Analyzing Multiple Data Sources with Multisource Data Federation and In-Memory Data Blending

Presented by: Trishla Maru
Agenda

• Overview

• MultiSource Data Federation
  • Use Cases
  • Design Considerations

• Data Blending
  • What is Data Blending
  • When to use Data Blending?
  • Benefits of Data Blending

• Demo

• MultiSource Option Vs. Data Blending

• Performance Gain with New Dashboarding Engine

• Q&A
MicroStrategy Analyzes Data from Corporate System of Record and Personal Data Files

1. **Modeled Data**
   - Modeled Data is performed through MicroStrategy Architect
   - Designer architects global model which abstracts physical structure of data (tables and columns, aggregates etc.)

2. **Self Service Data**
   - This self-service data is achieved via the “Import Data” feature
   - User adds the data, and during the import process, supplies basic mapping of Attributes, Metric

![Diagram](image)
High Level Overview
Multisource Option and Data Blending

Blends data from any modeled and unmodeled sources

Combines data from modeled data sources

Report Writing Documents         Visual Insight Dashboard
Data Blending

MicroStrategy Intelligent Server
In-memory Cubes

Multisource Option
Multidimensional SQL Engine       MDX Generation Engine
Freeform SQL Engine               Data Import Engine

Relational Databases             MDX Cube Sources
Operational Databases             Excel, CSV, Text Files and Databases

#mstrworld
Agenda

• Overview

• MultiSource Data Federation
  • Use Cases
  • Design Considerations

• Data Blending
  • What is Data Blending
  • When to use Data Blending?
  • Benefits of Data Blending

• Demo

• MultiSource Option Vs. Data Blending

• Performance Gain with New Dashboarding Engine

• Q&A
What is MultiSource?
Provides a Single Multi-dimensional View Across Multiple Data Sources

A Single Relational Source

Relational and Multi-Dimensional Sources

- Capability to combine data from more than one relational or multidimensional sources.
- Merging data at the architect/report level, when business users only need to consume the data
- A multisource report can be added as a dataset to documents/dashboards.

#mstrworld
MultiSource Option

Use Case 1: Hub & Spoke Architecture with Conformed Dimensions

Single Unified Multi-dimensional Business Model

- Marketing-specific extensions
- Mfg - specific extensions
- Finance - specific extensions
- Sales - specific extensions

MySQL
DB2
Netezza
SQL Server
Sybase IQ

Marketing Data Mart
Mfg Data Mart
Enterprise Data Warehouse
Financial Data Mart
Sales Data Mart
MultiSource Option

Use Case 2: Balance Workload Across Databases

**MicroStrategy Analytics Platform**
- Multisource ROLAP

- **Multi-dimensional Business Model**
  - **Multi-source Engine**
  - **ROLAP SQL Engine**

- **High-level Query**
  - Result
  - **Fast Database for Aggregates**
    - RDBMS1

- **Detail Query**
  - Result
  - **Scalable Database for Details**
    - RDBMS2

#mstrworld
MultiSource Option

Use Case 3: Gradual Evolution from Islands of BI to Enterprise BI

Stage 1
Disparate Islands of BI
All Running on MicroStrategy BI

Stage 2
Merging Islands of BI
Using MicroStrategy Multisource

Stage 3
Consolidating Data
Re-pointing Metadata to the EDW
Agenda

• Overview

• MultiSource Data Federation
  • Use Cases
  • Design Considerations

• Data Blending
  • What is Data Blending
  • When to use Data Blending?
  • Benefits of Data Blending

• Demo

• MultiSource Option Vs. Data Blending

• Performance Gain with New Dashboarding Engine

• Q&A
MultiSource Option is Designed for High Performance

1. Push Down ROLAP Joining Minimizing Data Movement

Multiple Relational and Multi-Dimensional Sources

Multi-dimensional Business Model

Multi-source Engine

ROLAP SQL Engine

MDX Engine

Multi-pass SQL

Table "A"
Table "B"
RDMBS1

Table "C"
RDBMS2

Push-down Join

Insert Partial Result into Temp Table

Partial Result
(Temporarily stored in Intelligence Server memory)

