B18 DB2 OLAP: Some "Rules of Thumb "

Dave Collins, Senior Consultant/Instructor, ThinkFast Consulting

So, you've just completed "DW600: DB2 OLAP, Up and Running" and you're ready to get going. Well, before you head off to build that first cube, you should attend this session. In this session, you will slice your learning curve to size with some important "rules of thumb". We will cover what you need to know about the DB2 OLAP data load, retrieval, and consolidation processes. Once you understand how DB2 OLAP thinks, design is easy!

B18

DB2 OLAP: Some "Rules of Thumb"

Dave Collins, ThinkFast Consulting

IBM Data Management Technical Conference

Anaheim, CA **Sept 9 - 13, 2002**

RETURN TO INDEX

Presentation Agenda

✓ Introduction to ThinkFast

✓ Anatomy of a "Block"

- Dense vs. Sparse
- Reality Check

✓ Block Configuration "Impact"

- Data Loads
- Retrievals
- Consolidation

✓ Q & A… Whenever You Like!

Presentation Agenda

- ✓ Introduction to ThinkFast
- ✓ Anatomy of a “Block”
  - – Dense vs. Sparse
  - – Reality Check
- ✓ Block Configuration “Impact”
  - – Data Loads
  - – Retrievals
  - – Consolidation
- ✓ Q & A… Whenever You Like!

Who We Are…

- ✓ Founded in 1996
- ✓ Partnerships
  - Hyperion Platinum Partner
  - IBM Business Partner
  - Other “Best of Breed”
- ✓ Offices Nationwide:
  - Chicago (Corporate Office), Denver, Atlanta, Detroit, Dallas, Kansas City, and San Francisco
- ✓ 300+ Customers; 500+ Applications Implemented
- ✓ High client satisfaction
  - 60% of New Business from Client Base

ThinkFast Client Sampler

Michaels THE ARTS AND CRAFTS STORE

ZALE CORPORATION

PACTIV

Belk All For You!

AutoZone

WAL★MART

SOUTHWEST AIRLINES

SONIC

Proffitt's

McRAE'S

Younkers

CATERPILLAR

Dominick's

mannesmann

KRAFT

MEN'S WEARHOUSE

PARISIAN

Herberger's

Carson Pirie Scott & Co.

