The State of Data Sharing

John Gay
Manager & Technical Specialist, Sales Engineering
Topics

- Data sharing market trends
- Challenges and pitfalls of ETL processes
- Introduction to MIMIX Share
- Real-world case studies
- Q&A
Data Sharing
Market Trends
Vision Solutions 2017 State of Resilience Report

4 Surveys on HA/DR, migration, data sharing and cloud computing
1,598 IT professionals responded
86% plan, manage or administer IT systems

Titles ranged from CTO, CIO, System Architect, and Head of Computer Systems to Database Administrator, Director of Information Systems, Senior Network Engineer, Manager of Cloud Systems

Respondents represented a wide array of countries from around the globe
The State of Data Sharing
The State of Data Sharing: Prevalence

Does your organization rely on multiple databases?

- Yes: 83%
- No: 10%
- I don't know: 8%

- 83% of respondents rely on multiple databases
- 10% do not
- 8% did not know
The State of Data Sharing: Concerns about Inconsistency

70% of respondents with multiple databases had redundant data stored in them.

Do you have redundant data in your databases?

- Yes: 70%
- No: 30%
Almost half of respondents with redundant data in their databases indicated that decisions have to be delayed in order to reconcile conflicting information (44%).

Over one-third reported that spending a significant amount of time and effort reconciling inconsistent information (36%).

The State of Data Sharing: Business Decision Concerns
The State of Data Sharing: Prevalence

Do you share data between your databases?

- **73%** of respondents share data between databases
- **20%** do not
- **7%** did not know if they do
59% are consolidating data from multiple sources into a single data warehouse
54% perform Business Intelligence on data offloaded from the production database
51% report on data offloaded from the production database
49% integrate data used by applications/databases selected for different purposes
The State of Data Sharing Methods

Many respondents use in-house tools or manual methods for data sharing:
- A majority (54%) of respondents use FTP/SCP/file transfer to share data between databases
- 50% use manual scripting
- 43% use backup/restore processes (snapshots)
- 41% use in-house tools
- 28% use ETL processes

41% of our sample use third party replication software
The State of Data Sharing: Future Benefits

In the future, could your company benefit from using a real-time copy of a production database for any of the following tasks?

- Business Intelligence: 30%
- Reporting: 26%
- Database backup: 26%
- Development work: 22%
- Analytics: 21%
- Testing system upgrades: 18%
- Ad hoc queries: 14%
- We see no benefit: 12%
- I don't know: 11%

30% indicated their company would benefit from using a real-time copy of production data for BI.

26% would benefit from real-time data for reporting and backup.
The State of Data Sharing: Migration Initiatives

When asked about database migration or consolidation projects:

- 34% plan to migrate data to new databases
- 23% plan to migrate database to the cloud
- 17% plan to consolidate to fewer databases
- 17% plan to migrate to new DBMS or OS platforms
Traditional Methods for Obtaining Data for Reports

- Direct network access
  - Reporting on production servers across the network during business hours (32%)
  - **Issue:** Negatively impacts network and database performance – resulting in user complaints! (41%)
Traditional Methods for Obtaining Data for Reports

- Direct network access
  - Reporting on production servers across the network during business hours (32%)
    - *Issue:* Negatively impacts network and database performance – resulting in user complaints! (41%)

- Off-hours reports and extractions
  - Run reports off-hours (16%) or perform nightly ETL processes (32%) to move data to a reporting server
    - *Issue:* Business operates on aging data until next extraction
    - *Issue:* Difficult to find acceptable time to perform an extraction
Traditional Methods for Obtaining Data for Reports

- Direct network access
  - Reporting on production servers across the network during business hours (32%)
    - Issue: Negatively impacts network and database performance – resulting in user complaints! (41%)

- Off-hours reports and extractions
  - Run reports off-hours (16%) or perform nightly ETL processes (32%) to move data to a reporting server
    - Issue: Business operates on aging data until next extraction
    - Issue: Difficult to find acceptable time to perform an extraction

- In-house ETL (Extract-Transform-Load) Processes
  - FTP/SCP/file transfer processes or Manual scripts or Backup/restore or In-house tools
    - Issue: Periodic, not real-time, delivery of data
    - Issue: Labor intensive to create processes and tools
    - Issue: Expensive to develop and maintain
    - Issue: Prone to errors
In-House ETL Scripts and Processes Are Not Free

- **Upfront development costs**
  - Development of code to perform database extraction, transformation, and load
  - Additional requirements for additional pairings, schemas, etc.

