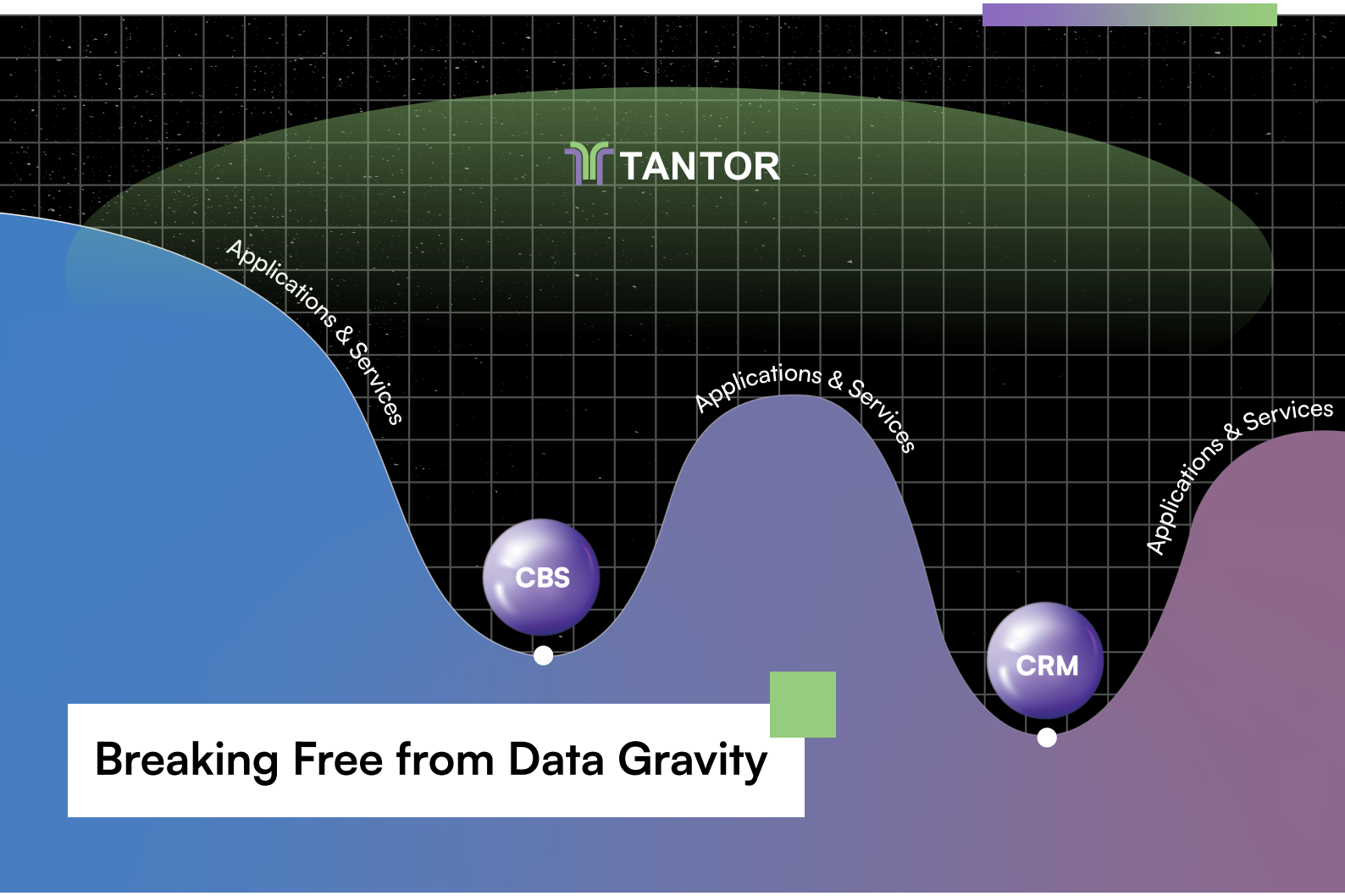



# Breaking Free from Data Gravity

As data accumulates, it becomes concentrated into multiple silos. It becomes harder for organizations to adapt to new technologies, scale effectively, or quickly leverage it to gain insights. Tantor Data Platform helps you break free from the limitations of data gravity, enabling your organization to unlock new levels of efficiency, agility, and innovation by connecting you to all your data.