CDW

Sprint

FedEx Express

SAKS FIFTH AVENUE

CHESAPEAKE Bagels, Bakery & Cafe

McDonald's

DAIMLERCHRYSLER

churchs CHICKEN

Popeyes CHICKEN & BISCUITS

HELZBERG DIAMONDS

The Walt Disney Company

TOYS R US

Seattle Coffee company

CINNABON

QUAKER

TARGET

GAP

Ahold

Ameritech Authorized Cellular & Paging Dealer

Torrefazione Italia

SEATTLE'S BEST COFFEE

beallsflorida

Abbott Laboratories

JCPenney

UsFreightways

Neiman Marcus

NORDSTROM

BOSCH

MAYTAG CORPORATION

IBM

© IBM Corporation 2002

ThinkFast CONSULTING

The Inc. 500 List - 2001

Inc 500

2001 Ranking of the Fastest-Growing Private Companies in America.

48. LexJet, Sarasota, FL

49.ThinkFast Consulting,Inc. Chicago, IL

50. CLT Meetings International, Orlando, FL

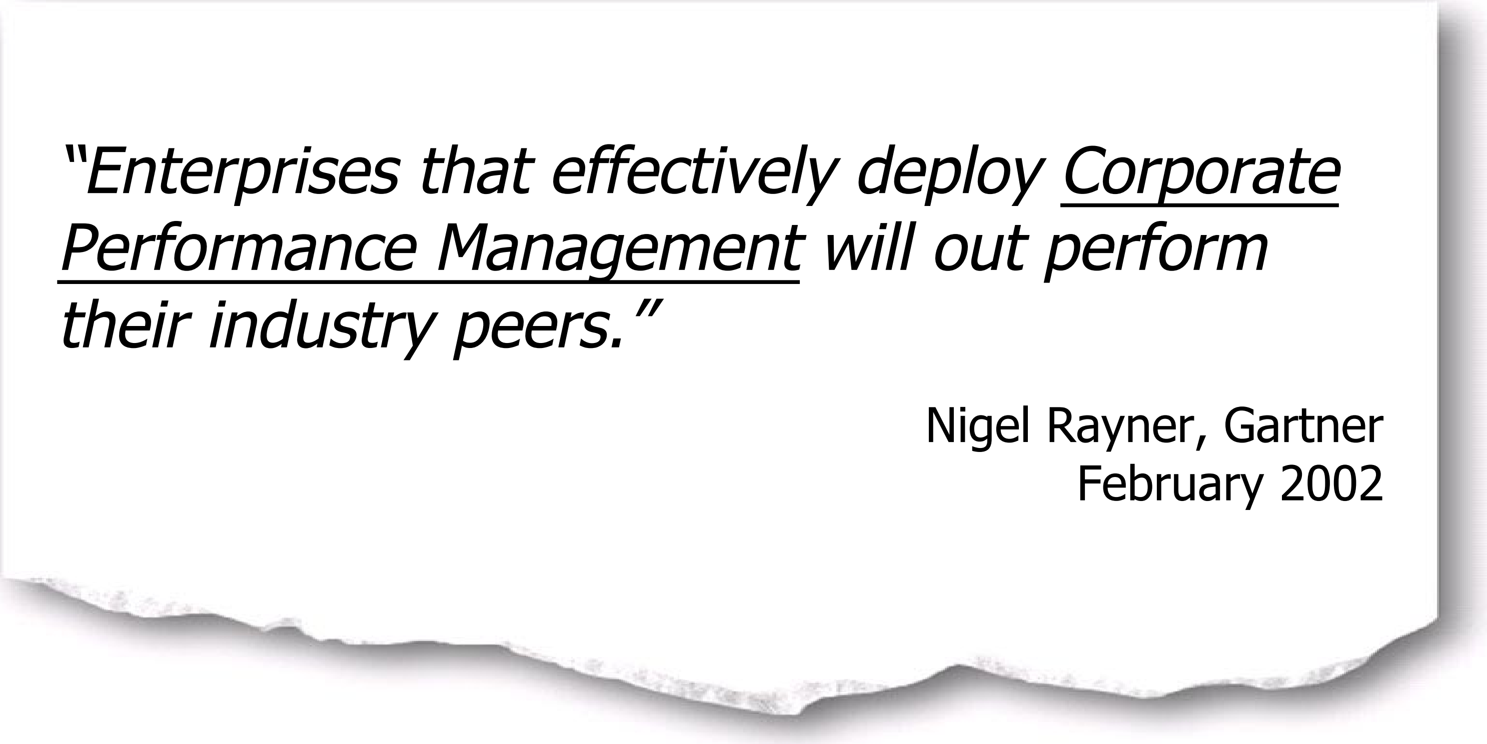
51. Apex Systems, Richmond, VA

52. Cargo Express, Yardley, PA

Source: Inc Magazine October 30, 2001

What We Do…

Business Performance Management is our focus… …it's what we do



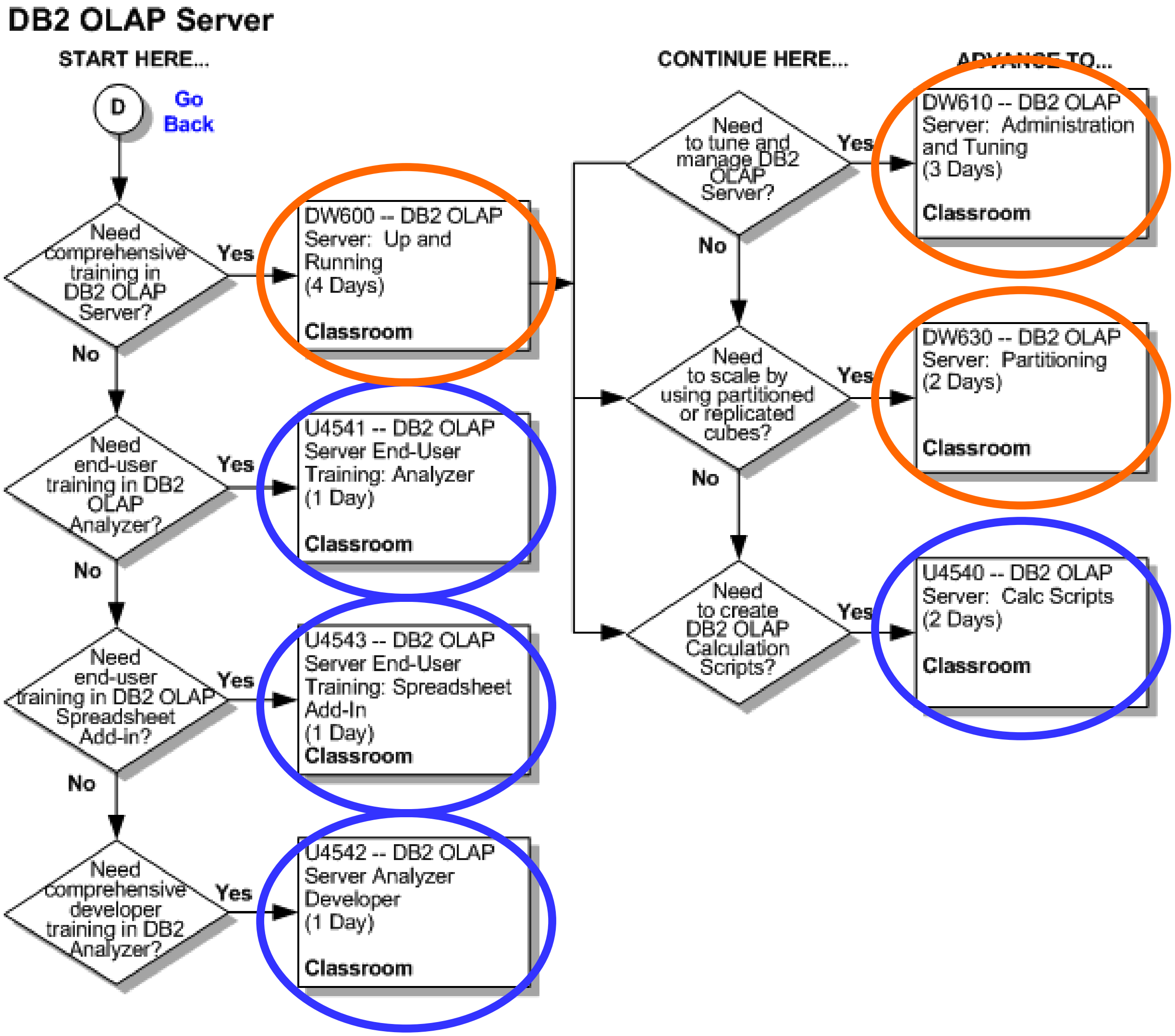
"Enterprises that effectively deploy Corporate Performance Management will out perform their industry peers."

Nigel Rayner, Gartner
February 2002

Why We Can Help…

- ✓ Customer Experience…
  - Enterprise Financial Management
  - Budgeting & Planning
  - Sales & Marketing Management
  - Data Warehousing
  - Education
- ✓ Partnerships…
  - IBM Business Partner
    - 5 IBM Business Intelligence Certified Professionals
  - Hyperion “Knowledge Leader” Award Winner
  - Hyperion Platinum Partner
    - 40 Hyperion Essbase Certified Professionals
- ✓ Proven Team of Experienced Professionals…
  - Consultants averaging 8+ years
  - Project Managers averaging 12+ years

IBM Learning Services Partnership



DB2 OLAP Server

START HERE...

D

Go Back

Need comprehensive training in DB2 OLAP Server?

Yes

DW600 -- DB2 OLAP Server: Up and Running (4 Days)

Classroom

No

Need end-user training in DB2 OLAP Analyzer?

Yes

U4541 -- DB2 OLAP Server End-User Training: Analyzer (1 Day)

Classroom

No

Need end-user training in DB2 OLAP Spreadsheet Add-in?

