DB2 UDB V8 - What's new for Business Intelligence

DB2 for Linux, UNIX and Windows

Bill O'Connell, IBM Toronto Lab

DB2. Data Management Software

e business software

Sample Key DB2 V8 Enhancements

- **Business Intelligence**
  - Multi-Dimensional Clustering
  - Cube Views
  - Query Sampling
- **Scalability**
  - Connection concentrator
  - Compression for nulls & defaults
  - Full 64 bit support, universal (32 or 64 bit) client
- **Availability**
  - On-line Utils, e.g., On-line table re-org, load, indexing
  - Dynamic configuration
  - Real-time loading and R/W to same tables
- **Ease of Use & Management**
  - Self Managing / Tuning
  - Health Center
  - New optional Distributed Tools
   - Performance Expert, Recovery Expert

8.1

- Availability
- Manageability
- Performance
- Business Intelligence
- Application Development
- Information Integration

DB2 V8…Getting SMARTer…

- **DB2 Turns Years of Experience into Seconds of Execution**
  - *Configuration Advisor* Eases Tuning
    - Sets dozens of major parameters
      - Each parameter modeled
    - Automatic HW detection
  - *Health Center* Keeps Things Running
    - Health Center, Automation, Monitors, Indicators, Thresholds, Beacons
    - Automatic, Rules-based Corrective Action or Notification
  - Recently Announced:
    - Recovery Expert
    - Performance Expert
- **And more…**

8,136 TPM **8,268 TPM**

Advisor

Human

98.4%

Actual Customer Experience*
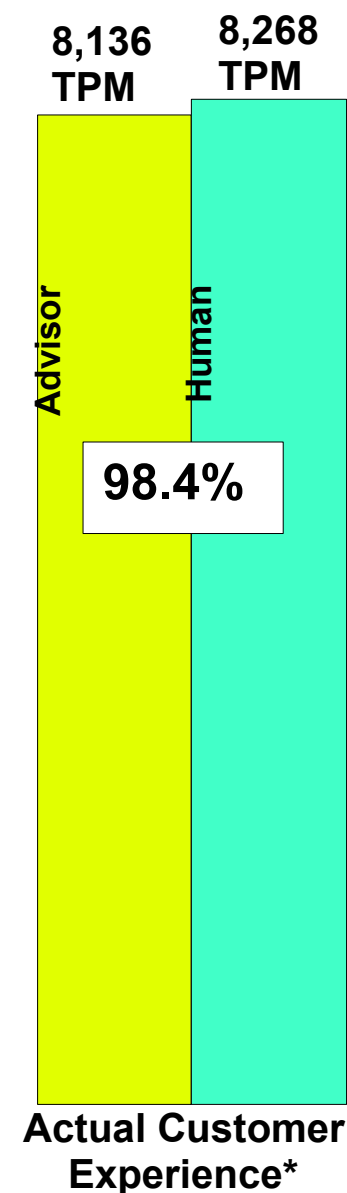
*Sample Results; Results may vary depending on many possible factors.

DB2 V8: Robust Foundation

- **Performance & Scalability Enhancements Continue Cross Workload, Cross Platform Leadership**

  - **Multi-dimensional Clustering** Speeds Queries
    - Clustering on Multiple Dimensions
    - Also Minimizes Reorgs & Index Maintenance

  - **Materialized Query Tables** Handle More Cases
    - Non-aggregated Joins, Nicknames
    - Incremental Maintenance

  - **Null & Default Compression** Minimize Disk Space

  - **Client Enhancements** Improve Scalability
    - *Connection Concentrator* Scales User Connections
    - *Common Client* Simplifies Communication Stack, Enhances Flexibility

Robust Foundation…

- ## Availability – Around the Clock
  - ### *Online Utilities* Minimize Planned Outages

    - ***In-place*** , Online Table Reorg
    - Online Index Creation & Maintenance
    - Online Load
    - Incremental AST Maintenance on Load Append
    - Dynamic Configuration Parameters
    - Dynamic Bufferpool Operations
    - Multi-dimensional Clustering Reorg Avoidance

