

The "Zero-Draft" Protocol: Orchestrating Data Fabric and Posit for the New Age of Clinical Delivery Harsh Negi, MD

connect·share·advance

PHUSE Mumbai SDE 2026

phuse.global



Agenda

1. Zero draft protocol
2. Data Fabric
3. Orchestration Posit





connect·share·advance

Zero-Draft Protocol



- an initial, machine-readable digital representation of a clinical study that exists before the first word of a traditional PDF document is written.
- It is a "living" repository that contains all the fundamental logic required to build, execute, and report a trial.
- Unlike a traditional draft (which is a rough version of a text document), the Zero-Draft is a ****data asset**** that serves as the "Digital Twin" of the study design.

It is called "Zero" for three reasons:

1. ****Zero-Day Readiness:**** It allows programming and site-build activities to start on "Day 0" because the logic is already machine-readable.
2. ****Zero Manual Entry:**** It eliminates the need for data managers and programmers to manually "interpret" a PDF to build an EDC or SDTM map.
3. ****Zero-Latency Amendments:**** When the design changes in the Zero-Draft shell, the change propagates through the ****Data Fabric**** to all downstream systems (R code, EDC, Lab kits) instantly.



Anatomy of the "Zero-Draft" : Defining the Initial Shell

connect·share·advance

Structured Elements (Machine-Readable)	Unstructured Elements (Narrative)
Inclusion/Exclusion Criteria: Coded as boolean logic (e.g., <code><age> > 18</code>).	Overall Rationale: Clinical background and therapeutic justification.
Study Schedule: Metadata-tagged visit windows and assessments.	Medical Judgment: Investigator guidance on complex patient scenarios.
Statistical Methods: Defined endpoints and primary analysis parameters.	Efficacy Interpretation: Text-based discussion of expected results.
Endpoint Definitions: Coded derived variables for automatischer TFL generation.	Safety Narrative: Description of rare or nuanced Adverse Events.

Market presence*

Challengers **Contenders** **Strong Performers** **Leaders**

Stronger current offering

Weaker current offering

Weaker strategy Stronger strategy

Companies plotted:

- Denodo Technologies
- Talend
- Paxata
- IBM
- Oracle
- Cambridge Semantics
- Cloudera
- SAP
- Trifacta
- Hortonworks
- Informatica
- Hitachi Vantara
- TIBCO Software
- Syncsort
- Podium Data



Introduction to Modern Data Landscapes

connect·share·advance



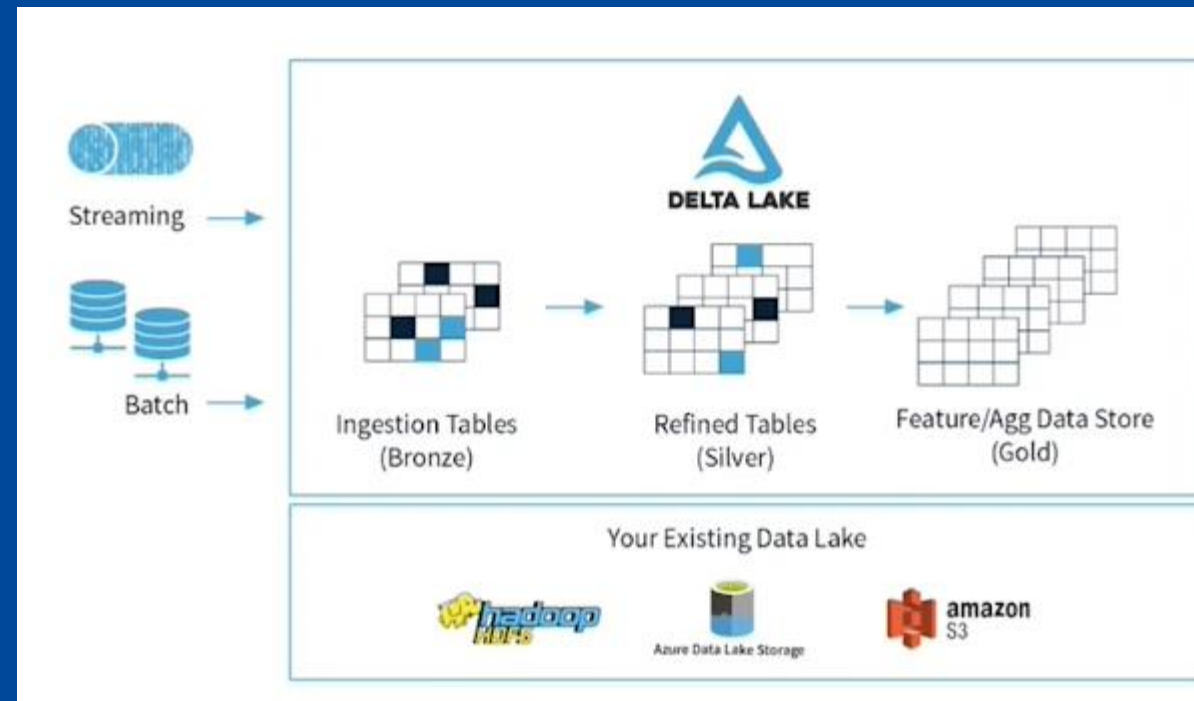
The Core Challenge: Most traditional tools require copying/moving data, leading to governance issues and data silos.