Yes

U4543 -- DB2 OLAP Server End-User Training: Spreadsheet Add-In (1 Day)

Classroom

No

Need comprehensive developer training in DB2 Analyzer?

Yes

U4542 -- DB2 OLAP Server Analyzer Developer (1 Day)

Classroom

CONTINUE HERE...

ADVANCE TO...

Need to tune and manage DB2 OLAP Server?

Yes

DW610 -- DB2 OLAP Server: Administration and Tuning (3 Days)

Classroom

No

Need to scale by using partitioned or replicated cubes?

Yes

DW630 -- DB2 OLAP Server: Partitioning (2 Days)

Classroom

No

Need to create DB2 OLAP Calculation Scripts?

Yes

U4540 -- DB2 OLAP Server: Calc Scripts (2 Days)

Classroom

Presentation Agenda

- ✓ Introduction to ThinkFast
- ✓ Anatomy of a “Block”
  - Dense vs. Sparse
  - Reality Check
- ✓ Block Configuration “Impact”
  - Data Loads
  - Retrievals
  - Consolidation
- ✓ Q & A… Whenever You Like!

First… Some Key Definitions

✓ OLAP…

- On-Line Analytical Processing
- OLAP vs. On-Line Transaction Processing (OLTP)

✓ IBM DB2 OLAP Server = Hyperion Essbase

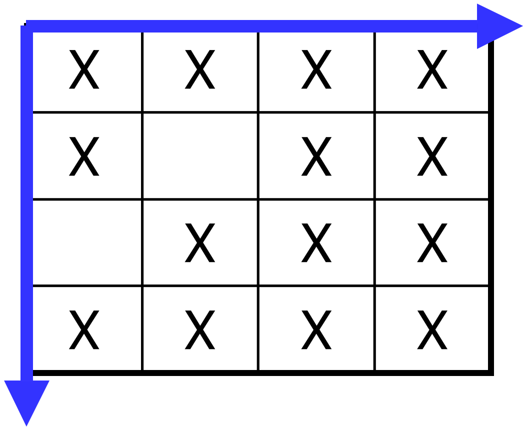
- DB2 OLAP 1.x = Essbase 5.x
- DB2 OLAP 7.1 = Essbase 6.0-6.2
- DB2 OLAP 8.1 = Essbase 6.5

✓ OLAP can be…

- MOLAP: Truly multi-dimensional
- ROLAP: RDBMS made to like MOLAP
- HOLAP: Mixture of MOLAP/ROLAP
- DOLAP: OLAP on one's desktop
- JOLAP: ~JDBC for OLAP

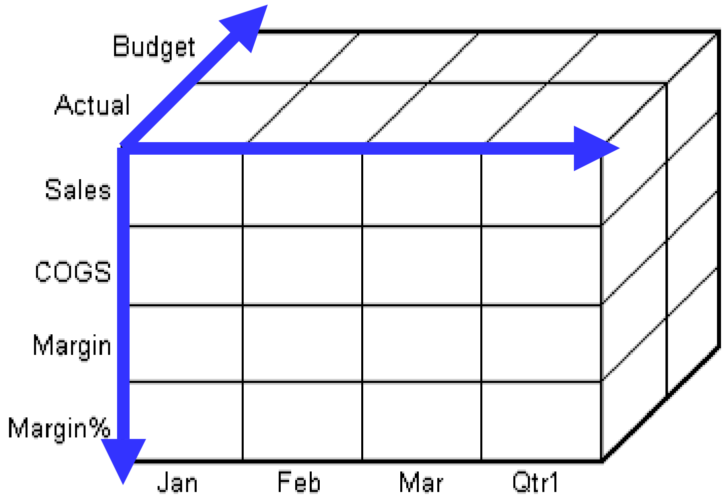
Multi-Dimensional Databases

- ✓ Dimension "Member" Combinations
  - – Created Automatically…
    - Excellent!
    - Or, Whoops!
- ✓ Essbase (aka DB2 OLAP) "Engine"
  - – Engineered to Anticipate Sparsity…
    - Dimensions may be Configured
      - – Dense
      - – Sparse
  - – Only the Relevant Combinations are Stored
    - "Blocks" of Data
    - "Index" Keeps Track
      - – Block Pointers



