DataHub

Enterprise AI Data Catalog That Eliminates Data Chaos

AI-powered discovery, governance, and observability unify across your data estate to deliver data quality, compliance, and AI readiness.

HashiCorp · dpg media · Etsy · chime · DOORDASH · FIS · Visa

25%[1]
ANNUAL PRODUCTIVITY SAVINGS

Find the right data in minutes. X-platform column-level lineage identifies sources, users, and dependencies across your stack instantly. Engineers ship faster, analysts cut report delivery time by 60%, and governance teams do more with less, freeing the equivalent of 80+ FTEs from busywork.

48%[1]
REDUCTION IN ANNUAL DATA INCIDENTS

Catch issues before dashboards or pipelines break. Proactive monitoring detects schema changes, freshness delays, and quality drift automatically. Column-level lineage shows the full blast radius instantly, cutting mean time to resolution by up to 55%.

30-35%[1]
REDUCTION IN RISK OF FAILED AUDIT

Stay audit-ready without the fire drill. DataHub continuously enforces governance standards and validates data quality across your entire stack. Automated workflows replace manual checklists, improving metadata coverage by 153%, assigning data owners, and tracing lineage end to end.

+119%[1]
AI/ML MODELS MOVED INTO PRODUCTION

Less than 20% of AI pilots reach production due to data quality gaps. DataHub provides AI-ready certification, feature-to-model lineage, and compliance controls so models train on trusted data. Governance guardrails and automated quality checks keep pipelines clean from experiment to deployment.

[1]Source: IDC White Paper, Sponsored by DataHub, Business Value of DataHub Cloud, March 2026

4.5 ★★★★½ on Gartner Peer insights

What You Can Do With DataHub

UNIFIED DISCOVERY

- Ask DataHub, conversational data discovery in Slack, Teams, DataHub UI
- Natural language search across 100+ sources
- Smart ranking by usage, quality, and freshness
- Chrome Extension for BI inline context
- AI-generated documentation at scale

CROSS-PLATFORM LINEAGE

- Automated column-level tracking
- Cross-platform impact analysis
- Change propagation & testing
- AI/ML feature lineage

GOVERNANCE CONTROL

- Business glossary
- Dynamic compliance workflows
- Certification & approval automation
- Cross-platform policy enforcement
- Audit-ready documentation

PROACTIVE OBSERVABILITY

- AI anomaly detection
- Smart quality assertions
- Incident management & alerts
- Unified health dashboards

CONTEXT MANAGEMENT

- Unified Context Graph
- Agent Context Kit
- Context Documents
- DataHub MCP Server

ENTERPRISE SCALE

- 99.5% uptime SLA
- SOC II compliant security
- Flexible deployment options
- Event-driven

| BUILT ON #1 OPEN SOURCE DATA CATALOG | 3M+ MONTHLY PyPI DOWNLOADS | 14,000+ COMMUNITY MEMBERS | 3,000+ ORGANIZATIONS |

---

chime

"My favorite part about DataHub is the lineage because this is one really easy way of connecting the producers to the consumers. Now the producers know who is using their data. Consumers know where the data is coming from. And it is easier to have accountability mechanisms."

SHERIN THOMAS
SOFTWARE ENGINEER, CHIME

NETFLIX

"DataHub gave us the extensibility features we needed to define new entity types easily and augment existing ones. During our evaluation, we assessed both functional and nonfunctional aspects, and DataHub performed exceptionally well in managing our traffic load and data volume."

AJOY MAJUMDAR
SENIOR STAFF ENGINEER, NETFLIX

See DataHub Cloud in Action