Delta Lake

connect·share·advance

- Data lake + Software layer
- ACID transaction
- Time travel(data versioning enable rollback , audit trail)
- Streaming and batch unification (using same tech)
- Schema enforcement
- Support Commands DELETE UPDATE and MERGE
- Performance improvement



<https://www.databricks.com>

/



What is a Data Fabric?

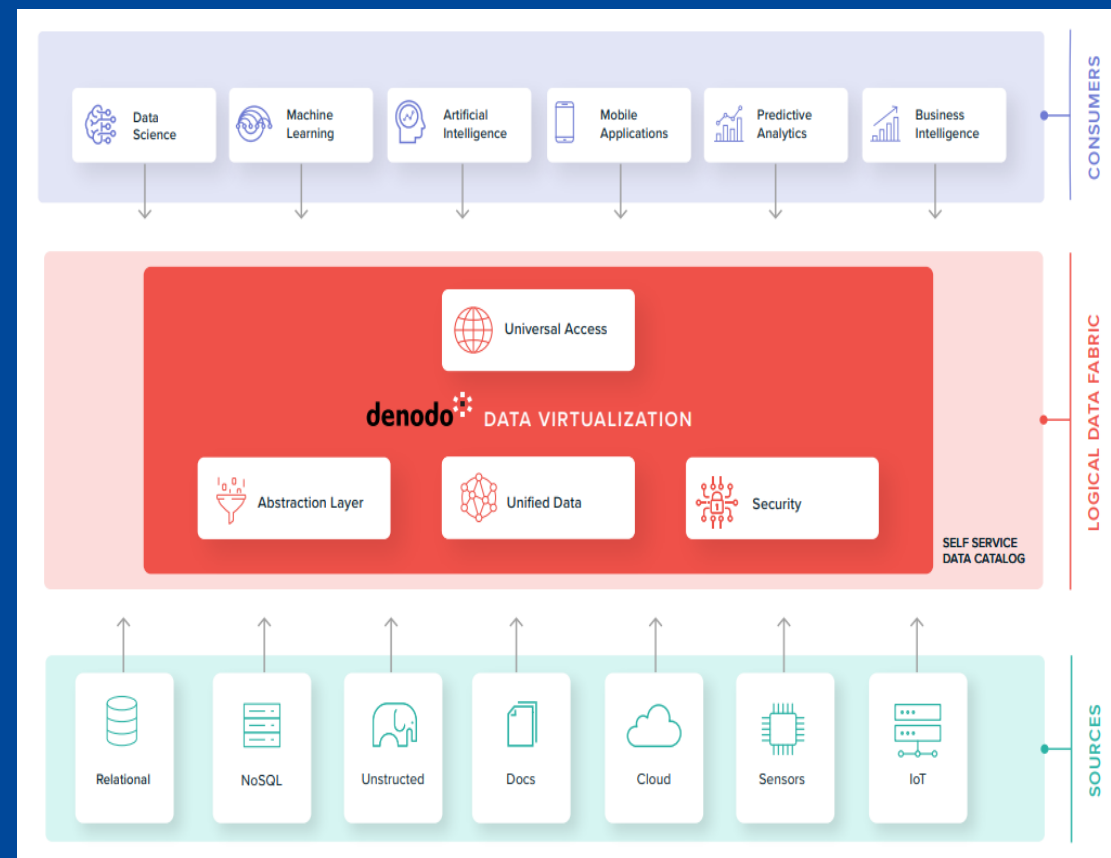
connect·share·advance

- **Definition:** An architectural approach and set of technologies designed to break down silos and deliver data to users.
- **Core Capabilities:** Enables accessing, ingesting, integrating, and sharing data across an enterprise in a governed manner.
- **Hybrid Environment Support:** Functions across on-premises systems and multiple public cloud environments.

Pillar 1 – Accessing Data

Pillar 2 – Managing the Data Lifecycle

Pillar 3 – Exposing Data for Use





Pillar 1 – Accessing Data

The Reality of Data Dispersion: Enterprises typically manage data across warehouses, lakes, RDMS and an average of 150+ SaaS applications

**Data
Virtualization**

**Integration &
ETL:**

- The Reality of Data Dispersion:** Enterprises typically manage data across warehouses, lakes, RDMS and an average of 150+ SaaS applications.
- Data Virtualization:** Accessing and aggregating data sources without the need to move or copy them into a new repository.
- Integration & ETL:** Using robust data pipelines and ETL tools when physical movement is necessary for latency or formal requirements.



Pillar 2 – Managing the Data Lifecycle

connect·share·advance

Governance and Privacy: Using active metadata to automate policy enforcement.

- **Security Measures:**

- Implementing role-based access control (RBAC).
- Automated masking and redaction of sensitive details.

- **Lineage and Quality:** Tracking data origin and transformations to assess overall quality.

Compliance: Ensuring adherence to global regulations like
GDPR,
CCPA,
