Prepared for market education · not a buying recommendation

AI Vendor Pulse

An advisory briefing on AI vendors selling into pharma & biotech

ISSUE 04 THE DATA & CLOUD FOUNDATION FOR CLINICAL AI

Snowflake
Data cloud + Cortex AI (Natoma)

Cloudera
Hybrid on-prem + cloud lakehouse

Databricks
Lakehouse + AI/ML platform

Palantir
Foundry ontology + AIP agents

Hyperscalers
AWS · Google · Azure health clouds

Some companies referenced may be current, former, or prospective advisory relationships of Avery & Riley; others are included independently for informational purposes. All data is publicly sourced unless otherwise noted.

WHY THIS ISSUE EXISTS

AI is only as good as the data layer beneath it — and that layer is repositioning.

Issue 01 mapped AI across the trial; Issue 02 the GxP backbone; Issue 03 the sensor edge. This issue goes to the **foundation** — the horizontal data clouds, lakehouses, ontologies, and hyperscaler health verticals that store, govern, and increasingly **reason over** clinical data. None are clinical-native; all are repositioning to host clinical AI.

WHAT SENIOR LEADERS SHOULD WATCH

01

The platforms are **converging on governed agentic AI** — every one now ships agents (Cortex, Mosaic, AIP, Bedrock). The race is governance, not capability; Snowflake's Natoma deal is the tell.

02

Horizontal breadth, variable clinical depth. These span most of the stack, but few own finished clinical apps or validated GxP fit. Know what is a building block versus a product.

03

The real questions are governance, residency, and validation. For regulated data, ask about lineage, audit, data sovereignty (cloud vs on-prem), and whether agentic access is inspectable.

THE MEASURING STICK

The Avery & Riley clinical data & AI platform stack

Coverage bars throughout this issue map each vendor against the ten layers below — a **directional, point-in-time read** of where a platform appears to operate, based on the latest publicly known data. These are not vendor-reported metrics and not a buying recommendation.

D01
Data ingestion & integration

D02
Storage & lakehouse

D03
Governance, lineage & access

D04
Health-data interoperability (FHIR / OMOP)

D05
Compute & analytics

D06
ML & MLOps

D07
Generative AI & agents

D08
Operational / application layer

D09
Security, compliance & validation

D10
Ecosystem, marketplace & data sharing



Directional reads from public materials & directory review — a snapshot, not vendor-reported metrics.

PLATFORM-STACK COVERAGE SNAPSHOT

Where each platform appears to play across the stack

Vendor	D01	D02	D03	D04	D05	D06	D07	D08	D09	D10	Coverage
Snowflake Data cloud + Cortex AI	■	■	■	■	■		■		■	■	8 /10
Cloudera Hybrid on-prem + cloud lakehouse	■	■	■		■	■	■		■		7 /10
Databricks Lakehouse + AI/ML platform	■	■	■		■	■	■			■	7 /10
Palantir Foundry Ontology + operational layer	■		■	■			■	■	■		6 /10
Hyperscalers AWS · Google · Azure health clouds	■	■	■	■	■	■	■		■	■	9 /10

Unlike Issue 03's focused biomarker vendors, these horizontal platforms span most of the stack — the question is depth and clinical fit, not breadth. ■ D07 generative AI & agents shown in gold: every platform is lit — the frontier they are all racing toward. A snapshot of the latest publicly known data and Avery & Riley records.

THE SENIOR-LEADER VIEW

Executive market map

VENDOR	PRIMARY CATEGORY	PROBLEM(S) TARGETED	PUBLIC OUTCOME SIGNAL
Snowflake	Data cloud + Cortex AI / governed agents	Siloed clinical / commercial / RWD; governed, identity-aware access for PHI & GxP; running AI on data without moving it.	**$10M+** productivity savings (AstraZeneca); **May 2026** Natoma → governed agentic access; **billions/day** data points (Komodo).
Cloudera	Hybrid (on-prem + cloud) data & AI platform	Data residency / sovereignty for regulated data; unifying on-prem + multi-cloud; running AI where data physically lives.	**25+ EB** under management; **2+ PB / 250** warehouses, 4× better eligibility prediction (IQVIA); **$22.8M** 3-yr NPV (pharma).
Databricks	Lakehouse + best-in-class ML / gen-AI	Unifying structured, unstructured & genomic data; large-scale ML on RWD / genomics; gen-AI on a governed source of truth.	**600×** genomic query speedup (Regeneron); **~60 min** to build an agent over 400k trial docs (AstraZeneca); **100×** faster genomics.
Palantir Foundry	Ontology + operational decision layer	Turning siloed data into a governed operational ontology; closing the loop from data → decision → auditable action.	**700+** lives saved / 83% faster placement (Tampa General); **100+/25k** trials/patients pooled; **60%** shorter IND/NDA authoring (Parexel).
Hyperscalers	AWS · Google · Azure health verticals	GPU-scale compute for genomics & clinical AI; FHIR/DICOM interoperability; foundation models & agents on PHI; global residency.	**55% / $750M–1B** infra savings (Pfizer/AWS); **2 days** to design COVID-19 sequence (Moderna/AWS); **2,500+** gen-AI use cases (Novo/AWS).