## What is Data Gravity?

Data, like stars and planets, have gravity. Data gravity refers to the phenomenon where large volumes of data attract applications, services, and other datasets towards itself, much like how stars and planets draw smaller objects toward them through gravitational pull. As the size of a dataset grows, so does its ability to "pull" related resources into its orbit, creating an ecosystem centered around that data.

Once data gravity builds up, its inertia makes the data harder to move and transform. As a result, organizations end up with multiple data silos, centered around each dataset.

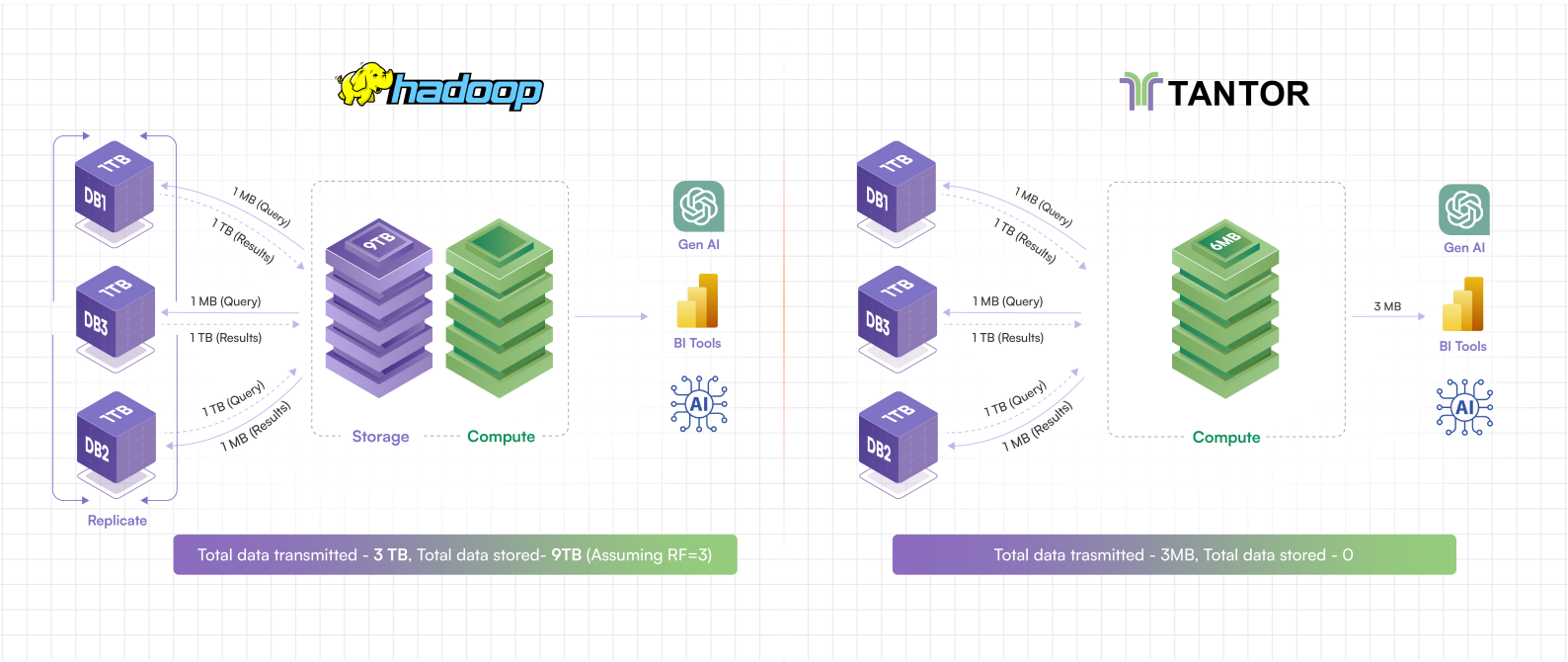
## Challenges of Data Gravity

For many years, organizations have turned a blind eye towards data gravity. They have spent millions of dollars on transferring data from many silos to create a single source of truth (SSOT) in the form of data warehouses (for structured data) and later data lakes (for unstructured data). Due to differing business use cases and data sources, different departments create their own data warehouses or lakes, leading to dilution of the SSOT concept.

