SmartData Fabric® (SDF)

aka Distributed Data and Master Data Virtualization and Management

Brief Technical Overview

March 2024

Copyright 2024 WhamTech, Inc.

Top ten data PROBLEMS facing almost all enterprises

1. **Data is everywhere** in multiple locations/countries, sources, source types, formats and platforms
2. **Cannot/will not copy/move ALL data to a single location**, e.g., multi-jurisdictional/data sovereignty and third-party ownership
3. **Some important data is not readily useable as is** and needs discovery, identification, classification, security, cleansing, transformation, standardization and analytics to support an **enterprise standard semantic layer**
4. **Some raw data needs to be aggregated to be useful to reporting, BI and analytics**
5. **Data needs to be connected within and across data sources**
6. **Data needs to be integrated through entity normalization/deduplication**, e.g., ELT, ETL to a DW, or MDM
7. **Some key performance indicators (KPIs) need to be calculated, monitored and acted on in near real-time**
8. **Many data sources do not have the query capabilities or load capacity** for external queries by conventional data virtualization/federated data access/query engines
9. **Large investments in existing systems and solutions**
10. **Need for rigorous enterprise-wide access control, and data governance, security and privacy**

Copyright 2024 WhamTech, Inc.

Conventional data solutions choice between a rock and a hard place

ROCK =
CONVENTIONAL DATA VIRTUALIZATION/ FEDERATION

Pros

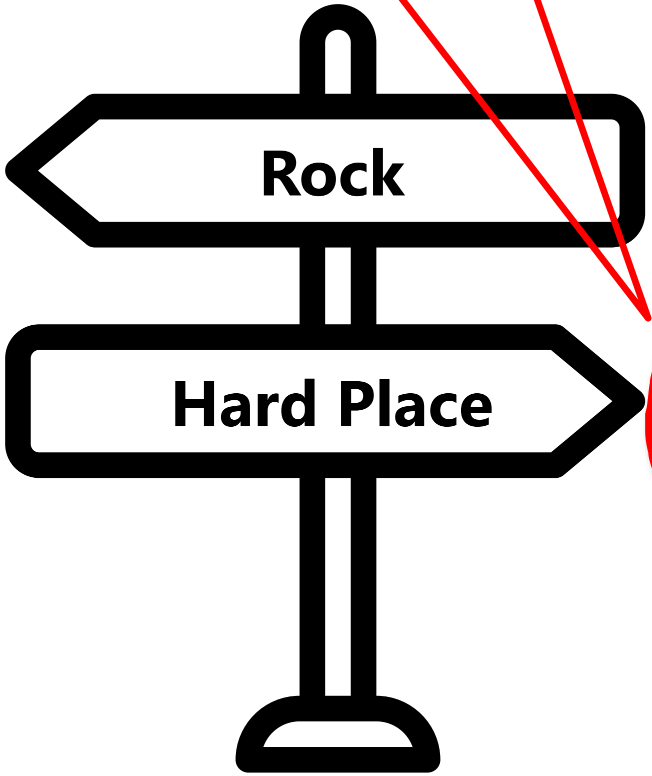
- Leave data where it is
- Real-time data access
- Easy to add or remove data sources
- Flexible schema on read
- Meet some compliance and on-soil data retention regulations

Cons

- Source data quality
- Source query capabilities
- Source query load
- Access control and data security
- Master Data Management (MDM) integration
- Two-tier architecture

100% dependent on data sources

ALL data needs to be copied and transformed



HARD PLACE =
DATA WAREHOUSE

Pros

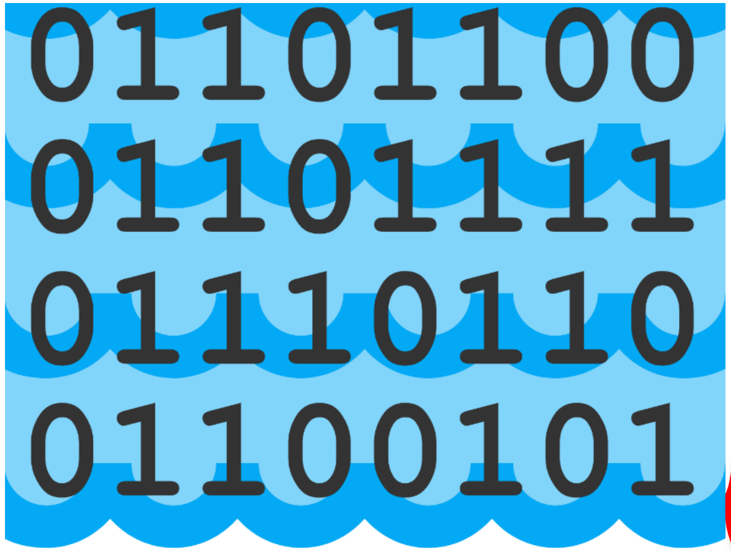
- Independent data quality
- Independent and high performing query capabilities
- Single database for queries
- Deduplicates/normalizes data

Cons

- All data needs to be copied
- Source schema transforms to a one-size-fits-all target schema during ETL
- Most introduce data latency
- Takes significant time and cost
- Inflexible schema on write
- Difficult to add or remove data sources
- Difficult to deal with personal data for CCPA/CCPR, GDPR, etc.
- Many cases, needs additional data marts
- Does not meet on-soil data retention regulations

Copyright 2024 WhamTech, Inc.

Data lake is in-between a rock and a hard pace



ALL data needs to be copied and transformed

IN-BETWEEN A ROCK AND A HARD PLACE = DATA LAKE

Pros

- All data in a single location or system
- Can leave schema and data as per data sources (some data lakes cannot)
- Helps with IT issues of access control, scalability, and query processing, performance and load
- Easy to add or remove data sources

Cons

- All data needs to be copied
- Does not help with data management => still requires ETL to a data warehouse/materialized views and then data marts, ELT or ***an additional data management system***
- Most introduce data latency
- New-meets-old market solutions of operational data store + ETL + data warehouse (+ data marts?), e.g., Cloudera Data Platform and Snowflake
- Difficult to deal with personal data for CCPA/CCPR, GDPR, etc.
- Does not meet on-soil data retention regulations

WHAMTECH®

SmartData Fabric® is plug-and-play **data management** that leverages, complements and is agnostic to existing systems

Almost ANY and MULTIPLE data sources on ANY and MULTIPLE platforms, including:

Unconventional data sources

- files
- logs
- office docs
- email
- Web docs

Conventional data sources

- mainframes
- databases
- applications/SaaS
- Big Data
- data lakes/warehouses/marts
- cloud storage

Unconventional data sources

- social media
- streaming
- IoT

DATA +

Forrester Zero Trust Data Security Framework

discovery
indexing
identification
classification
security
cleansing
entity extraction
transformation
standardization
governance
access control
lineage
updates in near real-time
virtualization
federation
[relationships/links]
master data management (MDM)
integration
supports standard applications, including reporting, BI and analytics
actionable data catalog
near real-time monitoring and event processing
interoperability/write-back
[virtual graph database that supports link analysis and graph visualization]

Indexing for data profiling, preprocessing and query execution **ONLY ON DATA AND/OR DATA SOURCES THAT NEED IT** - much/most data may not, e.g., transaction data

SmartData Fabric® solution

Distributed data virtualization and management software that:

- **Leaves data where it resides** in sources and/or local/regional data lakes/warehouses
- **Leverages the power of independent indexes with federated data adapters** to
  - discover, identify, classify, secure, cleanse, transform, standardize, analyze and connect data used to build and update indexes, indexed views for pre-aggregations, pre-calculations and pre-joins, and data and business objects
  - present multiple standardized views of data as an enterprise semantic layer
  - generate/use, and seamlessly and automatically integrate master data to enable data integration
  - execute uniform SQL queries from standard apps across all federated data adapters
  - either read data from indexes, or read pointers from indexes and use to read, cleanse, transform, standardize and secure raw results data from sources, and
  - provide integrated clean, standardized, secure, accurate and complete results data to standard apps
- **Uses open-source PostgreSQL** – keep costs low, access to latest developments and avoids vendor lock-in
- **Leverages, complements and agnostic to existing systems**, data sources, data virtualization software and query engines – avoids vendor lock-in
- **Enforces advanced enterprise-wide access control**, and data governance, security and privacy – Zero Trust Data Security

Addresses data issues BEFORE first query is made

Unique to SDF

Top twelve SmartData Fabric® DIFFERENTIATORS – most, unique

1. **Leaves data where it is, but PRE-PROCESSES STRUCTURED AND UNSTRUCTURED DATA** for discovery, profiling, identification, classification, security, quality, standards and analytics, and master data management, through indexes and indexed views in data source-specific federated adapters
2. **Overcomes any data source limitations** by enabling advanced queries and ABSORBING UP TO 100% OF QUERY LOAD on indexes and indexed views in federated adapters, and external joins in federation servers
3. **Enables high-performance queries** in open-source PostgreSQL, Trino and other query engines
4. **Provides an enterprise-wide semantic layer** with metadata, standard data views, data and business objects, and master and reference data
5. **Provides advanced and uniform-across-all-data-sources SQL query capabilities**
6. **Fully integrates its own automatic near real-time Master Data Management (MDM)/Master Patient Index (MPI) or can use separate third-party system as source and/or target**
7. **Supports operations as well as analytics, meeting the requirements of an ideal data fabric/mesh** (ref. Rick van der Lans, R20/Consultancy) – also **supports digital transformation** and the automation of existing, and creation of new, processes/workflows
8. **Supports "live query" mode vs. the much more common "data extract" mode** for reporting, BI, analytics and other apps
9. **Enables federation at multiple levels** that allows three or more tiers for performance, responsibility and accountability, supporting data mesh and local data management
10. **Allows development of FHIR and other APIs** for data sources that do not have them or are less capable
11. **Imposes advanced enterprise-wide data governance, access control, and data security and privacy, on all data sources – Forrester Zero Trust Data Security Framework**
12. **Distributed analytics, including Distributed AI (DAI) and virtual graph database and link analysis** (to be migrated in future from legacy database to PostgreSQL)

Copyright 2024 WhamTech, Inc.

Comparison between conventional data solutions and SmartData Fabric®

Ideal Feature	Conventional data virtualization/query engines, e.g., Denodo, Dremio, Presto/Trino	Data warehouse, e.g., Oracle, Snowflake and Teradata	Data lake, e.g., Hadoop and variants, incl. MapR	SmartData Fabric® true data fabric/mesh
Leave data in sources in original schema/format – ownership, compliance and cost	✓	✘	✘	✓
Clean, transformed and standardized data, views, and data and business objects	✘	✓	✘	✓
Uniform high-performance SQL queries on independent indexes and views	✘	✓	✓ or ✘	✓
Avoid high query loads on source systems	✘	✓	✓	✓
Advanced access control, data governance and security, regardless of source	✘	✓	✘	✓
Avoid additional ETL to a data warehouse/mart/views or ELT in data lake	✘	✓	✘	✓
Easy to add/remove data sources and schema-on-read flexibility	✓	✘	✓	✓
Avoids data latency	✓	✘	✓ or ✘	✓
Integrated and automated Master Data Management (MDM)	✘	✓	✘	✓
Pre-process and query unstructured data/text as part of SQL	✘	✓ or ✘	✘	✓
Actively monitor data sources, process events and support interoperability	✓ or ✘	✘	✘	✓
Data/entity relationship mapping, and virtual graph database views	✘	✘	✘	[✓]

