
**FEDERATED DATA SYSTEMS WITH EIQ SUPERADAPTERS™ VS.
CONVENTIONAL ADAPTERS WHITE PAPER**

REVISION 2.51

INTRODUCTION

WhamTech offers unconventional data integration, sharing and interoperability, adapter and sub-middleware products called External Index and Query (EIQ) Products. These are dataless hybrids that combine the capabilities and advantages, and overcome the disadvantages, of data warehouses, federated data systems with conventional adapters and enterprise search. EIQ Products provide unique combinations of capabilities that would either be difficult or prohibitively time-consuming and costly to implement using conventional approaches. Yet, EIQ Products can complement, and considerably enhance, conventional approach products and systems.

CONVENTIONAL DATA INTEGRATION, SHARING AND INTEROPERABILITY PRODUCTS

Conventional data integration, sharing and interoperability products fall into one of three approach categories: Data warehouses, federated data systems with conventional adapters and enterprise search (to a lesser extent). Data warehouses are over 25 years old and have proven successful to an extent, but they have disadvantages. Federated data systems with conventional adapters are almost 10 years old and have proven less successful in many cases, but they have advantages. Enterprise search engines are designed for content management and to locate information rather than data integration, sharing and interoperability, but they have advantages. Each approach has its advantages and disadvantages and until the recent introduction of EIQ Products, they were the only choices available for data integration, sharing and interoperability.

EIQ SUPERADAPTERS

One EIQ Product, **EIQ SuperAdapter**, directly compares to, and competes with, a conventional adapter in federated data systems, where data remains in a source and queries are submitted to the data source through a conventional adapter. **EIQ SuperAdapters** provide index and query processing layers that reside external to data sources, in almost any location: behind firewalls, local, regional, central or remote. Like conventional adapters, **EIQ SuperAdapters** leave data in sources. **EIQ SuperAdapters** transparently reside between data sources and applications, middleware or sub-middleware, assume the heavy load of maintaining clean, transformed and standardized indexes, and processing queries, and provide clean, transformed and standardized results. As a result of complete control over data quality, indexes, queries and results, query success is the same as, or better than, data warehouses. **EIQ SuperAdapters** have advanced capabilities to the extent that federated data systems with **EIQ SuperAdapters** can be considered virtual data warehouses that directly compete with physical data warehouses. **EIQ SuperAdapters** save considerable time and cost, and overcome many of the disadvantages of data warehouses, but in a federated data system. A Service Oriented Architecture (SOA) is a good example of a federated data system, where, for example, Web Services provide access to multiple data sources. In addition to standard data access, **EIQ SuperAdapters** offer advanced capabilities, including text search, entity extraction, other knowledge management, and link mapping and analysis.

THE POWER OF EIQ SUPERADAPTERS IN A FEDERATED DATA SYSTEM

EIQ SuperAdapters merge technologies from data warehouses, federated data systems with conventional adapters, and enterprise search, to provide a dataless hybrid of the three technologies. This has resulted in high-capability adapters that work in federated data systems, thus the name **EIQ SuperAdapters**. The benefits of data warehouses are retained, namely:

- Clean, transformed and standardized data
- Indexes: multiple types, including pre-aggregations and pre-calculations
- Query processing: multiple options
- Security: data and access

... at the same time, the primary benefit of federated database systems is retained:

- Data remains stored at source

... and at the same time, the primary benefit of enterprise search is retained:

- Find almost anything, regardless of where it resides

Other benefits include:

- Latest data
- Almost no index or query load on data source systems
- No schema transforms
- Almost any data source – structured, unstructured and semi-structured
- Actively monitor data sources – subscribed alerts/notifications – stored queries
- Rapid query response – scale as needed
- Denormalized indexes – accelerate query response – data mart-like
- Pre-aggregated and pre-calculated field indexes
- Link Indexes™ for performance and link analysis
- Highly flexible
- Row, column (and data element) security indexes
- User-level access to data sources
- Data sources only aware of low-level results; not queries

Figure 1 illustrates how EIQ SuperAdapters fit in an EIQ Products-based federated data system.