X	X	X	X
X		X	X
	X	X	X
X	X	X	X

Anatomy of a Data Block

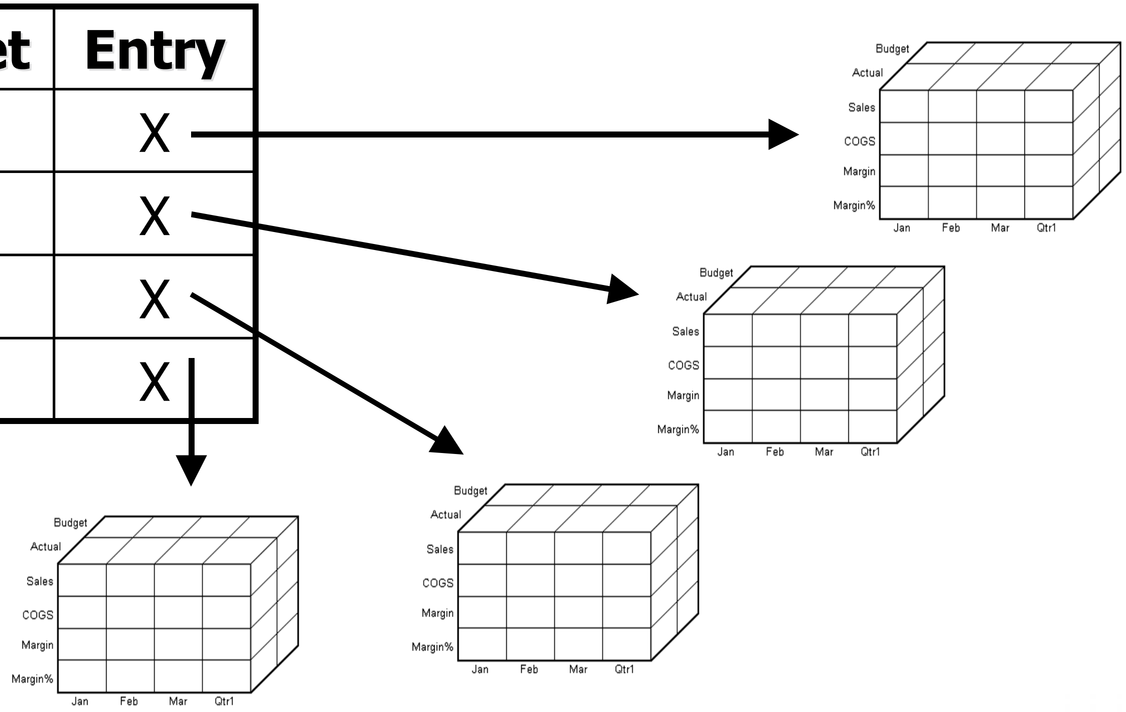


- ✓ Essbase stores data in “Blocks”
- ✓ Dimension’s are configured as Dense or Sparse
- ✓ The contents of a “Block” are defined by the dimensions marked as Dense
- ✓ The “Magic” is Understanding this Storage Method
- ✓ E.g. “Will I need these combinations?”

There are Many Blocks of Data

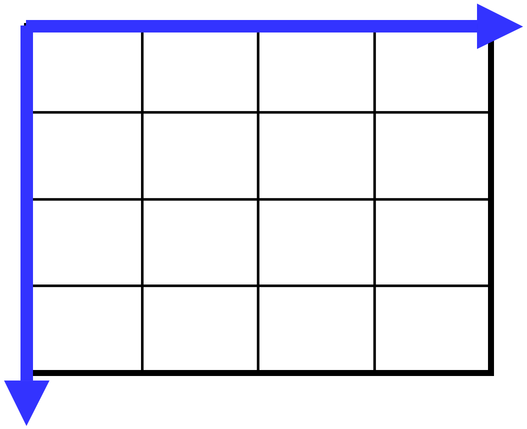
- ✓ One "Block" per Intersection of Sparse Dimension Members
- ✓ An "Index" of the EXISTING Sparse Dimensions Members Keeps Track

Product	Market	Entry
Cola	NY	X
Cola	IL	X
Diet Cola	IL	X
Diet Cola	FL	X



Why Dense and Sparse?

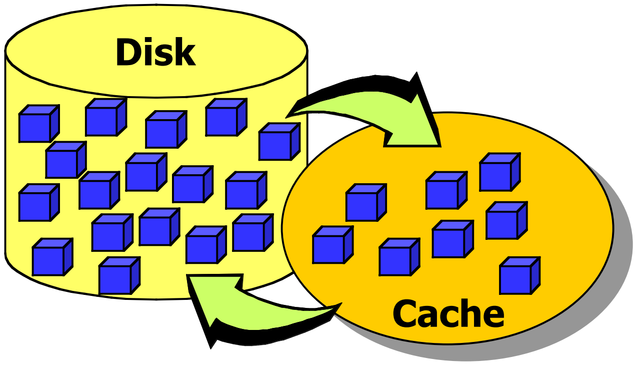
- ✓ Remember, with Multi-dimensionality…
  - – The BEST Thing…
    - Combinations are Automatic
  - – The WORST Thing…
    - Combinations are Automatic



- ✓ Why Build All Possible combinations?
  - – If it’s unlikely the combinations will ever exist… SPARSE
  - – If it’s likely the combinations will exist… DENSE
- ✓ Instead… Build What You Need
  - – When You Need It!
- ✓ Let’s Look some “Classic” Examples
  - – Demo: Multiple Years of Data

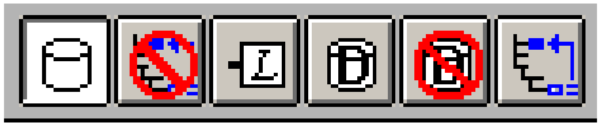
Block Size - Is it Important?

- ✓ Block Size
  - Dense Dimension Member Combinations
- ✓ Why is it ~~Important~~? It's CRITICAL
  - Impacts ALL Transactions
    - Data Loads
    - Calc Times
    - Retrieval Times
- ✓ Optimal Size
  - 8KB to 64KB
    - Per "Database Administrator's Guide"
  - Larger *Generally* Better Than Smaller
    - Reads/Writes Performed by Block



Block Size – How Can I "Optimize" it?

- ✓ Storage Properties (Dense Members)
  - → Store Data… Increase
  - → Never Share Data… Increase
  - ← Label Only… Reduce
  - ← Dynamic Calc… Reduce
  - → Dynamic Calc and Store… Increase
  - ← Shared Member… Reduce



- ✓ Let's have a look…
  - – Demonstration of Outline Optimization

DENSE vs. SPARSE… Reality Check

✓ Six Ways to Think about DENSE vs. SPARSE

1) Data Distribution
2) Block Size
3) Dimension Size
4) Business Function
5) Use of Attribute Dimensions
6) Use of Dynamic Calc

1) Data Distribution

Products

Markets

X				
			X	
X		X		
	X			
			X	X

Sparse:
Few combinations from
all possible ones exist

Measures

Time

X	X	X	X	
X	X	X	X	X
X	X	X		X
	X	X	X	X
X		X	X	X

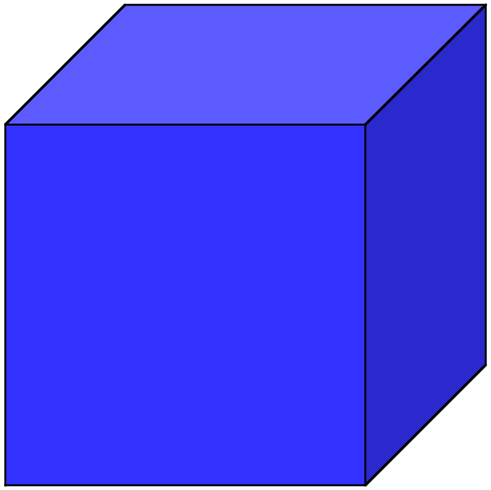
Dense:
Most combinations from
all possible ones exist

2) Block Size

✓Remember…

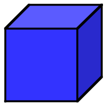
- Block Size
  - Dense Dimension Member Combinations
- Why is it Important?
  - Affects Data Load Times
  - Affects Calc Times
  - Affects Retrieval Times
- Optimal Size
  - 8KB to 64KB
  - Larger Better Than Smaller
    - Reads/Writes by Block

Too Big!