Data problems, and SDF differentiators and solution comparison

Data PROBLEMS facing almost all enterprises	SmartData Fabric® DIFFERENTIATORS	SmartData Fabric® solution
Data is everywhere in multiple locations/countries, sources, source types, formats and platforms	**Leaves data where it is – parallel, distributed/federated processing**	**Leaves data where it resides** in sources and/or local/regional data lakes/warehouses
Cannot/will not copy/move ALL data to a single location, e.g., data sovereignty and ownership		**Does not copy or move data from sources to a single location**
Some important data is not readily useable as is and needs discovery, identification, classification, security, cleansing, transformation, standardization and analytics	**Pre-processes and standardizes data** for discovery, profiling, identification, security, quality, standards and analytics, supporting an Enterprise Semantic Layer	**Leverages the power of independent indexes with federated data adapters** to - discover, identify, classify, secure, cleanse, transform, standardize, analyze and connect data used to build and update indexes, indexed views for pre-aggregations, pre-calculations and pre-joins, and data and business objects - present multiple standardized views of data as an enterprise semantic layer - generate/use, and seamlessly and automatically integrate master data to enable data integration - execute uniform PostgreSQL SQL queries from standard apps across all federated data adapters - either read data from indexes, or read pointers from indexes and use to read, transform, standardize and secure raw results data from sources, and - provide integrated clean, standardized, secure, accurate and complete results data to standard apps
Data needs to be connected within and across data sources	**Standardized data and pre-join views**, and, in the future, Link Indexes™ that support virtual graph database	
Data needs to be integrated through entity normalization/deduplication	**Pre-processed and standardized data enables high quality MDM/MPI in healthcare – automatic, distributed and integrated**	
Some raw data needs to be aggregated to be useful to reporting, BI and analytics	**Derived value indexes, and pre-aggregated, pre-calculated and pre-joined indexed views**	
Some key performance indicators (KPIs) need to be calculated, monitored and acted on in near real-time	**Business views based on KPIs can be maintained and monitored for event processing**, exposed as REST APIs, microservices, business process management and other apps	
Many data sources do not have the query capabilities or load capacity for external queries by conventional data virtualization/federated data access/query engines	**External processing for data, indexes, indexed views and queries** can be 100% independent of data sources - supports "live query" mode vs. much more common "data extract" mode, and multi-level federation	
Large investment in existing systems and solutions	**Data fabric/mesh adapters and federation servers are independently and individually configurable and accessible – highly flexible and fit where needed**	**Leverages, complements and agnostic to existing systems**, data sources, data virtualization software and query engines – avoids vendor lock-in
Need for rigorous enterprise-wide access control, and data governance, security and privacy	**Each adapter is a security gatekeeper to a data source, regardless of whether indexed or not, and has built-in data governance**	**Enforces advanced enterprise-wide access control**, and data governance, security and privacy – Zero Trust Data Security

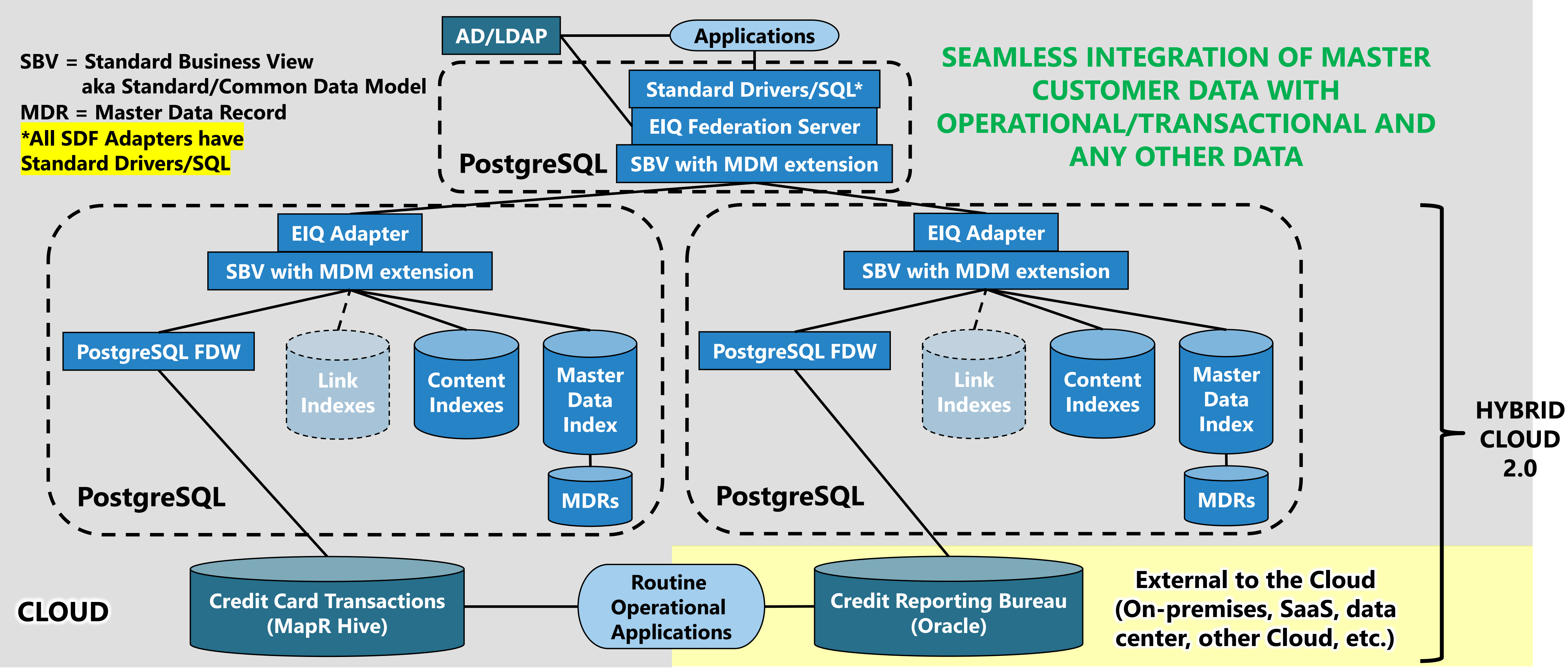
Before first query made

Unique to SmartData Fabric®

Example simple real-life cloud-based configuration

SBV = Standard Business View
aka Standard/Common Data Model
MDR = Master Data Record
*All SDF Adapters have Standard Drivers/SQL

SEAMLESS INTEGRATION OF MASTER CUSTOMER DATA WITH OPERATIONAL/TRANSACTIONAL AND ANY OTHER DATA



Copyright 2024 WhamTech, Inc.

Ideal Data Fabric by Rick van der Lans, R20/Consultancy

1. Data preparation, such as transformations, aggregations, filters and joins
2. Adaptable logic
3. high-performance
4. Data access by many data consumption forms
5. Access to all the data sources
6. Processing of all types of data
7. Data security and privacy
8. Real-time data access
9. Read and write data access
10. Data minimization
11. Event processing
12. Technical and business metadata management
13. Master and reference data management

SmartData Fabric® provides the above scalable data fabric capabilities, is agnostic to data virtualization software, query engines, data sources, types, formats, locations or platforms, and runs from a cloud, on-premises or hybrid cloud

Copyright 2024 WhamTech, Inc.

