tcgdigital

Data Lakehouse:
Creating the foundation for scaling AI-based innovation across the enterprise

mcube™
by tcgdigital

Data Lakehouse
powered by tcgmcube

Traditional data warehousing and data lake architectures have their limitations, particularly in navigating through diverse and voluminous datasets, making it extremely difficult for users to get to relevant, contextualized data.

The data lakehouse, which integrates the best features of both data lakes and data warehouses and adds a semantic layer for contextualization, emerges as a compelling solution.

Leveraging our end-to-end analytics platform, tcgmcube, organizations can create robust data lakehouses to streamline data management by integrating various data processing and analytics needs into one architecture.

The platform tcgmcube comes with mcube.data and mcube.ai, thus providing AI capabilities and data management on the same platform managed by common platform services. This makes it an extremely powerful platform for implementing the lakehouse and deploying analytical and AI applications on top of the lakehouse.

This approach helps avoid redundancies and inconsistencies in data, accelerates analysis throughput, and minimizes costs.

tcgdigital

Architecture

- AI capabilities and data management on the same platform managed by common platform services
- Distributed, fault-tolerant, and cloud-native architecture
- Cloud-agnostic platform that can make native cloud calls
- Highly interoperable - complements existing ecosystems
- Modular architecture - each module can scale dynamically

Data Ingestion

- Ingest structured, semi-structured, and unstructured data
- Options for real-time, near real-time, and batch ingestion
- Supports dynamic data pipelines
- Data transformation at various stages (ETL as well as ELT)
- Supports data collection and management at the edge - handles events through data caches and synchronization
- Semantic layer overlay

Pre-built Standard Connectors

- Connectors to various systems—relational databases, NoSQL databases, light repositories, big data stores, data warehouse appliances, JSON, XML, Excel, Text files, CSV files, RSS feeds, HL7 data, etc.
- Connectors to various instrument types: Balance, Beaker, DNA Sequencer, Gas Chromatograph, pH Meter, ThermoCycler, and Titrator
- Connectors to support OPC UA and OPC DA protocols, enabling the ingestion of near real-time tag data.

Data Management

- Data Quality management & data profiling
- Data Catalogues to identify schemas and transformations to be edited for addressing changing workload
- Data Lineage tracking
- Ontology Management and Knowledge Modeling capabilities - makes it "easy to get data out"

Data Governance & Access

- Role-based access control and fine-grained access policies (row, column, and object-level access control)
- Data encryption at rest and in transit
- Audit Logs for all data access and processing activities.
- Layered Security: Security can be defined at various levels; cluster, index, document, and field
- Metadata Management powered by knowledge graphs

Data Storage

- Hybrid storage options
- Raw Data Layer – stores data in its raw format with no transformations and facilitates source data processing with features to validate and catalogue the raw data
- Analytic Persistence layer – stores processed datasets for optimizing analytical queries and AI-driven processes
- Semantic Persistence Layer – stores fully processed and optimized data with knowledge graphs, private LLMs, and business metadata, providing semantic context

Scalability & Performance

- Optimally support hot/cold data repositories
- In-memory for rapid search and query response
- Distributed query processing for fast response
- Query result caching for faster performance
- Dynamic partitioning of data & built-in indexing
- High throughput data processing with sub-second query latency for real-time data analysis.