Some have tried to merge data warehouses and data lakes into data lakehouses. But such initiatives have come at great network and storage costs involved in transferring PBs of data across clouds and creating, storing and managing yet another copy of data.

Further, these so-called SSOTs, due to massive amount of data volume, exert their own data gravity and became data silo of their own.

## How Tantor Mitigates Data Gravity




With Tantor Data Platform, you can connect data across multiple clouds, applications, and sources in real-time without the need for complex integrations. It creates a single logical layer on top of all your data sources and allows you to access data irrespective of its location in real-time.

Traditional Big Data solutions have focussed on moving the data closer to the compute. This involves transferring TBs and even PBs of data from their source systems to Hadoop clusters and then storing and managing three (3) copies of the transferred data for redundancy. This does not include data stored (again in triplicates) in Silver and Gold Layers as per Medallion Architecture.

In Tantor's **Data Federation Architecture**, instead of consolidating data from multiple sources to a central repository, queries are sent from a central logical layer to multiple sources distributed across clouds, data bases and applications. The responses are consolidated from these systems and send back to the requester. The **Data Federation Layer** acts a unified consumption layer, without the need for physically consolidating or replicating the data. Data federation does not require data movement, reducing latency and storage costs while enabling dynamic access to data in its native location.

Furthermore, data federation makes it easier to integrate with existing systems without requiring extensive reorganization of the data infrastructure or building complex data pipelines. It is well suited for scenarios that demand real-time data access and decision-making across heterogeneous sources.

### What is Data Federation?

Since organizations have huge amounts of data coming from multiple different sources, accessing all of this data requires engineers to know what data is stored where, query them, and then consolidate them in a different place. Users need additional storage to store the replicated data. Data federation simplifies this process. It connects to different sources of data, and acts as a virtual, consolidated database. This virtual database does not duplicate or store data. It simply queries the data required and presents the results. As such, businesses can save on storage, time, and expertise and get near real-time data without worrying data gravity.

Data federation delivers a unified view of distributed data while preserving the autonomy of individual data sources, ensuring compliance with security and privacy regulations. Further, as there is not additional copy of data, you can have access to real-time data without worrying about data gravity.

# No-Code Data Federation

Create a logical layer on top of your sources to access all your data without additional storage.

## Move Results, Not Data

Reduce network overhead, simplify architecture, and reduce bandwidth needed by moving compute to the source instead of additional replication and storage.

