

Realtime Analytics – with Dramatically Simplified Data Stores

May 22, 2013 – IBM Smarter Analytics
Kyosti Laiho [Koppa] – Sales Lead, IBM Databases, Nordic



Themes of presentation

0. Background
1. Speed
2. Simplicity
3. Affordability



Smarter Analytics Should Be Your Goal

CIOs rank
Analytics as the
#1 factor
Contributing to an
organization's
competitiveness.¹

Organizations that
embrace analytics
are more than
2X
as likely to
outperform their
Peers.²



Financial
outperformers are
64%
more likely to
use analytics to
evaluate talent
supply and
demand on an
ongoing basis.³



Enterprises that
apply advanced
analytics have
33%
More revenue
Growth and
12X
more profit
growth.⁴

¹ IBM CIO Study 2009

² IBM IBV/MIT Sloan Management Review Study 2011

³ IBM CHRO Study 2010

⁴ IBM CFO Study 2010

The Analytic Enterprise

Optimization

BI Reporting and Ad-Hoc Analysis

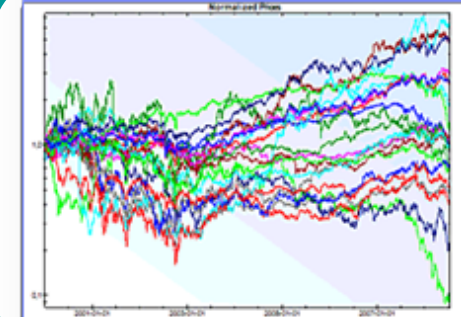
Predictive Analytics



- What happened?
- When and where?
- How much?



- What will happen?
- What will the impact be?



- What are the best choices?
- What actions do we take?

Organizations drive change typically by analyzing the following areas

1

Grow, retain and satisfy customers



Examples:

- Churn management
- Social media sentiment analysis
- Propensity to buy/Next best action

2

Increase operational efficiency



- Predictive maintenance
- Supply chain optimization
- Claims optimization

3

Transform financial processes



- Rolling plan, forecast and budget
- Financial close process automation
- Real-time dashboards

4

Manage risk, fraud & regulatory compliance



- Operational and financial risk visibility
- Policy and compliance simplification
- Real-time Fraud identification