Prepared for market education; not a buying recommendation.

Snowflake

PLATFORM-STACK COVERAGE
8 /10

DATA CLOUD + CORTEX AI / GOVERNED AGENTS

Targets: *Siloed clinical / commercial / RWD; governed, identity-aware access for PHI & GxP; running AI on data without moving it.*

WHAT THEY DO

- The AI Data Cloud — unified warehouse + Iceberg lakehouse across AWS / Azure / GCP, with a Healthcare & Life Sciences Data Cloud grounded in HL7 / FHIR, OMOP & GxP.[1]
- Cortex AI (managed LLMs, Cortex Agents, Snowflake Intelligence) + Horizon governance; the May 2026 Natoma acquisition adds governed agentic access via an MCP gateway.[2]

OPERATIONAL PROBLEMS ADDRESSED

- Unifying siloed clinical, commercial, R&D and real-world data into one governed platform.
- Governed, identity-aware access for GxP / PHI-sensitive workflows.
- Running AI/ML and gen-AI on governed data without moving it.
- Secure cross-org data sharing (pharma–payer, pharma–CRO) via Marketplace & clean rooms.

PLATFORM-STACK COVERAGE — AVERY & RILEY SNAPSHOT

D01 · D02 · D03 · D04 · D05 · D07 · D09 · D10

D01	D02	D03	D04	D05	D06	D07	D08	D09	D10
■	■	■	■	■		■		■	■

$10M+[3]
Productivity savings — AstraZeneca (named customer)

May 2026[2]
Natoma acquisition → governed agentic data access (MCP gateway)

Billions/day[4]
Data points in Komodo Health's Healthcare Map on Snowflake

SENIOR-LEADER READOUT

Broad data-cloud coverage — eight of ten layers — strong on governance, AI, and marketplace; the Natoma deal is the cohort's clearest move to make agentic AI inspectable.

Named pharma proof is solid (AstraZeneca, Sanofi, Komodo); quantified "time-to-insight" pharma metrics remain largely qualitative.

Public claims; not independently validated. Not a buying recommendation.

Cloudera

HYBRID (ON-PREM + CLOUD) DATA & AI PLATFORM

PLATFORM-STACK COVERAGE
7 /10

Targets: *Data residency / sovereignty for regulated data; unifying on-prem + multi-cloud; running AI where data physically lives.*

WHAT THEY DO

- The self-described only hybrid data & AI platform — the CDP open lakehouse (Apache Iceberg) running identically on-prem, private cloud, and public cloud.[5]
- Cloudera AI + AI Inference Service (NVIDIA NIM) and Agent Studio for private, on-prem gen-AI and agents; SDX for unified governance & lineage.[6]

OPERATIONAL PROBLEMS ADDRESSED

- Keeping regulated / PHI data resident on-premises for sovereignty & residency.
- Unifying siloed on-prem + multi-cloud clinical, lab and production data.
- Running AI/ML where the data physically lives — bringing models to the data.
- Federated learning across decentralized data without moving records.

PLATFORM-STACK COVERAGE — AVERY & RILEY SNAPSHOT

D01 · D02 · D03 · D05 · D06 · D07 · D09

D01	D02	D03	D04	D05	D06	D07	D08	D09	D10

25+ EB[7]

Of data under management on CDP (vendor claim)

2+ PB / 250[8]

Warehouses consolidated; 4× better trial-eligibility prediction (IQVIA)

$22.8M[9]

3-yr NPV; genome analysis "80 years → weeks" (global pharma)

SENIOR-LEADER READOUT

Coverage spans the hybrid data foundation and on-prem AI — seven of ten layers; the differentiator is data sovereignty (on-prem inference + federated learning) versus cloud-only peers.

Headline scale figures (25 EB; "9 of top 10 pharma") are vendor claims; IQVIA is the strongest named case.