## Reduced Data Exposure

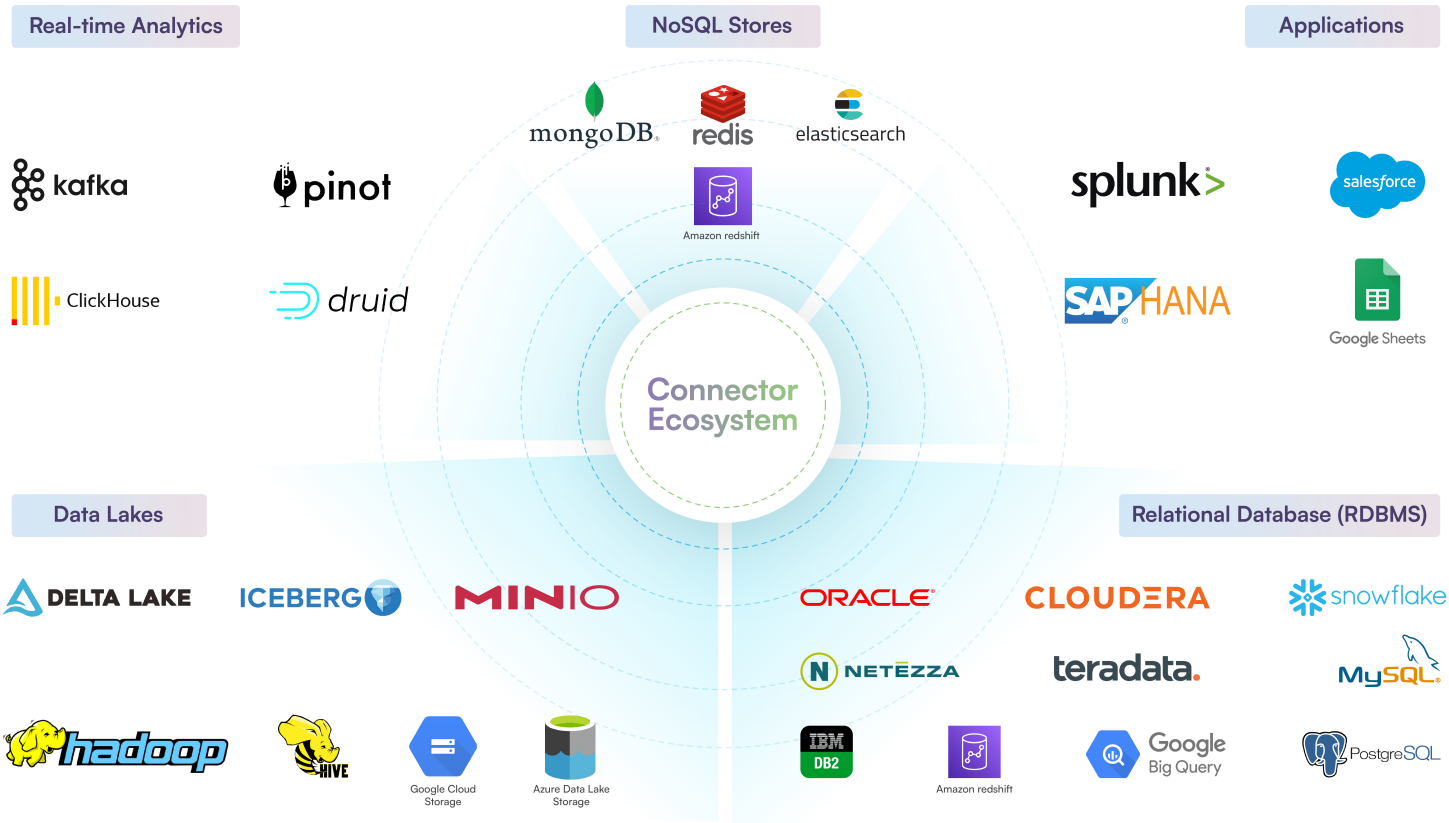
To minimize all possible exposures, data is not replicated. In addition, the federation layer can be used for data governance and restrict unauthorized access to your data.

## Full-Code Option

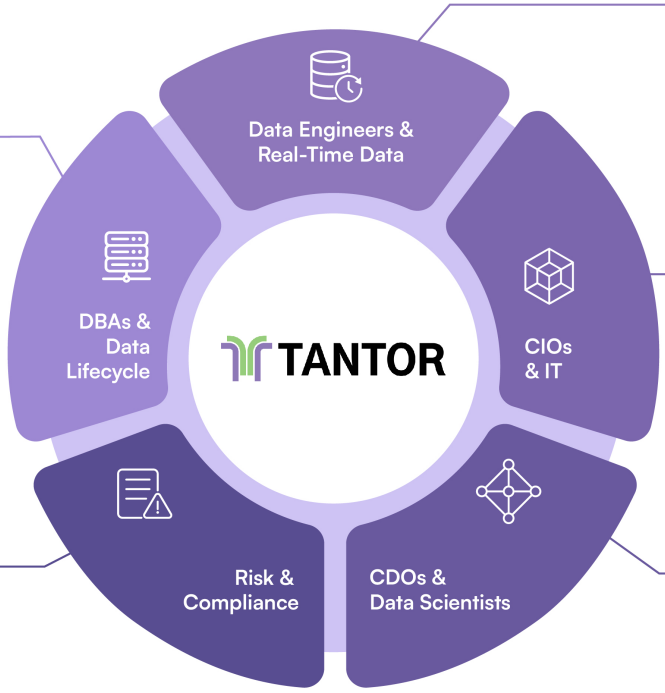
Data engineers and data scientists can leverage notebooks to develop and optimize code to use the platform to its full extent.

## Any Source, Any Target

Move data from anywhere to anywhere. With Tantor, you can connect to 50+ data sources.