SAMPLE: CUSTOMER ISSUES & CHALLENGES

Decreasing Loyalty

Consumerization of IT

Intensifying Competition

Social Networking

Soaring Customer Expectations

Increasing Transparency

Globalization

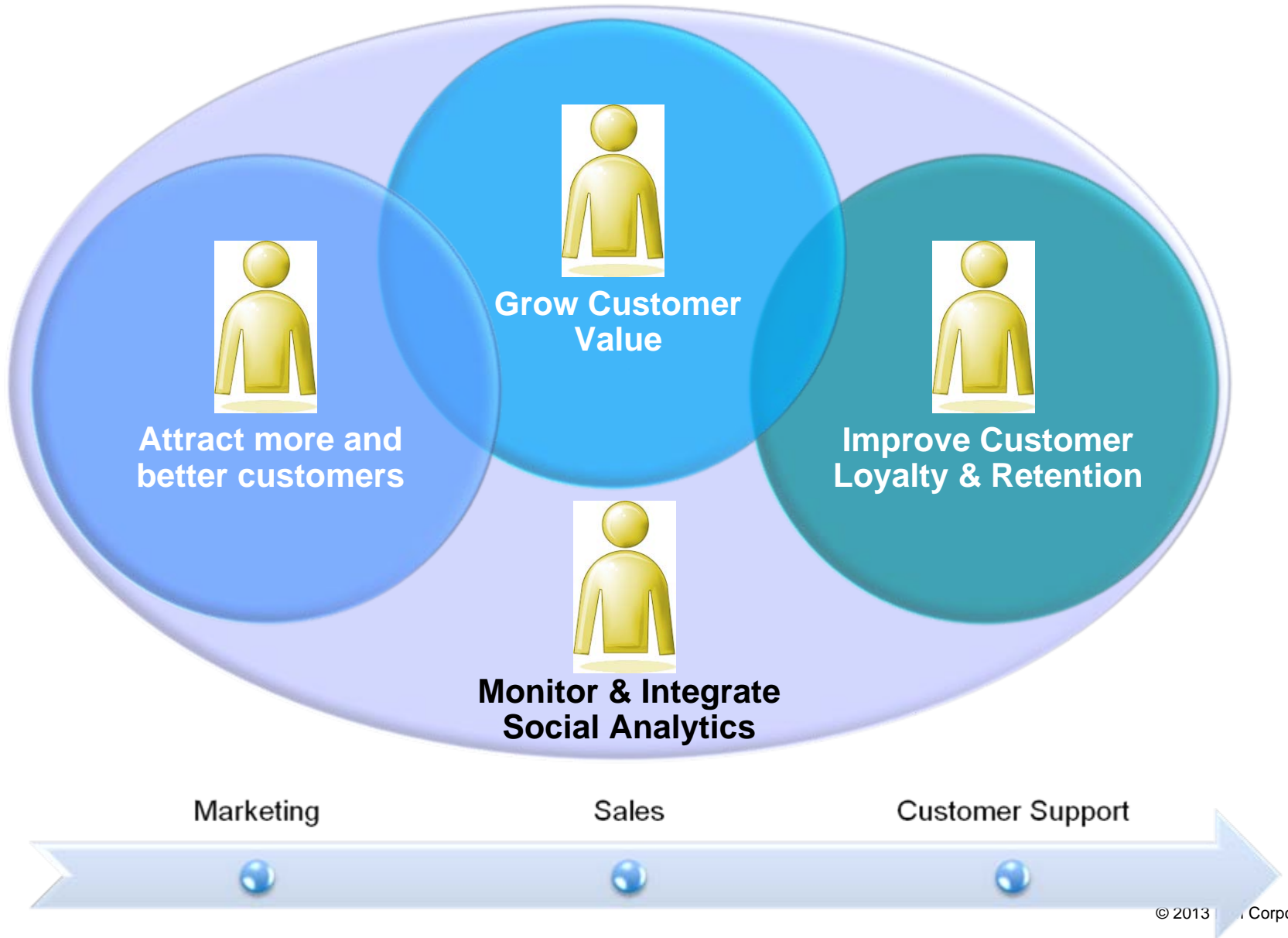
Channel Proliferation and Complexity

Mobile Commerce

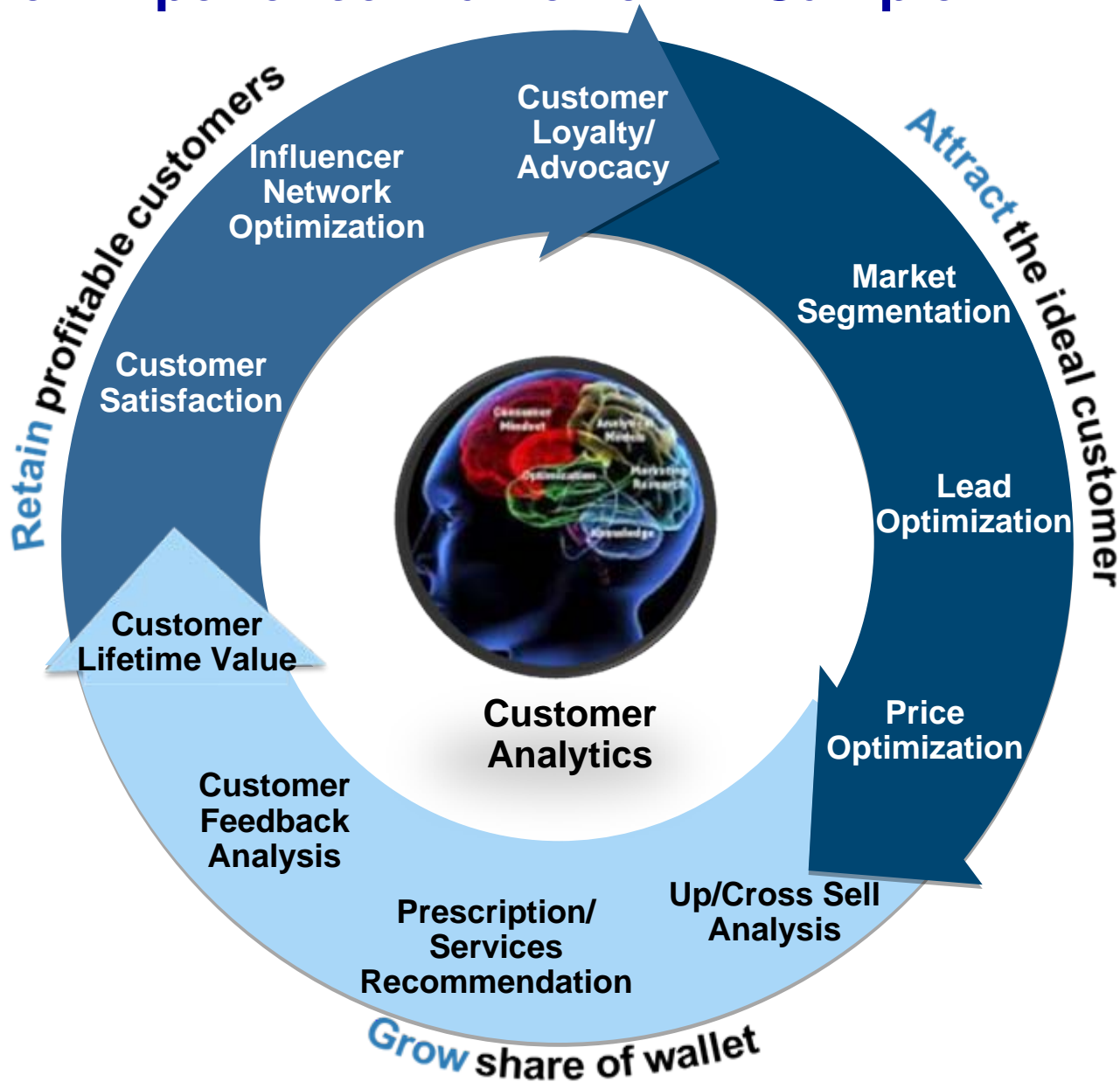
Shrinking Wallet Share



Facets Of Customer Analytics

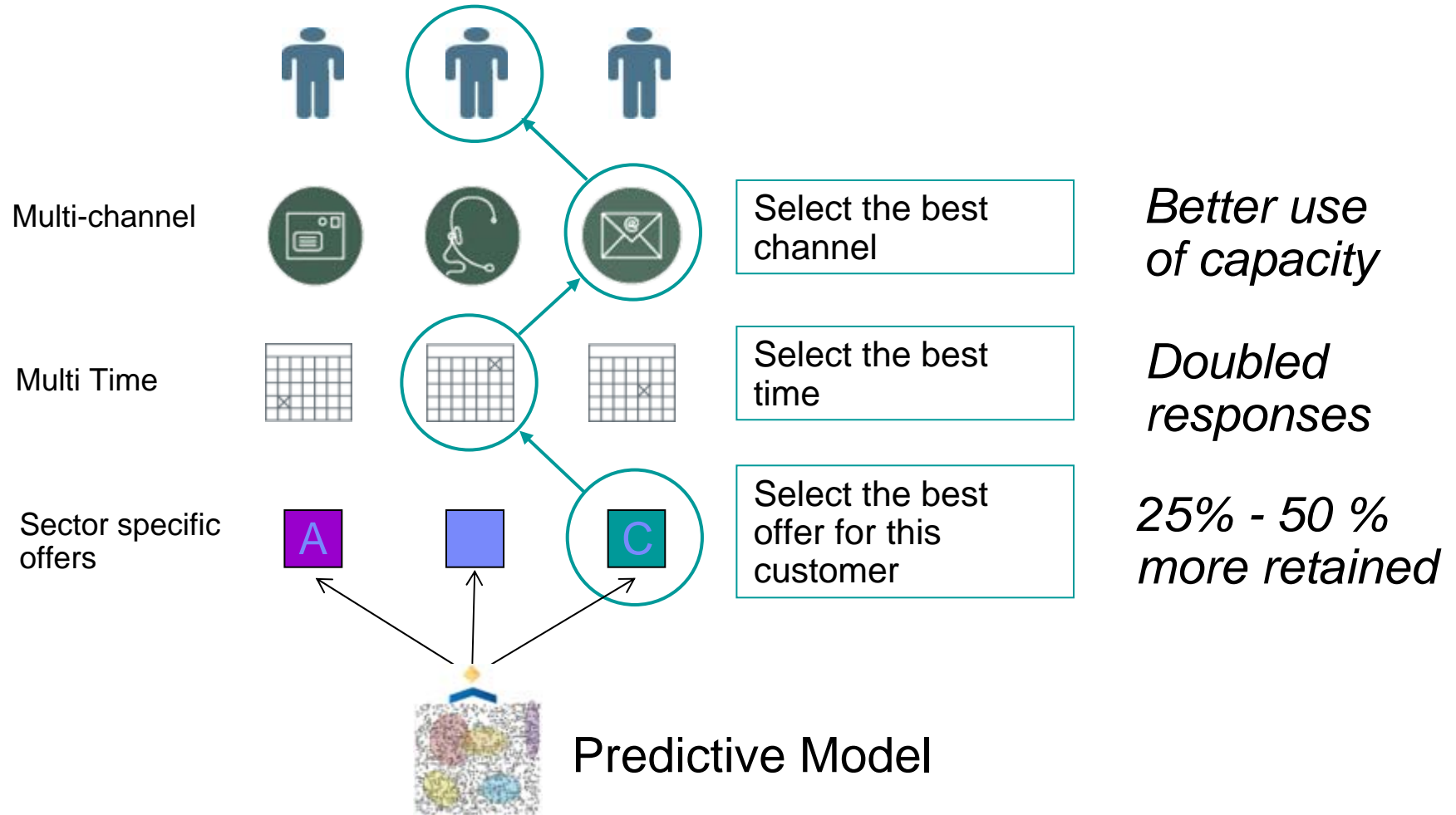


Customer Experience Framework – Sample



Customer Retention Offers: Marketing

Predict the needs, preferences and risks of individual customers



Overall Business Analytics portfolio sample

Business Analytics Scoping

Industry Solutions

<p>Financial Services</p>	<p>Public Sector</p>	<p>Distribution</p>	<p>Industrial</p>	<p>Communications</p>
----------------------------------	-----------------------------	----------------------------	--------------------------	------------------------------