Databricks

PLATFORM-STACK COVERAGE
7 /10

LAKEHOUSE + BEST-IN-CLASS ML / GEN-AI

Targets: *Unifying structured, unstructured & genomic data; large-scale ML on RWD / genomics; gen-AI on a governed source of truth.*

WHAT THEY DO

- The Data Intelligence Platform / lakehouse — Delta Lake storage, Unity Catalog governance, MLflow, and Mosaic AI / Agent Bricks; a vertical Lakehouse for Healthcare & Life Sciences.[10]
- Open accelerators for genomics (Glow, with Regeneron) and HL7 (Smolder); best-in-class large-scale ML / MLOps.[11]

OPERATIONAL PROBLEMS ADDRESSED

- Unifying structured, unstructured and genomic / EHR data into one governed lakehouse.
- Large-scale ML on clinical, RWD and biobank-scale genomic data.
- Building gen-AI / agents on the governed source of truth, not copies.
- Central governance of data + models + agents via Unity Catalog.

PLATFORM-STACK COVERAGE — AVERY & RILEY SNAPSHOT

D01 · D02 · D03 · D05 · D06 · D07 · D10

D01	D02	D03	D04	D05	D06	D07	D08	D09	D10
■	■	■		■	■	■			■

600×[12]
Genomic query speedup, 30 min → 3 sec (Regeneron, named)

~60 min[13]
To build an agent over 400,000+ trial documents (AstraZeneca)

100×[11]
Faster genomics vs prior solutions (vendor claim)

SENIOR-LEADER READOUT

Coverage centers on the lakehouse plus best-in-class ML and gen-AI — seven of ten layers; the deepest data-science / ML bench in the cohort.

Regeneron's 600× is the strongest named result; the 100× genomics figure is a vendor claim.

Public claims; not independently validated. Not a buying recommendation.

Palantir Foundry

PLATFORM-STACK COVERAGE
6 /10

ONTOLOGY + OPERATIONAL DECISION LAYER (AIP)

Targets: *Turning siloed data into a governed operational ontology; closing the loop from data → decision → auditable action.*

WHAT THEY DO

- Foundry — an operational data platform built on an Ontology (objects, links, and write-back actions): a "digital twin of the enterprise" where data, logic and decisions live together.[14]
- AIP layers ontology-grounded agents that retrieve typed objects and act only through governed Action Types — auditable, permissioned enterprise changes.[15]

OPERATIONAL PROBLEMS ADDRESSED

- Collapsing operational / data silos into one governed operational ontology.
- Closing the loop from data → decision → governed write-back action.
- Harmonizing historical trial data for pooled / meta analyses.
- Deploying agents that take auditable, permissioned clinical-ops actions.

PLATFORM-STACK COVERAGE — AVERY & RILEY SNAPSHOT

D01 · D03 · D04 · D07 · D08 · D09

D01	D02	D03	D04	D05	D06	D07	D08	D09	D10
■		■	■			■	■	■	

700+[16]

Lives saved; 83% faster patient placement — Tampa General (named)

100+/25k[17]

Immunotherapy trials / patients pooled for oncology analysis (US agency)

60%[18]

Shorter IND/NDA authoring — Parexel + Weave Bio (partner claim)

SENIOR-LEADER READOUT

Strength is the operational + action layer on a governed ontology — six of ten layers; it uniquely turns analysis into governed action, not dashboards.

Tampa General is the strongest named outcome set; the Parexel 60% is a partner claim. A Life-Sciences QMS supports GxP use.

Public claims; not independently validated. Not a buying recommendation.

Hyperscalers

AWS · GOOGLE · AZURE HEALTH CLOUDS

PLATFORM-STACK COVERAGE
9 /10

Targets: *GPU-scale compute for genomics & clinical AI; FHIR / DICOM interoperability; foundation models & agents on PHI; global residency.*

WHAT THEY DO

- AWS for Health (HealthLake FHIR, HealthOmics, Bedrock), Google Cloud (Healthcare API, Vertex AI, Med-Gemini / MedGemma) and Microsoft Azure (Health Data Services, Healthcare Agent Service, Azure OpenAI, Dragon / DAX Copilot).[19]
- Each wraps elastic compute + a FHIR / DICOM health-data service + a foundation-model / agent layer, sold as a life-sciences vertical.[20]

OPERATIONAL PROBLEMS ADDRESSED

- Elastic, GPU-scale compute for genomics, protein structure and clinical-AI training.
- FHIR / HL7 / DICOM interoperability to normalize fragmented clinical / EHR data.
- Foundation models + agents operating securely on PHI.
- Global infrastructure for data residency; HIPAA / HITRUST out of the box.

PLATFORM-STACK COVERAGE — AVERY & RILEY SNAPSHOT

D01 · D02 · D03 · D04 · D05 · D06 · D07 · D09 · D10

D01	D02	D03	D04	D05	D06	D07	D08	D09	D10

55% / $750M–1B[21]
Lower infra cost / annual savings — Pfizer PACT on AWS (named)

2 days[22]
To design the COVID-19 vaccine sequence — Moderna on AWS

2,500+[23]
Generative-AI use cases on Amazon Bedrock — Novo Nordisk

SENIOR-LEADER READOUT

The broadest stacks in the cohort — nine of ten layers — any modality, any workload, global residency; the foundation the other vendors often run on.