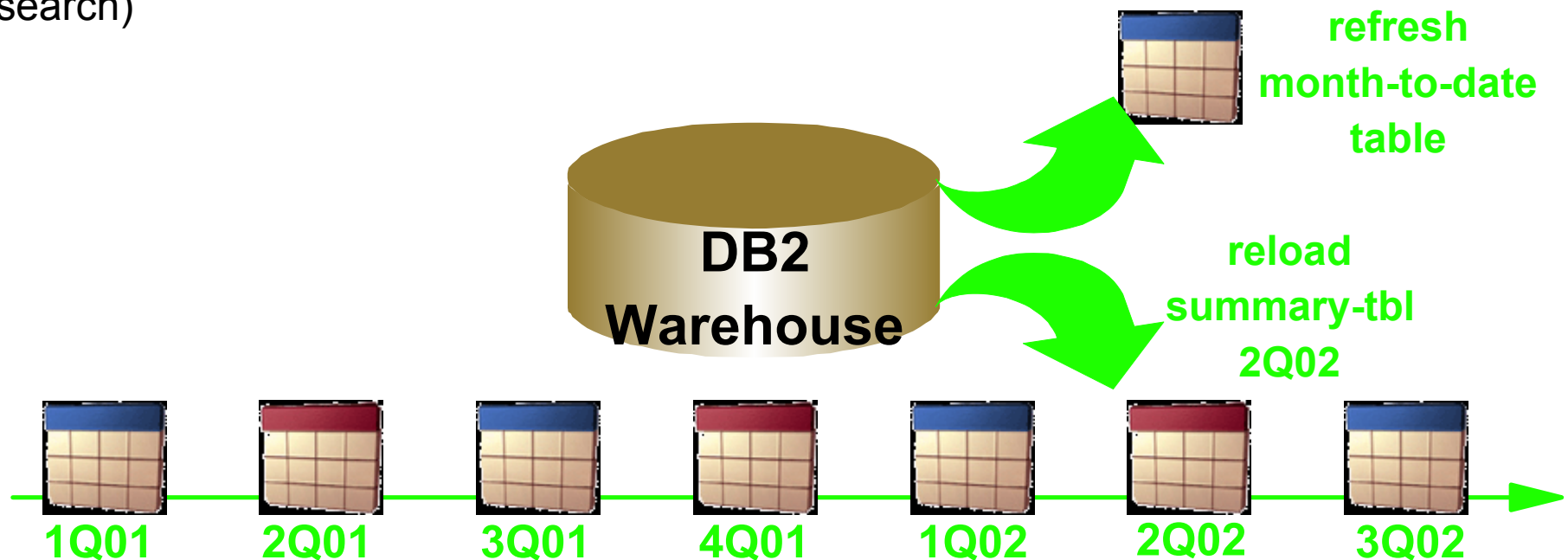 - ### *Numerous Enhancements* Minimize Unplanned Outage Impacts

    - Faster Tablespace Recovery
    - Order of Magnitude Improvements in Trace Performance
    - Type Two Indexes, Unlimited Active Log Space, Log Mirroring

  - ### And more…

New in 8.1 - Incremental MQT Maintenance

Inserted data visible when load completes (or constraints checked)

- MQTs refreshed incrementally via LOAD command
  - When base table changes, DB2 applies only the changes to the MQT summary
  - Tablespace available for full access during load and update
  - Existing table/index data available for read access
- Benefits:
  - Reduces nightly batch workload cycle
  - Eliminates DBA labor to maintain summaries
  - Higher data warehouse availability: no reload = less downtime
  - Summaries are always current

Table available for read access during load append

Ongoing Updates

division	country	city	product	date	amount

DB2 base table

MQT-x

**MQT-y**

ripple through

DB2 Warehouse

New in 8.1 - User Maintained MQTs

- DBA controlled summary table, Optimizer Uses it
  - Many warehouses have custom applications that maintain and load tables that are in reality user defined and maintained summary tables
  - It would be very beneficial to use them by having the optimizer recognize their existence and use them for query processing
- Benefits:
  - MQT summary performance boost valuable
  - Keep summary data static for "stable" analysis purposes (audits, research)

