

SmartData Fabric[®] (SDF) Power of Indexes

June 2021

SmartData Fabric[®] general solutions

Unique security-centric distributed indexed adapter-based data virtualization, data federation and data integration software for the following general solutions:

- Automated data discovery, profiling, quality, standardization, governance and relationships mapping
- Advanced data access and data security seen as a security solution
- Virtual data warehouse and/or virtual data mart
- Data lake + data management + master data management = clean and usable data reservoir
- Data provisioning for highly curated, self-serve reporting, BI and analytics
- Interoperability with write-back to data sources integrated data, not just app to app
- Seamless, automatic and near real-time updateable distributed master data management
- Virtual graph database and link analysis, and interactive graph/link visualization
- Hybrid Cloud 2.0 where data sources remain wherever they reside, but run all compute in the Cloud or data center
- Near real-time data source monitoring, event processing and Business Process Management (BPM)
- Embrace and enable STANDARDS such as ODBC, JDBC, REST APIs and SQL, and standard applications

➔ Discover, secure, access, integrate and deliver INTEGRATED structured, unstructured and semistructured data from almost ANYWHERE to almost ANYWHERE in almost ANY FORMAT, aka Actionable Data Catalog



Conventional data virtualization/federation vendors

- 1. Leave data in sources (some do not)
- 2. Virtualize the view of data in sources
- 3. Connect standard applications using standard drivers and SQL
- 4. Access and query data in multiple sources in parallel, usually through connectors or adapters
- 5. Combine results from multiple sources (some do not)
- 6. Cache results data for improved query performance and less data source load (some do not)
- 7. Build and maintain MDM (some do not)
- 8. Apply MDM to combined results data to provide integrated results to applications (some do not)

In step 4 above,

- ALL CONVENTIONAL DATA VIRTUALIZATION VENDORS ARE 100% DEPENDENT ON DATA SOURCES AND DATA IN SOURCES for data quality, data standardization, available indexes and indexed views, and query processing, which
- CAN IMPOSE SIGNIFICANT QUERY LOAD ON DATA SOURCES AND LEAD TO POOR QUERY PERFORMANCE
- HAS AN IMPACT ON DATA SOURCE ACCESS CONTROL AND DATA SECURITY and
- FAILS TO 100% DELIVER

→ Enterprises need to enable all/most data-related fundamentals AND address data-related issues to be successful – not just enable access to data

Source: #WhamTech SmartData Fabric Basic Overview



WHAMTECH SmartData Fabric[®] data virtualization/federation value-add

100% NOT dependent on data sources or data in sources, and addresses related issues upfront by

- a) pre-processing raw source data (and optionally storing processed data in indexes) while building and maintaining indexes and indexed views,
- b) processing queries against these indexes and indexed views, and

c) post-query processing raw results data read from sources (or read directly from indexes)

but only on data in sources that needs it!

- Most data in many systems is non-human-generated and does not need pre-processing or post-query processing, e.g., transaction systems, but may still need external indexing and query processing
- Some data needs pre-processing and post-query processing, such as customer, product, organization, etc., e.g., entities and associated attributes, usually for data quality, standardization and security, and master data management (MDM)

Enables all/most data-related fundamentals that enterprises continue to struggle with, mainly by addressing all/most issues with data sources, data in sources and access control

→ COMBINES THE BEST of conventional data virtualization, data warehousing, enterprise search and graph database, and OVERCOMES THE WORST of these approaches



Unconventional data virtualization

- Transparent virtual distributed data management layer that plugs-and-plays in existing IT infrastructures
- Complements and leverages existing IT systems, tools and applications
- Key differentiator: Federated adapters that Read, Transform and INDEX (RTI) data, wherever it resides, process queries against these indexes, and read and transform results data from sources
- Indexes enable UPFRONT semi-automated data discovery, security, quality, standards, MDM and other datarelated processes, BEFORE the first query made/application used
- Leaves and guards data in sources

Source: #WhamTech SmartData Fabric Basic Overview





INDEXES can reside anywhere and remain 100% contiguous

- Patient-centric smartphone app interacts with legacy data sources through new workflows developed and orchestrated by BPM software
- BPM workflows interact with data source through standard FHIR REST APIs provided as data services
- BPM workflows both read and write back to legacy data sources





New paradigm – Read, Transform and INDEX (RTI)



INDEXES ARE KEY to a complete understanding of data, enabling capabilities and driving value

Including being able to identify and access GDPR, HIPAA, PCI and other confidential, classified and risk data



Indexes key to understanding data, enabling capabilities and driving value



- 1. Use raw indexes for **DATA DISCOVERY** (metadata), build and maintain **DATA PROFILING, DATA MATCHING** within and across data sources, and **DEVELOPING AND TESTING DATA TRANSFORMS**
- 2. Support **FORRESTER ZERO TRUST DATA SECURITY FRAMEWORK** discover, INDEX, classify and secure GDPR, PCI, PHI, PII, etc.
- **3. PRE-PROCESS DATA** while building and maintaining production indexes to address data management fundamentals, e.g., cleansing, transformation, standardization and security data is usually discarded
- 4. Use LINK INDEXES[™] AS BASIS FOR MDM AND OTHER CAPABILITIES future development to use indexes exclusively for MDM match and merge
- 5. Provide **COMPLETE AND MULTIPLE VIEWS OF DATA** through queries on combined content, link and master data indexes
- 6. Provide **FULL DATA TRACEABILITY** as indexes and results contain unique pointers to data in sources data lineage, governance and audit
- 7. Enable **HIGH PERFORMANCE, DISTRIBUTED PARALLEL QUERY PROCESSSING** through standard drivers, APIs, Web/data services, SQL and other query languages
- 8. MONITOR DATA SOURCES for content and relationships in near real-time, and support EVENT PROCESSING
- 9. Enable VIRTUAL GRAPH DATABASE, link analysis and graph/link visualization
- **10. GENERATE RESULTS WITHOUT DATA SOURCE** when source is unavailable, for query optimization, or as storage, e.g., for IoT devices, as indexes are columnar and can be inverted and combined



Three types of indexes

Content indexes basis for other indexes

All indexes resolve to "record numbers" – internal to SDF, but correlated with external/data source references, and can be combined using Boolean operations on physical and virtual bitmaps





Eight types of content indexes





Index updates/changed data capture options





The End