Functional Solutions

Customer		Finance		Operations		Risk	
Customer Acquisition	Social Media Analytics	Budgeting & Forecasting	Financial Consolidation	Resource Optimization	Asset Management	Risk Identification	Risk & Control Assessment
Customer Lifetime Value	Customer Loyalty & Retention	Disclosure Management	Profitability Modeling & Optimization	Production Planning	Decision Management	Risk Mitigation Planning	Risk Aware Decisioning

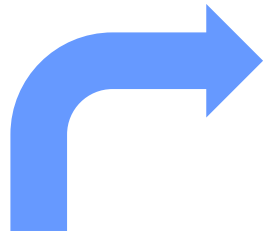
Core Capabilities

REPORT	ANALYZE	MODEL	PLAN	COLLABORATE	PREDICT
Visualize	Discover	Simulate	Govern	Contribute	
Forecast	Mine	Score	Survey	Decide	

Software Categories

<p>Business Intelligence</p>	<p>Predictive Analytics</p>	<p>Performance Management</p>	<p>Risk Analytics</p>
-------------------------------------	------------------------------------	--------------------------------------	------------------------------

BI is getting more complex, with richer sources of data



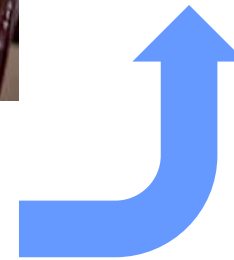
Packaged Reporting



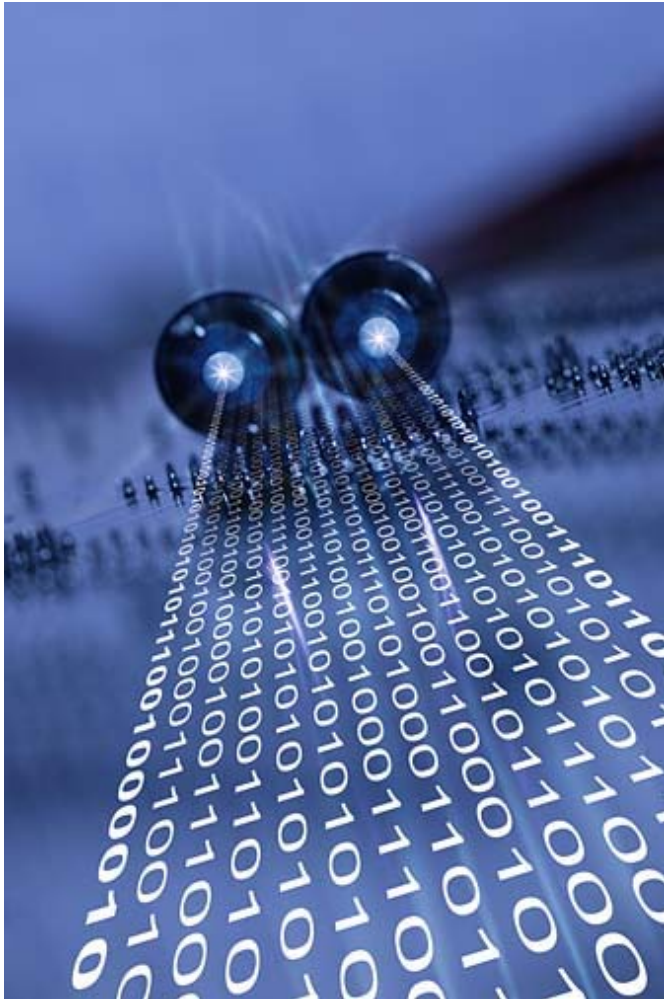
Self-Service BI
and Dashboards



Big Data and
Social Media



Summary on Analytics Requirements



Smarter Analytics has Broad Requirements

- A lot of functional areas to be covered
- Both overall view (dashboards) and detail data (optimize actions plans) is needed at the same time
- Ad-hoc needs are always present
- Social media and Big Data keeps adding to the complexity

Optimizing Performance is difficult

- There is - by definition - a lot of data to process
- Requirements for the data to be viewed, queried, joined etc keep varying rapidly
- Scanning data will happen, needing special expertise to develop faster responses
- Special skillsets and focus continuously needed to keep up with the balance between performance and changing requirements

Underlying Data Stores for Analytics

Simplicity, Flexibility, Choice

Samples of IBM Data Warehouse & Analytics Solutions

IBM PureData System for Analytics (Netezza)



Netezza Appliance

IBM DB2 with BLU Acceleration



Software Only

IBM PureData System for Hadoop



Hadoop Appliance

Distributed Warehouse Data Stores

Information Management Portfolio

(Information Server, MDM, Streams, etc)

The right mix of simplicity and flexibility

New: IBM DB2 with BLU Acceleration

DB2
WITH BLU
ACCELERATION



Super analytics
Super easy

- Operations
 - Simply **Load and Go**
 - Ease of evaluation and performs as advertised

- BI developers and DBAs – faster delivery
 - No configuration or physical modeling
 - No indexes or tuning – out of the box performance
 - Data Architects/DBA focus on business value, not physical design

- ETL developers
 - No aggregate tables needed – simpler ETL logic
 - Faster load and transformation times

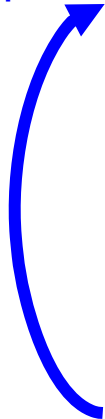
- Business analysts
 - Train of thought analysis – 10x to 25x faster
 - True ad-hoc queries – no tuning, no indexes
 - Ask complex queries against large datasets

Super fast, Super Easy – Create, Load, and Go!

Database Design and Tuning

1. Decide on partition strategies
2. Select Compression Strategy
3. Create Table
4. Load data
5. Create Auxiliary Performance Structures
 - Materialized views
 - Create indexes
 - B+ indexes
 - Bitmap indexes
6. Tune memory
7. Tune I/O
8. Add Optimizer hints
9. Statistics collection

Repeat



DB2 with BLU Acceleration

1. Create Table
2. Load data



DB2 BLU Test: Hardware and Software Details



- Server: POWER7+ 760
 - CPU: 48 cores @ 3.4GHz , 1TB RAM
 - Cognos/DB2 client LPAR: 23 cores, 384GB RAM
 - **DB2 server LPAR: 24 cores, 460GB RAM ***
 - 1 core, 4GB RAM dedicated to VIOS
 - Storage: V7000 with 1TB SSD and 4TB HDD
- Operating system: AIX 7.1
- DB2 versions:
 - DB2 10.1 FP2 Enterprise Server Edition
 - DB2 10.5 (Beta2) Advanced Enterprise Server Edition *
- Cognos Business Intelligence 10.2 (RP1)

*) DB2 10.5 config:

Bufferpool	50GB
Sortheap	16GB
Sheapthres_shr	110GB

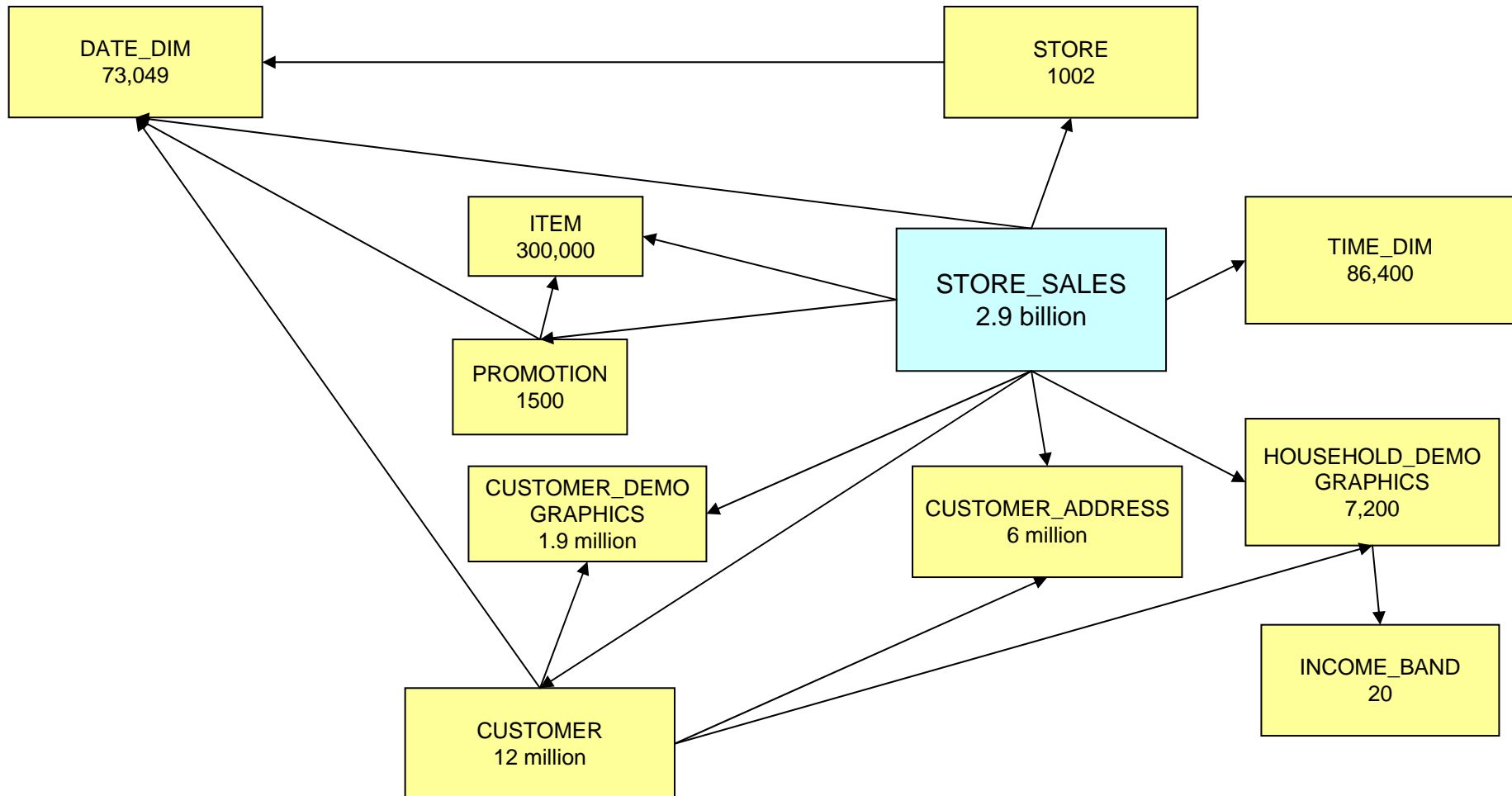
Full database characteristics

- 24 tables: 7 fact, 17 dimension
- Reports & queries utilize single star schema of STORE_SALES fact table with 10 dimension tables

Table name	Raw size	# cols	# rows	Avg row size
STORE_SALES	396 GB	23	2,879,987,999	69
CUSTOMER	1560 MB	19	12,000,001	109
CUSTOMER_ADDRESS	646 MB	13	6,000,001	165
CUSTOMER_DEMOGRAPHICS	77 MB	9	1,920,801	70
ITEM	82 MB	23	300,001	526
TIME_DIM	5.2 MB	10	86401	108
DATE_DIM	10 MB	31	73050	166
HOUSEHOLD_DEMOGRAPHICS	0.1 MB	5	7201	42
PROMOTION	0.2 MB	20	1501	201
STORE	0.2 MB	30	1003	353
INCOME_BAND	0.3 KB	3	21	24

TPC-DS STORE_SALES Star Schema

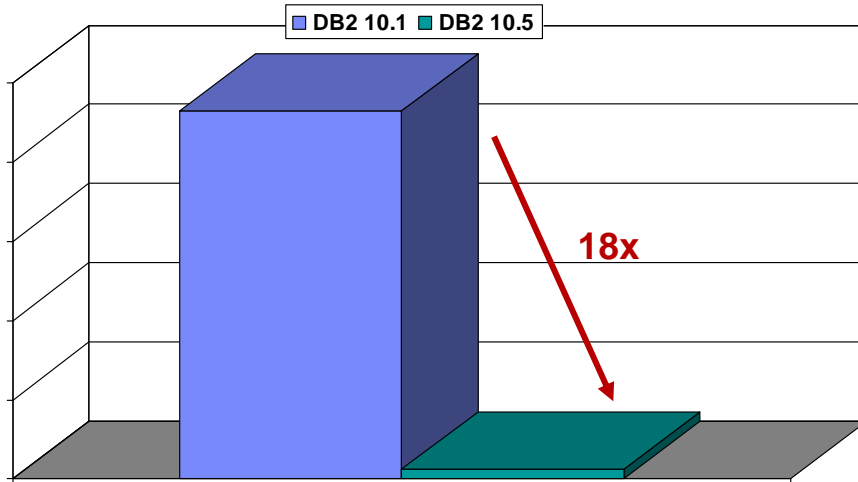
- TPC Benchmark™DS (TPC-DS) inspired database and workload with SF=1000 (ie. 1TB) *
- Cognos reports focus on STORE_SALES star schema



* TPC, TPC Benchmark, and TPC-DS are trademarks of the Transaction Processing Performance Council (www.tpc.org)

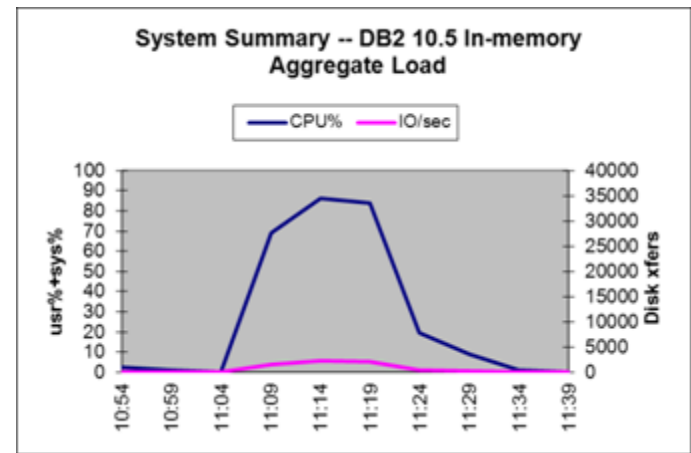
DB2 10.5 and Cognos BI 10.2 Dynamic Cubes on Power 7+