# Benefits




**No-Code CDC For Real-Time Data**

Change Data Capture

Ingest, Transform, & Load in Real-Time

Set Topics & Tables to Capture

Batch & Real-Time Analytics

**Infrastructure Modernization**

Seamless Cloud Adoption

Unified Data Fabric

Virtualization

**Low-Code Data Fabric For Unified Data Silos**

Connect to 50+ Databases

Logical Data Federation Layer

Low-code, Low Storage

Bypass Data Gravity & Silos

**Regulatory Compliance & Data Governance**

Role-based Access

Data Masking

Restrict Read and Write Access

**Automate Data Lifecycle Management**

Drop-Down Menus

Zero Code

Automatic Data Mapping & Transformation

Scheduler & Reports

### Increased Agility

Reduce time-to-market and create data products faster with a gravity-free single source of truth.

### Faster Insights

Get comprehensive insights drawn from all sources in near-real time with a logical layer instead of creating new lakehouses.

### Complete Control of Your Infrastructure

Create new sources and build data products wherever — multiple cloud providers or on-premises — without worrying about data gravity or fragmentation.

### Increased Data Accessibility

Tantor's no-code/low-code interface empowers all personas — including business users — to make data driven decisions.

- **Increase in ROI**
  Minimize total cost of ownership and enable all users to make data driven decisions to tap into new opportunities and higher returns.

- **Scalability**
  Virtualized platform which can scale across growing datasets and adapt without significant infrastructure changes.

- **Enhanced Data Governance**
  Centralized governance, tagging, and access control across diverse data sources to ensure compliance and data security.

## Tantor's Unique Selling Points

### Self-service

Tantor's data federation is made for all personas, including non-technical subject matter experts. Business users can leverage the no-code interface to build their own datasets and data products to get the most out of the platform. This enhances productivity for all users — business and technical alike.

### Federate, Move, & Archive in one platform

Connect to multiple source systems and query data from all of them without requiring additional storage or complex data pipelines.

DB migrations usually require senior DBAs with deep understanding of both the source and target. Seniors DBAs must analyse, match, and create the schema before migration. Tantor uses metadata to automatically perform these activities much faster and without errors, allowing you to save on time and effort.

Post-migration purging is greatly simplified by Tantor through its Safe Purging feature. Every Tantor Safe Purge run validates that the data being purged has been copied and accurate using hash value and checksum tests.

## Advanced Kubernetes with easy up scaling and down scaling

Tantor Data Platform is deployed on a Kubernetes cluster, making it agile and cloud native. Clusters can be automatically scaled up and down based on thresholds and utilization.

Similarly, upgrades to the cluster are simplified through Blue-Green upgrades. New version of an application (green) is deployed alongside the existing version (blue), enabling seamless traffic switching and rollback, minimizing downtime and disruption.

## Adaptive Cloud Friendly

Tantor supports seamless deployment across edge, cloud, and core. It provides a single pane of management regardless of where it is deployed and can connect to any source, irrespective of its location.

## Plug-and-Play Analytics

Tantor offers seamless integration with existing analytics tools (AI, BI and Generative AI), allowing users to easily deploy and use powerful ML models and existing BI solution without requiring extensive configuration or expertise.

## Robust Computational Governance

Data Governance is central to Tantor's operations. Using advanced metadata and data catalogues, Tantor ensures that all data and data operations adhere to regulations and security policies. This includes:

- Data Encryption (at rest and in motion)
- Audit Trails
- Data Lineage
- Role-based Access Control (RBAC) and Attribute-based Access Control (ABAC)
- Data Masking

# Key Features

| Feature | Description |
|---|---|
| Unified Data Access | Virtual access to structured, semi-structured, and unstructured data from diverse sources. |
| Real-Time Connectivity | Direct interaction with live data without replication. |
| Data Governance | Centralized tagging, lineage, and security controls across all data sources. |
| Kubernetes Deployment | Containerized, scalable architecture for deploying on-premises, cloud, or hybrid environments. |
| Integration | Connects with BI tools, ETL pipelines, and cloud platforms. |
| Real-Time Analytics | AI-driven dashboards and insights based on live queries. |
| Collaboration | Shared workspaces for data scientists, analysts, and engineers. |