Too Small!

Just Right!

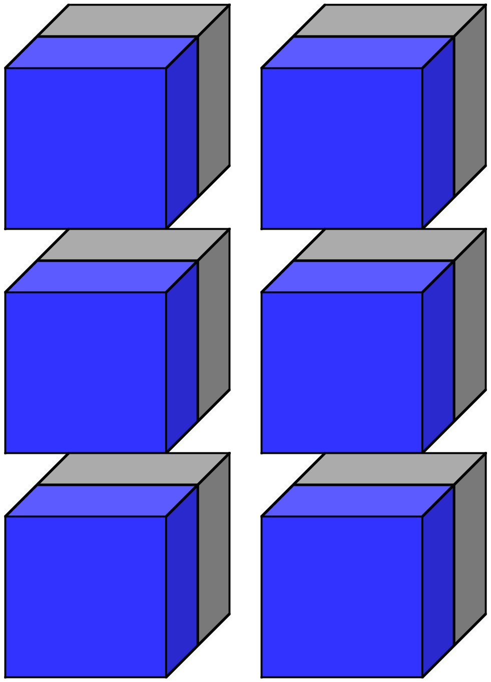


3) Dimension Size

- ✓ The Block Can Only be SO Big…
  - – Combinations!
    - 1000 Accounts BY 12 Months = 12,000 Cells
      - – 96,000 Bytes (96KB) per Block
    - 1000 Accounts BY 12 Months BY 4 Scenarios = 48,000 Cells
      - – 384,000 Bytes (384KB) per Block
- ✓ What If You Have…
  - – 3,000 Products BY 5,000 Customers
    - Data Distribution MAY Be "Dense"
    - Number of Combinations PROHIBITIVE
      - – 15,000,000 Cells = 120,000,000 Bytes (120,000KB)

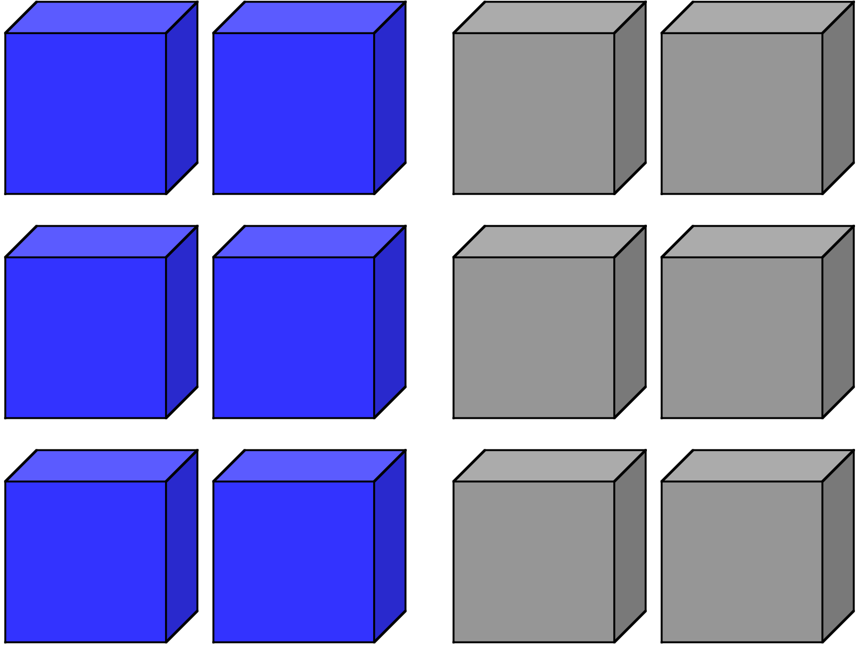
4) Business Function

Actual/Budget - Scenario is DENSE



Vs.

Actual/Budget - Scenario is SPARSE



Why? Organizes I/O per Business Function/Use.

5) Use of Attribute Dimensions

✓Attribute Dimensions

- Boolean: Two_Door_True
- Text: Color_Yellow
- Numeric: MPG_30
- Date: Intro_1/1/2000

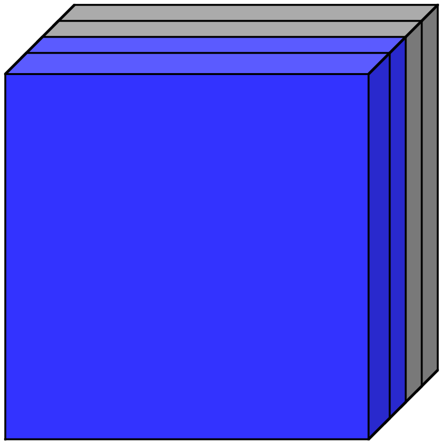
✓Property of a Base Dimension

- SPARSE Dimensions

✓Allows for OLAP Aware Attribute Statistics

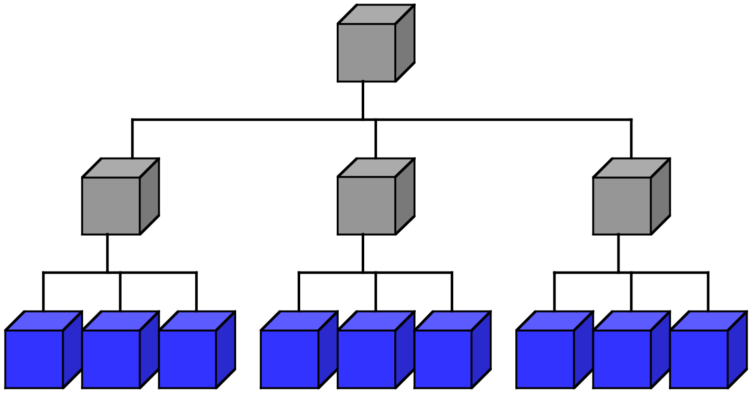
- Sum
- Min
- Max
- Average
- Count

6) Use of Dynamic Calc



✓Dynamic Calculation on a DENSE Member

- I/O Same as Before



✓Dynamic Calculation on a SPARSE Member

- Requires More I/O
- Fetches Descendant Blocks

Presentation Agenda

✓ Introduction to ThinkFast

✓ Anatomy of a “Block”

