EXTERNAL INDEX AND QUERY (EIQ) SERVER

REVISION 3.21

Introduction

EIQ Server™ is MIDDLEWARE that externally INDEXES, QUERIES, retrieves, integrates and shares data and information from multiple databases, files, documents and e-mail, from multiple platforms, in near real-time, WITHOUT moving data, creating a data warehouse, or relying on queries native to the source data. EIQ Server supports complex structured database queries, unstructured text searches, and a powerful combination of the two on all data sources.

EIQ Server is NOT a data warehouse, federated database, enterprise search, or an Extract, Transform, and Load (ETL) tool. EIQ Server is NOT conventional technology.

Background

WhamTech and its predecessors developed unique high performance index and query processing technologies over the last 25 years as part of its own relational database management system (RDBMS), called Thunderbolt. WhamTech separated out the relational index and query management system (RIQMS) from the database. As the RIQMS is so fast, WhamTech developed unstructured text search technologies, as part of its search engine products for the Web, enterprise, and specialized search, which by definition, operate on external files, documents, including Web pages, and e-mail.

With EIQ Server, WhamTech applies the power of its RIQMS and unstructured text search experience to the ubiquitous challenge of real-time data and information integration and sharing across multiple data sources.

EIQ Server Enables Radically Improved/Changed Processes (Not Just Better Technology)

Not conventional technology

WhamTech's EIQ Server product is not conventional, as it takes a very different approach to problems facing almost any large organization: *How to integrate and share data in near real-time without (a) creating additional infrastructure, as with data warehousing, (b) overloading existing systems, as with federated database, and (c) losing the ability to execute structured database queries, as with enterprise search?* EIQ Server technology offers the best of the above three approaches without the negatives associated with each of them.

Turbo charges and extends query processing

EIQ Server not only enables data integration and sharing in and between organizations, but also "turbo charges" existing databases and adds significant features such as complex query processing, heuristic data mining, value indexes for performance and aggregation, querying across and between databases, files, documents and e-mail, and link analysis and mapping (connecting seemingly disparate data).

Middleware with security and privacy controls, and query rules

EIQ Server is a highly flexible middleware product that transparently resides between applications and data sources. Security and privacy access controls can be applied, and rules can be applied to enable the best combination of data to be queried and passed back to applications, providing true intra and inter-enterprise data integration and sharing. Figure 1 illustrates where EIQ Server resides in the application/data source architecture.

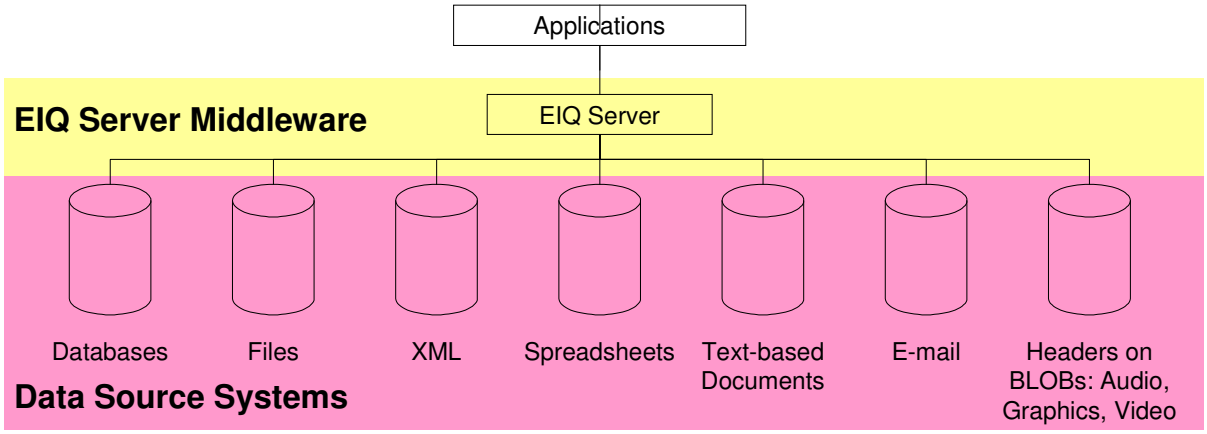


Figure 1: EIQ Server transparently resides between applications and data sources

EIQ Server Benefits

Detailed benefits are listed in Appendix 1 and some of the technical features that support them are listed in Appendix 2; the high-level benefits are as follows:

- INTERNAL DATA INTEGRATION AND SHARING: Externally index, query, retrieve, integrate and share data from multiple sources on multiple platforms within an organization. Source data remains in place. EIQ Server operations do not interfere with existing systems. Provides a single-point, universal and uniform system where a consistent approach is taken and results are automatically integrated and prioritized
- EXTERNAL DATA INTEGRATION AND SHARING: Enabling others outside the core organization controlled capability to query, retrieve and integrate data and information, for example, partners, supply chain management, and government agencies. A security and privacy access profiles module enables secure inter-organization communication
- CONSISTENT INDEXES AND QUERIES ACROSS MULTIPLE DISPARATE DATA SOURCES: Allows complete control over how data is indexed, the number of indexes, and types of queries that can be run. This ensures consistent and accurate results, and that all possible results would be obtained; this “widens the mouth of the funnel”. Improved indexes and query processing can also serve to “narrow the spout of the funnel” by allowing more query terms and/or using query terms to rank and prune results
- LEGACY SYSTEMS: Accelerate legacy systems and enable advanced and complex queries. Use as means to transition/migrate legacy data and applications to modern systems
- QUERY/SEARCH ALL DATA AND INFORMATION: Regardless of the source – structured databases, unstructured text-based documents (HTML, word processing, e-mail), or semi-structured files (XML, spreadsheets)

