

SMARTARRAYS®

ARRAY TECHNOLOGY FOR BUSINESS ANALYTICS

Innovative Approaches To High-Performance Analytic Data Applications

March 27, 2007





Agenda

- Introduce SmartArrays
- Case Studies using SmartArrays technology
- Describe the concepts behind SmartArrays technology
- Show why it works so well
- Demonstrate Ease of Implementaion



SmartArrays, Inc.



Jim Brown: IBM veteran and inventor of APL2



James Wheeler: Array technology head at Manugistics



Pat Buteux, founding executive at Manugistics

Incorporated in 1999



Customer Quotes

“SmartArrays is an impressive technology. Anybody with web applications that do number crunching will find it useful. SmartArrays can be used as a **‘powerful back-end analytics server.’**”

Clients should use different tools like DB2 for the transactional data store, and a snazzy web page formatting tool for the glitzy front end, but **put the heart of the analytics for the application in SmartArrays.**”

**Mike Van Der Meulen,
IBM, WebAhead Technologies.**



Customer Quotes

“The TDD Tool application that SmartArrays developed for DOD, based on their SmartArrays technology, **far exceeded our expectations for the project.”**

With the new information we learned from our data and we have recommended changes in procedures .”

**Deb Bennett, Office of Supply Chain
Integration,
Department of Defense**



Customer Quotes

“We are delighted with the astounding performance of SmartArrays.

“SmartArrays has worked wonders for us and we view it as one of the keys to our success.”

**Clive Verrall, BNP Paribas
London**



SmartArrays Success Formula

Find applications where canned solutions don't work

+

Very fast analytics on very large data sets are required

+

Leverage the SmartArrays technology to develop a
custom solution

=

**Cost Effective, quickly developed, analytic data
applications that meet customer's needs**



Finding the applications

- A burdened shared resource where unacceptable response times is expected.
- Times series or continuous-measure data
- Pushing the limits of SQL, things it can't do
- Needs custom functions
- RDBMS or OLAP too expensive
- Budget cycle will take too long

Alternative: Create custom application using a virtual database in SmartArrays Analytic Data Toolkit classes to perform the analysis



Alternative to BI and OLAP solutions

- Use your own developers, or external resources skilled in SmartArrays
- Adopt agile development principles – fast production ready prototypes, constant collaboration with actual users, quick change cycles

Yield: The right solution, in less time, for less money, with superior performance.



Benefits of SmartArrays Solution

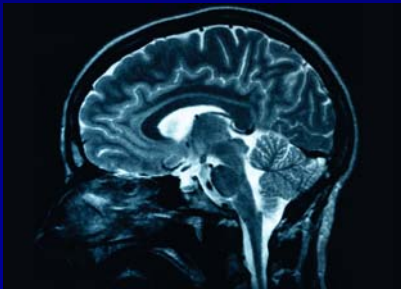
- Unique solution
- Prototype to production is seamless
- Prototype on client side and go to server for production
- Rapid development – using APIs
- Easy to change to meet users needs
- Significantly less cost
- Doesn't interrupt IT architecture
- Create virtual database – separate from production database

Yield: The right solution, in less time, for less money, with superior performance.



Why SmartArrays?

- Help developers think differently about data.
- Accelerate software development with high-level data APIs.
- Accelerate applications with high-performance code.



Accelerate Thought



Accelerate Development



Accelerate Computing



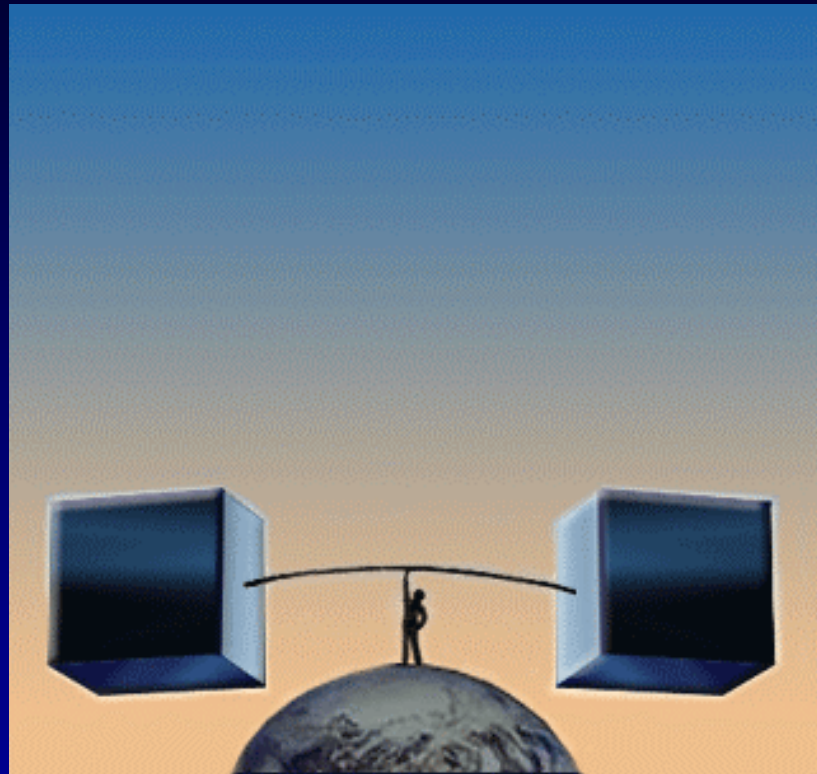
SmartArrays Business Model

- SDK is sold at “per-seat” fee.
- Deployment license for embedding the SmartArrays tools and array engine:
 - Per-server pricing option
 - Per-application pricing option
 - Can be tailored to an independent software vendor’s business model.
- Evaluation licenses are free.
- Attractive pricing for consultants who learn the product



SmartArrays

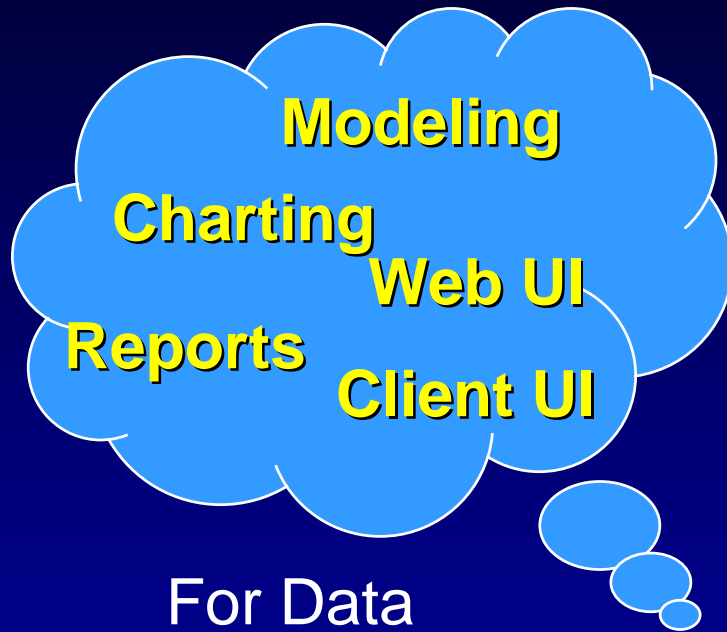
Analytic Data Toolkit



Analytic Application Development
Made Easy



Enterprise Developers and their Tools



For Data
Presentation
and Editing



For Bulk Data
Selection and
Calculations



Issues with Using a RDBMS for Intensive Analytics

- Slowwwwwwwwwww.
- Burdens a shared resource.
- Inappropriate data model for time-series or continuous-measure data
- Limitations of SQL language – not powerful enough for expressing analytical computations

Instead:

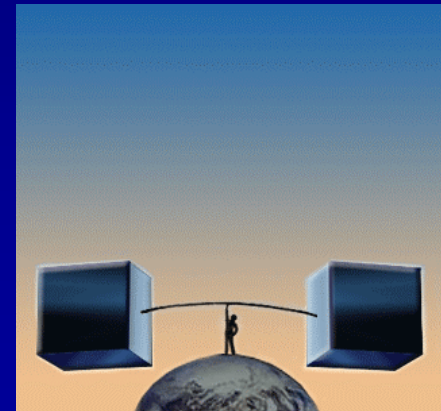
- *Hold data in SmartArrays Analytic Data Toolkit classes to complement the database.*



SmartArrays

Analytic Data Toolkit

- Easy to integrate into IT
- Easy to deploy
- Easy to acquire data
- Rapid development of exact fit to user requirements
- Appropriate for both server and client applications





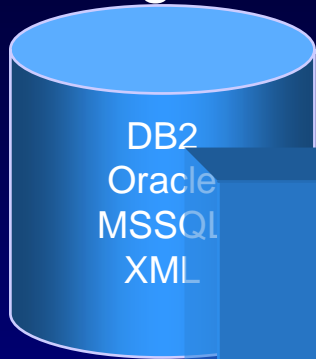
The SmartArrays SDK Provides:

- An **array manager** for allocating and managing memory-resident arrays and memory-mapped files
- An **array engine** with more than 250 highly optimized bulk-data operations
- An object-model **class library** for expressing data handling and analytical computations as intuitive high-level data abstractions.