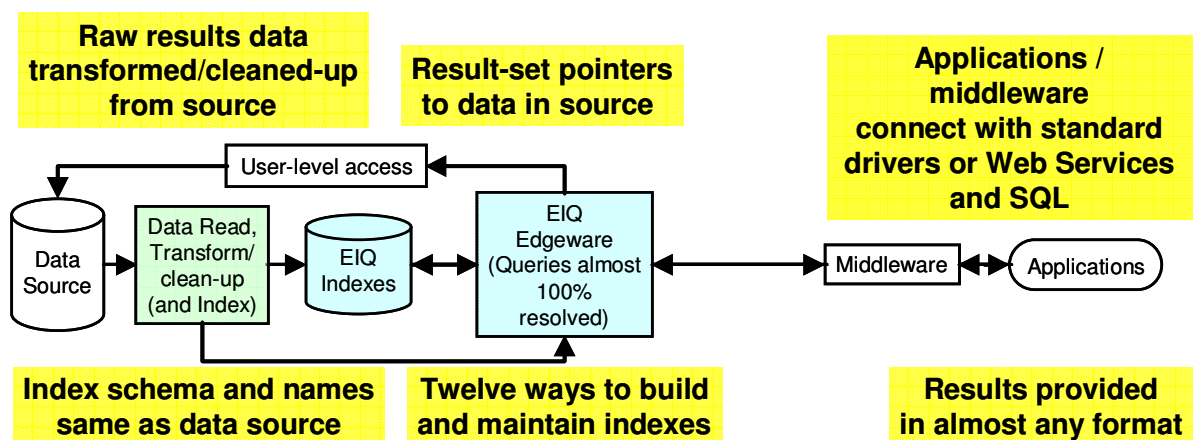


Figure 1: EIQ SuperAdapters are data source-specific edgeware EIQ Products - middleware could be SOA or server-based

Unique or unusual capabilities

EIQ SuperAdapters provide unique or unusual capabilities typically not available with data warehouses or federated data systems:

- Full text search
- Entity extraction and other knowledge management options
- Active data source monitoring with optional alerts/notifications
- Link mapping and analysis

New and unique applications

Based on some of the above capabilities, WhamTech is developing new and unique applications for **EIQ SuperAdapters** include:

- Extending text search capabilities to include advanced knowledge management to enable a better integration of unstructured text and structured data, including sense-making
- Real-time operational business and other intelligence where **EIQ SuperAdapters** constantly monitor key performance indicators and update dashboards and/or publish significant changes to subscribers - also many military applications (situational awareness, effects based operations, etc.)
- Combining link mapping and analysis with structured content queries and text search for scalable intelligence gathering, enabling virtual applications for virtual CRM and CDI-MDM, etc., data mining and other applications

THE ADVANTAGES AND DISADVANTAGES OF DATA WAREHOUSES

Data is extracted, transformed from multiple data sources and loaded (ETL) into a separate database, called a data warehouse. Figure 2 illustrates a typical data warehouse system.

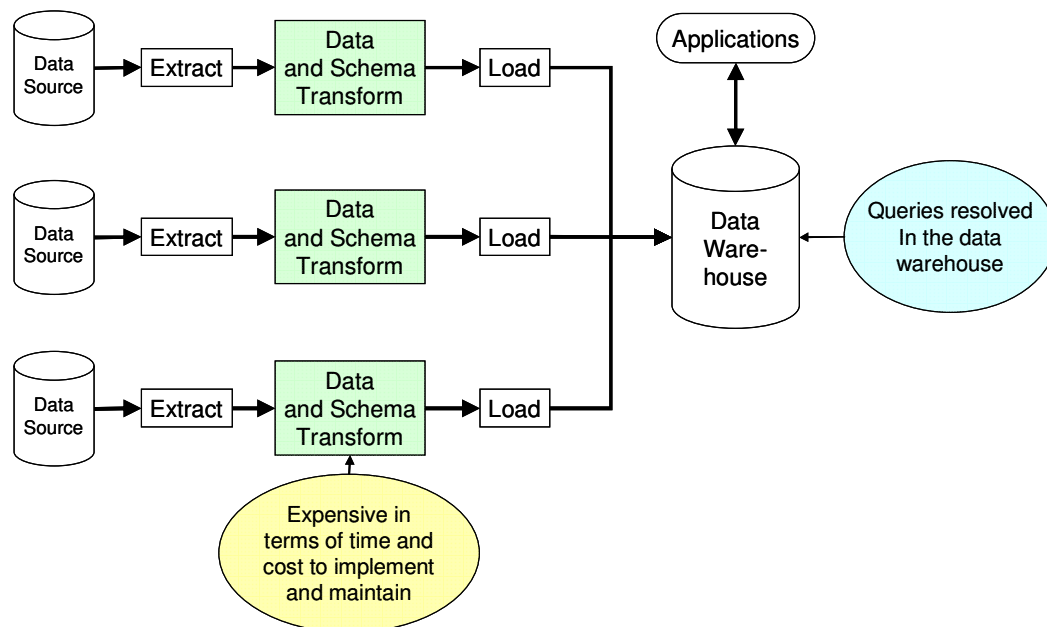


Figure 2: Data warehouse system

Advantages

Data warehouses tend to have a very high query success as they have complete control over the four main areas of data management systems:

