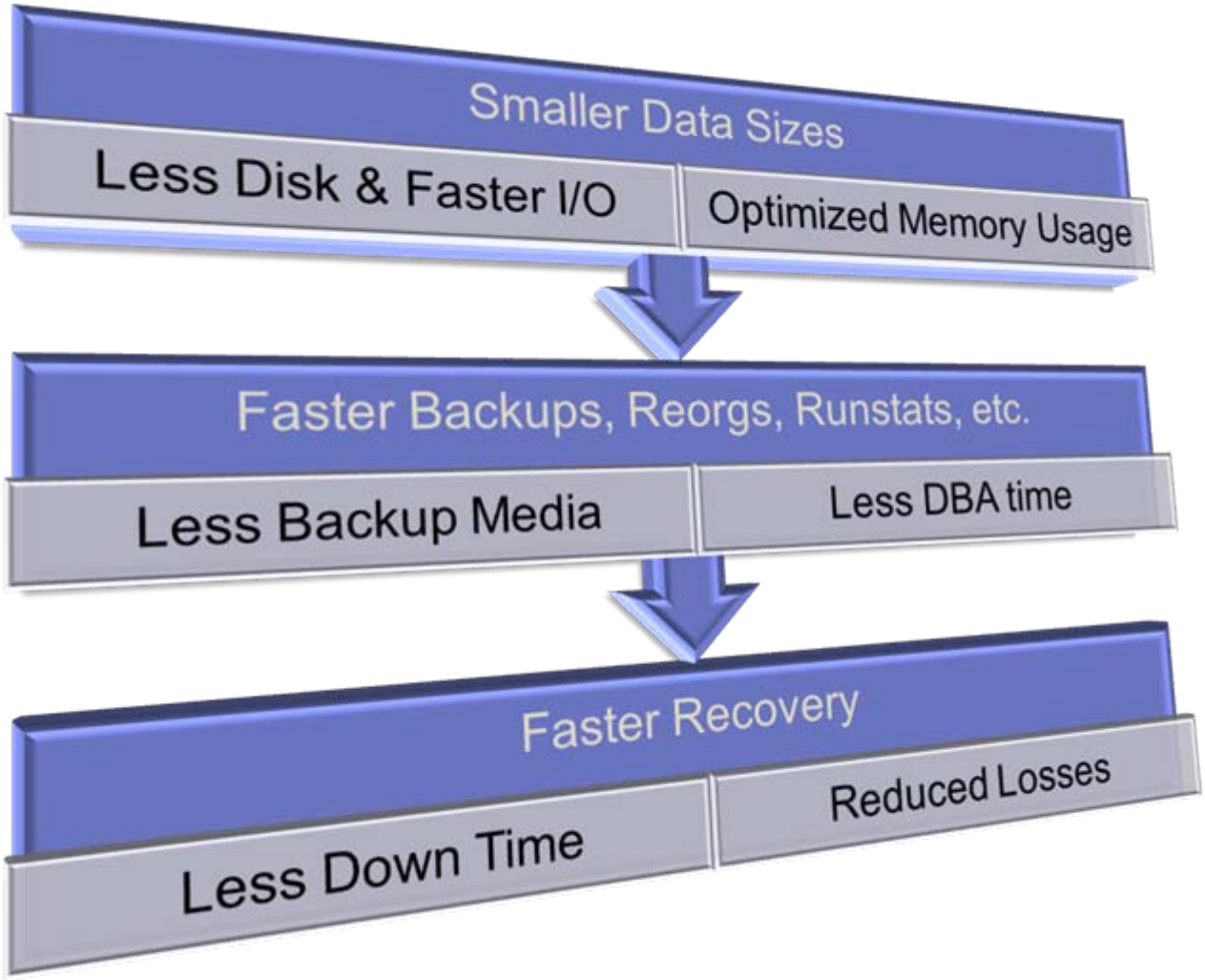
DB2 – Delivering Data Faster



DB2 – Delivering Data Faster



Compression Benefits




Thank
You