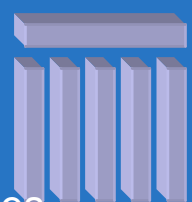


Building Analytical Web Apps

Existing databases



Relational Classes
(tables, columns)

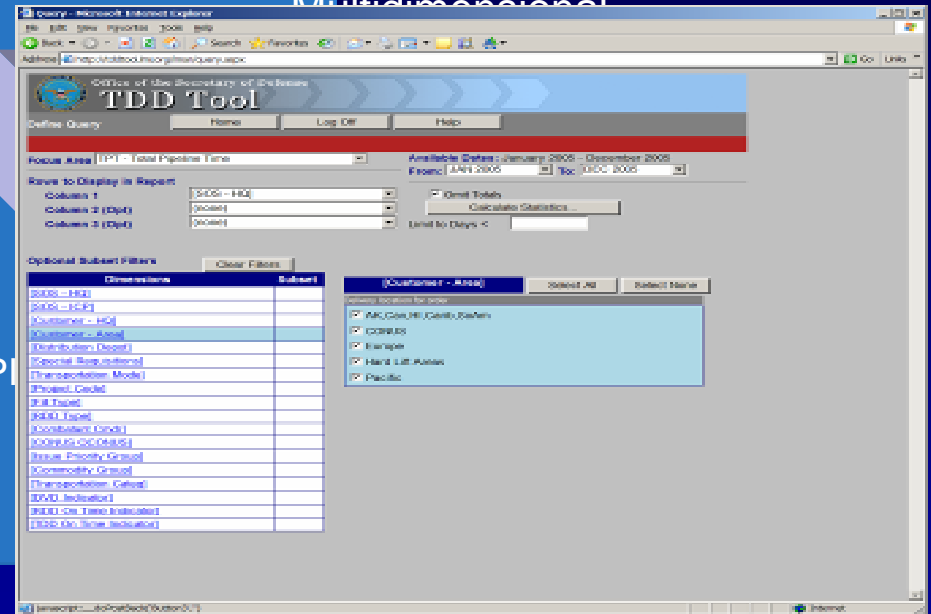


Array API

Data Import Tools

SmartArrays SDK Components

Multidimensional



Rapidly Developed, Ultra-Fast Web Applications

SMARTARRAYS®

Case Study Citibank Global Portfolio Management





GPM Analytics “Web Console”

Web console

Source Data | Import | **Run Company** | Peer Analysis | Templates | Formulas

Company Search and Selection (5688 items found)

3COM CORPORATION
 3COM CORPORATION
 3I GROUP PLC
 3M COMPANY
 7-ELEVEN INC
 724 SOLUTIONS INC
 A O SMITH CORPORATION

Company Scenario Selection

MAIN

Time Periods and Data Source Selection:

<input checked="" type="checkbox"/> FY05	EXTEL
<input checked="" type="checkbox"/> FY04	COMPUSTAT
<input checked="" type="checkbox"/> FY03	AU
<input checked="" type="checkbox"/> FY02	COMPUSTAT
<input checked="" type="checkbox"/> FY01	COMPUSTAT
<input type="checkbox"/> FY00	AU

Annual Quarterly (Single Quarter)

3COM CORPORATION

Chart of Accounts: C&I GRANS
 Industry: Technology, Media & Comms
 Sub Industry: Hardware
 Primary Data Source: EXTEL
 Country: UNITED STATES
 Currency: US dollar

Available Data Reports:

- Editable GRANS Workbook
- GRANS Workbook
- Raw Extel Data
- Editable GRANS w/Formulas
- Quarterly GRANS

Request Feedback and Status:

Run Report | Reload Company



Functionality

- Web user interface in “CitiVision” portal.
- SmartArrays-powered web service for the “heavy lifting”.
- Uses data from disparate data feeds
- Normalizes into standardize accounts
- Calculates analytics and peer-group reports
- Output as Excel files



Citigroup Analytics Software Architecture



XML
Queries to
External
Data Feed



Database holds
financial data, business
rules, report templates

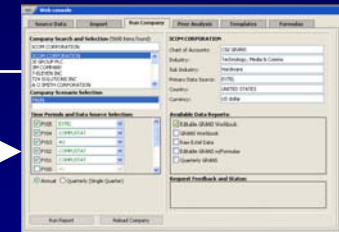
SmartArrays Powered Web Service

SmartArrays
Data Tables
-Formulas
- Financials

SmartArrays
Formula
Engine

ExcelWriter
Reporting
Engine

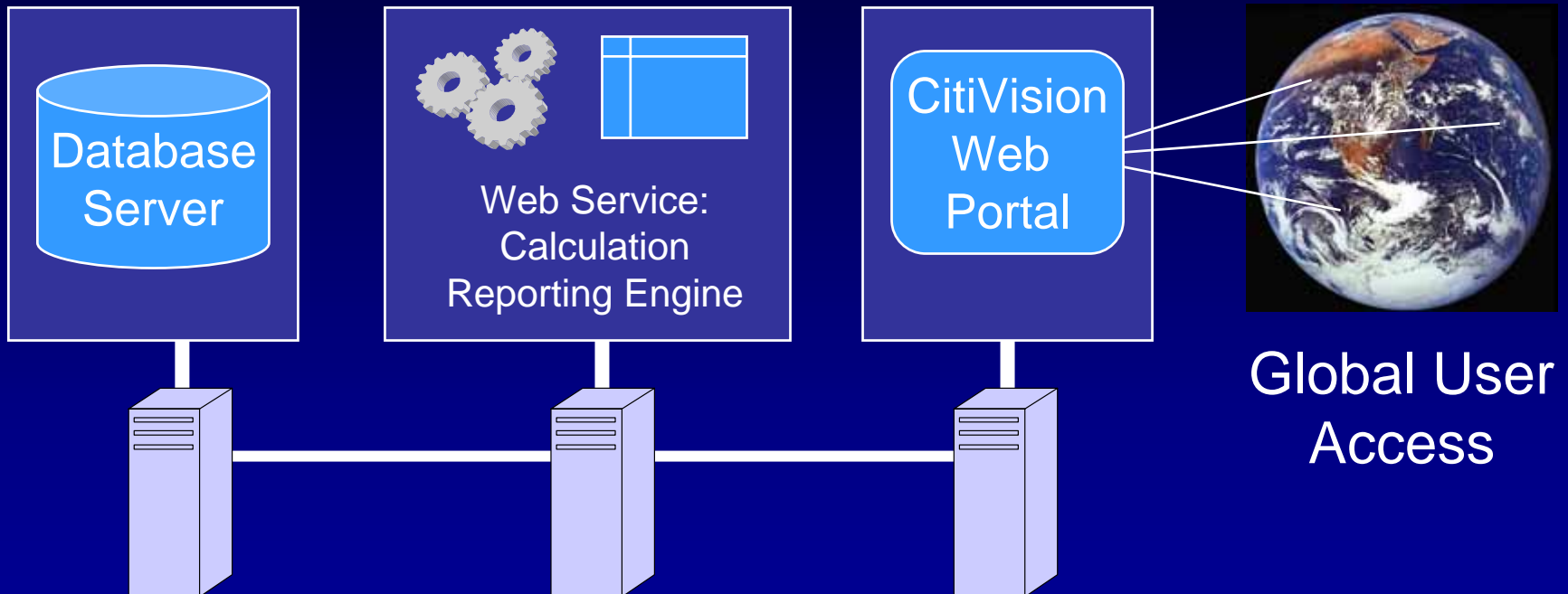
Web User
Interface in
CitiVision Portal



Users request
calculated reports



Citigroup Analytics Hardware Architecture



Windows Servers in New York

SMARTARRAYS®

Case Study – Property & Casualty Insurance





Case Study – Insurance

- Leading property & casualty insurance provider
- 35 years of claims history, 120 million records
- Existing data in VSAM
- Existing Loss Development application runs on mainframe.