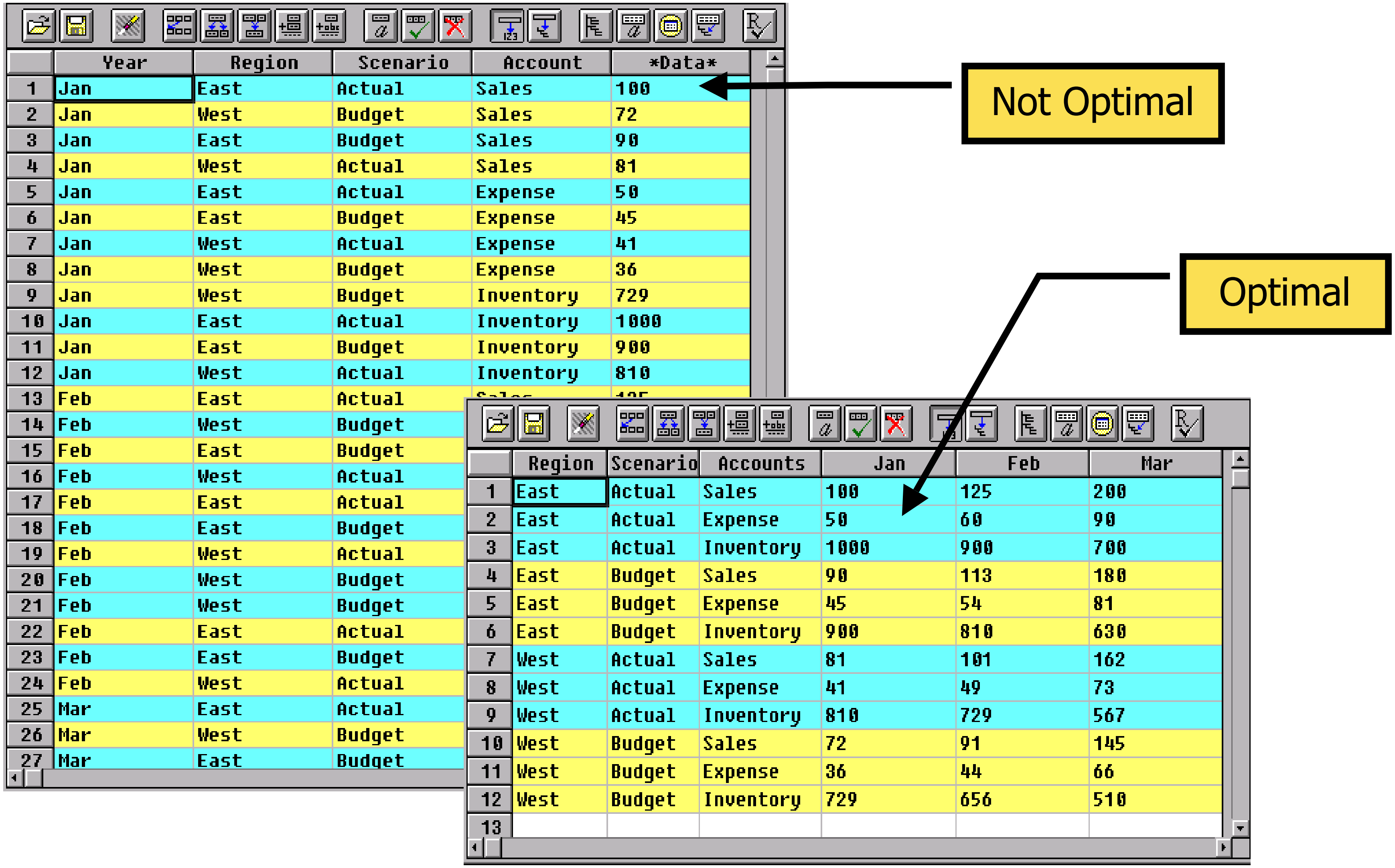
  – Dense vs. Sparse
  – Reality Check

✓ Block Configuration “Impact”

  – Data Loads
  – Retrievals
  – Consolidation

✓ Q & A… Whenever You Like!