- HIGH PERFORMANCE: Enables high performance from legacy database systems and large modern database systems that suffer from performance issues associated with complex queries, n-way table joins, range queries, and/or large number of users
- NEAR REAL-TIME UPDATES: Enables near real-time systems, which are becoming increasingly necessary
- LOW COST AND IMPLEMENTATION TIME: As EIQ Server works with existing systems, does not require any schema transforms, and uses existing tools and drivers, implementation costs are a lot less than conventional approaches in time and resources
- ADDITIONAL QUERIES: Enables additional query features not provided by many data sources, such as combined structured queries and unstructured searches, aggregations, text searching, spatial and temporal queries, and simple data mining

EIQ Server and Databases

Introduction

The EIQ Server technology is best illustrated through a description of how it works with databases:

WhamTech's legacy database technology uses physical row numbers internally, a.k.a. record numbers or ROWIDs, which act as pointers from indexes to data in a database; these pointers are represented by integers in WhamTech's indexes. When WhamTech's indexes are used for other databases, these internal ROWIDs are mapped to external ROWIDs or other means of direct source data access.

Each database vendor has their own form of ROWIDs or unique primary keys that can be used for identifying rows within tables; WhamTech makes use of each database's form of ROWIDs. ROWIDs or primary keys are acknowledged to be the fastest route to data in a database. If a database does not have accessible ROWIDs or primary keys, a unique combination of query-able fields can be used.

Indexing

EIQ Server indexes data at two optional levels, database schema level and results level. Database schema level assumes knowledge of the underlying database schema, and allows for near real-time updates to indexes using replication software from database transaction logs, change logs, triggers, and other methods. Results level indexes avoid the need to know anything about the underlying database schema, but do not allow for near real-time updates to indexes and can impose higher loads on database systems. Both levels allow for batch or incremental updates. One of EIQ Server's underlying strengths is the capability to index large volumes of data and make the index immediately available to queries in near real-time. Indexes are built using generic definitions for tables and fields to overcome reserved words and character issues.

Query Processing

EIQ Server processes almost all queries virtually in the indexes, including table-joins and range queries. Only when a final query result-set is isolated, is the actual raw data in the database retrieved through ROWIDs, primary keys, or a unique combination of query-able fields. This has many benefits including minimizing contact between EIQ Server and the source database, resource usage, performance, multi-user support, and scalability.

Superschemas™

SQL requires that database tables and fields be specified in a query. As there are usually multiple disparate data sources attached in an EIQ Server system, each with different underlying data source schemas, WhamTech introduced a concept called Superschemas. Superschemas are virtual results level schemas that can be relational or a single table, and consist of standard metadata fields. Standard metadata field names with associated defined attributes can be obtained from XML definitions, such as e-Gov and other standards.

For Superschemas to work, each data source has a simple local mapping file created and maintained by the local database administrator (DBA), which maps database field names that need to be indexed to standard metadata field names, and the primary and foreign keys used to access one table from another. Much, if not all, of this information can be obtained automatically through an ODBC driver. Also, the local DBA may have to define a query execution plan, depending on how complex the underlying data source schema is. WhamTech is developing an automatic query processing module that will generate the queries needed to meet the query conditions and metadata fields, requested in the Superschema after security and privacy access profiles have been applied; this may avoid the need for the DBA to specify a query execution plan.

There may be differences in attributes between data source fields and standard fields; however, much of these transforms can be taken care of in the ETL or replication process and the same rules applied when source data is retrieved. Ideally, these transforms should take place at the server level, but in some cases, such as dealing with a non-EIQ Server data source in a federated database manner, mapping and transforms would have to be performed at a higher EIQ Server level.

Superschemas allow applications to work with data regardless of where it is and the format it is in. EIQ Servers can work with multiple Superschemas using the same metadata dictionary, and in the future, will be able to work with multiple metadata dictionaries.

EIQ Server Index Update Options

EIQ Server indexes can be updated as follows:

1. Batch or incremental
2. Near real-time, using replication from transaction (or redo) logs, change logs or database triggers
3. Real time, using two-phase commit or an application-aware API
4. Results level, where regular queries are used to generate result-sets, and the result-sets themselves are indexed. Queries are used to check for the latest data. This is particularly useful when accessing:
    a. An older legacy, possibly mainframe-based, system that has very simple index and query processing capabilities
    b. A very complex or unknown data structure that does not lend itself to being easily indexed
    c. An uncooperative and possibly hostile data source owner
    d. A remote data source that is only accessible through a query interface
    e. A proprietary or secret data source schema.

Compatible Databases

WhamTech is concentrating on developing EIQ Server Components to common databases and has initially identified Access, ADABAS, DB2, Oracle, MySQL, SQL Server and Teradata. Other databases and flat-file systems will be included over time.

Distributed, Parallel Processing

Scale through distributed, parallel processing and edge servers

EIQ Server scales through distributed, parallel processing on edge servers, which usually reside behind the data owner's firewall due to security and privacy reasons, but can reside anywhere. EIQ Servers cascade queries throughout an EIQ Server system for eventual execution on edge servers, and the results passed back up through the system. Results may be ranked and pruned according to any retrieval rules passed down with queries.