Case Study – Insurance

SmartArrays-Powered Replacement

- Runs in WebSphere on inexpensive Windows server.
- Web user interface (JSP).
- Users can produce loss-development triangles from any combination of data by 7 attributes.
- SmartArrays holds the full database.
- Data is mapped to memory on demand.
- Response time: **Less than 1 second.**
- Customer was astonished!

SMARTARRAYS®

Case Study –
Office of the Secretary of Defense
Supply Chain Integration





Case Study – Office of the Secretary of Defense

Problem:

- Large database of processing history on 70,000,000 DoD requisitions
- Each record a time series with more than 200 columns and more than 3K of data
- Goal: Set performance criteria based on measuring past results
- Track and measure improvements in how fast orders are processed.



Case Study – Office of the Secretary of Defense

Technical Challenges:

- Need to focus in on data selected on a variety of criteria (“dimensions”)
- Logistics Management Institute had tried database analytic services and OLAP tools
- None could handle the volume of data or produce the statistical analysis needed.



Case Study – Office of the Secretary of Defense

Challenges:

- Approx 30 different “dimensions” (attributes) needed to select subsets of data.
- Percentiles must be calculated – so pre-aggregation used in many OLAP products won’t work



Case Study – Office of the Secretary of Defense

SmartArrays Based Solution:

- All the data online
- Handles all the dimensions easily
- Recalculates in a few seconds
- User-friendly web site built with ASP.NET



On-Demand Analysis

User chooses:

- Type of interval data
- Range of dates
- Dimensions to report
- Optional filters to drill-in on specific dimensions



On-Demand Analysis

Processing the request

- For each “slice” of data –
 - Apply the user’s filter selections and date range
 - Extract a subset of data values (days to process) from 70 million orders
 - Calculate the count, total dollar value, mean, standard deviation, percentiles, etc on that subset
 - Result becomes one row of the report



LRTT Tool Report

Any combination
of dimensions

Statistics calculated
on each “slice” of data
(Days to process orders)

			Days									
[Customer HQ]	[Item Type]	Count	Mean	Std Dev	Min	25%	50%	75%	85%	90%	95%	Max
Air Force	Consumable	1,742,471	13	60	0	0	1	7	11	19	51	2,764
	Depot Level Repairable	502,393	6	37	0	0	0	2	5	8	20	2,638
	Field Level Repairable	168,004	10	54	0	0	0	4	7	11	32	2,772
Marine Corps	Consumable	184,302	36	119	0	7	13	30	56	81	130	3,781
	Depot Level Repairable	890	75	160	3	10	15	50	110	193	371	1,010
	Field Level Repairable	5,240	42	91	1	9	14	38	69	100	167	3,477




Live Demo

LRT Tool Login - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Recycle Bin Mail Print Internet Options

Address <http://lrttool.lmi.org/lrt/login.aspx> Go Links »



OFFICE OF THE SECRETARY OF DEFENSE

LRT Tool

Logistics Response Time

Select Date Range

LRT Data for most recent 12 months

LRT Data for Oct 2001-Sept 2002

LRT Data for Oct 2002-Sept 2003

LRT Data for Oct 2003-Sept 2004

LRT Data for Oct 2004-Sept 2005

About LRT Tool



This tool measures requisition delivery performance by computing key statistics on the total pipeline time (TPT) and various other nodes that make up the TPT, selected by user-defined criteria and broken out by user-selected characteristics.

All nodes statistics are computed only for those requisitions that have passed through the whole logistics pipeline, and a valid Total Pipeline Time has been computed by DAASC. LRT statistics for a given month may change with each monthly update, because not all customer receipts are reported to DAASC in the month they occur. To ensure that delivery times are reported in the proper month, the LRT Tool derives the requisition's closing Year and Month from the Customer Receipt Date, not from the date that DAASC received the corresponding transaction.

PLEASE NOTE: As of October 2004, LMARS non-military customer records are included in the tool statistics.

Click the Help button for instructions.

OSD LRT Tool Version 3 - Last Modified February 2006
Developed and Hosted by LMI Government Consulting
SmartArrays version 3.0.6.0 Oct 21 2004 10:25:40
Using Data Cubes: 2006_2_27