SmartData Fabric® value-add (1 of 2)

1. **Leaves data where it is in operational and other systems**, and continue to be used and updated by enterprise existing and new processes and applications
2. **Compute anywhere**, e.g., 100% in the cloud or multiple clouds, and/or on-premises, in data centers or third-party systems (with conventional adapters)
3. **Addresses data and data source issues upfront**, including data quality, standards, governance, highly advanced security and privacy, MDM and reference data, and query processing capabilities, performance, uniformity and load – on any and all data sources, regardless of their capabilities
4. **Provides high-performance integration capabilities** normally associated with data warehouses, by leveraging seamless and automatic MDM, and avoiding costly, time-consuming and inflexible ETL processes, and data latency/detachment from operations
5. **Lowers implementation and maintenance time and costs**, resulting in lower TCO and higher ROI

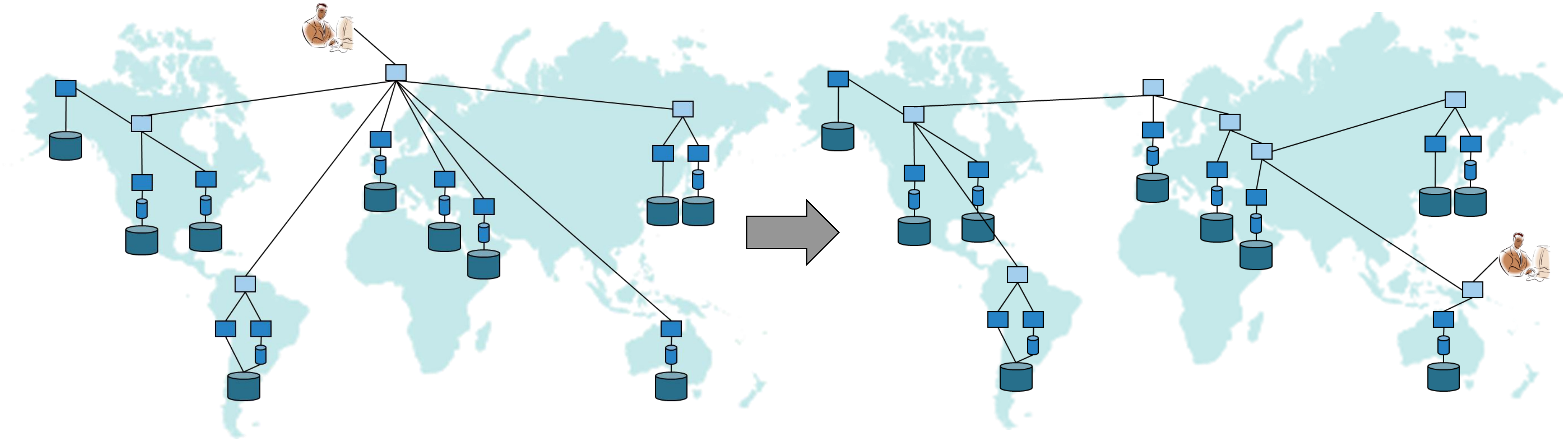
Copyright 2024 WhamTech, Inc.

SmartData Fabric® value-add (2 of 2)

6. **Provides near real-time data access**, for operations, reporting, BI, analytics and AI, and feedback from analytics to operations – an ideal, highly flexible data fabric
7. **Leverages, complements and agnostic to existing and future systems** – avoids vendor lock-in and future-proof
8. **Uses open-source PostgreSQL**, and other open-source query engines and tools
9. **Supports advanced, high-performance and uniform SQL across all data sources**, and multiple other query languages through PostgreSQL, including Python, R and GraphQL
10. **Provides unique capabilities** such as "live query" mode for reporting, BI, analytics and AI, and virtual graph database, link analysis and graph visualization

Copyright 2024 WhamTech, Inc.

SmartData Fabric® is highly configurable and changeable



Federation Server

Adapter

INDEXES

Data source

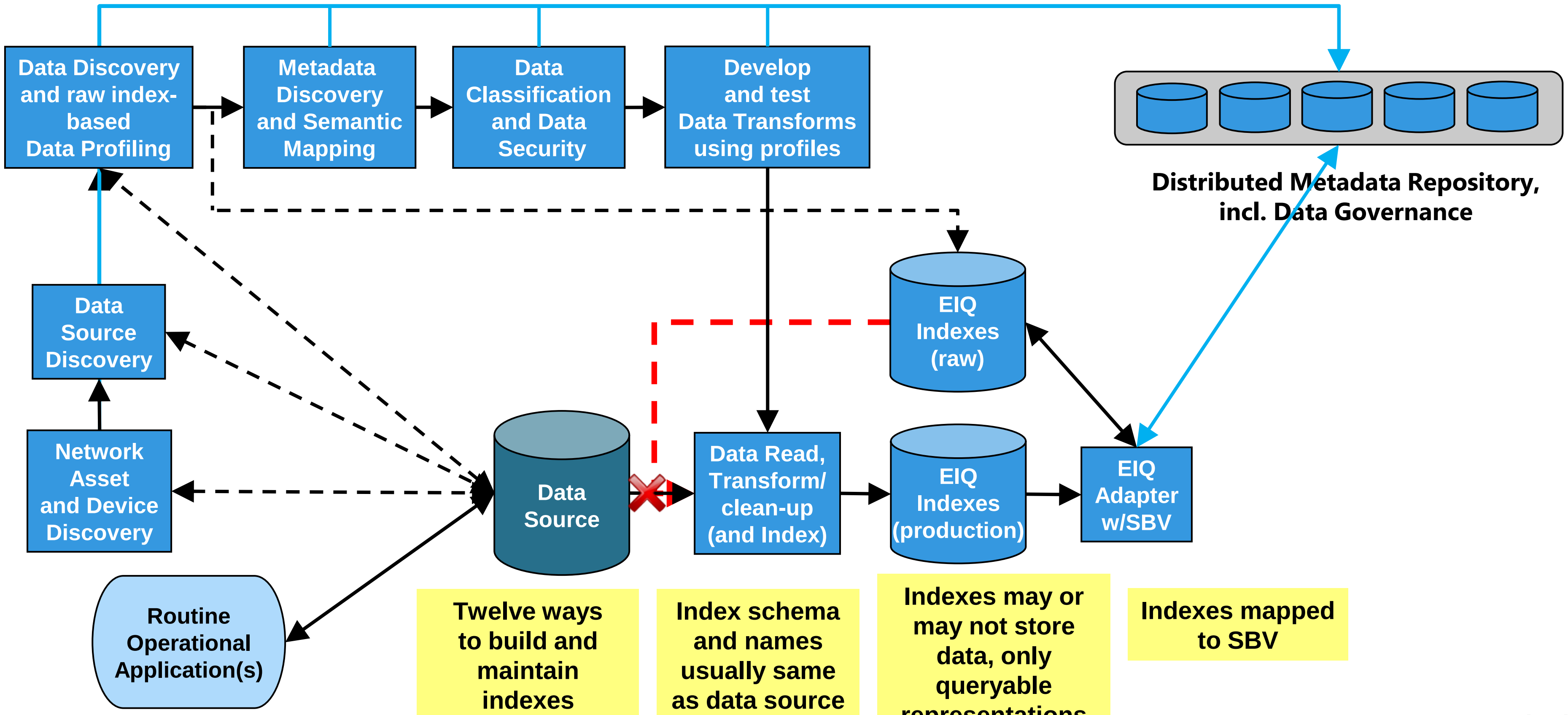
- Each Federation Server and Adapter is independently configurable, manageable and accessible PostgreSQL instance
- Runs anywhere, on almost any platform and in multiple locations
- Leverages, complements and agnostic to existing systems