Cognos Aggregate Cache Load Elapsed Time



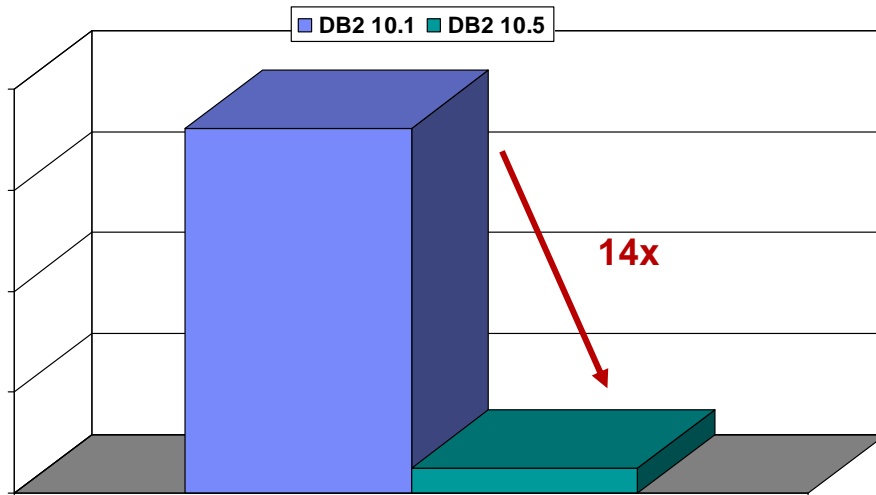
12 concurrent threads used in Dynamic Cube loading:

DB2 10.1: 9 hrs 29 mins
DB2 10.5 (beta2): 29 mins



DB2 10.5 and Cognos BI 10.2 Dynamic Cubes on Power 7+

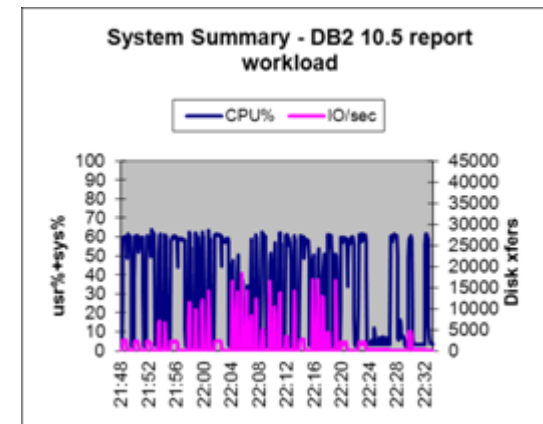
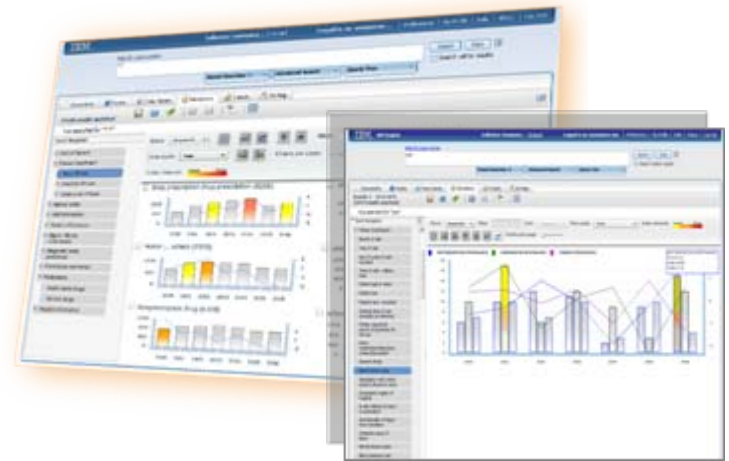
Report Workload Elapsed Time



30 reports – SQL statement execution time – ‘warm runs’:

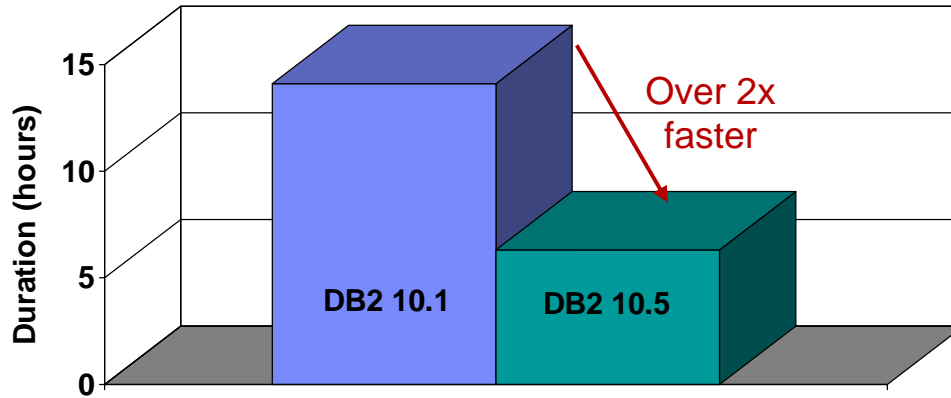
DB2 10.1: 3 hrs

DB2 10.5 (beta2): 13 mins

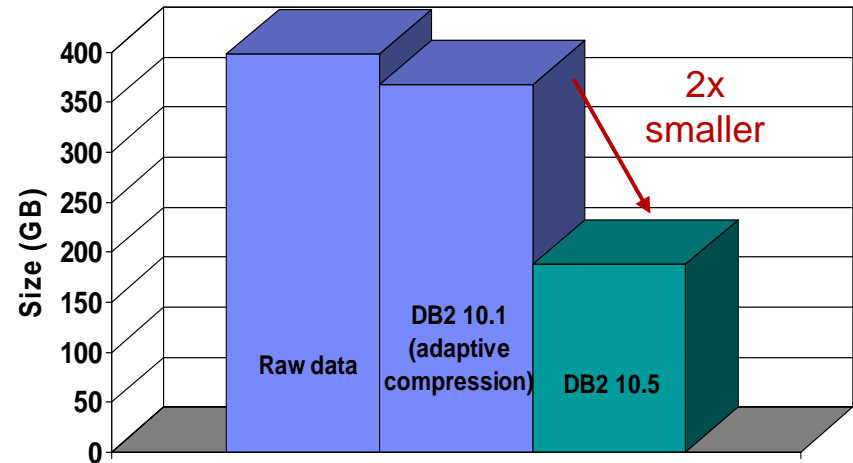


DB2 10.5 on Power 7+

Database creation time for STORE_SALES star schema



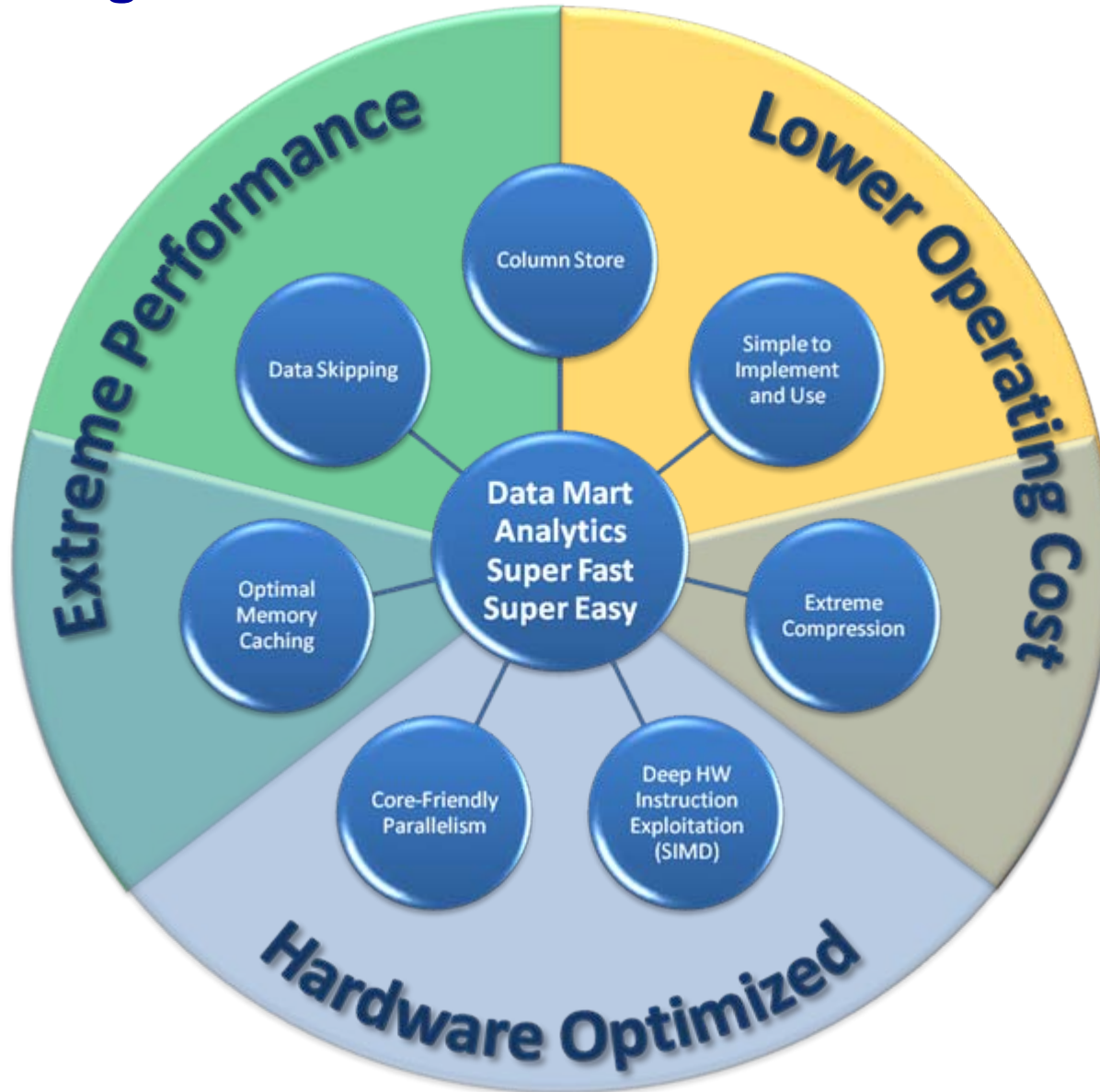
Database size for STORE_SALES star schema



Video Clip on DB2 BLU and Cognos BI

DB2 with BLU Acceleration Technology 'What's Under the Hood'

The Seven Big Ideas of DB2 with BLU Acceleration



7 Big Ideas: **1** Simple to Implement and Use

- **LOAD and then... run queries**
 - No indexes
 - No REORG (it's automated)
 - No RUNSTATS (it's automated)
 - No MDC
 - No MQTs or Materialized Views
 - No partitioning
 - No statistical views
 - No optimizer hints

- **It is just DB2!**
 - Same SQL, language interfaces, administration
 - Reuse DB2 process model, storage, utilities

7 Big Ideas: **1** Simple to Implement and Use

- One setting optimized the system for BLU Acceleration
 - Set `DB2_WORKLOAD=ANALYTICS`
 - Informs DB2 that the database will be used for analytic workloads

- Automatically configures DB2 for optimal analytics performance
 - Makes column-organized tables the default table type
 - Enables automatic workload management
 - Enables automatic space reclaim
 - Page and extent size configured for analytics
 - Memory for caching, sorting and hashing, utilities are automatically initialized based on the server size and available RAM

- Simple Table Creation
 - If `DB2_WORKLOAD=ANALYTICS`, tables will be created column organized automatically
 - For mixed table types can define tables as `ORGANIZE BY COLUMN` or `ROW`
 - **Compression is always on – No options**











- Easily convert tables from row-organized to column-organized
 - `db2convert` utility

7 Big Ideas: **2** Compute Friendly Encoding and Compression

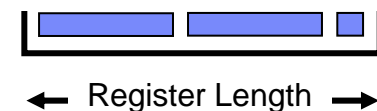
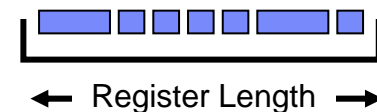
- Massive compression with approximate Huffman encoding
 - More frequent the value, the fewer bits it takes

- Register-friendly encoding dramatically improves efficiency
 - Encoded values packed into bits matching the register width of the CPU
 - Fewer I/Os, better memory utilization, fewer CPU cycles to process

LAST_NAME Encoding

Johnson	
Smith	
Smith	
Smith	
Smith	
Johnson	
Smith	
Gilligan	
Sampson	
Smith	

Packed into register length



7 Big Ideas: 2 Data Remains Compressed During Evaluation

- Encoded values do not need to be decompressed during evaluation
 - Predicates and joins work directly on encoded values

```
SELECT COUNT(*) FROM T1 WHERE LAST_NAME = 'SMITH'
```