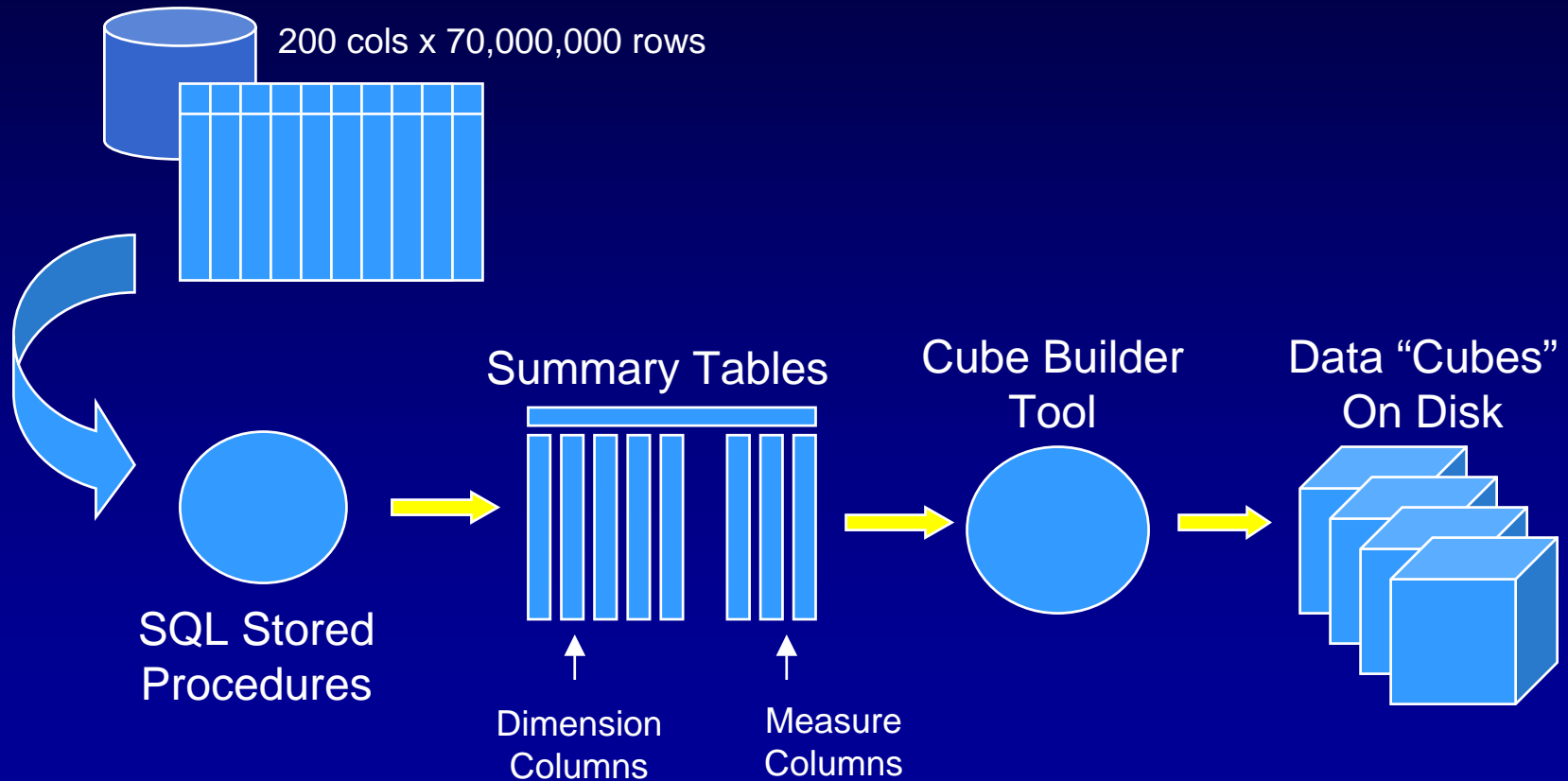


POWERED BY SMARTARRAYS

Done Internet

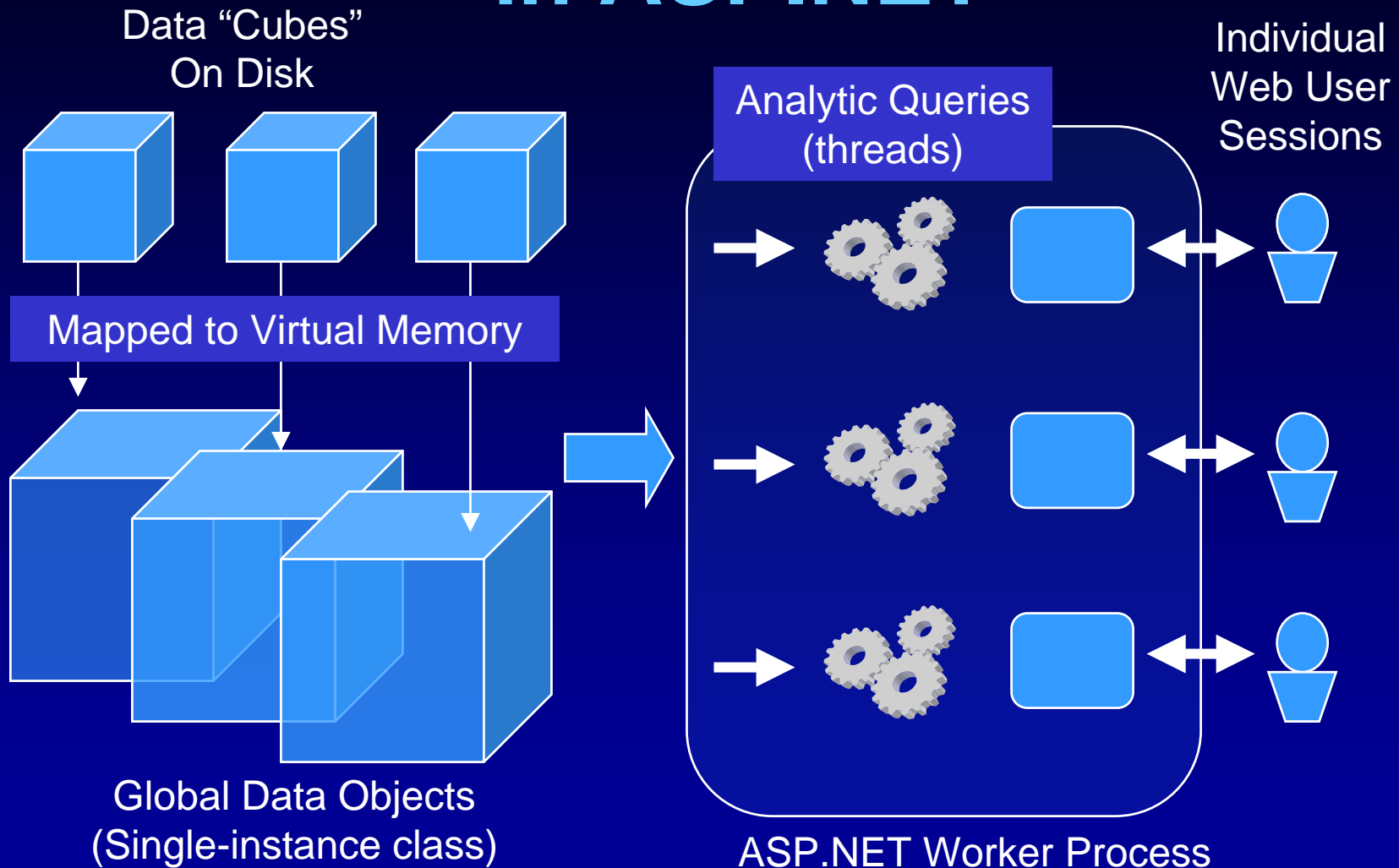
LRTTool Data Loading

DAASC Order Detail Database
(updated monthly)



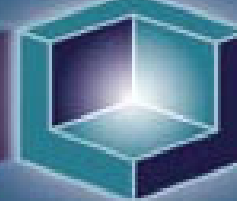
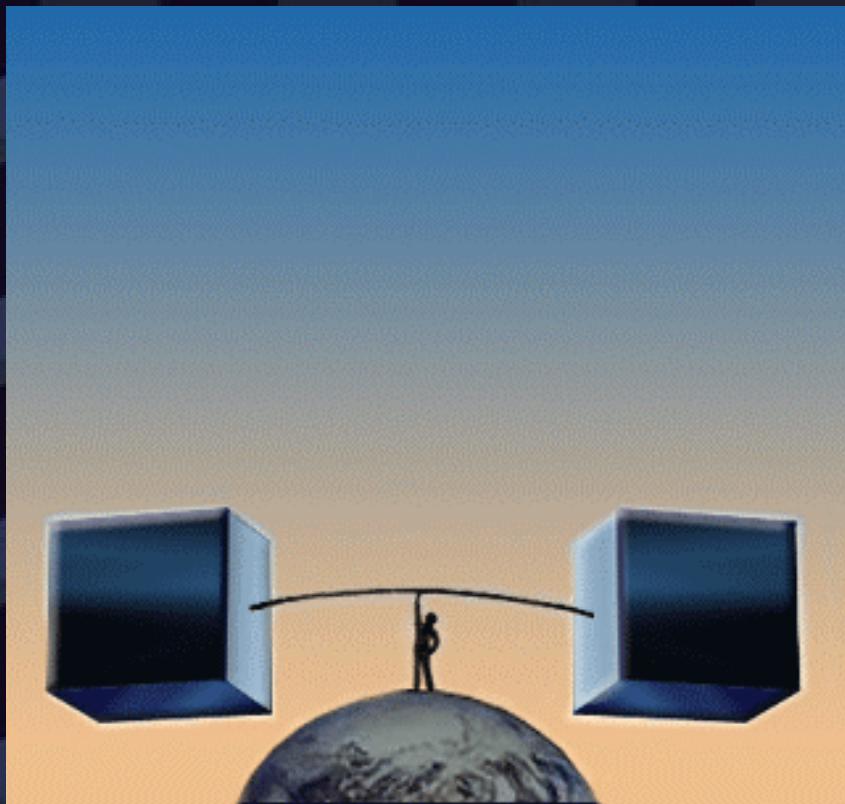


LRTTool Online Analytics in ASP.NET



SMARTARRAYS®

Analytics Data Toolkit





SmartArrays

Analytic Data Toolkit

Benefits

- Leverages experience and best-practices of SmartArrays developers.
- Makes it easy for you to learn SmartArrays.
- Faster execution, better use of memory.
- Simplifies importing external data.
- Provides an initial data “backbone” for application architecture.



SmartArrays

Analytic Data Toolkit

Features

- Dynamic table and column objects that mirror databases but are based on arrays
- Column-oriented data -- the right structure for analytics
- Selection filter classes
- Relation classes for connecting tables

Provided as high-performance objects controlled by your own code.



