
SMARTDATA FABRIC® BEST USE CASES

REVISION 1.4

Reference: WhamTech SmartData Fabric Basic Overview.

All data virtualization vendors follow the general steps in approximately the same order:

- DV1. Leave data in sources (*some do not*)
- DV2. Discover and virtualize the view of data in sources
- DV3. Connect standard applications through standard drivers and SQL
- DV4. Access and query data in multiple sources in parallel, usually through federated adapters
- DV5. Combine results from multiple sources (*some do not*)
- DV6. Build and maintain MDM (*some do not*)
- DV7. Apply MDM to combined results data to provide integrated results data to applications (*some do not*)

In step DV4 above, **ALL CONVENTIONAL DATA VIRTUALIZATION VENDORS ARE 100% DEPENDENT ON DATA SOURCES AND DATA IN THE SOURCES** for data quality, data standardization, available indexes and indexed views, and query processing - much of which has an impact on data source access security and data security. If data sources and/or data in sources have issues, conventional data virtualization queries will provide incomplete, incorrect or no results. This is in addition to imposing a sometimes-heavy query load typical of reporting, BI or analytics queries on data sources that are not designed for such, e.g., transaction systems.

SmartData Fabric® addresses data source and data-in-the-source issues through index-based adapters, by externally:

- SDF1. Discovering, profiling, identifying, classifying and mapping raw source data
- SDF2. Reading, cleansing, transforming, standardizing and indexing data, including building indexed views
- SDF3. Optionally discarding raw data or storing curated data in indexes and indexed views
- SDF4. Processing queries against these indexes and indexed views
- SDF5. Reading, cleansing, transforming and standardizing raw results data from data sources if data is not stored in indexes and indexed views

ONLY ON DATA THAT NEEDS IT – NOT DATA THAT DOES NOT. If a data source does not have issues and can absorb the load of external queries, and some/most/all data does not have issues, a conventional federated adapter can be used for that data – either in part, in conjunction with an index-based adapter (aka Hybrid Adapter), or in whole. The raw results from a conventional federated adapter can be transformed, standardized and combined with results from other adapters, including index-based adapters, in SmartData Fabric®.

Best Use Cases for Index-based Adapters or Hybrid Adapters for Data Sources that Need:

- UC1. Automated data discovery, profiling, identification, quality, standardization and governance that (a) can be acted on directly, and (b) updates complete metadata/data profiles as indexes are updated
- UC2. Data to be cleansed, transformed, standardized, masked, tokenized and/or encrypted, e.g., personal, sensitive, MDM, other entity, “dirty” and incomplete data
- UC3. Data/entity link/relationship mapping for MDM, virtual graph database and other uses
- UC4. Seamless, automatic and optionally distributed MDM with near real-time updates for integration within and across multiple data sources
- UC5. Standard data views, business objects and knowledge graph across all data
- UC6. Standard drivers of ODBC, JDBC and APIs, and/or SQL query processing, e.g., mainframe files, file systems, IoT devices, office docs, email, Web pages and other unstructured/semi-structured data
- UC7. An external indexing and query processing layer that can absorb the load of external queries
- UC8. Integration of unstructured data with structured data through text analytics, e.g., entity extraction, and search
- UC9. Advanced access control within and across domains, e.g., AD/LDAP, IAM, SSO, RBAC, ABAC/RLS and CLS, regardless of data source support for any of these – also applies to conventional federated adapters
- UC10. Data monitoring, event processing and BPM workflows, including writing back, in near real-time, e.g., for operational reporting, BI and analytics
- UC11. Hybrid Cloud 2.0 deployment options where compute is in the Cloud, but data sources remain remote on-premise, in data centers, SaaS, third-parties and multi-Cloud, and Hybrid Cloud 1.0