What is Hybrid Cloud 2.0?

1. **Data is everywhere - leave it where it is**: On premise, on mainframes, in data centers, cloud(s), SaaS, third-parties, Web, social media, etc.
2. **Avoid deploying remote compute near or on data source system platforms**, as is the case with Hybrid Cloud 1.0
3. **Deploy SmartData Fabric® unconventional data virtualization 100% IN THE CLOUD**, leveraging index-based and conventional federated adapters for data-related pre-processing and query processing in and/or from the cloud – no need to install and run anything elsewhere, as is the case with Hybrid Cloud 1.0
   - Establish index update process through changed data capture (CDC)
   - Multiple CDC options, including near real-time (NRT)
4. **Focus on data that needs processing** for quality, standardization, security, relationship mapping and master data management (MDM) – various options for the rest of the data
   - Enable data management fundamentals
   - Address data, data source and access control issues
5. **Multiple configuration options**, including (a) some data indexed and the rest stays in the source, (b) all data indexed and stored in indexes, and (c) no data indexed and all queries on data source, with other options in-between
6. **Avoid incomplete or incorrect query results, query load and/or poor query performance of conventional data virtualization/federation**, i.e., avoid dependence on data sources, data in sources or data source own access control
7. **Immediate short-to-medium-term implementation**
8. **Facilitates medium-to-longer-term transition-migration to the cloud**

Copyright 2024 WhamTech, Inc.

Initial EIQ Indexed Adapter config, index build and business view mapping


Data Discovery and raw index-based Data Profiling
Metadata Discovery and Semantic Mapping
Data Classification and Data Security
Develop and test Data Transforms using profiles
Distributed Metadata Repository, incl. Data Governance
EIQ Indexes (raw)
Data Source Discovery
Network Asset and Device Discovery
Data Source
Data Read, Transform/ clean-up (and Index)
EIQ Indexes (production)
EIQ Adapter w/SBV
Routine Operational Application(s)

Twelve ways to build and maintain indexes

Index schema and names usually same as data source

Indexes may or may not store data, only queryable representations

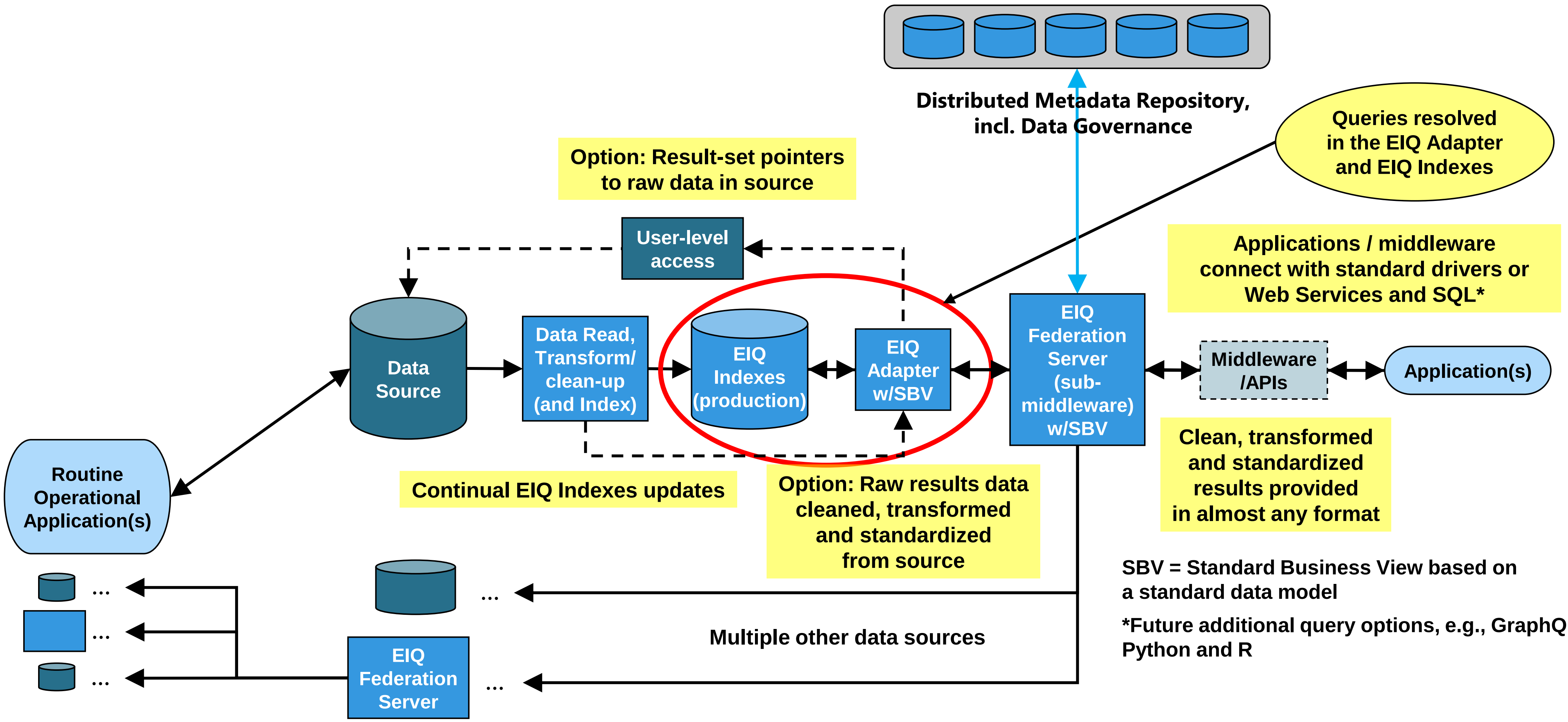
Indexes mapped to SBV

SBV = Standard Business View based on a standard data model

Alternate use of raw indexes to initially build production EIQ Indexes

Copyright 2024 WhamTech, Inc.