SmartArrays

Analytic Data Toolkit

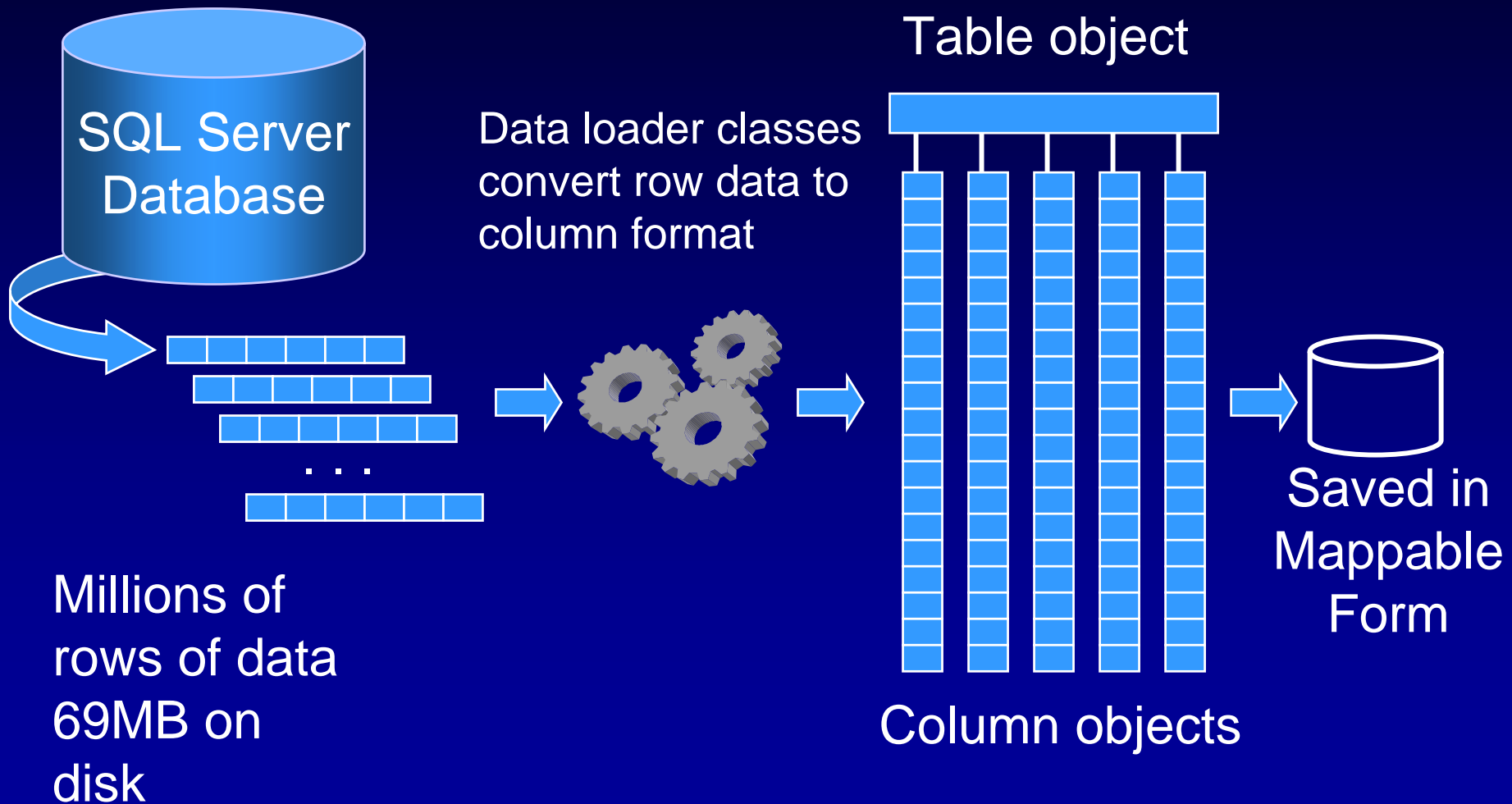
More Features

- External database adapter
- Bulk data import tools
- Date and time handling
- Statistical regression, correlation

Plus, hundreds of array-processing functions for performing any data manipulation or calculation.



Data Import Example





SmartArrays

Analytic Data Toolkit

Demonstrate Data import

SmData Import Wizard

Connect Tables Config Import View

Connection Type:

- ODBC
- MS SQL Server
- Oracle

Connection String:

server=(local);database=auto_sales_demo;Integrated Security=SSPI

Test Connection

Cancel < Prev Next >



SmartArrays

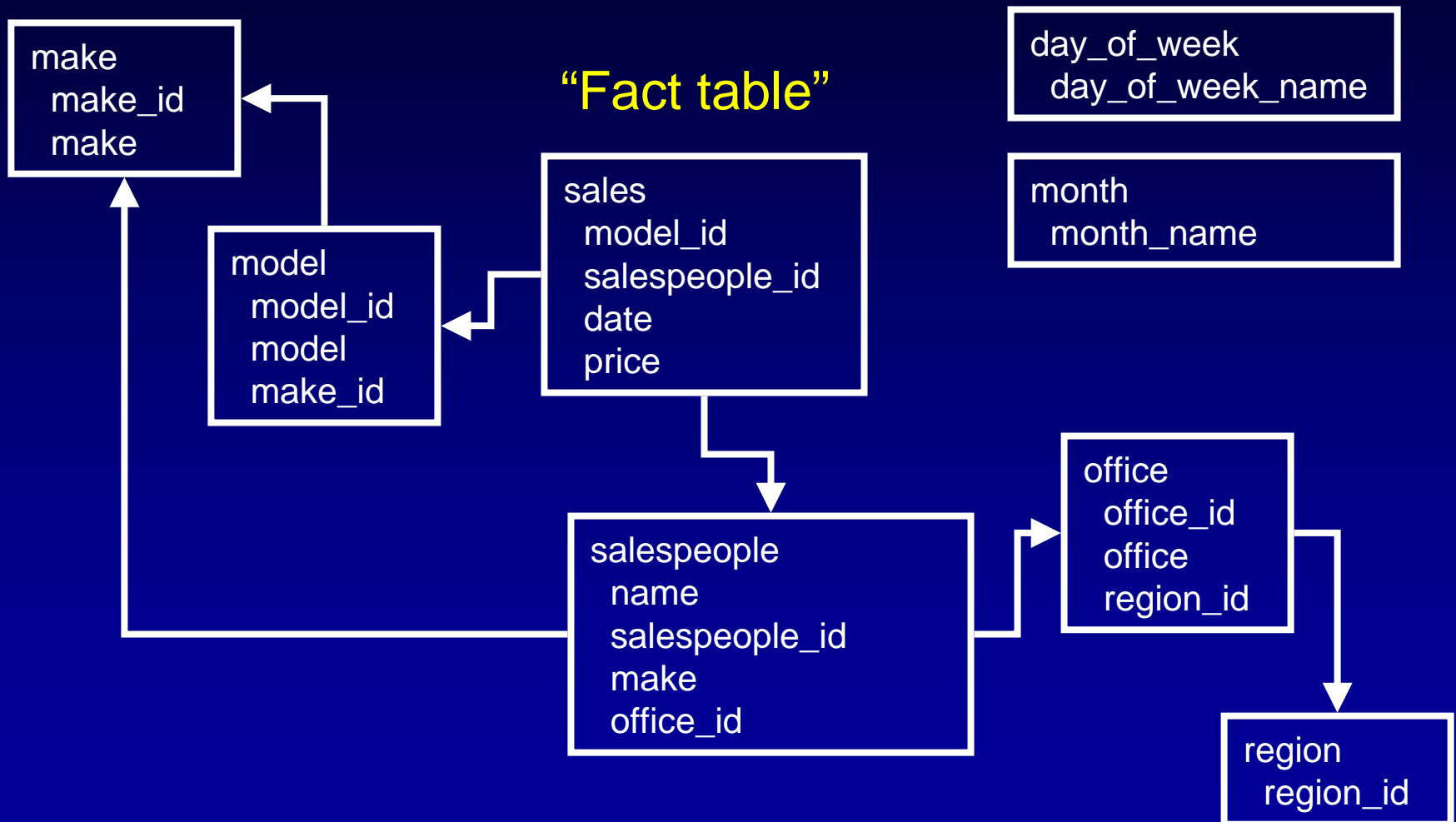
Analytic Data Toolkit

Sample Web Application Illustrating:


- Memory-mapping: From disk to memory in an instant
- Filtering
- Cross Tabulation
- Statistics
- Time slicing with dynamic buckets



Sample Data Set



Demo Analytical Web App



SMARTARRAYS[®]
ARRAY TECHNOLOGY FOR BUSINESS ANALYTICS

Map Data 8 tables mapped in 00:00:00.1402016 seconds

Transform Data Denormalized Table created. Time = 00:00:02.4435136 seconds

Analytical Queries

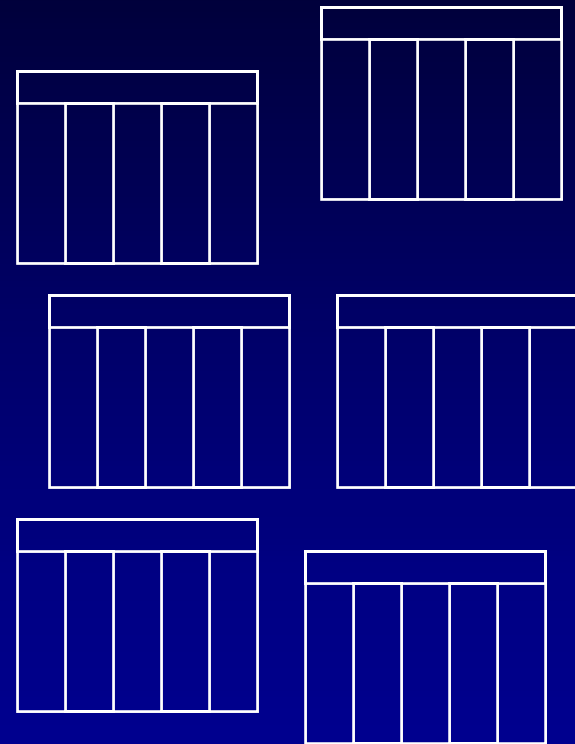
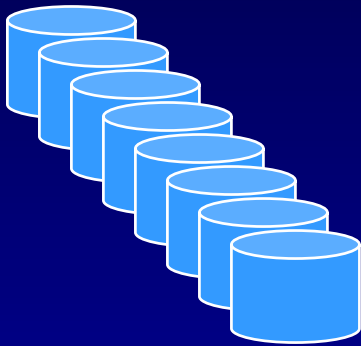
price	salesperson	region	office	make	model	date
30806	Henry	Mid-Atlantic	Washington DC	Hudson	Hornet	6/28/2004 12:00:00 AM
34109	Jose	Mid-Atlantic	Dover	LaSalle	Cabriolet	3/27/2005 12:00:00 AM
43729	Henry	Mid-Atlantic	Washington DC	Packard	Touring 810	5/8/2004 12:00:00 AM
29317	Tom	New England	Providence	LaSalle	Roadster	8/2/2004 12:00:00 AM
27067	Frank	New England	Boston	LaSalle	Cabriolet	1/31/2004 12:00:00 AM
26271	Bob	New England	Boston	LaSalle	L16	10/21/2004 12:00:00 AM
36996	Henry	Mid-Atlantic	Washington DC	Packard	Victoria	2/5/2005 12:00:00 AM
33452	Bob	New England	Boston	Packard	LeBaron	3/21/2004 12:00:00 AM
43022	George	Mid-Atlantic	Philadelphia	LaSalle	L16	3/9/2004 12:00:00 AM
29203	Mark	New York	Buffalo	LaSalle	Cabriolet	3/3/2005 12:00:00 AM
29854	Henry	Mid-Atlantic	Washington DC	LaSalle	Cabriolet	1/4/2004 12:00:00 AM
27401	Muhammed	Mid-Atlantic	Philadelphia	LaSalle	Town Sedan	5/12/2004 12:00:00 AM
25299	Bob	New England	Boston	Hudson	H112	6/8/2005 12:00:00 AM
33648	Francis	New England	Boston	LaSalle	Roadster	8/4/2005 12:00:00 AM
35956	George	Mid-Atlantic	Philadelphia	LaSalle	L12	2/5/2005 12:00:00 AM
32042	Francis	New England	Boston	Packard	LeBaron	1/19/2005 12:00:00 AM
26432	George	Mid-Atlantic	Philadelphia	Packard	Phaeton	10/27/2004 12:00:00 AM
37471	Matthew	Mid-Atlantic	Philadelphia	LaSalle	Cabriolet	10/21/2005 12:00:00 AM
25221	Ed	New York	Buffalo	Packard	Phaeton	11/11/2004 12:00:00 AM
29132	John	New England	Hartford	Packard	LeBaron	6/7/2004 12:00:00 AM
33307	George	Mid-Atlantic	Philadelphia	Packard	LeBaron	6/20/2005 12:00:00 AM

Done Local intranet



Memory-Mapping Data

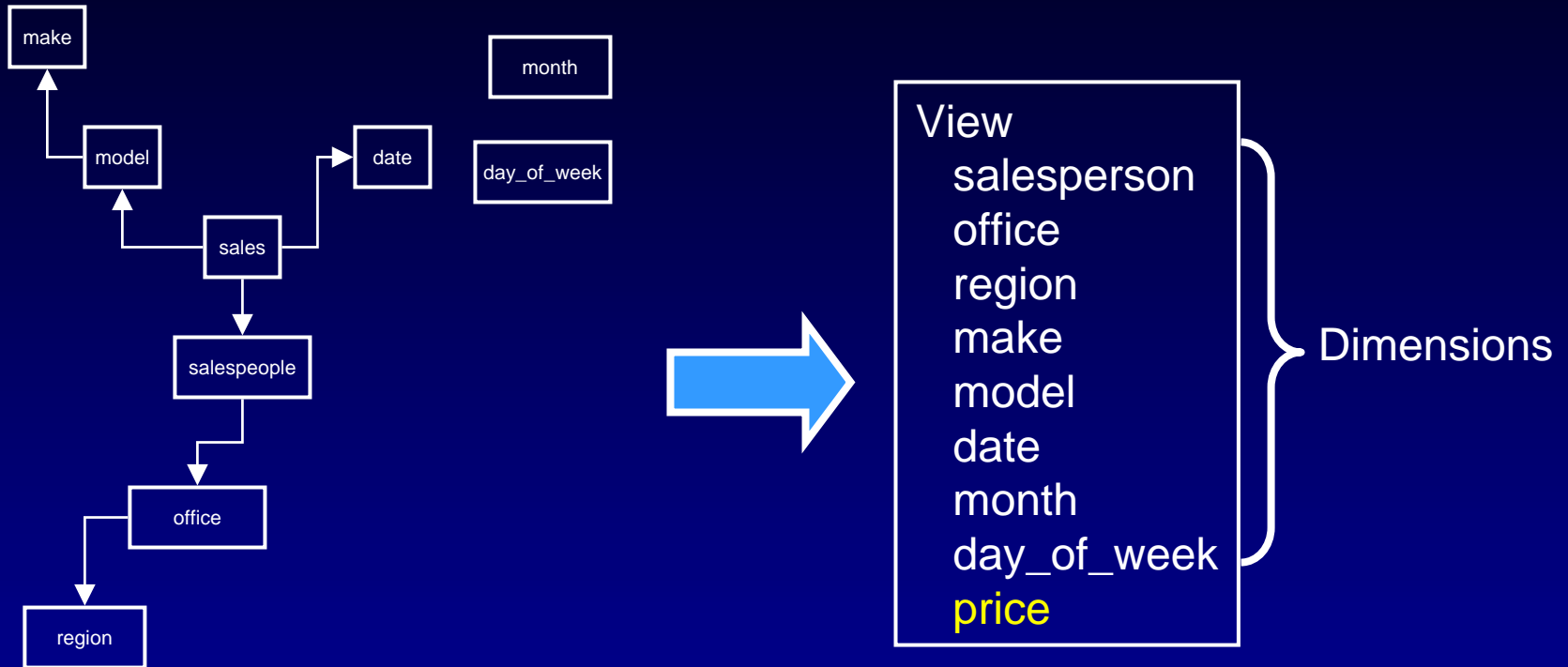
Tables saved as
XML and data files



Tables are instantiated in memory using
virtual-memory file-mapping operations



Transformed Data Set



Denormalized data table is simpler and faster for analysis in many situations.