- **Test system expenses**
  - Hardware and storage resources
  - Database licenses for test systems
  - Add-on products, e.g. gateways

- **Maintenance costs**
  - Ongoing enhancements for altered schemas, additional platforms
  - Testing new database and OS releases
  - Cross training and documentation to reduce turnover risk

- **Lost opportunity costs for other initiatives**
MIMIX Share
Overview
MIMIX Share for Easy, Automated Data Sharing

- Breaks down barriers between databases
  - Same or different database management systems
  - Same or different operating systems
  - Physical, virtual or cloud platforms
  - Across any distance

- Makes data sharing easy
  - Replicates database changes in real time
  - Transforms and enhances data during replication
  - Supports leading database and operating systems
  - Offers a variety of replication architectures
  - Easy graphical UI – no programming required!

- Quickly returns your investment
  - Stronger decision making
  - Greater business productivity
  - Ability to choose more cost-effective infrastructure
  - Frees IT to focus on other business initiatives
Change Data Capture (CDC) for Real-Time Replication

- Change Data Capture (CDC) captures database changes immediately and quickly replicates them to another database(s)
- Only changed data is replicated to minimize bandwidth usage
- Automatically extracts, transforms and loads data into target database without manual intervention or scripting
- Ensures write order consistency and guaranteed delivery
- Ensures data integrity with conflict resolution and collision monitoring
- Enables tracking and auditing of transactions for compliance
Replaces Manual Processes

- Point & click graphical user interface
- Single view of data across databases and operating systems
- Simple, model-based configuration
- Automatically creates target tables from the source table definition
- 80+ pre-built, click-and-go data transformations
- Transformations can be added through Java-like scripting
- No programming required
Supports a Broad Range of Platforms

<table>
<thead>
<tr>
<th>Leading Operating Systems</th>
<th>Leading Databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>• IBM i</td>
<td>• IBM DB2 for i</td>
</tr>
<tr>
<td>• IBM AIX</td>
<td>• IBM DB2 for LUW</td>
</tr>
<tr>
<td>• HP-UX</td>
<td>• IBM Informix</td>
</tr>
<tr>
<td>• Solaris</td>
<td>• Oracle</td>
</tr>
<tr>
<td>• IBM Linux on Power</td>
<td>• Oracle RAC</td>
</tr>
<tr>
<td>• Linux SUSE Enterprise</td>
<td>• MySQL*</td>
</tr>
<tr>
<td>• Linux Red Hat Enterprise</td>
<td>• Microsoft SQL Server</td>
</tr>
<tr>
<td>• Microsoft Windows, including Microsoft Azure</td>
<td>• Teradata*</td>
</tr>
<tr>
<td></td>
<td>• Sybase</td>
</tr>
</tbody>
</table>

* Target only
Flexible Replication Options

Choose a topology or combine them to meet your data sharing needs.
High-Level Architecture

SOURCE

MIMIX SHARE

EDMM
Metadata

Change Data Capture

Source Database

TARGET

MIMIX SHARE

EDMM
Metadata

Target Database

LAN/WAN

MIMIX SHARE

Administrative Console
Did you know?

MIMIX Share can leverage published Log or Journal standards to identify and capture the change before copying to the MIMIX Share Queue.

The MIMIX Share Queue ensures that data integrity is maintained and zero data loss occurs in the event of a dropped connection during file transmission.