refresh month-to-date table

DB2 Warehouse

reload summary-tbl 2Q02

1Q01   2Q01   3Q01   4Q01   1Q02   2Q02   3Q02

V8.1 - Multidimensional Clustering - Benefits

- **Faster Query speed**
  - Access only the data pages necessary - no searching
  - Slice / dice (a.k.a., partition) elimination
  - Get all the data in a few accesses versus 100s of page reads
  - Reduces CPU & disk I/O use for other users to leverage
- **Reduces index size** --Saves disk, faster queries
  - One index entry covers entire data page, not one per row
- **Reduces DBA reorg's** --row clustering managed by DB2
- **Faster Deletes** --just drop a few data pages
- **Faster Inserts** --store the record, but rarely insert an index entry
- Simple & Intuitive --Multidimensional keys & Star Schema friendly
- Perfect for OLAP style hierarchical analysis
  - Executive reports, Summaries, etc.
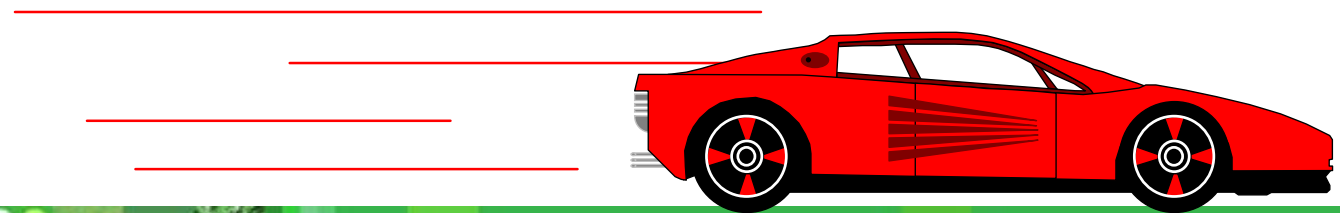
Before V8.1 - Single dimensional data clustering

- **Benefits:**
  - physically cluster data on insert according to order of single 'clustering' index
  - improves performance of range queries and prefetching
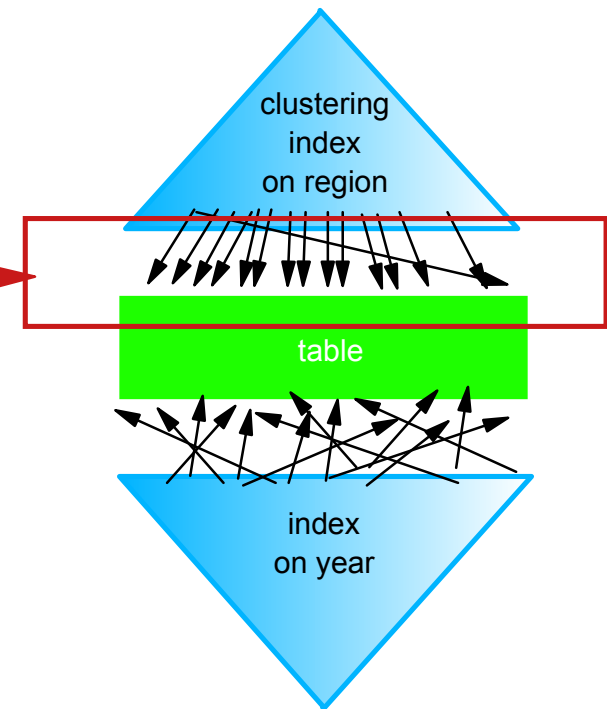
Clustering Index

Data Page 1

SKU	Store	Date	Qty	Amt
101	21	04/02	1	1.50
101	21	04/02	1	1.50
101	7	04/02	2	3.00
101	7	04/01	6	8.11

Data Page 2

SKU	Store	Date	Qty	Amt
101	7	04/02	1	1.50
101	21	04/02	3	4.10
101	7	04/01	2	3.00

clustered →

clustering index on region

table

index on year

New with V8.1 - Multidimensional Clustering

Old page layout without MDC

Clustering Index

Data Page 1

SKU	Store	Date	Qty	Amt
101	21	04/02	1	1.50
101	21	04/02	1	1.50
101	7	04/02	2	3.00
101	7	04/01	6	8.11