EIQ Indexed Adapter index update, query and results retrieval


Distributed Metadata Repository, incl. Data Governance
User-level access
Data Source
Data Read, Transform/ clean-up (and Index)
EIQ Indexes (production)
EIQ Adapter w/SBV
EIQ Federation Server (sub-middleware) w/SBV
Middleware /APIs
Application(s)
Routine Operational Application(s)
...
...
...
EIQ Federation Server
...
...
Multiple other data sources

Option: Result-set pointers to raw data in source

Queries resolved in the EIQ Adapter and EIQ Indexes

Applications / middleware connect with standard drivers or Web Services and SQL*

Continual EIQ Indexes updates

Option: Raw results data cleaned, transformed and standardized from source

Clean, transformed and standardized results provided in almost any format

SBV = Standard Business View based on a standard data model

*Future additional query options, e.g., GraphQL, Python and R

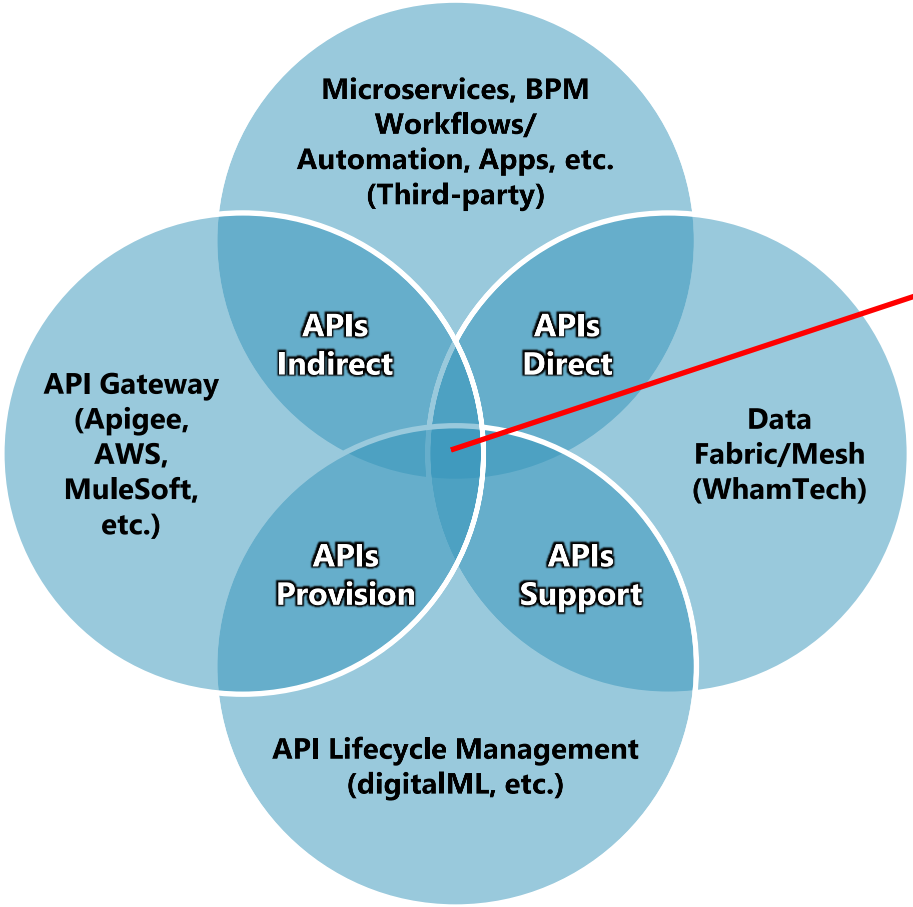
Copyright 2024 WhamTech, Inc.

Digital transformation is driven by access to the **latest clean, standardized and secure data** in **near real-time** through a **semantic layer**

Additional components may be **standard APIs** made available through an **API catalog/gateway**, **microservices with orchestration**, **standard apps and workflows** and **hybrid architectures**

Copyright 2024 WhamTech, Inc.