HIPAA (Healthcare), and
FCRA (Finance).



Pillar 3 – Exposing Data for Use

- **Self-Service Access:** Providing data to
 - business analysts, data scientists, and developers via enterprise search catalogs.
- **Interoperability:**
 - Supporting multiple vendors and business intelligence (BI) platforms.
 - Integrating open-source technologies like R | Python and Spark.
- **Trustworthy AI:**
 - Operationalizing machine learning via ML Ops and monitoring results for bias and fairness and explainability.



connect·share·advance

Orchestration layer Posit





The orchestration : Posit R studio

connect·share·advance

Challenge in RWE	Data Fabric Solution	R-Programming Interaction
Data Fragmentation	Data Virtualization	dbplyr queries against a virtual layer.
Non-Standard Data	AI-Driven Metadata	{metatools} mapping to automated Fabric tags.
Compliance/Privacy	Federated Execution	Running R scripts in containerized "Safe Havens."

- **New Age Trials** and **RWE efficiency**—the **Data Fabric** is the architectural "connective tissue" that allows you to analyze Real-World Data (RWD) without the traditional nightmare of manual data moving and cleaning
- (<https://medium.com/@alexendrascott01/what-is-data-fabric-explained-with-real-world-examples-e5f427a32e89>)



Integrating the "Zero-Draft" with Data Fabric for Real-Time Delivery

connect·share·advance

Integration Step	What Happens?	The "Delivery" Advantage
1. Metadata Ingestion	The Fabric "reads" the Structured Draft (e.g., XML/Dataset-JSON).	Eliminates Manual Specs. No re-interpreting protocol logic.
2. Active Virtualization	The Fabric creates virtual views of EHR/EDC data based on the draft.	Zero Data Movement. Programming starts on Day 3 against live sources.
3. Contextual Retrieval (RAG)	AI pulls narrative elements (Unstructured Draft) when a programmer runs an analysis.	"Trust-but-Verify" Validation. Code logic is instantly traceable to the medical text.