Data Page 2

SKU	Store	Date	Qty	Amt
101	7	04/02	1	1.50
101	21	04/02	3	4.10
101	7	04/01	2	3.00

New - with MDC

SKU	Store	Date	Qty	Amt
101	21	04/02	1	1.50
101	21	04/02	1	1.50
101	21	04/02	3	4.10

SKU	Store	Date	Qty	Amt
101	7	04/01	6	8.11
101	7	04/01	2	3.00

keys

SKU	Store	Date	Qty	Amt
101	7	04/02	2	3.00
101	7	04/02	1	1.50

- A CREATE TABLE option - simple syntax
- Puts rows with same key values in same data page (extent)
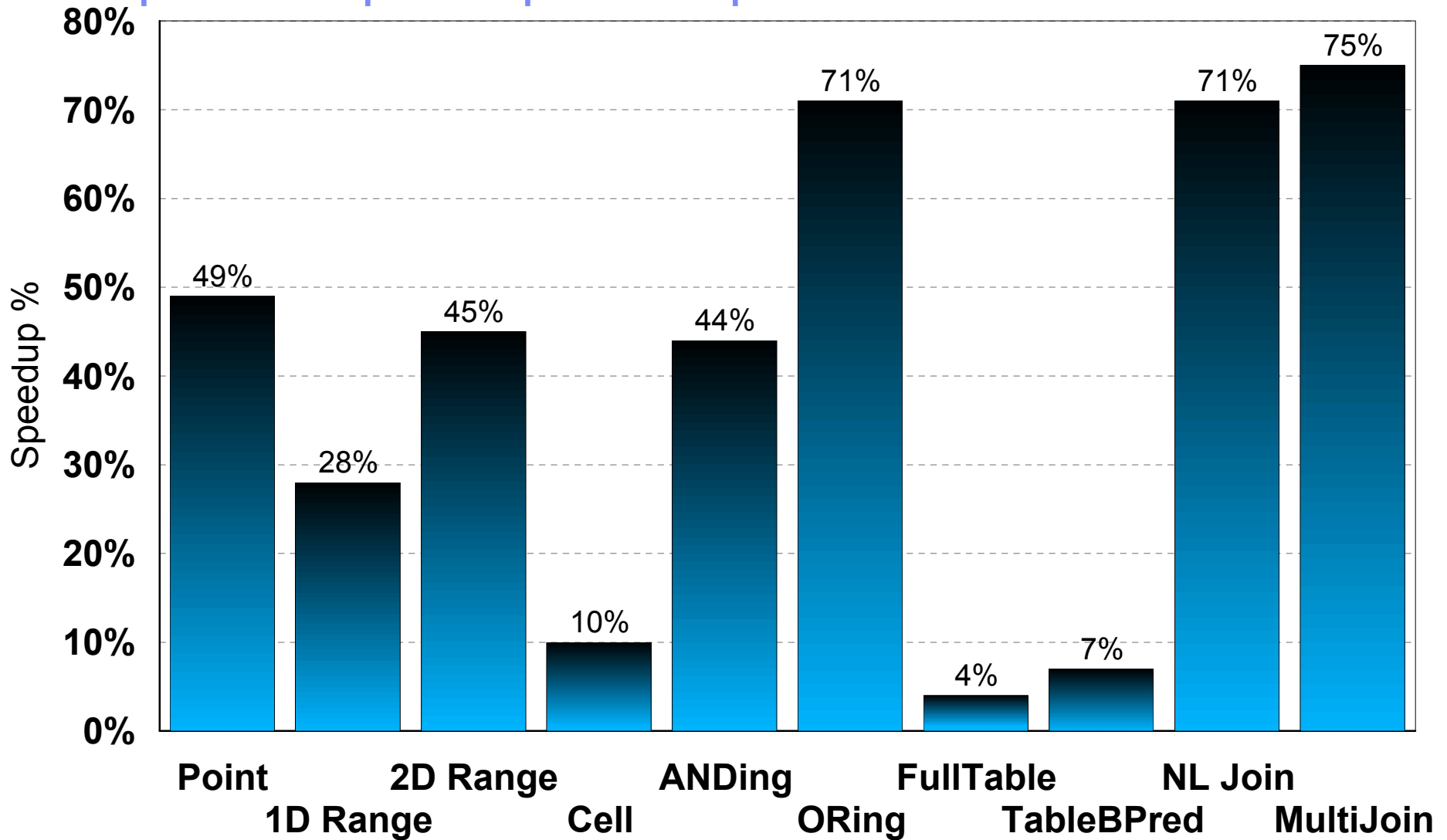- Transparent to applications and end user tools