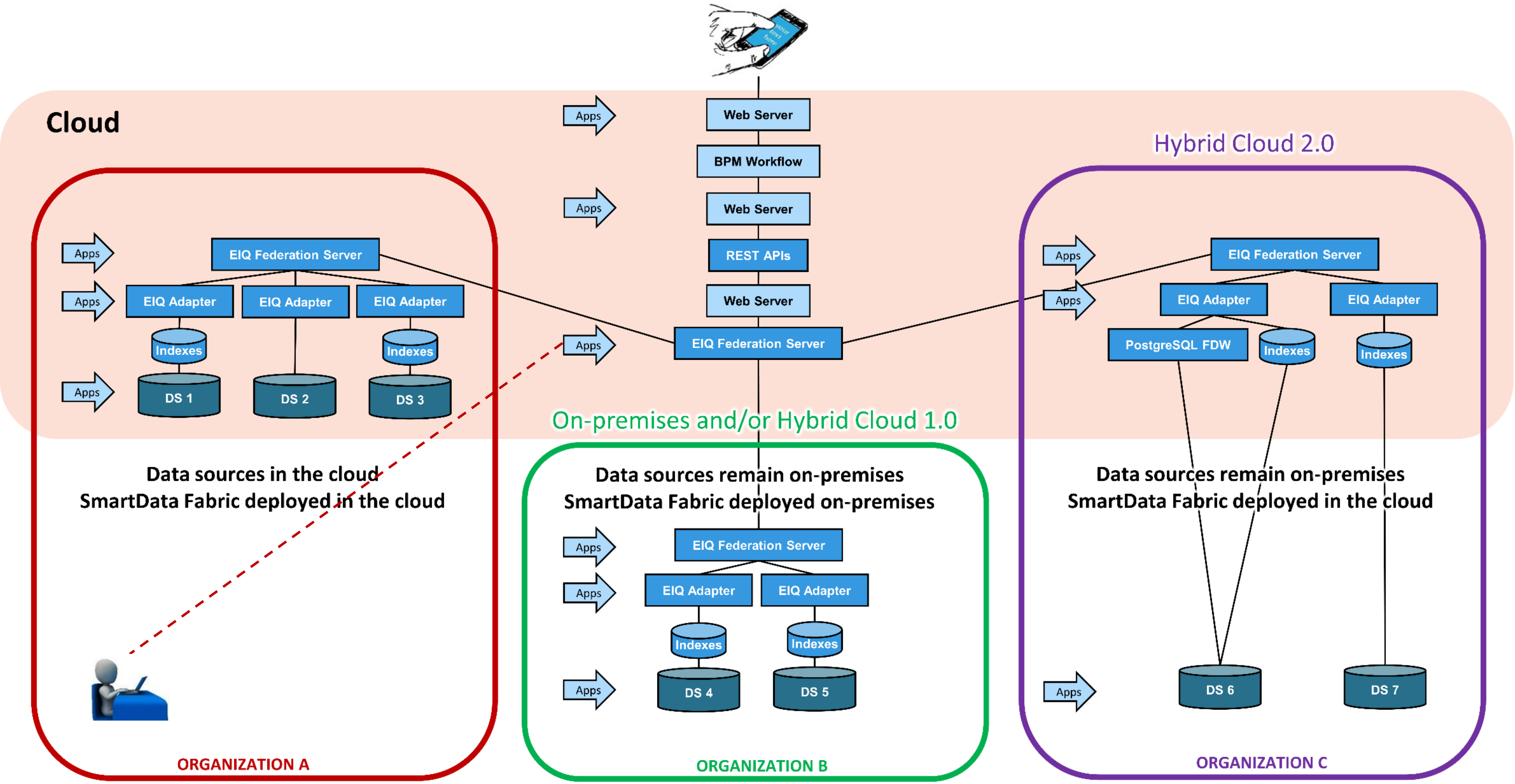
Complete API-based combination solution


Microservices, BPM Workflows/ Automation, Apps, etc. (Third-party)
API Gateway (Apigee, AWS, MuleSoft, etc.)
Data Fabric/Mesh (WhamTech)
API Lifecycle Management (digitalML, etc.)
APIs Indirect
APIs Direct
APIs Provision
APIs Support

Complete API Solution

Copyright 2024 WhamTech, Inc.

Example real-life smartphone CRM app interacting with BPM workflows with write-back to legacy systems through standard APIs/data services



Copyright 2024 WhamTech, Inc.

Example target markets

- Healthcare
- Banking
- Insurance
- M&A in all of the above

Copyright 2024 WhamTech, Inc.

SmartData Fabric® general use cases (1 of 2)

- **Data discovery, profiling, quality, standardization, governance and relationships mapping**
- **Advanced data access and data security** – seen as a security solution by defense contractors, AD/LDAP-based IAM, SSO, RBAC, TLS, RLS, CLS and multilevel classification for ANY data source regardless of its support for any of these security protocols
- **Multi-jurisdictional data and/or data residency compliance solutions**
- **Process improvement, re-engineering and automation, leading to digital transformation**
- **Virtual data warehouse and/or virtual data mart** - accesses and integrates multiple disparate data sources on multiple platforms in multiple locations in near real-time
- **Data lake + virtual data management + master data management** = clean and usable virtual data reservoir
- **Data provisioning for highly curated, self-serve reporting, BI and analytics** – bridge the gap between operations, reporting and BI, and analytics and other apps
- **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** - avoid centralization and conform to on-soil data retention regulations
- **Interactive REST APIs** on data sources that do not have or support them

Copyright 2024 WhamTech, Inc.

SmartData Fabric® general use cases (2 of 2)

- **Actionable data catalog or semantic layer** that can work with API and workflow catalogs – built-in metadata, data lineage and data governance, combined with AD/LDAP-based access control, data governance, security and privacy
- **Hybrid Cloud 2.0** where data sources remain wherever they reside, but run all compute in the cloud or data center - immediate short-to-medium-term implementation and an optional medium-to-longer-term transition-migration to the cloud
- **Other cloud use cases** – distributed data lake, multicloud, cloud + other data sources, and cloud-based M&A, access control and data security
- **Near real-time data source monitoring, event processing and Business Process Management (BPM)** to support digital transformation/process automation, develop new workflows, drive operational dashboards and support interoperability among apps
- **CCPA/CCPR, GDPR and future ADPPA solutions**, including multiple "erase" options
- **Virtual graph database and link analysis**, and interactive graph/link visualization
- **Enable STANDARDS** such as ODBC, JDBC, REST APIs and SQL, and run standard applications on data sources that may not support these
- **Versatile and flexible true data fabric/mesh**

➔ Discover, secure, clean, transform, standardize, query and deliver INTEGRATED structured, unstructured and semi-structured data from almost ANYWHERE to almost ANYWHERE in almost ANY FORMAT

Copyright 2024 WhamTech, Inc.

Example use case in banking

A large bank, which has a number of data lakes and data warehouses, stated that they want to:

a. **Leave data where it is and develop REST APIs that address data and master data management**, e.g., in data lakes, and data warehouses as well as eventual operational/transactional systems = no more or limited data lakes or data warehouses

b. Provide an **enterprise data catalog** for analysts and business users

c. Access data through **centralized data governance, and permissions-based access control and data security**

d. **View data and business objects that are standardized, understood and integrated through an API catalog**, to support customer 360-degree and other entity views, and business user-level self-serve reporting, BI and analytics

e. Implement a solution that is **scalable and high-performance**, with the goal of running most of the compute in the cloud – data remains where it resides, aka Hybrid Cloud 2.0, but, in addition, allow for Hybrid Cloud 1.0