- Clean data: cleansed, transformed and standardized as data is copied from source systems
- Indexes: multiple types and numbers
- Query processing: multiple options and speed
- Security: data and access

Disadvantages

However, there are considerable disadvantages involved in copying data from multiple, often highly disparate data sources to a single data warehouse, that translate into long implementation time, high cost, lack of flexibility, dated information and limited capabilities:

- Major data schema transforms from each of the data sources to one schema in the data warehouse can represent more than 50% of the total data warehouse effort
- Data owners lose control over their data, raising ownership (responsibility and accountability), security, privacy and legal issues
- Long initial implementation time and associated high cost
- Adding new data sources takes time and associated high cost, and can be difficult to integrate with existing data
- Limited flexibility of use and types of users – tends to require multiple separate data marts
- Typically, data is static and dated
- Typically, no data drill-down capabilities
- Difficult to accommodate changes in data types and ranges, data source schema, indexes and queries
- Typically, cannot actively monitor changes in data in anything approaching real-time

THE ADVANTAGES AND DISADVANTAGES OF CONVENTIONAL ADAPTERS IN FEDERATED DATA SYSTEMS

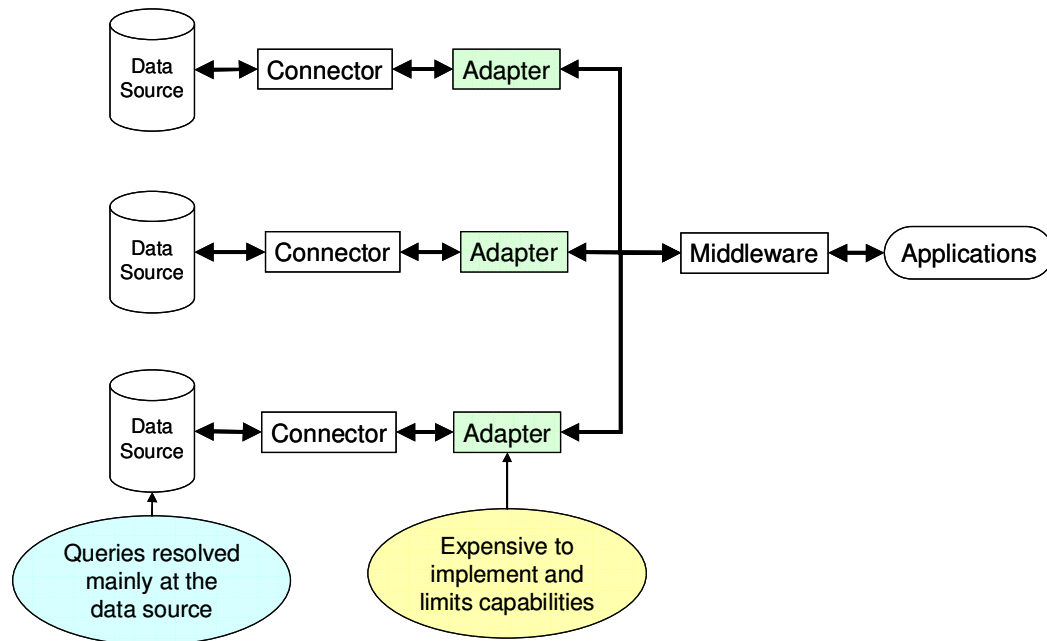


Figure 3: Conventional adapters in a federated data system

Advantages

Conventional adapters in federated data systems were pursued in an attempt to overcome some of the disadvantages of data warehouses by providing the following primary benefit:

- Data remains at source

The above benefit overcomes many of the data warehouse challenges of:

- Complex ETL process (time and cost)
- Data ownership issues
- Static and dated data
- No drill-down capabilities

Disadvantages

However, there are considerable disadvantages of conventional adapters in federated data systems that counter-match data warehouse advantages, as they have to cope with data sources directly:

- Dirty data “as is” – typos, missing, incorrect format and non-standard
- Limited indexes – not consistent across data sources and not flexible
- Limited query processing
- Query load on data source system
- Query performance
- Security
- Data source owners aware of queries (intelligence/security/confidentiality-related)

To accommodate the translation between an application or middleware and any particular data source, conventional adapters are developed, typically, over a significant period and at large cost to cover basic requirements. In fact, it typically costs 300 to 500% of the initial adapter purchase cost to customize conventional adapters to cover basic requirements. One advantage conventional adapters has over the data warehouse ETL process is that schema transforms are not as difficult; however, query processing (transforms and execution) and subsequent results processing are significantly more complex.