"Partition elimination"

V8.1 Multidimensional Clustering - how it works

- In multidimensional clustered (MDC) tables, data is organized along extent boundaries according to dimension (clustering) values

Extents making up an MDC table

with dimensions region and year

Dimension Block Index on Region

region

East, 1993
East, 1996
North, 1996
North, 1997
North, 1998
South, 1999

year

Dimension Block Index on Year

Data stored
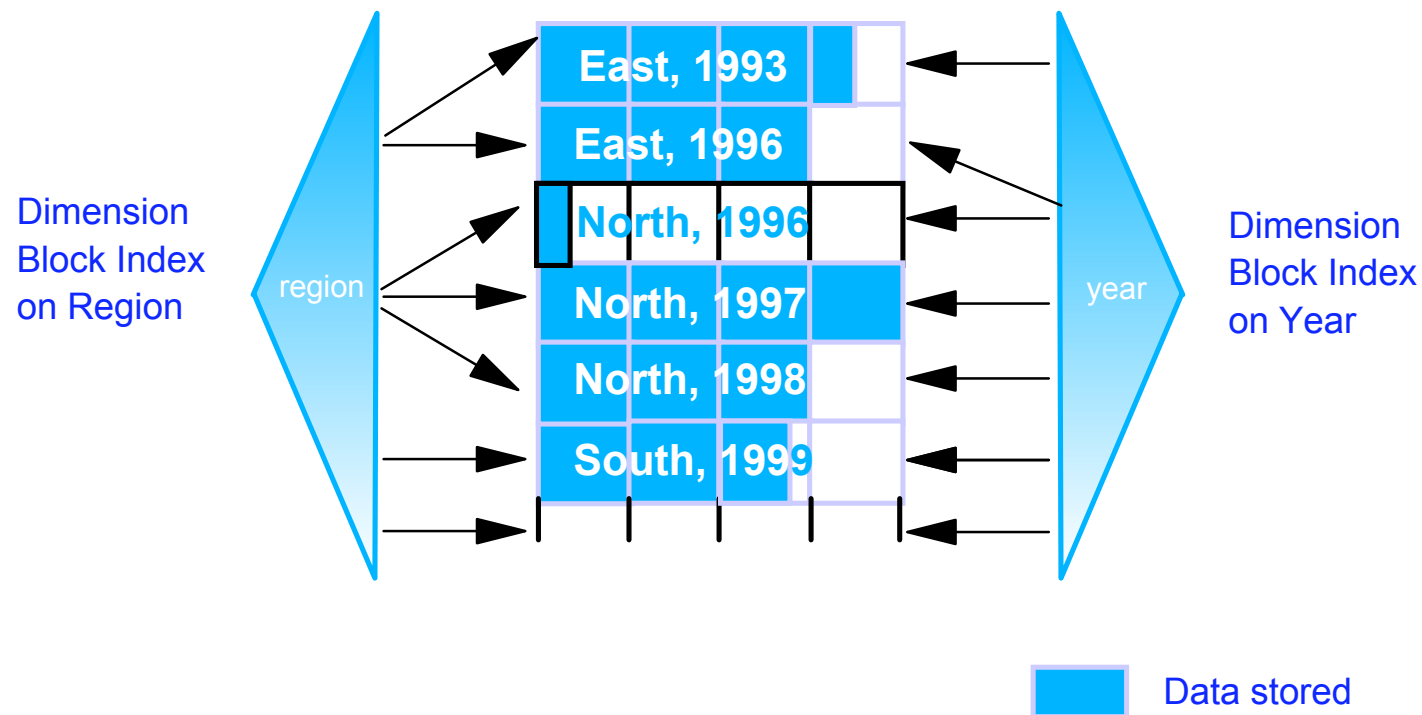
MDC queries Speedup - sample tests