Data Load Efficiency



Not Optimal

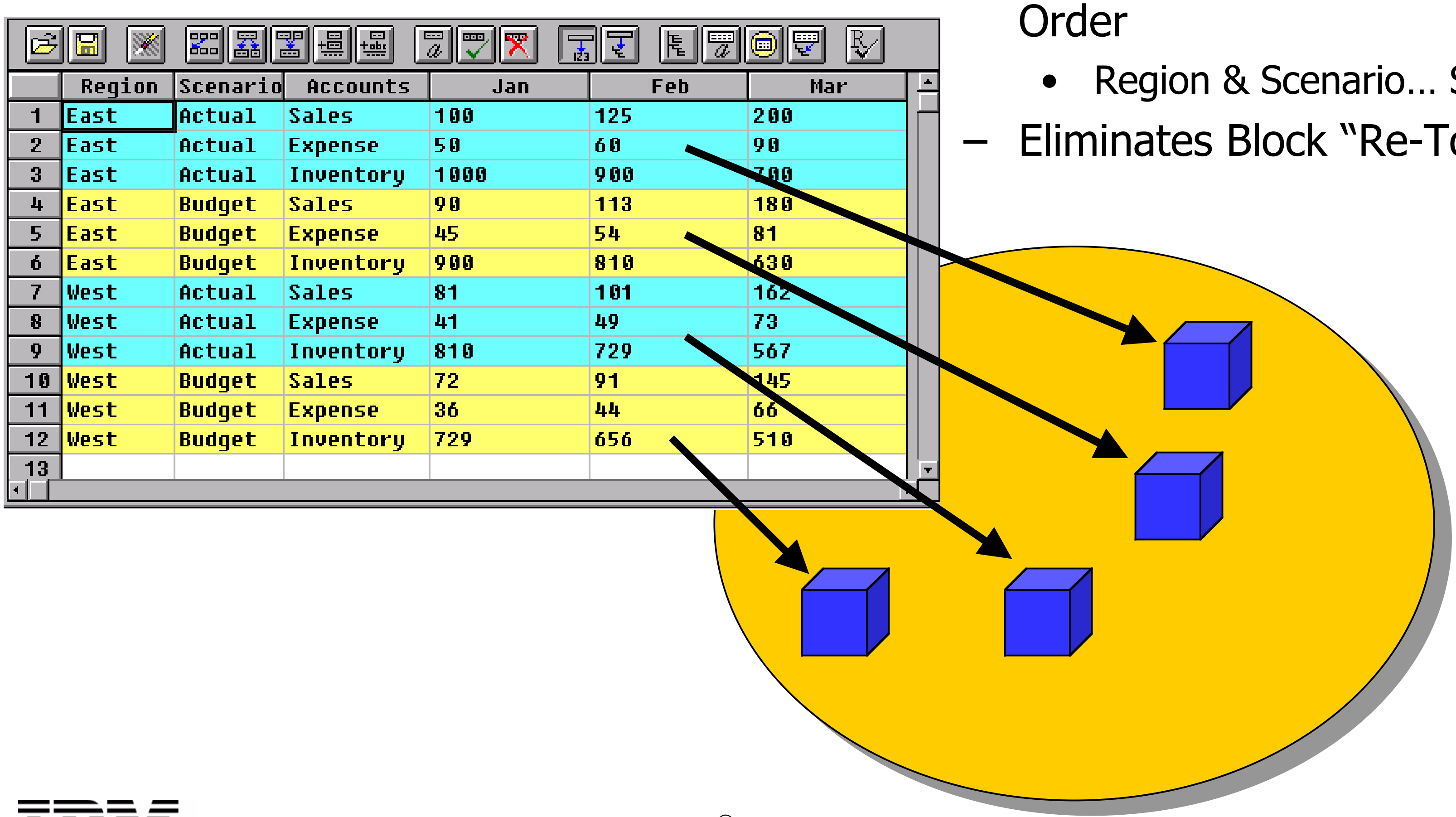
	Year	Region	Scenario	Account	*Data*
1	Jan	East	Actual	Sales	100
2	Jan	West	Budget	Sales	72
3	Jan	East	Budget	Sales	90
4	Jan	West	Actual	Sales	81
5	Jan	East	Actual	Expense	50
6	Jan	East	Budget	Expense	45
7	Jan	West	Actual	Expense	41
8	Jan	West	Budget	Expense	36
9	Jan	West	Budget	Inventory	729
10	Jan	East	Actual	Inventory	1000
11	Jan	East	Budget	Inventory	900
12	Jan	West	Actual	Inventory	810
13	Feb	East	Actual	Sales	125
14	Feb	West	Budget		
15	Feb	East	Budget		
16	Feb	West	Actual		
17	Feb	East	Actual		
18	Feb	East	Budget		
19	Feb	West	Actual		
20	Feb	West	Budget		
21	Feb	West	Budget		
22	Feb	East	Actual		
23	Feb	East	Budget		
24	Feb	West	Actual		
25	Mar	East	Actual		
26	Mar	West	Budget		
27	Mar	East	Budget		

Optimal

	Region	Scenario	Accounts	Jan	Feb	Mar
1	East	Actual	Sales	100	125	200
2	East	Actual	Expense	50	60	90
3	East	Actual	Inventory	1000	900	700
4	East	Budget	Sales	90	113	180
5	East	Budget	Expense	45	54	81
6	East	Budget	Inventory	900	810	630
7	West	Actual	Sales	81	101	162
8	West	Actual	Expense	41	49	73
9	West	Actual	Inventory	810	729	567
10	West	Budget	Sales	72	91	145
11	West	Budget	Expense	36	44	66
12	West	Budget	Inventory	729	656	510
13						

Optimal Block Input/Output

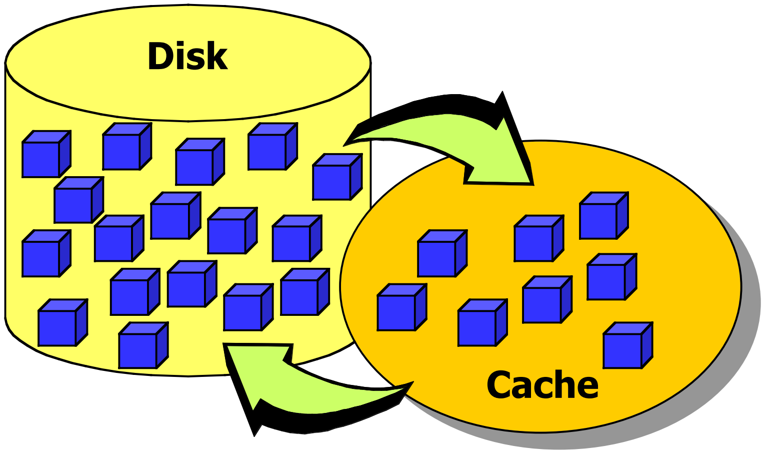
- ✓ It's Easy!
  - – Sort Data by Sparse Members
  - – Translation: Block-by-Block Order
    - Region & Scenario… SPARSE
  - – Eliminates Block "Re-Touch"



	Region	Scenario	Accounts	Jan	Feb	Mar
1	East	Actual	Sales	100	125	200
2	East	Actual	Expense	50	60	90
3	East	Actual	Inventory	1000	900	700
4	East	Budget	Sales	90	113	180
5	East	Budget	Expense	45	54	81
6	East	Budget	Inventory	900	810	630
7	West	Actual	Sales	81	101	162
8	West	Actual	Expense	41	49	73
9	West	Actual	Inventory	810	729	567
10	West	Budget	Sales	72	91	145
11	West	Budget	Expense	36	44	66
12	West	Budget	Inventory	729	656	510
13						