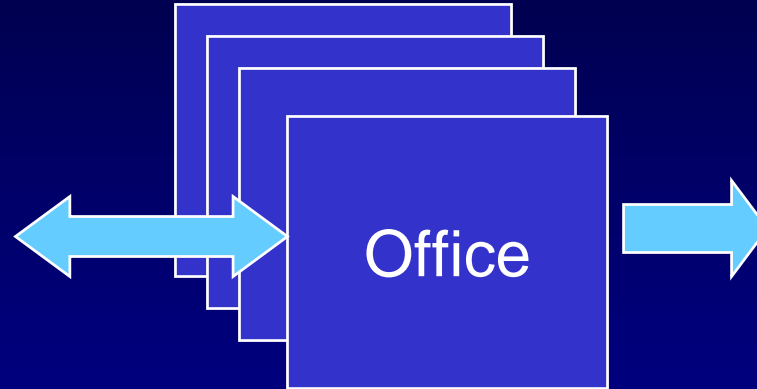


Filtering Dimension Columns

Filter Objects –
One for each dimension

Select Values

- Albany
- Boston
- Buffalo
- Dover
- Harrisburg
- Hartford
- New York
- Philadelphia
- Providence
- Washington DC



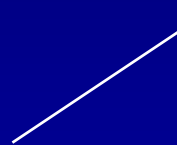
1
1
0
0
1
1
0
0
1
1

Bit vector indicating which
Rows of table fit filter selection.
Used to select data subset



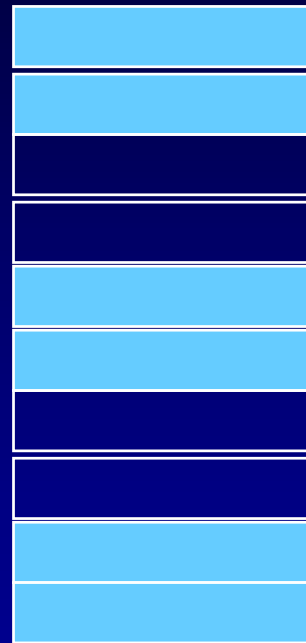
Calculating Statistics

Filters are combined with “AND” to get a composite bit vector.

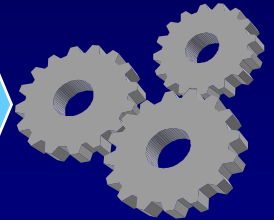
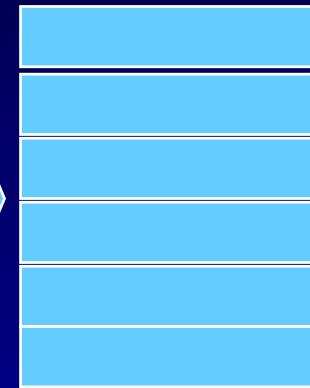


Bit vector - one bit for each row in data table

Table



Subset Table



Statistical calculations

SMARTARRAYS®

High Performance
Analytical
Techniques





Example:

Compact Data Representation

Original Table
In Database



SummaryLast12_CIT

```
[YRMO] [int],
[YR] [smallint],
[MO] [tinyint],
[SOS - ICP] [nvarchar] (50),
[SOS - HQ] [nvarchar] (50),
[Project Code] [nvarchar] (3),
[Customer - HQ] [nvarchar] (50),
[Fill Type] [nvarchar] (50),
[RDD Type] [nvarchar] (6),
[Class of Supply] [nvarchar] (15),
[CONUS OCONUS] [nvarchar] (1),
[Customer - Area] [nvarchar] (20),
[Issue Priority Group] [float],
[Commodity Group] [nvarchar] (255),
[Distribution Depot] [nvarchar] (50),
[Transportation Mode] [nvarchar] (255),
```

16 dimension
columns

```
[LRT] [int],
[Reqs] [int]
```

2 "fact" columns



Example: Compact Data Representation

Original Table
In Database



Imported into SmartArrays
Table Object and Saved to Disk



- 15x denser information.
- Much faster to access.
- Fits in memory



Shrinking Data to Its Essence

“Customer” Column in the database:

- 4 million rows of varchar() data
- 100 MB on disk

“Customer” Column in SmartArrays

- 2 MB *in memory*



SmBitmapColumn

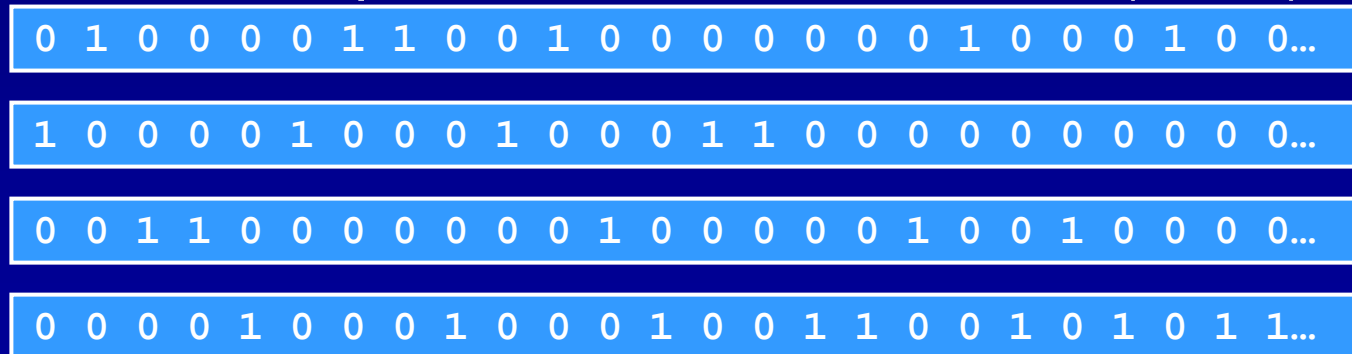
Customer column has 4 million rows but only 4 distinct values: “Air Force”, “Army”, “Navy”, “Marines”

Represent with Arrays:

- Vector of 4 strings: “Air Force”, “Army”, “Navy”, “Marines”