Result
MultiSource Option is Designed for High Performance

2. Optimized Source Selection Across Databases

Multiple Relational and Multi-Dimensional Sources

Multi-dimensional Business Model

Multi-source Engine

ROLAP SQL Engine

High-level Query

Result

Fast Database for Aggregates

RDBMS1

Detail Query

Result

Scalable Database for Details

RDBMS2

Multiple Relational and Multi-Dimensional Sources

Multi-dimensional Business Model

Multi-source Engine

ROLAP SQL Engine

High-level Query

Result

Fast Database for Aggregates

RDBMS1

Detail Query

Result

Scalable Database for Details

RDBMS2
MultiSource Option is Designed for High Performance

3. Optimized SQL and/or MDX for Each Data Source

Multiple Relational and Multi-Dimensional Sources

Tuneable VLDB Settings
Allows DBA’s to Tailor the automatic SQL Generation for each Data source Technology to Deliver Maximum Performance

Optimized SQL or MDX
MicroStrategy ROLAP SQL Engine Creates Data source-specific SQL or MDX

#mstrworld
Agenda

- Overview

- MultiSource Data Federation
  - Use Cases
  - Design Considerations

- Data Blending
  - What is Data Blending
  - When to use Data Blending?
  - Benefits of Data Blending

- Demo

- MultiSource Option Vs. Data Blending

- Performance Gain with New Dashboarding Engine

- Q&A
What is Data Blending?

Blend data from modeled and unmodeled sources on the fly in a dashboard.

Personal Data Sources
- Excel
- Hadoop
- SAP BW

Relational Databases
- IBM Informix
- Microsoft SQL Server
- Oracle
- Sybase

Multi-Dimensional Sources
- SAP Analysis Services
- TM1

Map Reduce Databases
- Hadoop
- Cloudera
- Aster Data
- Impala

#mstrworld
Data Blending

Link Data from Different Sources with Right-Mouse-Click and Drag-and-Drop Actions

“School Name” Column from ‘Attendance’ data set is…

… mapped to “Name” column from ‘Ranking’ dataset
Data Blending Among Two Modeled Datasets

1. Modeled Dataset 1: Revenue
   - Common ‘Schema Attribute’ Country

2. Modeled Dataset 2: Forecast
   - The new Dashboarding Engine automatically links ‘common’ attributes using the modeled schema
   - No ‘Manual Linking’ is allowed between different modeled attributes
   - If users need to manually link different attributes, this can be accomplished using Architect

Join Behavior
- Relationship Found = Full Outer Join
- No Relationship Found = Cross Join
Data Blending Among Modeled and Unmodeled Data

More Than One Dataset in one VI Analysis

Auto-Linking
- RWD Engine tries to identify a link between columns using ‘Header Name’ and ‘Data Type’
- Users have the flexibility to Unlink the auto-link

OR

Manual-Linking
- Users can manually link imported/unmodeled attributes to other imported/modeled attributes

Data Blended from Two Datasets

Join Behavior
- Relationship Found = Full Outer Join
- No Relationship Found = Cross Join

#mstrworld
When to use Data Blending?
On the fly Blend data in a Dashboard or Document

Blend data from any number of sources with simple drag and drop actions!

Add any number of datasets to your analysis!
When to use Data Blending?
Blend Modeled data with Unmodeled data in a Single Visualization

Instantly merge data from modeled and unmodeled data sources. Blend and analyze without any IT dependency.

Performance Score comes from an Excel file

City Data comes from EDW

EDW

#mstrworld
When to use Data Blending?
Blend data from Multiple In-Memory Cubes in a Dashboard

• Blend data from multiple In-memory cubes from one or more source.

• **In-Memory Cubes as datasets speeds up the dashboard execution by 60%,** tested in real customer cases.

• Dynamic Selections afterwards (filtering, grouping etc.) speed up by 70%.

Multiple In-memory cubes acting as datasets to your analysis!
Benefits of Data Blending
High Scalability

- **Add and Blend Data** from as many number of cubes and sources in your VI analysis

- **Analyze over two billion rows limitation** by dividing a single cube into multiple smaller cubes and blend them with data blending

- **Multiple cubes and Multiple sources** in one visualization!