Query refinement and result-sets combination

EIQ Server's underlying technology has always been able to work with multiple attached databases and this concept has been extended to EIQ Server. When EIQ Server queries are executed, EIQ Server indexes (NOT the actual data sources) are queried in parallel. The total number or records and the number of records found in each data source can be determined before result-sets are retrieved from the data sources, and can be subject to retrieval rules and/or allow a degree of interaction for query refinement. Result-sets from all data sources are combined, and can be (again) ranked and pruned on the EIQ Server initiating the query.

Single-point, Universal and Uniform Index, Query, Retrieval and Integration

Database and search technologies convergence

WhamTech sees database and search technologies converging, whereby databases will support more search functions and search engines will support more database functions (to a lesser extent). It is estimated that 85% of corporate data and information resides outside of structured databases. Even the 15% of data that is estimated to reside within structured databases can be in separate databases with dissimilar data schemas and contain unstructured text.

Multiple data source retrieval options

To cover the wide range of data sources, different indexing approaches and result-set data and information retrieval options are required:

- As mentioned earlier, for databases, ROWID or primary key access can be used to directly retrieve data records using SQL. This is the fastest way to retrieve data from a database.
- For files, there is usually an API that accepts a unique record key.
- For semi-structured data, links similar to documents in combination with other referential data can be used, e.g., row and column for a spreadsheet, row and tag (or position) for an XML document.
- For documents, a URL (uniform resource locator), UNC (universal naming convention), or other file reference can be used, e.g., on a document server.

- For e-mail, a unique ID can be used.

One-stop middleware

EIQ Server is "one-stop" middleware that appears to applications as a database driver and can be used to access ALL structured, semi-structured, and unstructured data sources, using structured queries and unstructured text search on ALL data sources. Structured metadata and/or extracted entities from unstructured data sources would have to be available for structured queries.

Two-way gateway

EIQ Server is also a two-way gateway for data and information integration and sharing, as external EIQ Servers and data sources can be configured, as well as internal EIQ Servers and data sources. This is one of the "buy-ins" of the system aimed at data source owners and DBAs, as they are not only allowing others controlled, secure access to their data sources, but they are also providing a means to access others' data sources.

EIQ Server Deployment Architecture

EIQ Servers can call on other EIQ Servers, and different EIQ Server configuration files can be used for different applications, security access levels, etc. Also, of note, EIQ Servers do not need to conform to a fixed hierarchical structure; lower-level data sources can be direct connected to higher-level EIQ Servers, bypassing intervening layers. EIQ Server deployment is illustrated in Figure 2, below. A more detailed diagram is available in Appendix 3.

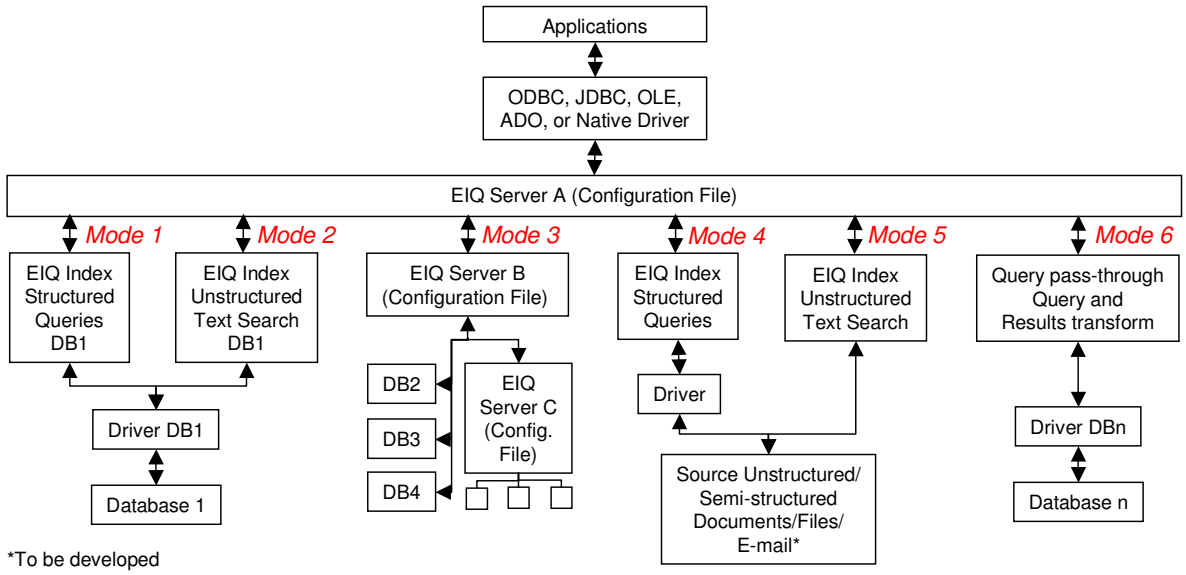


Figure 2: EIQ Server deployment with multiple databases, lower-level EIQ Servers, files and documents

Comparison of Conventional Approaches to EIQ Server

Three Conventional Approaches to Enterprise Data and Information Sharing

(a) **Universal data warehouse approach:** Load data into a data warehouse; designed to accommodate the most requested data and information, probably, de-normalized or a large flat-file system. This data can be further loaded into data marts and OLAP cubes for specific analysis.