THE ADVANTAGES AND DISADVANTAGES OF ENTERPRISE SEARCH

Advantages

Enterprise search has the following advantages, including:

- Find almost anything, regardless of where it resides
- No need for detailed information on content or format
- Simple, keyword or phrase input

Disadvantages

However, there are considerable disadvantages with enterprise search, including:

- Unstructured approach with limited structured capabilities
- Limited capabilities to clean data
- Typically, does not work with SQL
- Typically, not real-time updates
- Centralized indexes

FEDERATED DATA ACCESS WITH EIQ SUPERADAPTERS OFFER LOW COST, RAPID DELIVERY, SCALABILITY AND ADVANCED CAPABILITIES

As there are no major schema transforms, complex query transforms or complex results processing to manage, **EIQ SuperAdapters** take less time and less cost for (a) initial implementation, (b) subsequent addition of data sources and (c) data source changes. **EIQ SuperAdapters** contribute to systems that enable true data integration, sharing and interoperability with the flexibility to evolve as the customer's needs change. **EIQ SuperAdapters** allow for early and rapid sequential deliverables that demonstrate incremental successes.

Appendix 1 is a summary of the major differences between WhamTech **EIQ SuperAdapter** and a conventional adapter in federated data systems.

Appendix 2 is a comparison between EIQ Products and other approaches ranked by advantage to EIQ Products.

For additional information on this material or any of the other EIQ Products, please contact:
Mark Armstrong, President, mark.armstrong@whamtech.com, (972) 991-5700 x206
Gavin Robertson, CTO and Sr. VP, gavin.robertson@whamtech.com, (972) 991-5700 x208

Appendix 1: Summary of the major differences between WhamTech EIQ SuperAdapter and a conventional adapter in federated database systems:

Feature	WhamTech EIQ SuperAdapter	Conventional Adapter
Clean and useable data	Yes, data is cleaned, transformed and standardized before being read to build indexes, and then discarded Raw results data are retrieved from data sources and in turn, cleaned, transformed and standardized before being provided to the calling application	Only as available in the data source – clean or unclean – no control Queries can be “optimized” in an attempt to find unclean data, and then results data can be post-query processed.
Indexes	100% in SuperAdapter - totally under control - consistent across multiple, heterogeneous data sources	In data source – no control
Add index algorithms	100% in SuperAdapter - totally under control - consistent across multiple, heterogeneous data sources Additional indexes can be added and indexes can be created from indexes	No – only what’s available in data source – no control
Query execution plan	100% in SuperAdapter – automatic and optimized as complete info on data and data source exists in SuperAdapter, e.g., counts	In adapter or associated middleware - tends to be manual to overcome any known data issues and data source limitations, and can involve schema transforms
Query processing	100% in SuperAdapter - totally under control – can scale - consistent across multiple, heterogeneous data sources	Mainly in data source - some post-query processing on results data in adapter, as queries tend to be simplified for data source
Unstructured text	Yes – text indexes as complex as needed	No – only what’s available in data source
Query optimization for large, highly normalized systems	Yes - through JOIN or denormalized indexes	No – only what’s available in data source
Row and column-level, and data element security	Yes – through specialized row and column security indexes, and embedded value indexes	Maybe – what’s available in data source and/or can be imposed in query terms
Use metadata as part of query, e.g., date-time stamp, data element security, alias and other tokens	Yes – through embedded value indexes, created field indexes or composite field indexes	No – only what’s available in data source

Feature	WhamTech EIQ SuperAdapter	Conventional Adapter
Built-in data profiling	Yes - indexes can be run in a pre-implementation mode – data transforms can be tested against resultant histograms – index trees are built-in data profiles that can be constantly referred to	No – need additional tools
Single-point update indexes, transforms, and metadata	No and Yes: No, as currently implemented at the SuperAdapter level; yes, plans are to enable a single-point update in the future	No – other type of individual adapter work needed