Blend data from as many number of cubes with negligible overhead!
Benefits of Data Blending
Faster Cube Load and Refresh

- Quicker load and refresh of the data with multiple Intelligent cubes.
- Improve the dashboard execution times by 60%

**MicroStrategy 9.3.1**
One big cube feeding the dashboard

**MicroStrategy Analytics Platform**
Multiple cubes feeding the dashboard

- Faster Auto Refresh than for one big cube
Agenda

• Overview

• MultiSource Data Federation
  • Use Cases
  • Design Considerations

• Data Blending
  • What is Data Blending
  • When to use Data Blending?
  • Benefits of Data Blending

• Demo

• MultiSource Option Vs. Data Blending

• Performance Gain with New Dashboarding Engine

• Q&A
Demo
Blend Data Instantaneously!

Analyze FIFA World Cup data from various files and blending the data together instantly!

- FIFA World Cup Roster Data
- Player Information
- FIFA World Cup Winners Data
Agenda

• Overview

• MultiSource Data Federation
  • Use Cases
  • Design Considerations

• Data Blending
  • What is Data Blending
  • When to use Data Blending?
  • Benefits of Data Blending

• Demo

  • MultiSource Option Vs. Data Blending

• Performance Gain with New Dashboarding Engine

• Q&A
# Multisource Option Vs. Data Blending

<table>
<thead>
<tr>
<th>Multisource Option</th>
<th>Data Blending</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeled approach to join data</td>
<td>On the fly blending data</td>
</tr>
<tr>
<td>Join data from various <strong>modeled sources</strong> at the schema/report level</td>
<td>Join data from various <strong>modeled and unmodeled sources</strong> at the dashboard/document level</td>
</tr>
<tr>
<td>Join is pushed to the underlying database</td>
<td>Join occurs in MicroStrategy’s Intelligent Server</td>
</tr>
<tr>
<td>Developed for <strong>Architects</strong></td>
<td>Developed for <strong>Business Users</strong></td>
</tr>
<tr>
<td>Allows the user to have more than one modeled schemas in a single project</td>
<td>Removes the limitation of one cube per VI analysis</td>
</tr>
</tbody>
</table>
Agenda

• Overview

• MultiSource Data Federation
  • Use Cases
  • Design Considerations

• Data Blending
  • What is Data Blending
  • When to use Data Blending?
  • Benefits of Data Blending

• Demo

• MultiSource Option Vs. Data Blending

• Performance Gain with New Dashboarding Engine

• Q&A
Performance Gain with New Dashboarding Engine
40% Higher Throughput for Dashboards

- **Kilocycle improvements** of at least 40% for dashboards
- **Dashboards faster by at least 70%** with no modeling or design changes.
- **Simple Upgrade to MicroStrategy Analytics Platform** and leverage these performance improvements

*Kilocycle is 1000 user request per hour. A commonly used unit for job throughput is requests per minute (rpm)*
Performance Gain with New Dashboarding Engine
Dashboards Faster by more than 50%

- Cubes as datasets speeds up the dashboard execution by 60%, tested in real customer cases.
- Dynamic Selections afterwards (filtering, grouping etc.) speed up 70%.

![Bar Chart]

- Normal Reports as Datasets in 9.3.1
  - Normal Reports as Datasets in MicroStrategy Analytics Desktop
  - Cubes as Datasets in MicroStrategy Analytics Desktop
Performance Gain with New Dashboarding Engine

Lower Memory Footprint

- New storage format reduces the data size of the dashboards by 50%
- 40% faster selections observed on customer dashboards
- Get it all with an easy upgrade to MicroStrategy Analytics Platform!

**Lower Memory Footprint**

9.3.1 MicroStrategy Analytics Platform

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Metrics</th>
<th>Sub-Totals</th>
<th>Filters</th>
<th>Graphs</th>
<th>4.6 MB</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Selections</th>
<th>Groupings</th>
<th>Filtering</th>
<th>Manipulations</th>
<th>2.2 sec</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Selections</th>
<th>Groupings</th>
<th>Filtering</th>
<th>Manipulations</th>
<th>1.6 sec</th>
</tr>
</thead>
</table>

#mstrworld
Agenda

- Overview

- MultiSource Data Federation
  - Use Cases
  - Design Considerations

- Data Blending
  - What is Data Blending
  - When to use Data Blending?
  - Benefits of Data Blending

- Demo

- MultiSource Option Vs. Data Blending

- Performance Gain with New Dashboarding Engine

- Q&A
Questions?