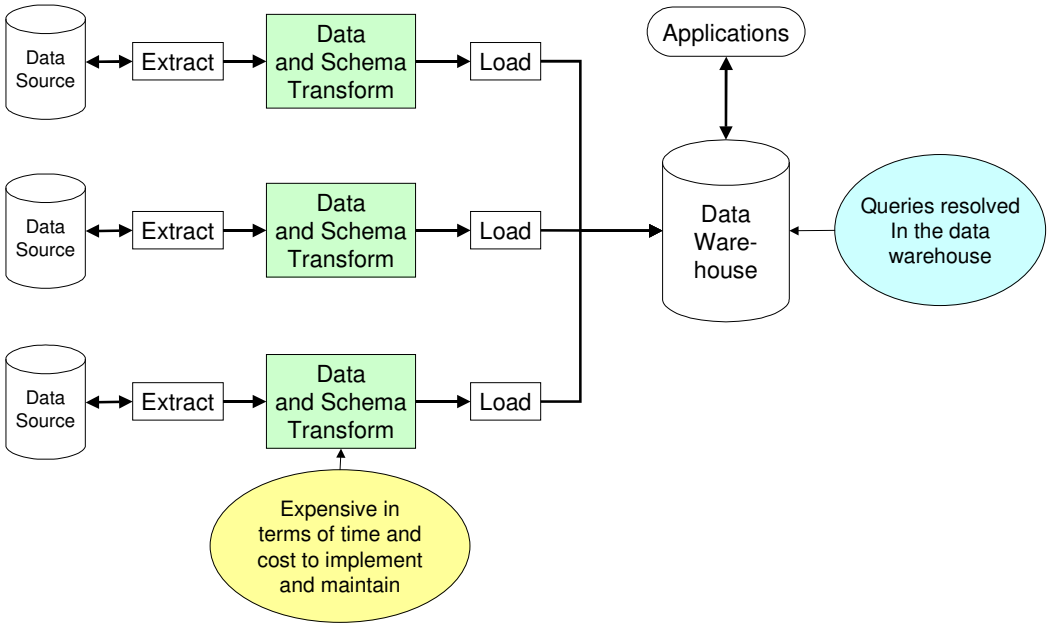


Figure 3: Universal data warehouse approach

- Pros:
  - Data "clean" and usable "as is"
  - Very fast query responses
  - Minimum existing system interference
  - Good for archive
  - "Better the devil you know, than don't know" – known, established systems
- Cons:
  - Data moved from data source – ownership, security and privacy issues
  - Major ETL needed (up to 50% of the work involved in data schema transforms)
  - Additional systems and storage
  - Generally, does not allow for detail drill-down
  - Generally, not updated in near real-time
  - One size fits all schema different from transactional and operational databases – difficult to relate back

- Database -> data warehouse - > data marts/OLAP cubes is a long, involved process, i.e., expensive
- Only a small group of highly trained staff can use
- Specialized data mining and business intelligence tools required

(b) **Federated database approach**: Queries are submitted to a common standard data schema and then adapted to the correct syntax for individual databases, individual database queries are executed, and individual database results are sorted and presented in a universal format.

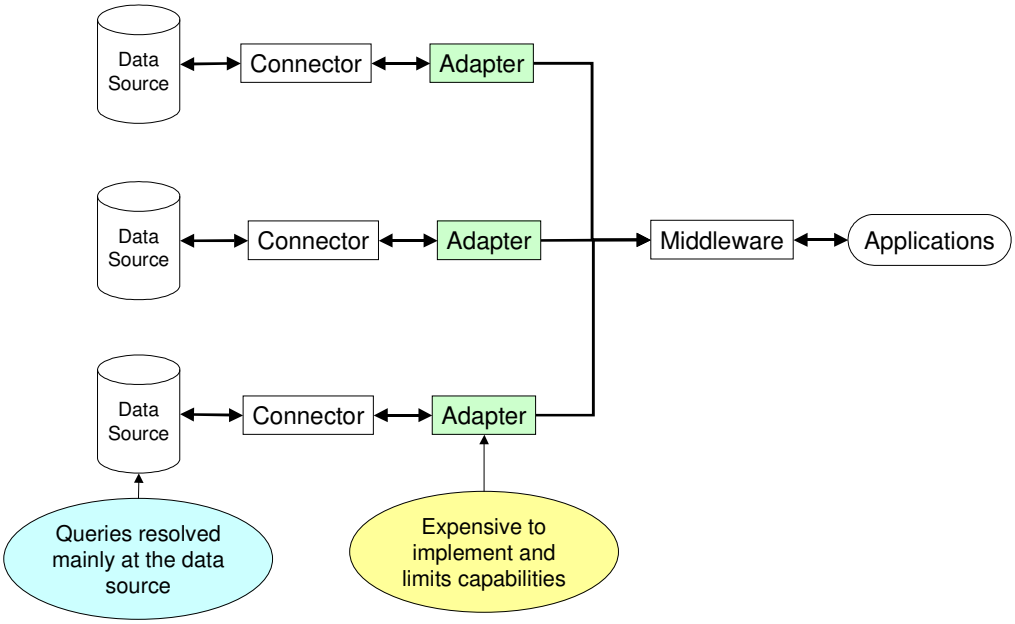


Figure 4: Federated database approach

- Pros:
  - Data not moved
  - Little to no additional storage
  - Updates as fresh as the data source
  - Good for standard applications and related data sources, e.g., SAP, or modern and flexible systems
  - ”Better the devil you know, than don’t know” – known, established systems
- Cons:
  - Have to build adapters that tend to take time, be very limited, are expensive - complete understanding of databases, associated indexes, and query execution plans required
  - Imposes a heavy load on data source systems and adapters - data source queries tend to be “dumb-downed”, large result-sets are then filtered for final result-sets
  - Variable query performance: System as fast as the slowest individual database

- One size fits all front-end schema
- Limited to databases
- Low-level data only – no high-level summaries or aggregations
- May be difficult to execute complex queries across multiple different database schemas and fields
- Difficult to use data and information from one data source to find data and information in another – external joins ACROSS data sources
- Difficult to merge results – queries and data not the same across databases

(c) **Search engine approach**: Creates an index, which is searched, and metadata and the source document link provided as a result, e.g., a Web search engine approach.