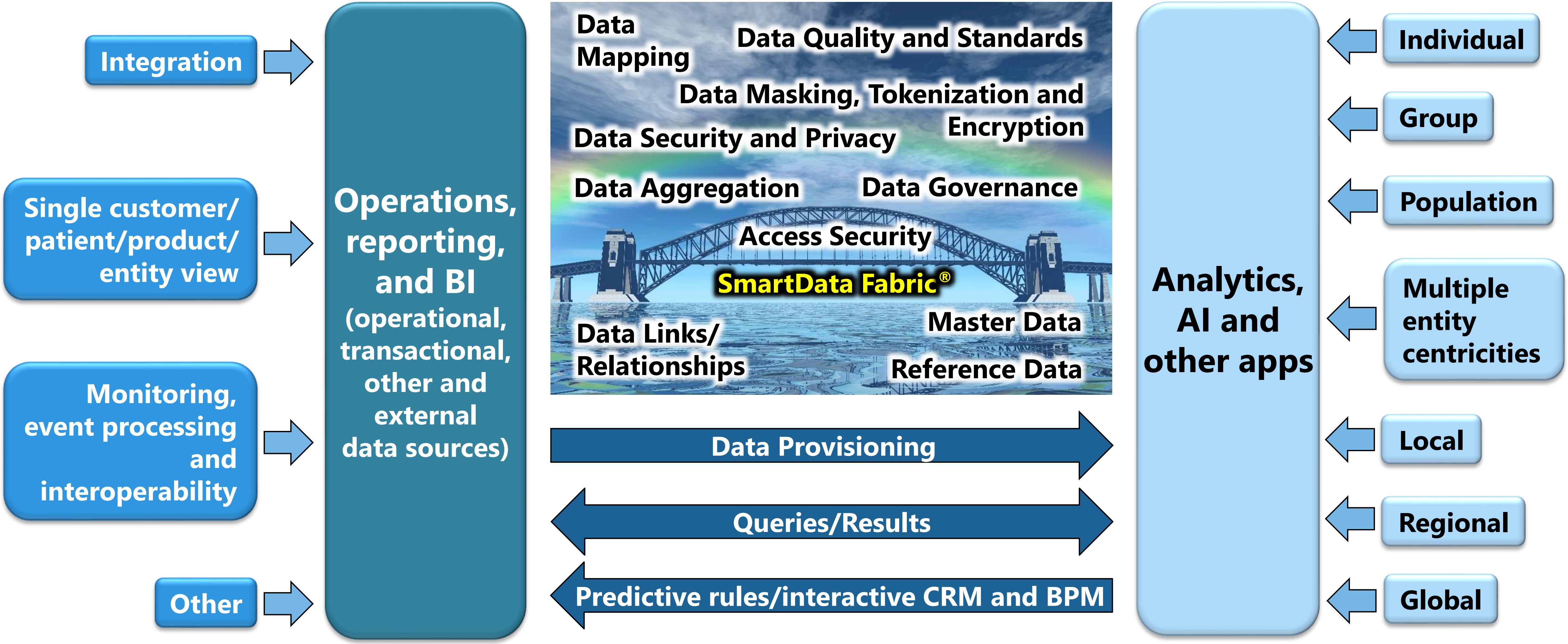
Copyright 2024 WhamTech, Inc.

Example use case in M&A

- Discover data sources available on network and associated data
- Semi-automatically profile, identify, classify and map data to standard business data names and objects/views
- Develop, typically, hybrid adapters to data sources that consist of indexed and non-indexed, aka conventional, adapters
- Make adapters available through REST APIs
- Run MDM across adapters to data sources and/or include existing MDM
- Enable an enterprise access control, data governance and data security layer, by leveraging AD/LDAP domain controller(s) – there may be more than one
- Make metadata available to users through an enterprise semantic layer accessed with AD/LDAP-based user credentials and associated security permissions
- Allow end-users to run reporting, BI and analytics apps in “live query” or “data extract” modes
- Allow workflows to interact with data sources through REST APIs using a combination of microservices and larger-scale workflows, e.g., apps and BPM software
- Integration with other M&A tools, VDRs and M&A platforms
- Incorporate User Behavior Analytics (UBA) and other cyber-related software into the SmartData Fabric

Copyright 2024 WhamTech, Inc.

SDF enhances operations, reporting, BI, analytics, AI and other apps, and bridges the gap between



Integration

Single customer/ patient/product/ entity view

Monitoring, event processing and interoperability

Other

Operations, reporting, and BI (operational, transactional, other and external data sources)

Data Mapping

Data Quality and Standards

Data Masking, Tokenization and Encryption

Data Security and Privacy

Data Aggregation

Data Governance

Access Security

SmartData Fabric®

Data Links/ Relationships

Master Data

Reference Data

Data Provisioning

Queries/Results

Predictive rules/interactive CRM and BPM

Analytics, AI and other apps

Individual

Group

Population

Multiple entity centricities

Local

Regional

Global

Copyright 2024 WhamTech, Inc.

Implementation steps

1. Identify a difficult and representative problem, e.g., for compliance
2. Develop a 60-day limited POC (30 days to gain access/prepare (may be reduced if on a cloud) and 30 days to implement and demo) – evaluation, development and test licensing cost-free
3. Install a single Windows VM on a remotely accessible development/test environment and access to data for POC – production versions run on Linux and Windows, containers, clouds, on-premises, etc.
4. POC -> prototype -> production – rapid and incremental implementation
5. Customer basic training 1.5 to 2 days, depending on skill level and experience
6. Team with a customer-preferred implementation partner (PIP)
7. Extend agile development environment to customer and PIP to expedite feedback and support
8. Production subscription licensing preferred, with various pricing models

Copyright 2024 WhamTech, Inc.

The End

Copyright 2024 WhamTech, Inc.

Topics covered in this presentation

- Top ten data problems facing almost all enterprises
- Conventional approaches to data management
- What is SmartData Fabric® (SDF)?
- SDF value-add
- SDF is close to an ideal data fabric/mesh that supports ALL ASPECTS of data management from operations to analytics, and near real-time interaction between them
- What is Hybrid Cloud 2.0?
- SDF indexed adapter processes
- SDF supports digital transformation – REST APIs, microservices, BPM/process automation software, enterprise semantic layer, etc.
- SDF is deployable at all levels, from (i) transparent single data source adapter/API, through (ii) enhancement to existing systems and applications, e.g., Snowflake and data lakes, to (iii) enterprise data management
- SDF target markets
- SDF use cases
- SDF enhances operations, reporting, BI, analytics, AI and other apps, and bridges the gap between

Copyright 2024 WhamTech, Inc.