LAST_NAME Encoding

Johnson 

Smith 

Smith 

Smith 

Smith 

Johnson 

Smith 

Gilligan 

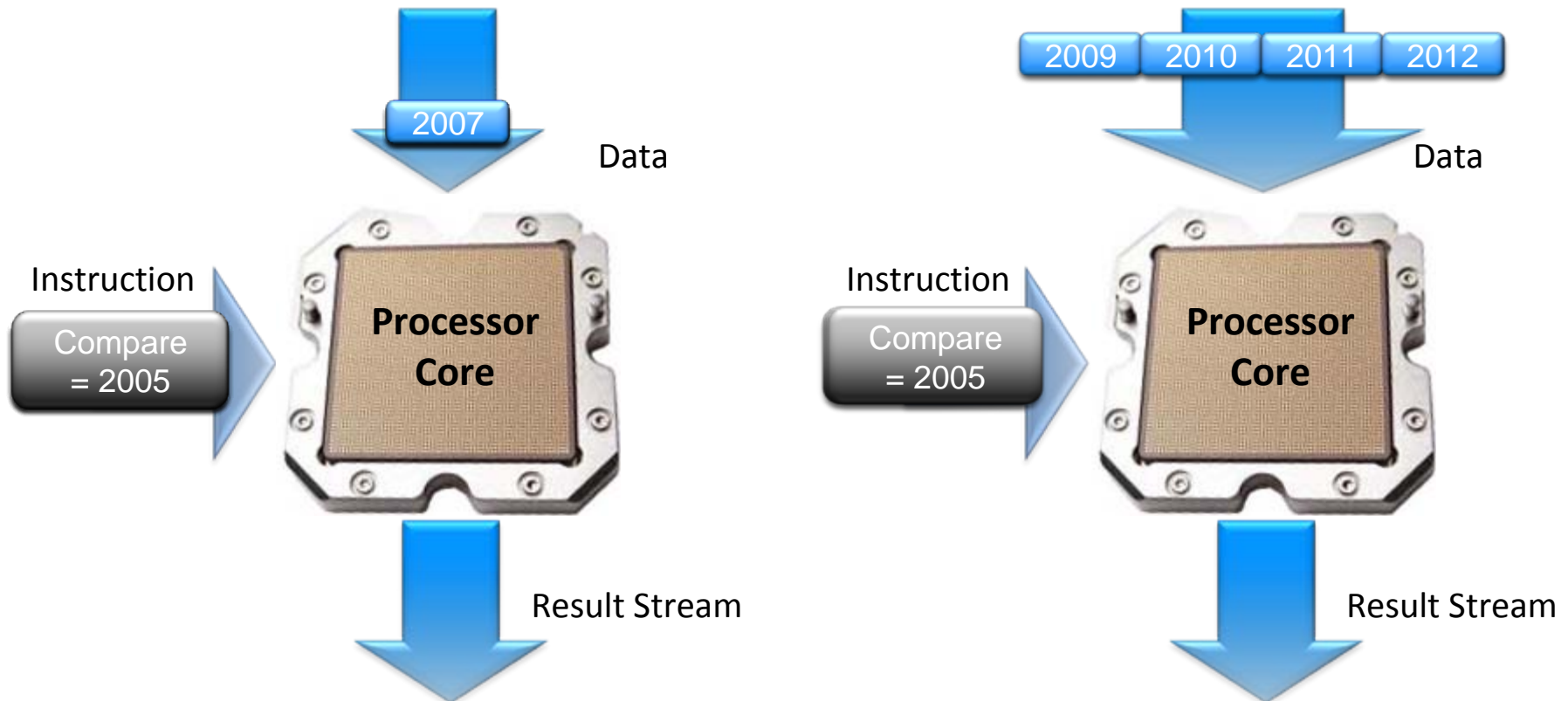
Sampson 

Smith 



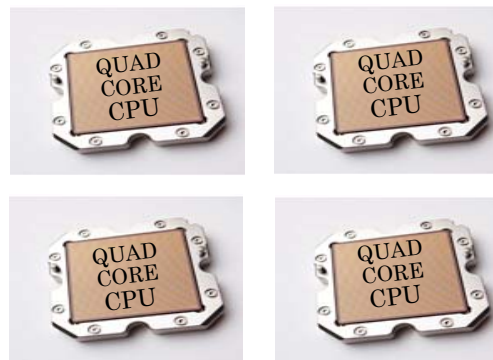
7 Big Ideas: 3 Multiply the Power of the CPU

- Performance increase with Single Instruction Multiple Data (SIMD)
- Using hardware instructions, DB2 with BLU Acceleration can apply a single instruction to many data elements simultaneously
 - Predicate evaluation, joins, grouping, arithmetic



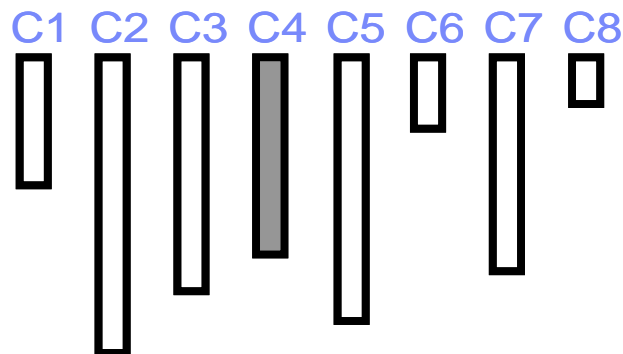
7 Big Ideas: 4 Core-Friendly Parallelism

- Careful attention to physical attributes of the server
 - Queries on BLU Acceleration tables automatically parallelized
- Maximizes CPU cache, cacheline efficiency



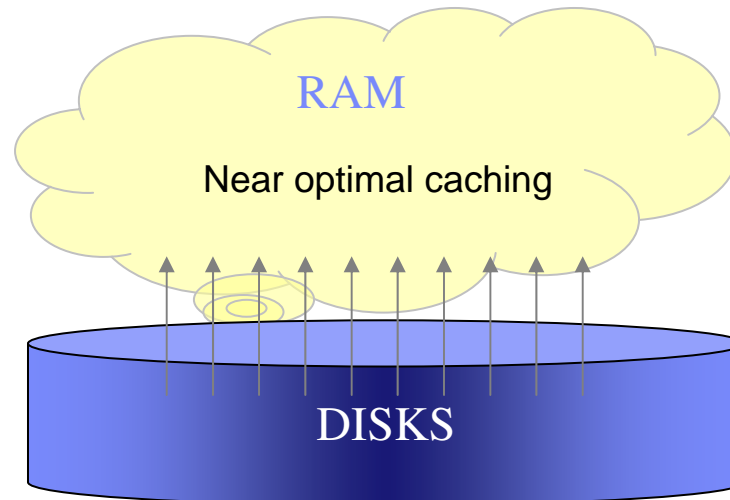
7 Big Ideas: 5 Column Store

- **Minimal I/O**
 - Only perform I/O on the columns and values that match query
 - As queries progresses through a pipeline the working set of pages is reduced
- **Work performed directly on columns**
 - Predicates, joins, scans, etc. all work on individual columns
 - Rows are not materialized until absolutely necessary to build result set
- **Improved memory density**
 - Columnar data kept compressed in memory
- **Extreme compression**
 - Packing more data values into very small amount of memory or disk
- **Cache efficiency**
 - Data packed into cache friendly structures



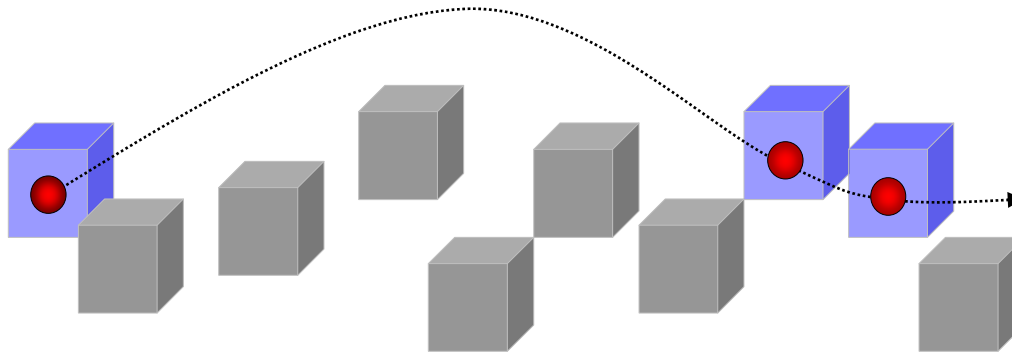
7 Big Ideas: ⑥ Scan-Friendly Memory Caching