"The Data Fabric makes the 'Zero-Draft' actionable. It bridges the gap between the medical blueprint and the programming execution, enabling a single point of truth across the global trial lifecycle."



Specifics of the Data fabric capability :1

connect·share·advance

Data Virtualization	Role	Keynote angle
Electronic Health Records (EHR) in one hospital, insurance claims in a payer's cloud, and pharmacy records in another.	A Data Fabric uses Data Virtualization to create a single logical view of these sources.	Instead of waiting 6 months for a "Data Dump" (ETL), your R programmers can query a "Virtual SDTM" layer that pulls live data from disparate sources on demand.
Active Metadata & Automated Mapping	Role	Keynote angle
Traditional RWD is "messy" and doesn't follow CDISC standards. A Data Fabric uses Active Metadata (AI-driven) to identify patterns in raw clinical data and suggest mappings.	It automatically tags data attributes (e.g., "SBP" in a physician's note vs. "SYS_BP" in a lab result) and aligns them to a common data model like OMOP or SDTM.	R packages like {metacore} can interface with the Data Fabric's metadata to automate the generation of mapping specifications.



Specifics of the Data fabric capability :2

connect·share·advance

Federated Governance and Privacy-Preserving Linkage	Role	Keynote angle
With strict regulations like India's DPDP Act (2023) and GDPR, you cannot always move patient data across borders.	The Fabric enables Federated Analytics. Your R-based analysis code is sent to the local data environment, executed there, and only the aggregated results (non-identifiable) are sent back.	This is how you achieve "Efficiency" in 2026—by bypassing the month long legal battle for data transfer through secure, federated linkage.



Specifics of the Data fabric capability :3

connect·share·advance

Semantic Knowledge Graphs for Hybrid Trials	Role	Keynote angle
A Data Fabric creates a Knowledge Graph that links a patient's "Trial Data" (the RCT part) with their "Real-World Journey" (the RWE part).	It provides the longitudinal context (e.g., the patient's history 5 years before the trial) that a simple database cannot	Use this to explain how we can now build Synthetic Control Arms with higher confidence, as the Data Fabric ensures the RWD patients are truly "comparable" to the RCT ones through linked clinical attributes.



The Role of USDM in the Integration Architecture..cont

connect·share·advance

Use case	Role	Improvement
Automated "Data Contract" Generation	USDM allows the "Zero-Draft" trial shell to export a precise Data Contract . This contract tells the Data Fabric exactly which variables are required for which visits.	Instead of a programmer manually telling the Data Fabric, "I need Blood Pressure from the EHR," the USDM file automatically "instructs" the Fabric's semantic layer to provision those specific virtual views.
Mapping RWE to Standards (USDM + Biomedical Concepts)	USDM integrates with Biomedical Concepts (BC) . This means a "Hemoglobin" assessment in your protocol is linked to a globally unique ID.	When your Data Fabric looks at "messy" Real-World Data, it uses the USDM/BC ID to find the correct match automatically. It bridges the semantic gap between a hospital's local code and your SDTM LBTESTCD without human intervention.



The Role of USDM in Your Integration Architecture

connect·share·advance

Use case	Role	Improvement
Real-Time Traceability for Risk-Based Validation	USDM creates a "Digital Thread" from the Endpoint in the protocol to the specific R code in Posit	this is the "Compliance Holy Grail." If a regulator asks, "Why was this RWE patient included?", your R session in Posit can trace the logic back through the Data Fabric directly to the Structured Eligibility Criteria in the USDM file



THANK YOU!