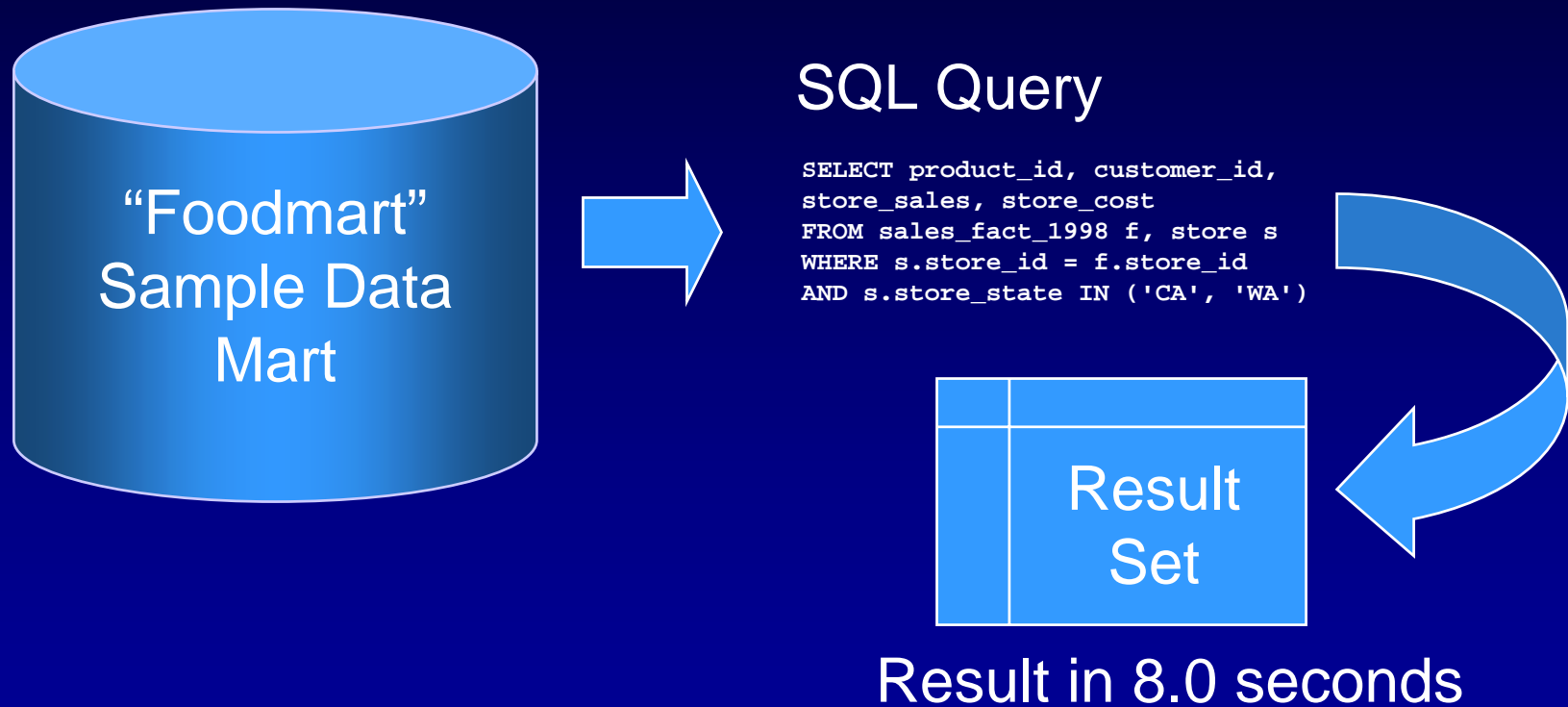


- Vector of 4 bitmaps, each with 4 million bits (500K)



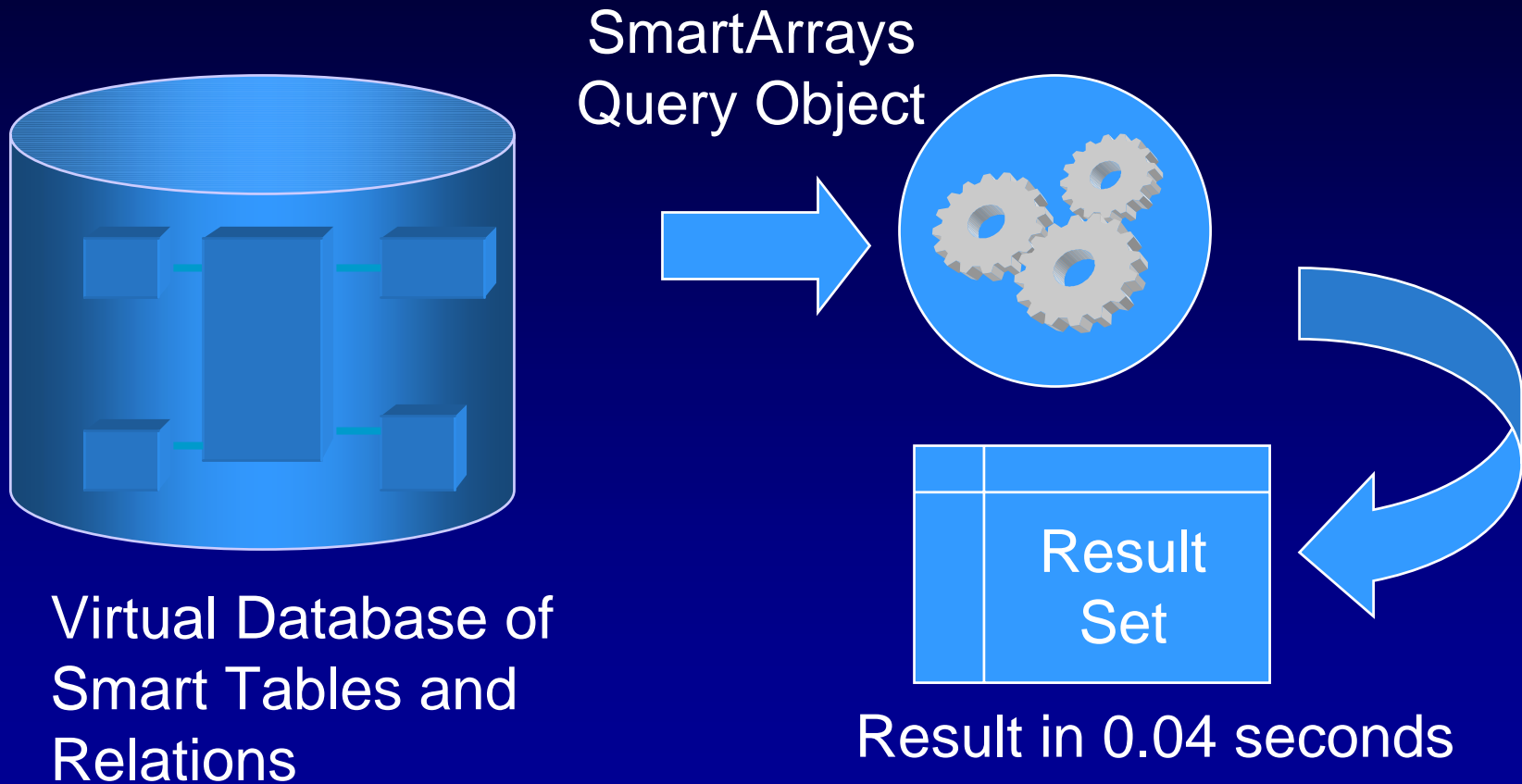


Example: Query with Database





Example: Query with SmartArrays





What Makes It So Fast?

Database Server

8.0 sec

Smart Tables

0.04 sec

200 times faster

- Data is already in memory.
- Data is in local process – no network traffic
- Data is column-sliced – structured for analysis
- Optimized array algorithms



Memory-Mapped Table Objects

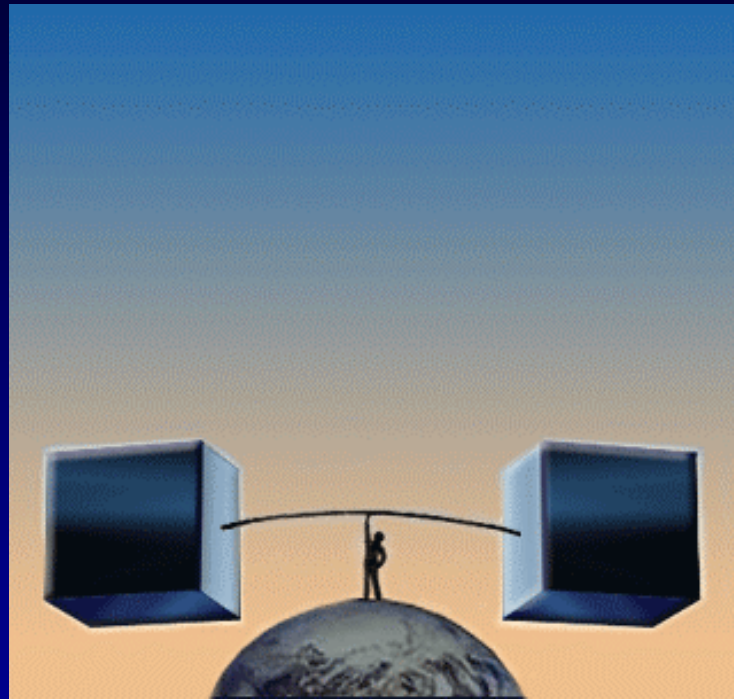
Maps a 10-million-row table to memory in
0.1 second:

```
SmTable t = new SmTable();  
t.fromPath( directory );
```

Data paged into memory when actually
used.



Question and Answer



Thank You

For more information or to be included in future
Web Conferences:

<http://www.smartarrays.com>

Email: sales@smartarrays.com

Phone: +1 703 326 0066