- New algorithms cache in RAM effectively
- High percent of interesting data fits in memory
 - We leave the interesting data in memory with the new algorithms
- Data can be larger than RAM
 - No need to ensure all data fits in memory
 - Optimization for in memory and I/O efficiency



7 Big Ideas: 7 Data skipping

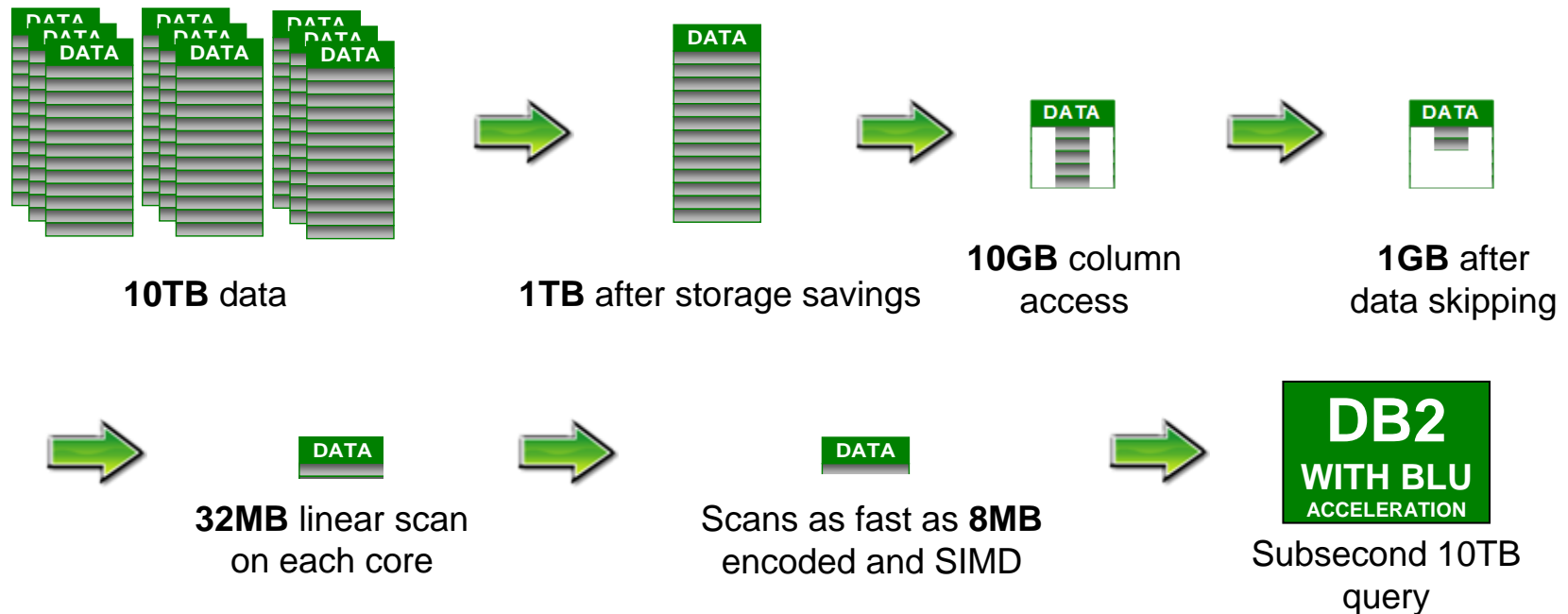
- Automatic detection of large sections of data that do not qualify for a query and can be ignored
- Order of magnitude savings in all of I/O, RAM, and CPU
- No DBA action to define or use – truly invisible
 - Persistent storage of min and max values for sections of data values



7 Big Ideas: How DB2 with BLU Acceleration Helps

~Sub second 10TB query – An Optimistic Illustration

- The system – 32 cores, 10TB table with 100 columns, 10 years of data
- The query: `SELECT COUNT(*) from MYTABLE where YEAR = '2010'`
- The optimistic result: sub second 10TB query! Each CPU core examines the equivalent of just 8MB of data

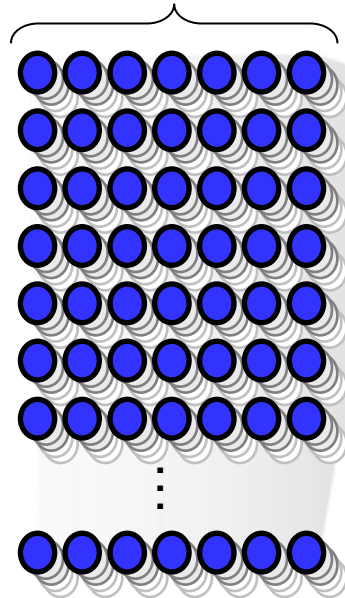


Unlimited Concurrency with “Automatic WLM”

- DB2 10.5 has built-in and automated query resource consumption control
- Every additional query that runs naturally consumes more memory, locks, CPU, and memory bandwidth. In other database products more queries means more contention
- DB2 10.5 automatically allows a high level of concurrent queries to be submitted, but limits the number that consume resources at any point in time
- Enabled automatically when `DB2_WORKLOAD=ANALYTICS`

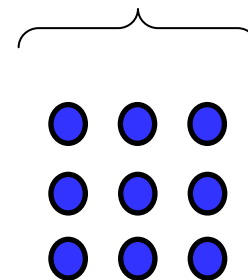
Applications and Users

Up to tens of thousands of SQL queries at once



DB2 DBMS kernel

Moderate number of queries consume resources



Workloads Where You Would Consider BLU Acceleration

- Data Mart Analytic Workloads
 - Characterized by grouping, aggregation, range scans, etc.
 - Typically selecting more than 1% of the data
 - Typically accessing < 1/4 of the table columns in any single query
 - Typically Star / Snowflake Schema

- SAP Business Warehouse solution
 - DB2 BLU is specifically optimized for SAP BW

- Software only offering
 - Initially AIX or Linux
 - COTS hardware (recent x86 / AMD or Power7 recommended – SIMD support)
 - Virtualization is supported

Themes of presentation

1. Speed
2. Simplicity
3. Affordability



DB2 10.5 BLU Acceleration Affordability

- **Low cost of ownership**
 - Just load the tables and go – less design costs
 - Workgroup Edition – lower license costs
 - Terabyte Licensing – lower license costs
 - COTS Hardware / virtualization – lower HW costs
 - Higher compression rates - less storage costs
 - No indexes ETC – smaller db – less storage costs
- **Significantly Better Performance**
 - Be more productive – get the job done
- Can you afford anything else?



Questions?

Thank You