Appendix 2: Comparison between a federated data system with EIQ SuperAdapters, data warehouses and federated data systems with conventional adapters:

Nr.	Feature	Federated data system with EIQ SuperAdapters	Data Warehouses	Federated data system with conventional adapters	Comment
1	Minimal implementation time	✓	✗	✗	Unique to EIQ SuperAdapters
2	Quickly add new data sources	✓	✗	✗	
3	Flexibility of use and users	✓	✗	✗	
4	Actively monitor data sources	✓	(✗) ^[1]	✗	EIQ SuperAdapters advantages over data warehouses and conventional adapters
5	Full text search	✓	(✗) ^[2]	✗	
6	Unlimited query options and performance	✓	(✓) ^[3]	✗	
7	Denormalized views	✓	(✓) ^[4]	✗	
8	Link mapping and analysis/data mining	✓	(✓) ^[5]	✗	
9	No major schema transforms	✓	✗	(✓) ^[6]	
10	Can write to data sources	✓	✗	(✓)	
11	Row, column and data element security	✓	(✓) ^[7]	(✗) ^[8]	
12	Data source changes readily made	(✓) ^[9]	✗	(✗) ^[10]	
13	Clean and useable data	✓	✓	✗	

Nr.	Feature	Federated data system with EIQ SuperAdapters	Data Warehouses	Federated data system with conventional adapters	Comment
14	Consistent and multiple indexes and types	✓	✓	x	conventional adapters and same as data warehouses
15	Almost any data source	✓	✓	x	
16	Do not install anything on data source systems	✓	✓	(x) ^[11]	
17	Pre-aggregated and pre-calculated fields	✓	✓	x	
18	Data remains at source	✓	x	✓	EIQ SuperAdapters advantages over data warehouses and same as conventional adapters
19	User-level access to data sources	✓	x	✓	
20	Latest data available	✓	(x) ^[12]	✓	
21	Drill-down capabilities	✓	(x) ^[13]	✓	
22	No index or query load on data source systems	(✓) ^[14]	✓	x	Data warehouses advantages over EIQ SuperAdapters and conventional adapters
23	Data source owners not aware of queries	(✓) ^[15]	✓	x	
24	Archive	(x) ^[16]	✓	x	
25	Good for standard application data sources	(✓) ^[17]	x	✓	Conventional adapters advantages over EIQ SuperAdapters and data warehouses
26	No need for data or index update process	(x) ^[18]	x	✓	
27	No additional system cost	(x) ^[19]	x	(✓) ^[20]	

See next page for footnotes.

- ^[1] Real-time data warehouses only
- ^[2] Typically, data warehouses do not have full text search
- ^[3] Typically, additional databases or data marts are used for unlimited query options and performance
- ^[4] Typically, additional databases or data marts are used for denormalized views
- ^[5] Typically, additional databases or data marts are used for link analysis/data mining
- ^[6] No major schema transform if flat front-end schema used, e.g., GJDXDM and NIEM in government
- ^[7] Data owners relinquish control over their data. Only a few DBMS vendors provide this level of security
- ^[8] Only if data sources provide this level of security (see footnote 7)
- ^[9] EIQ SuperAdapters can accommodate some changes, e.g., indexes can be used to create new indexes; however, fundamental changes may require reindexing
- ^[10] Conventional adapters can accommodate minor changes, but not to the extent that EIQ SuperAdapters can
- ^[11] Many conventional adapters require specialized connectors and/or special access that requires installation of software on data source systems
- ^[12] Only for real-time or active data warehouses
- ^[13] Can be implemented in data warehouses, but not typical
- ^[14] Small overhead on data source system when retrieving final result-set data (about 5% of a typical query)
- ^[15] Data source system receives low-level request for specific records only – not the query that resulted in them
- ^[16] EIQ SuperAdapters can (a) store results data, (b) be configured to work with mirror-image copies of original data sources that act as archives; this complies with Sarbanes-Oxley and other regulatory requirements – note: data warehouses are not considered original data source copies, and (c) maintain an index to archives
- ^[17] Can use application vendor or third-party change data capture (CDC) capability or results level indexes
- ^[18] Only index update process, but not data
- ^[19] Typically, separate CPU for query processing and storage for indexes for EIQ SuperAdapters, but not storage for data or separate DBMS to maintain as for data warehouses
- ^[20] Typically, separate CPU and minimal storage for conventional adapters