# Capabilities

## Data Sources

**Relational Databases**

- Generic (JDBC)
- IBM DB2 (JDBC): 8, 9, 10
- MS SQL*Server (JDBC, ODBC): 2008R2, 2012, 2014, 2016, 2017
- MySQL (JDBC): 5, 8
- Oracle (JDBC): 11g, 12c, 18c, 19c
- Oracle E-Business Suite (JDBC): 12
- PostgreSQL (JDBC): 8, 9, 10, 11, 12
- Sybase Adaptive Server Enterprise (JDBC): 12, 15
- MS Access (ODBC)
- BigQuery
- Cassandra - Cassandra JDBC driver.
- ClickHouse - ClickHouse JDBC driver.
- Druid - Druid JDBC driver.
- Elasticsearch - Elasticsearch JDBC driver.
- Exasol - Exasol JDBC driver.
- Hive - Apache Hive JDBC driver.
- Ignite - Apache Ignite JDBC driver.
- MariaDB - MariaDB JDBC driver.
- MongoDB - MongoDB JDBC driver.
- MySQL - MySQL JDBC driver.
- OpenSearch - OpenSearch JDBC driver.
- Oracle - Oracle JDBC driver.
- Phoenix - Apache Phoenix JDBC driver.
- PostgreSQL - PostgreSQL JDBC driver.
- Redshift - Amazon Redshift JDBC driver.
- SingleStore - SingleStore JDBC driver.
- Snowflake - Snowflake JDBC driver.
- SQL Server - Microsoft SQL Server JDBC driver.
- Vertica - Vertica JDBC driver
- Oracle 12c In-Memory v

### Engineered Systems

- Oracle Exadata (JDBC): X8,X9,X10
- Oracle PCA (DBs on Oracle PCA)
- Oracle Database Appliance

### Cloud data warehouse

- Amazon Redshift (JDBC)
- Amazon Athena (JDBC)
- Amazon Aurora
- Amazon DynamoDB
- MongoDB
- Snowflake

### Big Data

- Apache Hive (JDBC): 0.12, 1.1.0, 1.1.0 for Cloudera, 1.2.1 for Hortonworks, 2.0.0
- Impala (JDBC): 2.3
- Spark SQL (JDBC): 1.5, 1.6, 2.x, 3.x
- PrestoDB (JDBC)
- PrestoSQL
- Databricks Delta 2.x

### NoSQL

- MongoDB
- Cassandra
- HBase

### Data Lake storage

- S3
- Azure Data Lake Storage
- Azure Data Lake Storage Gen 2
- Azure Blob Storage
- Google Cloud Storage
- Parquet/Avro
- Delta/Iceberg

## Semantic Repositories

### Supported Protocols

- REST
- Open API

### Publishing Options

- SAML
- SSL

### Security

- Support for encrypted metadata import/export.
- Integration with advanced security features like Kerberos, SAML, and OAuth

### Integration

Compatible with various semantic storage formats (e.g., Delta, Iceberg).

### Web Services

REST (XML, JSON)

### Flat and Binary Lines

- CSV, pipe-delimited, regular expression-parsed
- All files can be local or in remote filesystems, through FTP/ SFTP/FTPS, and in clear, zipped and/or encrypted format.

### Active Directory as Source or leveraging Security

- LDAP v3
- Microsoft Active Directory 2003, 2008

### Message Queues

Kafka

### Publishing Options

- REST
- Open API
- GraphQL
- SAML
- SSL

## Data Catalog

### Integration

- Connectivity to relational databases, big data systems, cloud warehouses, and NoSQL solutions (e.g., MySQL, Snowflake, MongoDB).
- Integration with REST APIs and file-based data (CSV, JSON, Avro, Parquet).

### Optimization

- Massively parallel processing integration.
- Multi-mode caching (full, partial, incremental, or total refresh).
- Pushdown support for projections, selections, and joins

### Support for Data Movement

## Performance Optimization

- Massively parallel processing integration
- Multi-mode caching: full, partial, incremental, or total refresh, event-based or scheduled, configured at the view level, incremental queries for SaaS sources
- Pushdown of selections/projections/joins/groupby operations also on federated views
- Multiple join strategies
- Simplifying partitioned unions (Partition pruning)

### Data movement

- Redis
- Iceberg
- Hive
- Impala

### Data Pipelines

Tantor scheduler

### Third Pary MPP Options

- Impala
- Presto
- Spark 1.5, 1.6, 2.x
- Databricks 2.x

## Data Governance

### Security

- Comprehensive support for data in motion (SSL/TLS) and at rest (encryption mechanisms).
- Fine-grained access control at row and column levels.
- Advanced masking options for sensitive data.
- Authentication support via LDAP/Active Directory, Kerberos, SAML, and OAuth 2.0.