Example: "2 Dimensions range query is 45% faster with MDC"

V8.1 Availability Enhancements - On-line Load

- **On-line load**
  - Tablespace online, table online for read
  - Other tables in tablespace remain operational
- **In Addition:**
  - The Load utility now directly supports partitioned databases
    - V7 Autoloader integrated into standard V8.1 Load utility
    - V7 Autoloader still supported for backward compatibility
  - Generated Columns are populated during Load
  - Multiple input files for same table now supported by Load utility
 - Load wizard now part of DB2 control center to set up Load

V8.1 Availability Enhancements - Online Reorganization

- Online Table Reorganization - for High Availability
  - A table is reorganized "in-place" --no requirement for large temporary space
  - Table stays available to all users for read & write
  - Status monitoring + break points for pause & resume
    - You can *resume* reorg at the partition level vs. ALL partitions being restarted
  - Reorganizes table data objects only; not indexes, long fields or LOBs
- Benefits:
  - Table is never "offline" from user or batch access
  - Performance improves as data pages are consolidated
  - Saves disk space by returning fragmented pages for reuse
  - Reduces DBA labor -- no more long weekend reorgs!

**end user update**

reorg in progress

DB2 table
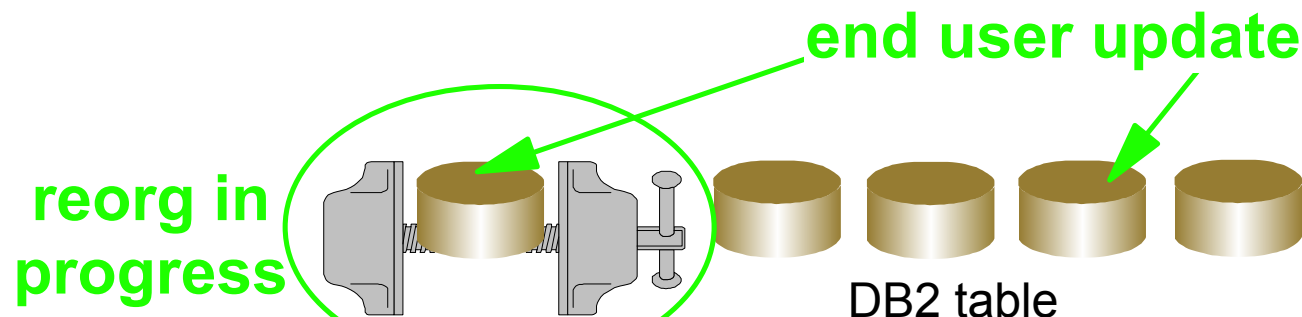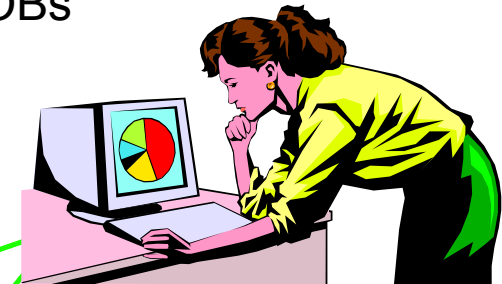
V8.1 Availability Enhancements - Online Index Maintenance

- Full read/write access to tables during index creation
- New "REORG Index" command
  - *Shadow* index is created concurrent with normal database updates
  - When index creation is complete, old index is swapped out for the new
    - final synchronization via DB2 log file
  - Full read/write access to table & index during index reorganization
  - Temporary space required to build or reorganize the new index
- Benefits
  - Table is never "offline" from user or batch access
    - no more weekend reorgs to ensure end users always have access
  - Compacted index pages improves Performance