Strongest named proof sits with AWS (Pfizer, Moderna, Novo Nordisk); depth in finished clinical apps (D08) is where Palantir differentiates.

Public claims; not independently validated. Not a buying recommendation.

CROSS-VENDOR PATTERNS FOR SENIOR LEADERS

Three patterns to carry into your diligence

01

Everyone is now an agentic-AI platform

The differentiation has moved from data storage to governed agent access — the gold column shows all five lit on generative AI & agents.

02

Breadth versus depth

Hyperscalers and data clouds are broad; Palantir goes deep in the operational / clinical layer; Cloudera owns the on-prem sovereignty wedge.

03

Governance is the moat

Lineage, identity-aware access, and auditability are what make a horizontal platform usable for regulated clinical AI — not raw capability.

THE BOTTOM LINE FOR ADOPTION

"These are the rails, not the train. The question is not which platform is most capable — it is which can host clinical AI that is governed and validated enough to trust with patient data."

HOW TO USE THIS ISSUE

From market read to executive diligence

EXECUTIVE DILIGENCE QUESTIONS

- Is this a foundation (data + governance + compute) or a finished clinical application — and which do you actually need?
- How is data governed — lineage, identity-aware access, audit — and is agentic access inspectable?
- What is the residency / sovereignty story — cloud-only, hybrid, or on-prem — for your regulated data?
- Can the platform's AI be validated for GxP use, and what is the named-customer (not vendor-claim) evidence?

SUGGESTED CATEGORIES FOR FUTURE ISSUES

- Real-world data & evidence platforms
- Clinical data management & EDC
- Federated learning & privacy-preserving analytics
- MLOps & model governance for regulated AI
- Data interoperability & FHIR / OMOP tooling

Source standard. *All data is publicly sourced unless otherwise noted. Public claims are summarized from vendor websites, customer stories, press releases, analyst reports, or industry-publication summaries and should not be treated as independently validated performance guarantees.*

REFERENCES & SOURCES

References

Reference numbers are assigned by first appearance in the body and shown as sequential superscripts throughout this issue.

1 Snowflake, "Healthcare & Life Sciences AI Data Cloud," company website.

2 Snowflake, "Snowflake to acquire Natoma — governed agentic access," blog & press release (May 27, 2026).

3 Snowflake, AstraZeneca customer case study (productivity savings).

4 Business Wire / Snowflake, Komodo Health "Healthcare Map" launch.

5 Cloudera, "Hybrid data & AI platform" / CDP open lakehouse, company website.

6 Cloudera, AI Inference Service (NVIDIA NIM) & Agent Studio blog posts.

7 ISG analyst perspective, "Cloudera evolves its open data lakehouse" (25+ EB).

8 Cloudera, IQVIA customer case study.

9 Cloudera, global-pharmaceutical-company case study ($22.8M NPV).

10 Databricks, "Lakehouse for Healthcare & Life Sciences," company website.

11 Databricks, genomics / Glow materials & "100× faster genomics" press release.

12 Databricks, Regeneron customer case study (600× speedup).

13 SiliconANGLE / Databricks DAIS 2026 coverage (AstraZeneca agent).

14 Palantir, "Foundry Ontology overview," documentation.

15 Palantir, "AIP overview," documentation.

16 PR Newswire, Tampa General Hospital & Palantir partnership / Sepsis Hub results.

17 Palantir, "Clinical Data R&D" impact page (pooled oncology trials).

18 Applied Clinical Trials, Parexel–Palantir / Weave Bio collaboration (IND/NDA authoring).

19 AWS for Health, Google Cloud Healthcare & Life Sciences, and Microsoft Azure Health Data Services product pages.

20 Vendor health-vertical overviews — HealthLake / HealthOmics, Cloud Healthcare API / Vertex AI, Microsoft Cloud for Healthcare.

21 AWS, Pfizer "PACT" generative-AI case study (infrastructure savings).

22 AWS, Moderna case study (COVID-19 vaccine sequence design).

23 AWS, Novo Nordisk generative-AI on Amazon Bedrock case study.

Disclosure: Some companies referenced may be current, former, or prospective advisory relationships of Avery & Riley; others are included independently for informational purposes. Platform-stack coverage reads use Avery & Riley's vendor framework compiled from public materials and directory review; they are a snapshot as of the latest publicly known data and may change as vendors add, discontinue, or reposition capabilities. "Hyperscalers" combines AWS, Google Cloud, and Microsoft Azure healthcare/life-sciences verticals.