### Global Policies

- Tag-based policies and attribute-based access control (ABAC).
- Customizable security constraints and external policy server integration.

### User and Role Management

- Role-based and attribute-based user access management.
- Integration with external directories (e.g., Active Directory, LDAP).

## Security

### Data in Motion — secure channels

- Using SSL/TLS
- Available for all protocols (JDBC, ODBC, ADO.NET and WS)

### Data at Rest - secure storage

- Cache: third party database. Can leverage its own encryption mechanism
- Swapping to disk: serialized temporarily stored in a configurable folder that can be encrypted by the OS

## Encryption/Decryption

- Support for custom decryption for files and web services
- Transparent integration with RDBMs encryption
- Encrypted metadata import/export
- User and Role Based including integration with AD/ LDAP
- Row and Column level authorization
- Advanced customizable masking
- Custom policies for specific security constraints and integration with external policy servers

## Global Policies

- Tag-based security policies
- Support for RBAC and ABAC
- Column and row restrictions, multiple masking options, deny execution
- Authentication
- Native and LDAP/Active Directory
- Base64
- OAuth, OAuth 2.0 (JWT)
- SAML
- SSL

## Advanced Semantics

### Capabilities

- Rich semantic mapping using REST, Open API.
- Advanced query optimization techniques (e.g., partition pruning).

### Interoperability

- Compatibility with third-party query engines

### Security

- Semantic metadata encryption during import/export.

### Extensibility

- Ability to support domain-specific extensions.

## Data Quality

### Capabilities

- Library of transformation, filtering, and matching functions.
- Quality rules for validating, cleansing, enriching, standardizing, matching, and merging data.

### Extensibility

- Support for custom functions and integration with external data quality tools.

### Optimization

- Multi-mode caching for incremental or event-based refresh.

### Integration

- Compatibility with flat files, binary files, and cloud storage solutions

## Operating System

Linux (all versions >= 8.x)

## Deployment Patterns

- On-premises, private cloud, public cloud
- Basic single server configuration
- HA cluster with load balancing (Active-Passive and Active-Active)
- Shared or distributed local cache
- Containerization support through Docker
- Public cloud
- Auto-scaling support both in AWS, Azure, OCI, and GCB

# Tantor 1.0 Editions

| Feature | Sub Feature | Standard | TANTOR Advanced | Enterprise |
|---|---|---|---|---|
| Platform | Deploy on Any Cloud | ✓ | ✓ | ✓ |
| Platform | Auto-scaling | ✓ | ✓ | ✓ |
| Platform | Embedded Kubernetes | ✓ | ✓ | ✓ |
| Platform | Low/No-code Interface | ✓ | ✓ | ✓ |
| Ingestion | Maximum Concurrent Data Sources | 05 | Unlimited | Unlimited |
| Ingestion | Data Virtualisation | ✓ | ✓ | ✓ |
| Ingestion | ETL | ✗ | ✓ | ✓ |
| Ingestion | CDC | ✗ | ✓ | ✓ |
| Security | AD/LDAP Integration | ✓ | ✓ | ✓ |
| Security | DAM/PAM/SAML Integration | ✗ | ✗ | ✓ |
| Security | Data Encryption at Rest | ✓ | ✓ | ✓ |
| Security | Data Encryption in Transit | ✓ | ✓ | ✓ |
| Governance | Data Catalog | ✓ | ✓ | ✓ |
| Governance | Data Masking | ✓ | ✓ | ✓ |
| Governance | Data Lineage | ✗ | ✗ | ✓ |
| Engineering | Data Quality | ✓ | ✓ | ✓ |
| Engineering | Data Transformation | ✓ | ✓ | ✓ |
| Data Products | Archive | ✓ | ✓ | ✓ |
| Data Products | Purge | ✓ | ✓ | ✓ |
| Data Products | Data Lifecycle Management | Add-on | ✓ | ✓ |

## Tantor helps you

- **Acquire** and securely move data from any source, in batch or stream with ease.
- **Federate** isolated silos for a unified source to build AI and data products.
- **Govern** data with role-based access, data masking and encryption.

TANTOR

www.tantor.ai | info@tantor.ai