- Pros:
  - Very fast
  - Very comprehensive – multiple file formats
  - Little knowledge needed of content and structure – parsers and universal storage format take care of that
  - Can accommodate very large volumes
  - Can accommodate very complex and ad hoc Boolean-type searches
- Cons:
  - Unstructured text search only
  - Additional storage for indexes
  - Source data needs processing
  - Unstructured
  - Can be stale – depends on refresh rate
  - Cannot accommodate complex database-type queries and probably not range queries

EIQ Server Approach

…combines the best of the above three conventional approaches in a "one-stop", flexible solution that should meet most of the needs, and at the same time, overcome most of the problems and hurdles to implementing alternative solutions.

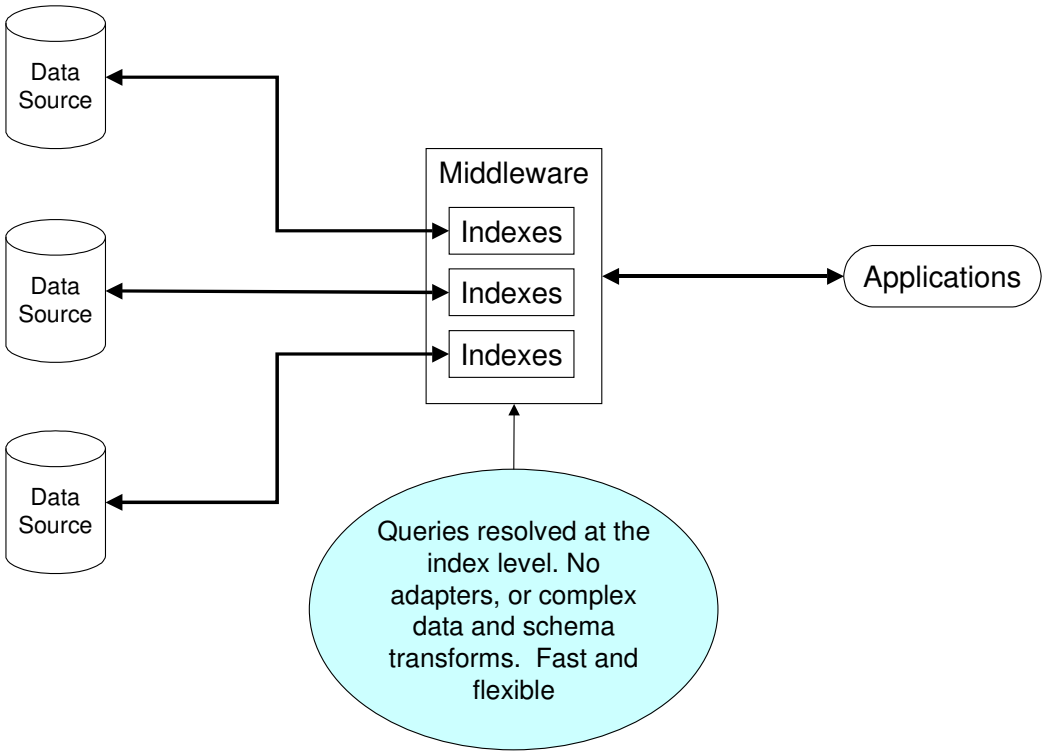


Figure 5: EIQ Server approach

For completeness, the following are the pros and cons of the EIQ Server approach:

- Pros:
  - Data not moved
  - No major ETL (no schema transforms; some data transforms)
  - No adaptors
  - Consistent and multiple indexes across multiple disparate data sources
  - Highly flexible indexes and queries
  - Near real-time updates
  - Any data source – structured, unstructured, and semi-structured
  - Simple to add new data sources
  - High performance
- Cons:
  - Index update process
  - Additional storage for indexes

For EIQ Server, everything is relational database, whether structured queries or unstructured text search, as they are executed in a similar manner – structured queries and unstructured text search each has their advantages and disadvantages. Some of the real benefits of EIQ Server are when both structured database queries and unstructured text search are used in combination, in the same SQL statement.

For a more detailed comparison, please see the separate WhamTech document entitled "EIQ Server Versus Conventional Approaches".

Real-life Examples of EIQ Server Solutions

Example 1: Real-time Security System Involving Multiple International Organizations and Multiple Departments Within Organizations

Typically, departments and organizations are very protective of their data and information, and sharing is not common. EIQ Server enables advanced query capabilities and controlled access to data and information WITHOUT imposing an additional load on existing systems AND without relying on the native (or lack of) query processing of these systems. All queries are executed "virtually" within EIQ Server, only final result-sets requesting specific data and information are retrieved from the data source, and results integrated within EIQ Server. Security and privacy access profiles are established for organizations, individual users within organizations and applications. Access rights should be down to the row and field-level, and controlled by the data source owner.

The Security System is designed with six Lines of Defense (LODs) to STOP terrorists from:

LOD1. Obtaining visas for the country

LOD2. Stepping on a plane/ship bound for the country

LOD3. Entering the country

LOD4. Activities in the country

LOD5. Leaving the country

LOD6. Conducting activities abroad (restricting money flow, extradition, sanctions, military action and war)

Each of these LODs involves data and information sharing between different agencies and organizations reporting to (a) government departments, (b) state and local authorities, (c) private industry, and (d) foreign authorities. Similar data and information sharing requirements are needed at each LOD, and the same system should be used by different agencies and organizations. For the system to be effective, data and information must be available in near real-time.