1. Use of transaction logs or triggers eliminates the need for invasive actions on the DBMS.
2. Selective extracts from the logs and a defined queue space ensures data integrity.
3. Transformation in many cases can be done off box to reduce impact to production.
4. The apply process returns acknowledgment to queue to complete pseudo two-phase commit.
Guarantees Information Accuracy

Ensures ongoing integrity
- Changes collected in queue on source
- Moved to target only after committed on source
- Ensures write-order-consistency retained
- Queues retained until successfully applied
- No database table locking

Ensures failure integrity
- Automatically detects communications errors
- Automatically recovers the connection and processes
- Alerts administrator
- No data is lost
Accurate Tracking & Data Auditing

D.detects and resolves conflicts
  ▪ Maintains data integrity

Model verification
  ▪ Validates date movement model

Audit Journal Mapping tracks all updates and changes
  ▪ Records
    ▪ Before and after values for every column
    ▪ Type of transaction
    ▪ Type of sending DBMS
    ▪ Table name
    ▪ User name
    ▪ Transaction information
  ▪ Records to flat file or to database table
  ▪ Can assist with SOX, HIPPA audit requirements
Lets You Share Exactly WHAT You Need

Filters determine what data gets moved

- Select specific column and table
- Select specific rows and table
Let's You Transform the Data Exactly HOW You Need To

Transforms data into useful information

- 80+ built-in transformation methods
- Field transformations, such as:
  - DECIMAL(5,2)
  - nulltostring(ZIP_CODE,'00000')
- Table transformation, such as:
  - Column merging
  - Column splitting
  - Creating derived columns
- Custom lookup tables
- Create custom data transformations using powerful Java scripting interface
Real-World Use Cases
Use Case: Offload Reporting from Production Database

Many cost effective tools available on MS SQL server platform for query reports

Data Warehouse load

Data is already partially ‘scrubbed’ and available for loading data warehouses and data marts without performance impact on production system

Production System Offload Query System

IBM System i DB2 Lawson M3 (Movex)

MS SQL Server

Real time CDC replication with transformation

Query reports

Reduce CPU and I/O overhead on production system improve user response times
Use Case: Centralized Reporting

Customer loyalty
Amounts paid
Amounts won
Time at the table
Time at the machine

Real time CDC replication with transformation

Single Data Warehouse Database
Windows Cluster
MS SQL Server

Casino 1
IBM System i
DB2

Casino 2
IBM System i
DB2

Casino 3
IBM System i
DB2

Casino 4
IBM System i
DB2

Casino 5
IBM System i
DB2

Casino 6
IBM System i
DB2

Business intelligence
Use Case: Business Process Integration

Orders are placed at HQ and processed at the plant. When the network does down, so does the plant. MIMIX Share keeps them both up and running.

Orders are placed at HQ and processed at the plant. When the network does down, so does the plant. MIMIX Share keeps them both up and running.

Manufacturing Company
Use Case: Database Migration

Old System
- IBM i
- DB2 for i
- JDE (standard)

Real time CDC replication with transformation

New System
- IBM i
- DB2 for i
- JDE (Unicode)

Manufacturing Company
Use Case: Database Replatforming

**Old System**
- IBM i
- DB2

**Transformation between different OS and database platforms**

**New System**
- Sun
- Oracle RAC

- **Two-way Active-Passive replication to enable application server switching**
- **Near-zero downtime for cutover to new systems**

Users are moved to new server in phases over a period of time.
Use Case: Gradual Database Replatforming

Old System

IBM i
DB2

Transformation between different OS and database platforms with completely different schemas

New System

Windows
SQL Server

Active-Active replication eliminated need for hard cutover and enabled partners to move back and forth between systems

True zero downtime for migration to new systems

100s of partners moved to new server after training and at their own pace

America II Corp
Market trends show data sharing methods lagging
Summing It Up

1. Market trends show data sharing methods lagging

2. ETL processes are not free and create challenges
Summing It Up

1. Market trends show data sharing methods lagging

2. ETL processes are not free and create challenges

3. MIMIX Share provides easy, automated data sharing
Summing It Up

1. Market trends show data sharing methods lagging
2. ETL processes are not free and create challenges
3. MIMIX Share provides easy, automated data sharing
4. MIMIX Share addresses a variety of data sharing needs
Ready to Learn More?

