WHAMTECH®

SmartData Fabric® (SDF) Summary Comparison with Informatica®, Similar to Other Conventional Data Virtualization Vendors

January 2019

Comparison between SDF and Informatica® (1 of 3)

Feature	SmartData Fabric®	Informatica®, similar to other conventional data virtualization vendors
DISCOVERY: Device/host, data source and data discovery, and data profiling, classification and security	✓ Automated SmartData Discovery and Classification (ASDAC) following Forrester Zero Trust Data Security Framework – involves indexing, profiling (using raw indexes) and metadata mapping – uses open source network discovery tool	✓ Extensive data discovery, profiling and repository – derived data stored in Model Repository and Profiling Warehouse with Search Index on top of metadata (but not the source data)
DATA QUALITY: Clean usable data in indexes and results	✓ Data quality addressed BEFORE any data is queried and read from source – data profiles used to develop data cleansing, transformation, standardization, and for standard data view mapping and data security	× No control - data quality addressed only on results AFTER a query on a data source is executed, limiting the success of the originating query and potentially requiring a wide-ranging query or multiple queries in an attempt to overcome data quality
ADVANCED QUERY PROCESSING: Consistent and multiple indexes and index types	✓ Full control over indexes and index types allowing a comprehensive range, including structured and unstructured, extracted entities, fuzzy match, and master data indexes, Link Indexes™ capturing data relationships within and across multiple sources, indexed views, etc.	× No control over indexes – limited by availability in data sources – not consistent over multiple data sources
QUERY LOAD: Almost no index or query load on data sources	✓ Own indexing, indexes and query processing separate from data sources absorb almost 100% of the load – small disk load when sequentially reading raw results data from data sources – distributed parallel processing – can enable extremely high scalability and performance	× 100% query load on data sources – typical BI, reporting and analytics queries impose a high load on mainly operational/transactional data sources not designed for such queries
QUERY PERFORMANCE: Unlimited query options and performance	✓ Almost all query processing in multi-tiered architecture of indexes, adapters and federation servers – can segment indexes, scale servers and use tiered media, e.g., SSDs, for index storage and query performance	× Completely dependent on data sources for query execution – cache attempts to overcome this bottleneck, however, introduces other problems as described later
QUERY ACCELERATION: Pre-aggregated, pre-calculated and pre-joined views	✓ Own indexed views provide pre-aggregated, pre-calculated and pre-joined views – can be materialized or virtual – can also be hierarchical – can also be maintained in near real-time – can also be manually or automatically generated at any time – can also be automatically used in a query with no need to specify	× No own indexed views – limited by availability in data sources – not consistent over multiple data sources

Comparison between SDF and Informatica® (2 of 3)

Feature	SmartData Fabric®	Informatica®, similar to other conventional data virtualization vendors
MULTIPLE DATA SOURCES: Almost any data source	✓ Almost any including mainframes, databases, files, logs, office docs, applications, email, Web docs, social media, Big data, streaming, Cloud and IoT	✓ Wide range of data sources
UNSTRUCTURED DATA: Full text search	✓ Advanced text search developed as part of work with Web and internal search engines and an eDiscovery tool on both structured and unstructured data – can combine with structured SQL queries	× Only if supported by data source
TEXT ANALYTICS	✓ Uses third-party/open source entity extraction, categorization and other analytics	×
ACTIVELY MONITOR DATA SOURCES AND EVENT PROCESSING	✓ Index updates allow, whether near real-time or less frequently – can combine with BPM workflows and can monitor KPIs that drive near real-time operational dashboards	× Have to constantly or regularly poll data sources
LINK MAPPING AND ANALYSIS/DATA MINING	✓ Built-in graph database/link mapping and link analysis that provides other benefits such as Master Data Management (MDM) support, query acceleration and ontological/semantic model representation	(✓) Has some built-in graph database/link mapping and link analysis capabilities – unclear as to what role beyond just visualization
MDM: Seamless and automatic integration of master data with data access	✓ MDM is an integral part of the SmartData Fabric – application queries can be automatically directed to associated indexes and master data can be automatically substituted for associated operational/transactional data – master data can be stored anywhere, but SmartData Fabric allows it to be distributed to adapters for security, privacy and performance reasons	✓ MDM is a centralized system that exists in the data infrastructure layer – it is unclear as to how seamless and automatic the integration with data access it is, but it appears to be well integrated – unclear as to how MDM would operate across multiple, distributed Informatica instances throughout the enterprise
RESULTS WHEN DATA SOURCES UNAVAILABLE	✓ If data is indexed, indexes can be inverted to reconstruct result rows of data as per the last index update – can be an automatic default when data sources are not available – can be used as a query optimization – indexes can be used as compressed, standards-based, SQL-compatible storage for IoT data on the edge, for example, which can be further enhanced by just storing aggregated data	× Only results data in cache

WHAMTECH®

Comparison between SDF and Informatica® (3 of 3)

Feature	SmartData Fabric®	Informatica®, similar to other conventional data virtualization vendors
DATA SOURCES MAINLY UNAWARE OF QUERIES	✓ Important in a secure environment and/or to prevent others knowing what data is being sought	×
ARCHIVE SOLUTIONS	✓ Can enable changed data storage in addition to live/operational data access or use indexes as storage as described above – can also persist all or just changed data in any target system, e.g., data warehouse, Big Data and Cloud	✓ Can persist data in any target system
MINIMAL IMPLEMENTATION TIME	✓ As can have complete control over data, indexes and query processing, it is fairly straightforward to deploy adapters – time consuming efforts are usually for data quality and changed data capture, but adapter templates and shared metadata repositories improve enterprise-wide implementations	× Conventional federated adapters do not take long to deploy, however, query optimization to overcome data source limitations can take considerable time
ROW, COLUMN AND DATA ELEMENT SECURITY	✓ In more secure environments, this can be important	✓ In more secure environments, this can be important and presume that Informatica® can address
NO ADDITIONAL STORAGE	× Indexes do require storage	× Query results cache and the data infrastructure layer do require storage – unclear as to how much compared to the SmartData Fabric
NO NEED FOR INDEX UPDATE PROCESS	×	✓
CACHE: REFRESH RATES	✓ Near real-time updates to indexes with options to account for latency	× Latency and dirty reads for results data
CACHE: DATA SECURITY AND PRIVACY	✓ Leave data in sources and impose gatekeeper RBAC security, even on data sources that do not have indexes, e.g., EIQ ConventionalAdapters	× Multiple copies of potentially high risk data (PHI, PCI, etc.)
CACHE: COMPLEXITY	✓ Relatively simple approach, as access to data, indexes, queries and results are 100% under the control of the adapters	× Everything becomes more complex, starting with query optimization and then cache optimization

The End