end user update

DB2 table   table index   shadow index   **DB2 log**
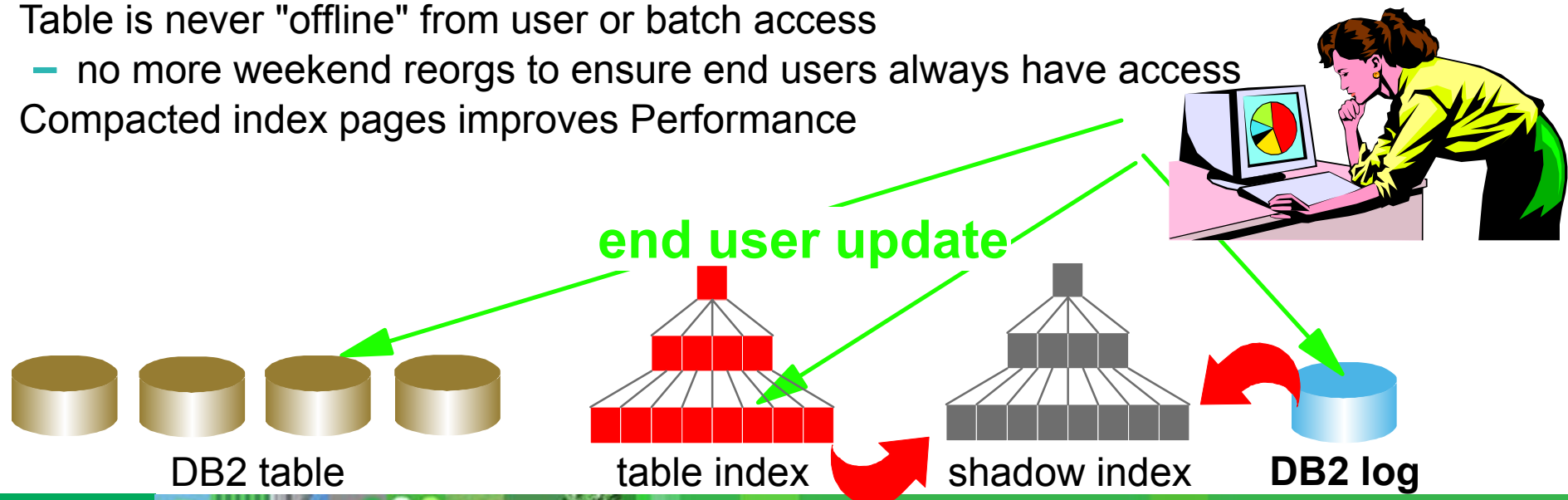
V8.1 Storage Enhancements - Null and Default Value Compression

- Eliminate storage of nulls and system default column values
  - Eligible datatypes: Numeric, Char, Varchar, DBCS (fixed and variable), BLOB
  - does not apply to site specific defaults
- Benefits:
  - **10-50% reduction in disk storage costs!**
  - Performance Improvements
    - More rows packed per page = fewer I/Os for same # of records
    - Fewer reads, less busy I/O channels, more data in memory simultaneously
  - **Fewer data blocks = Faster Backups & Restores**

Table-Y

compressor

results
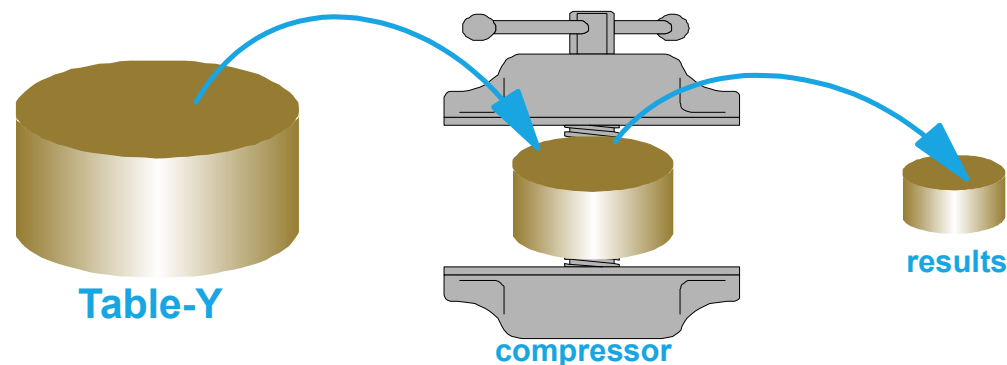
Sampling - Quick Approximate Answers