- Visit our website at www.visionsolutions.com
- Read a whitepaper or case study
- Watch an on-demand demo
- Connect on social media
- Request a live demo
- Give us a call!

+1-800-683-4667
+1-949-253-6500
Connect with Vision!

**Website:** visionsolutions.com

**Twitter:**
twitter.com/VSI_Power
@VSI_Power

**Facebook:**
facebook.com/VisionSolutionsInc

**YouTube:**
youtube.com/c/VisionSolutionsInc

**LinkedIn:**
linkedin.com/company/vision-solutions

**Blog:**
http://www.visionsolutions.com/blog
Thank You!
The State of Data Sharing

Key Takeaways

Results indicate a clear need for a paradigm shift in database management and data sharing.

Businesses are at risk when faulty data is used or decisions are delayed because time is spent reconciling data inconsistencies.
The State of Data Sharing

Key Takeaways

Results indicate a clear need for a paradigm shift in database management and data sharing.

- Businesses are at risk when faulty data is used or decisions are delayed because time is spent reconciling data inconsistencies.

- The tools and processes used to manage data are lagging, so organizations aren't moving quickly enough to capitalize on its strategic value.
The State of Data Sharing

Key Takeaways

Results indicate a clear need for a paradigm shift in database management and data sharing.

Businesses are at risk when faulty data is used or decisions are delayed because time is spent reconciling data inconsistencies.

IT professionals noted that they had business directives to integrate data for real-time accuracy across their organizations in order to achieve competitive advantage.

The tools and processes used to manage data are lagging, so organizations aren’t moving quickly enough to capitalize on its strategic value.
The State of Data Sharing

Key Takeaways

Results indicate a clear need for a paradigm shift in database management and data sharing.

- Businesses are at risk when faulty data is used or decisions are delayed because time is spent reconciling data inconsistencies.

- The tools and processes used to manage data are lagging, so organizations aren’t moving quickly enough to capitalize on its strategic value.

- IT professionals noted that they had business directives to integrate data for real-time accuracy across their organizations in order to achieve competitive advantage.

- Companies must take aggressive steps to align their data sharing schemes strategically and operationally with business directives.
Additional Use Cases

- eCOMMERCE & WEB PORTALS
  - Customer Orders
  - Payment Details
  - Product Catalogue
  - Price List

- ERP SYSTEM

- Outside Vendor
  - DATA EXCHANGE WITH OUTSIDE VENDOR (FLAT FILE)

- TEST & AUDIT ENVIRONMENT

- DR / BACKUP
The State of Data Sharing: Perceived Needs

Indicate your agreement with the following statements on a scale of 1-7
(7 = Strongly Agree; 1 = Strongly Disagree)

- My organization is interested in ways to use our data to better provide my company with a competitive advantage. (5.2)
- My organization has business directives today to integrate data so it is closer to real-time accuracy across the business. (5.0)
- My organization would benefit if data currently isolated in departmental applications was shared across the business. (4.8)
- There have been times when my organization had to make decisions or projections based upon data that was less than 100% accurate/current, and complete. (4.6)
- The diversity of the types of databases and operating systems in my shop continues to grow. (4.8)
- For the most part, we allow our business users to determine what application software they need to best do their job. (4.4)

#1 - My organization is interested in ways to use our data to better provide my company with a competitive advantage (5.2)
#2 - My organization has business directives today to integrate data so it is closer to real-time accuracy across the business (5.0)
#3 – My organization would benefit if data currently isolated in departmental applications was shared across the business (4.8)