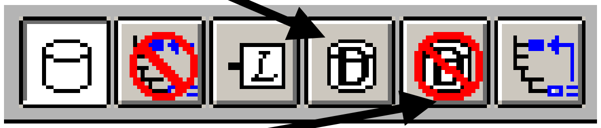
Calculation Efficiency

- ✓ Touch the Blocks ONCE!
- ✓ Calculate in Stages
- ✓ Group Like Calculations Together
  - – Look at Groups vs. Block Configuration
- ✓ Calculate Dense BEFORE Sparse
  - – Sparse Member Calc's Increase Blocks
  - – Dense Member Calc's "Fill-in" Blocks
- ✓ Use "Dynamic" AMAP
  - – As Much As Possible
  - – Generally on Dense Members
    - More on that soon…



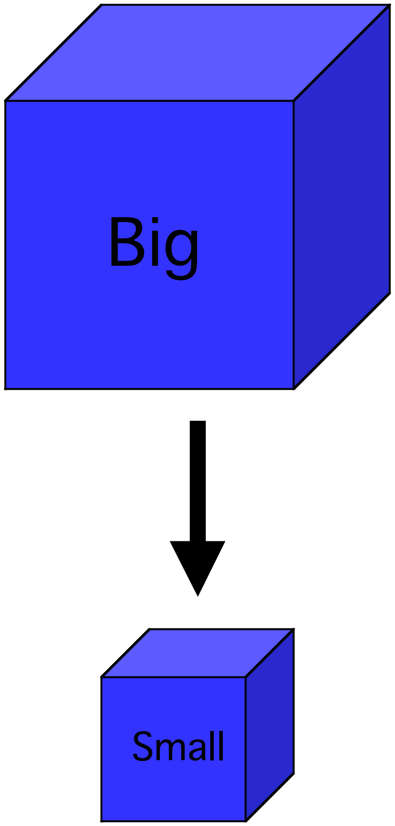
Use Of Dynamic Calculation

- ✓ Dynamic Calc and Store
  - – Request then CALC then Store
  - – No Reduction in DB/Block Size
    - Data is Stored
  - – “Sticky” Store
    - If Details Change… Out of Sync!
    - Use When Data is “Static”



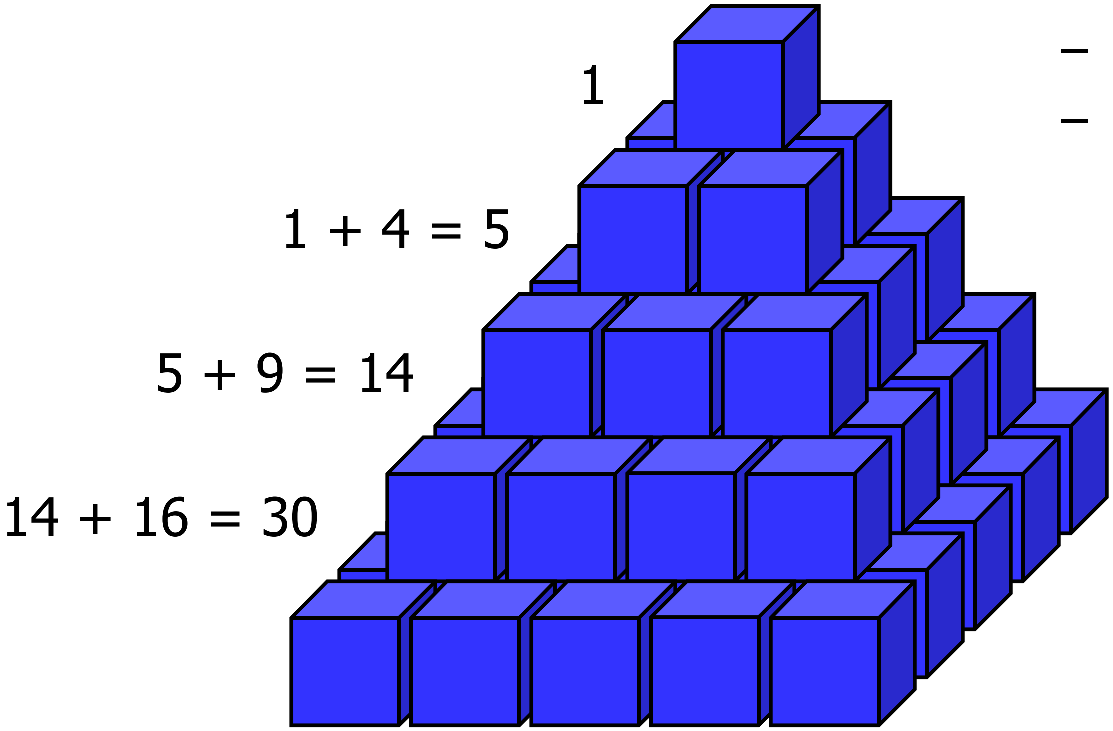
- ✓ Dynamic Calc
  - – Request then CALC… Every Time
  - – Reduction in DB/Block Size
    - Data is NOT Stored
    - If Details Change… No Problem!
  - – Best Used on Dense Dimension Members

Why Dynamic Calc DENSE Members?



- ✓ Any "Calculated" DENSE Member
  - Consolidation Point
  - Formula
- ✓ Eliminate Block "Re-Touch"
  - Multi-PASS Calc Solution?
- ✓ Pseudo-Normalization
  - Store DETAILS
  - On-the-Fly CONSOLIDATION
- ✓ No I/O Increase
  - Same Block of Data Read as Before!
- ✓ Smaller Blocks… Smaller Cube
- ✓ Remember 6)

Take Care on SPARSE Members



- ✓ Higher "Level" Sparse Dimension Members
  - i.e. Tree Tops
- ✓ Keep to a Small "Fan-Out"
  - The Deeper You Go…
  - The More Block I/O You Create

Questions?

To download the latest version of this or any other presentation, please visit:

www.ThinkFast.com/Presentations.html

To reach me directly, please email me at:

DCollins@ThinkFast.com

RETURN TO INDEX