- **Error estimation (combine sampling + SQL)**

```
SELECT loc.country AS country, year(t.pdate) AS year,
    sum(ti.sales) / :samp_rate AS est_sales,
  sqrt((1e0/:samp_rate)*((1e0/:samp_rate)-1e0)*
    sum(sales*sales)) AS std_err
  FROM trans t TABLESAMPLE SYSTEM(100 * :samp_rate),
      transitem ti, loc loc
  WHERE t.transid = ti.transid AND loc.locid = t.locid
GROUP BY loc.country, year(t.pdate)
```
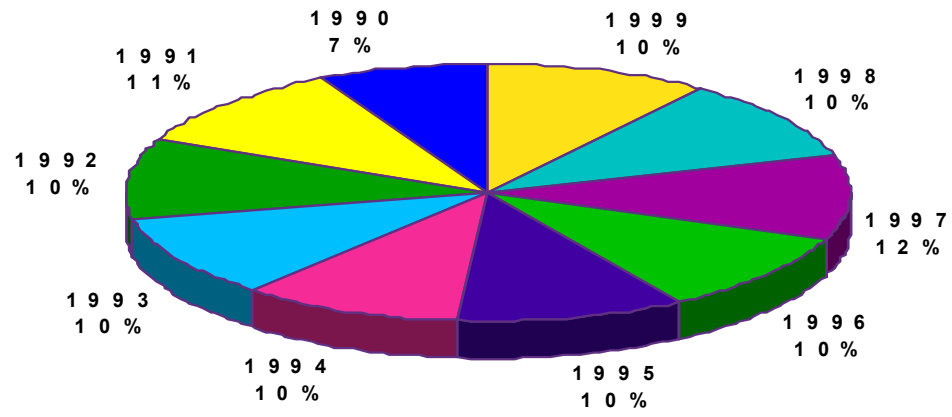
country	year	est_sales	std_err
USA	1998	127505	1326.09
USA	1999	236744	2133.17
GERMANY	1998	86278	961.45
GERMANY	1999	126737	1488.66
. . .			

```
Large-table results:
--------------------------
within +/- 1.96 SE's: 95%
within +/- 2.58 SE's: 99%
```
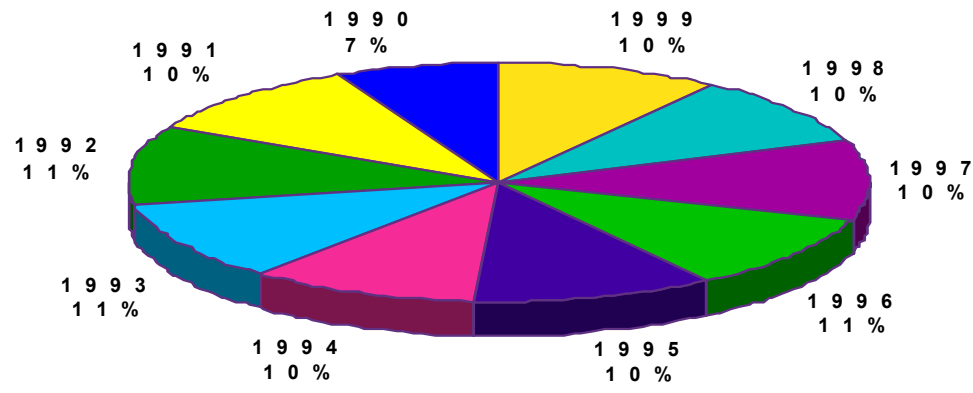
Can YOU tell the difference?

All data

1% sampling

1% sampling

Further Reference - IBM Redbooks

- DB2 UDB ESE: Partitioning for Performance in an e-Business Intelligence World
  - (SG24-6917)
- DB2 UDB's High-Function Business Intelligence in e-business
  - (SG24- 6546)

www.redbooks.ibm.com

Reference - DB2 UDB V8.1 - Sample Other new features

- Connection Concentrator
- Full 64 bit support
- Index and statistics on temp tables
- Partitioned catalog caching
- Dynamic config parms
- Type 2 indexes
- Online runstats
- Increased log size
- Database Container flexibility
- Insert through Union All
- Informational constraints
- Health Center and Memory Visualizer
- SQL in UDFs
- UDF Builder
- Utility throttling