If the system is properly implemented, it should ease travel rather than impede travel, as perhaps as many as 90% of passengers could be quickly eliminated from detailed scrutiny. It would make travel safer, as there would be more selective interviews and searches made, and less inconvenienced passengers.

Example 2: Government Agency Seeking Data and Information From Education Institutes

EIQ Servers can be used to index and query data and information from each education institute in a non-intrusive and low-impact way by either installing locally or remotely. Only certain significant data and information need to be indexed regularly/continuously by the EIQ Servers. The EIQ Servers are used to (a) risk score the data and information coming from the education institutes and send alerts to the government agency, or (b) process specific queries from a higher-level government agency EIQ Server. In the case of (a), specific applications could be run on high-level EIQ Servers to risk score and send alerts.

The power of such a system would be when the indexed data and information is used in conjunction with indexed data and information from other systems. In the event an education institute does not have an associated EIQ Server:

(i) A native query can be made to the education institute and then mapped to EIQ Server standards on an EIQ Server (some knowledge of the education institute data sources would be required) – federated database approach, or

(ii) The education institute undertakes to provide the data and information requested by the government agency in a prescribed format – simple data and information sharing, for example, XML

Note that in all the above scenarios, the education institute would have 100% control of access to its own data sources, and the source data and information would stay with the education institute.

Example 3: Transition/Migration Tool

A legacy system consists of a flat-file database and many stand alone applications. The goals are:

- In the short-term, to externally index and link multiple legacy data sources, enable advanced queries and fast query response, and open up these legacy data sources to modern applications
- In the longer-term, to use EIQ Server as a transition/migration tool while legacy data and eventually, legacy applications, are moved to a modern system.

See Figure 3 on the next page for an illustration of the above.

Some of the features needed are a combination of structured database queries and unstructured text searches on databases, records from one legacy system connected in a one-to-many manner to other systems through link mapping, and combining database queries and searches with other unstructured documents. These features may still be needed after migrating legacy systems over to modern systems.

EIQ Server's functionality can change over time by applying different business rules in EIQ Server middleware layer. No changes in the application or the source data are required. This provides tremendous flexibility and minimizes impact on systems.

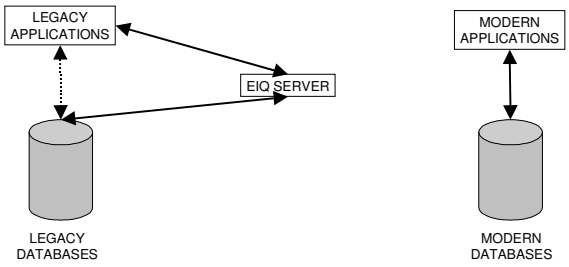
Notes:

(a) Potentially no need to see or understand applications – may need to know the type of queries currently being made and desired in the future

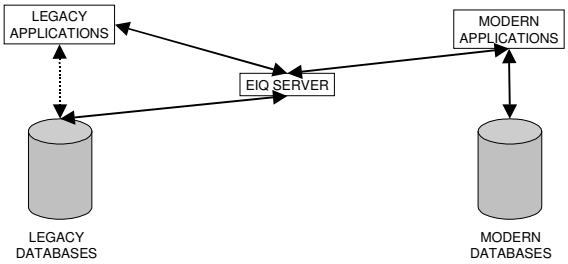
(b) Multiple legacy and/or modern data sources can be externally indexed, queried and integrated simultaneously

EIQ Server can de-normalize modern relational systems (virtual data warehouse) for legacy applications and normalize (to a limited extent) legacy flat-file systems for modern applications

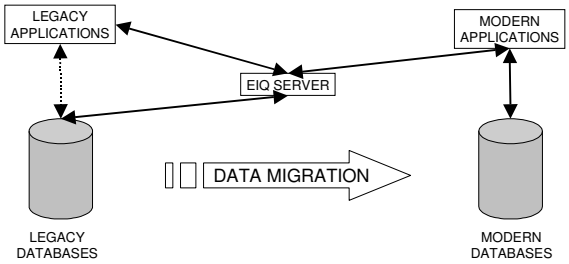
STAGE 1: Initially, no interconnectivity between legacy applications and modern databases and between modern applications and legacy databases



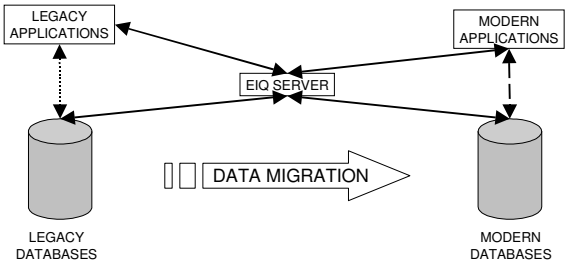
STAGE 2: EIQ Server improves performance on legacy system and adds capabilities, which results in less direct communication between legacy applications and legacy databases



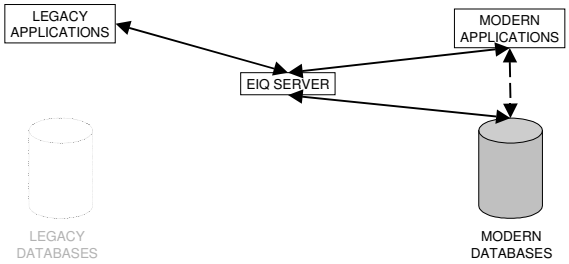
STAGE 3: EIQ Server provides modern applications access to legacy databases



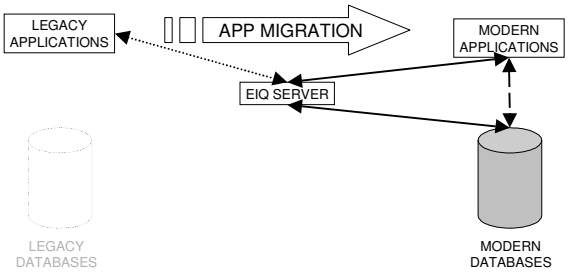
STAGE 4: Over time, data is migrated from legacy databases to modern databases, which may continue as replication



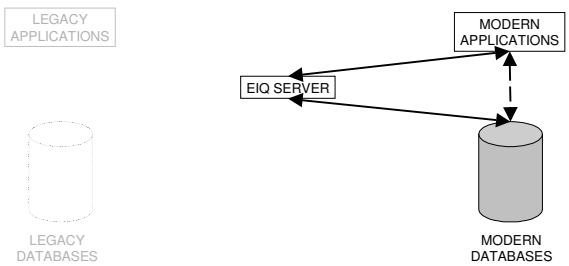
STAGE 5: Legacy applications start accessing data on modern databases - Modern applications may start accessing modern databases through EIQ Server for performance and additional capabilities



STAGE 6: Eventually, legacy databases are no longer supported



STAGE 7: Legacy applications are rewritten or replaced by modern applications



STAGE 8: Ultimately, legacy applications and legacy databases are replaced by modern applications and modern databases

Figure 6: EIQ Server as Transition/Migration Tool

Example 4: Human Resources Tool

An organization needs to access multiple legacy data systems to run payroll and other HR systems, and eventually migrate legacy data over to a modern database system for use by modern applications; however, these multiple legacy data systems are multiple types, platforms, locations, schemas, and field names. Similar to Example 3, there is an immediate, short-term need for the payroll system to have a unified view of the disparate legacy data, and a longer-term goal of migrating legacy data over to a modern database.

A solution would be a combination of the multiple data and information sharing solution (as in Example 1) and the transition/migration tool solution (as in Example 3).

The proposal could be implemented in other organizations, as we understand the same situation exists almost everywhere. An additional thought would be to enable higher-level payroll and other HR systems to be run against lower-level systems for a better overview.

Example 5: Merger of Multiple Company Systems

A large company has grown through developing separate lines of business units (LOBUs), which were in the past allowed total freedom on IT matters, resulting in 32 separate systems. Many customers are customers of more than one LOBU, in some case, a large number of LOBUs.

In an effort to create a single company-wide view of a customer, EIQ Server can be used to process queries against all 32 LOBUs and their respective systems. For some single LOBUs, more than one system may need to be involved in the process. The alternative is a data warehouse, with all the associated issues.

EIQ Server middleware offers a non-intrusive, low-impact means of gaining the latest collective view of a customer, without the huge effort required to build and maintain a data warehouse.

Appendices

APPENDIX 1: EIQ Server Benefits and Functions

APPENDIX 2: EIQ Server Technical Features of Indexing and Query Processing

APPENDIX 3: EIQ Server Deployment with Multiple Databases, Documents and Lower-level EIQ Servers

For technical information, please contact:

Gavin Robertson, CTO
(866) 360-4645 x223 (toll free)
(972) 380-4645 x223
(972) 380-1409 (fax)
gavin.robertson@whamtech.com

For sales information, please contact:

sales@whamtech.com

For company information, please contact:

Mark Armstrong, President

(866) 360-4645 x222 (toll free)
(972) 380-4645 x222
(972) 380-1409 (fax)
mark.armstrong@whamtech.com

APPENDIX 1: EIQ SERVER BENEFITS AND ASSOCIATED OPTIONS

BEN1. Universal and uniform near real-time external indexing, query processing, and integration of ALL structured, unstructured and semi-structured data and information in multiple databases, files, documents, and e-mail

- Connects like a database driver
- Single-point access
- Intra-organization and inter-organization
- Can combine database queries and unstructured text searches
- Queries are processed "virtually" in the indexes with no temporary or interim tables for table-joins or range queries
- Can use data and information from one system to find data and information in another, e.g., table-joins ACROSS databases – heuristic data mining ACROSS databases and other data sources
- Extremely simple SQL statements to EIQ Server – "rocket science" on the back-end – do not need to specify data sources (but can) or how table-joins or queries are processed
- Use of standards for field names
- Can set up Superschemas which create custom views (tables) of standard data and information fields – can be relational
- Brings some structure to unstructured data and information and MORE structure to less-structured data such as flat-file systems

BEN2. Universal and uniform access to different platforms and locations

- Use native drivers/access where possible
- Use secure LANs, WANs, VPNs, and IP (Internet, intranets, and extranets)
- Secure logins, passwords and access levels to specific agencies, data sources within agencies, and fields within data sources

BEN3. Leave source data in place in original format

- Use third-party replication/ETL tools to update indexes from transaction logs for databases – spiders/crawlers for other data
- Only indexes conform to standards – data remains unchanged
- ROWIDs, primary keys or other unique identifiers used to directly retrieve final result-set data
- Mapping tables between standard field names and actual data source field names, and security access established and maintained by the data source owner – no loss of control of data

BEN4. Continue to use legacy applications (and data), but enable modern application access to legacy data, and legacy application access to modern databases

- Also use as transition/migration tool from legacy to modern systems
- Can "virtually" normalize legacy flat-file systems to a certain extent and "virtually" "flatten" modern relational databases for legacy applications

BEN5. Scale through multi-tiered access to independently maintained indexes in different agencies and organizations

- Can use secure SOAP
- Can lead to eventual (secure) XML and Web Services
- Able to submit queries and integrate responses from non-EIQ Server data sources – a federated database technique

BEN6. Extremely fast query processing – typically, 10 to over 100 times faster than other systems, particularly in a multi-user environment

BEN7. Real-time updates that are immediately available to queries

BEN8. Value Indexes allow additional data and information to be stored in the indexes – the original data source is not altered in any way

- Improve query performance – faster table-joins (not join-indexes)
- Add value to original data
- Connect or group data and information
- Create virtual data warehouses/data marts – de-normalization

BEN9. Link analysis and mapping, which externally provides and maps links between, or groups, disparate data and information

BEN10. Significant benefits compared to a federated database approach. Federated systems are:

- As fast as the slowest data source
- Limited to the data indexed in individual systems and HOW it is indexed
- Additional query loads on systems
- In need of a detailed understanding of the system – indexes, resource requirements, etc.
- Not necessarily able to allow multiple indexes on the same data
- Generally, limited to databases; not files, documents, e-mail, etc.
- Unable to add external tables, data and information to original data source accessible through indexes
- Unable to add value to the original data in the indexes like Value Indexes
- Difficult to use data and information from one data source to find data and information in another – heuristic data mining ACROSS data sources
- Difficult to merge and work with result datasets, depending on metadictionaries and mapping

APPENDIX 2: EIQ SERVER TECHNICAL FEATURES OF INDEXING AND QUERY PROCESSING

1. **Extremely fast query response times**. WhamTech's query processing is executed "virtually" using indexes and Boolean operations on query result-sets. As an example of WhamTech's complex query processing, Thunderbolt achieves sub-second responses to complex queries on a live one billion record database – see www.billionrecords.com for an online demo and details. WhamTech owns www.sub-second.com, as that is WhamTech's goal for most queries.
2. **Very large databases (VLDBs)**. WhamTech is first and foremost a VLDB technology company that deals with VLDB issues; in particular, the combination of high performance of complex queries by a large number of users on VLDBs. EIQ Server brings many of WhamTech's database technology benefits to other mainstream database technologies.
3. **Real-time indexes with extremely fast update rates**. WhamTech's indexes allow INSERT/UPDATE/DELETE rates of up to 10s of 000s of records per second on a single server. Queries can be made on the indexes immediately following the brief moment they are updated. One proof-of-concept achieved a single-term query and insert rate of 80,000 records per second in a 60 GB, 300 million-record database on a dual-933 MHz Intel server with 4 GB RAM and local SCSI, 7200 RPM, RAID 5 disks.
4. **Disproportionately high number of users per server**. WhamTech's RDBMS, Thunderbolt, and EIQ Server support a disproportionately large number of users per server due to:
    a. Extremely fast index and query processing
    b. Virtual query processing, where data in the database is not accessed until final result sets are isolated
    c. User channel reuse, where channels (or threads) to the database engine are available to other users before and after a query execution
5. **Storage required for indexes only**; data remains in the source database. WhamTech's indexes are very efficient and usually require a lot less space than conventional indexes. Storage can actually be reduced overall if WhamTech's indexes replace some of the source database indexes.
6. **Unique WhamTech commands** that allow multiple ways to present JOIN data, and a heuristic SELECT that performs link analysis and data mining functions that are traditionally dealt with through OLAP.
7. **Immediate record counts**…are available at every point in a query. These counts are automatically tracked at the data value level and are also available in result-sets. The data itself does not need to be counted.
8. **Value Indexes**…are a separate form of indexes and can be used to accelerate access to low-level data, by storing higher or same-level data in low-level indexes. Value Indexes can be used to avoid or minimize extensive table-joins, compute aggregation statistics and aggregation data on the fly, e.g., SUM, AVERAGE, MAXIMUM, MINIMUM, MEAN, STD. DEVIATION, etc., and simplify complex queries.

Value Indexes and normal indexes can be combined for ad hoc aggregations that would not be possible with other database index and query technologies.

9. **Spatial and temporal queries**…are the reason WhamTech was formed as a company to develop the database technology, which has exceptional range query processing. EIQ Server indexes are well suited to processing range queries, such as GIS and period data, without creating interim or temporary tables. WhamTech developed an oil and gas pre-processing utility that cuts down by several orders of magnitude the time it takes to organize raw seismic data, in particular, multi-component, 3D seismic data before processing and interpretation. This feature avoids the need to create OLAP-style PERIOD tables for subsequent analysis.

10. **Listpick**…is a means of only retrieving data that is needed at a specific time using the indexes or result-set pointers. For instance, if a user queries and isolates millions of records from a VLDB, the entire millions of records do not to be read and sent across a network to the user at one time, as this would take a long time, slow down the network and overwhelm the user. Instead, the user is only shown a limited number of records at any given time, e.g., 25, and is provided controls to navigate the list of data using pointers rather than actual data. This allows users to quickly page through huge numbers of records and jump to different points in the list of records without paging.

APPENDIX 3: EIQ SERVER DEPLOYMENT WITH MULTIPLE DATABASES, DOCUMENTS AND LOWER-LEVEL EIQ SERVERS