Data Access & Delivery

- Microservices API for data access - Data APIs, Application APIs, APIs for AI/ML/Gen AI, etc.
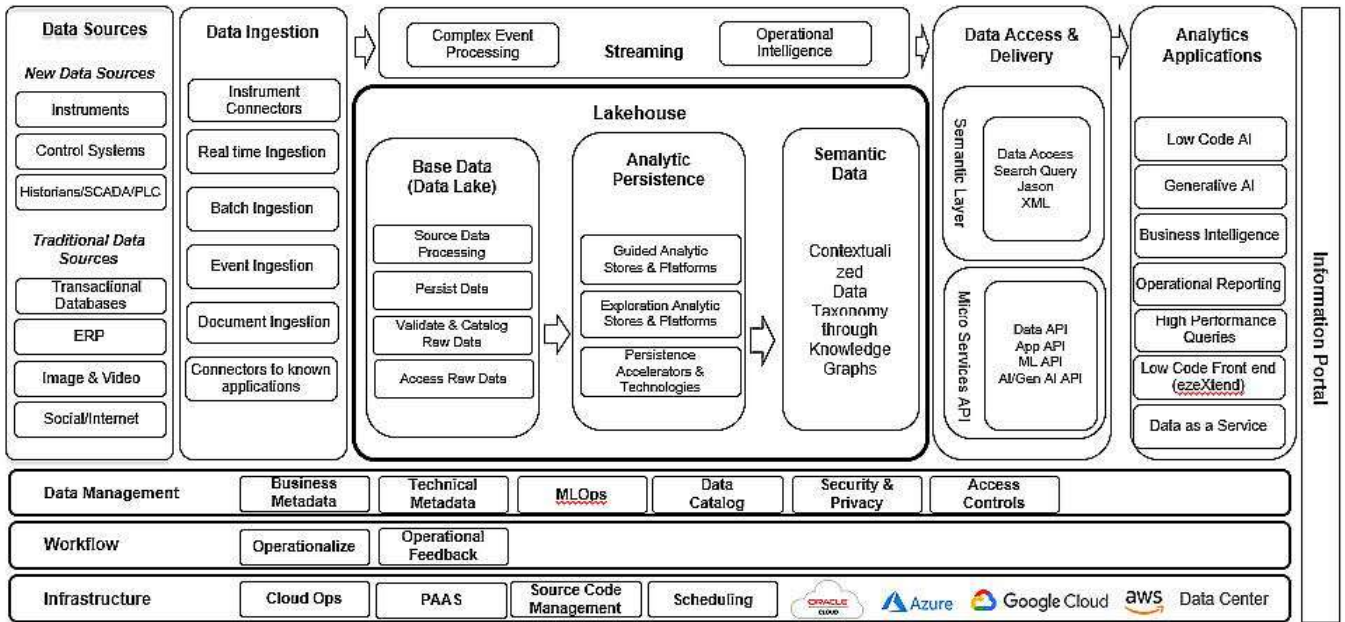- Semantic layer for domain contextualization for ease of data access

Options for Information Disseminations

- Insights dissemination with multiple user interfaces such as dashboards, operational reports, and low-code "upgrade safe custom screen painting"
- Action dissemination inputs to automated operational processes (alerts, recommendations, action triggers, etc.)
- Traditional AI at scale – 1000+ statistical, AI/ML, and optimization algorithms
- Comprehensive Generative AI algorithms covering traditional LLM, multimodal LLM, and RAG models for fast information retrieval and complete source traceability

Deployment & Monitoring

- Deployment on cloud or on-premise
- Cloud-agnostic container based deployment with Kubernetes for container management and orchestration
- Complete platform monitoring features

REFERENCE ARCHITECTURE OF THE DATA LAKEHOUSE

tcgdigital

::::mcube™
by tcgdigital

TCG Digital is the digital & AI arm of The Chatterjee Group (TCG), a multi-billion dollar conglomerate with a diverse portfolio in Pharmaceuticals, Biotech, Petrochemicals, and Real Estate across the US, EU, and South Asia. Our umbrella includes companies such as LabVantage, Lummus Technology, and TCG LifeSciences. At TCG Digital, we are driven by our mantra of delivering "Velocity to Value", helping enterprises transform faster and smarter. Our AI platform tcgmcube is at the heart of these transformations. We enable organizations unlock the full potential of their data, and by seamlessly integrating AI/ML capabilities into their business processes, we empower businesses to accelerate their digital transformation journey, enhancing agility and driving impactful results

www.tcgdigital.com

tcgmcube is TCG Digital's flagship Data, AI and Analytics platform. Built with a domain driven design at the cross-roads of industry knowledge and digital prowess, our architecture is designed to handle the most disparate data landscapes with AI 2.0 being at the heart of it combining powerful and advanced models to solve the most complex business problems. The platform integrates mcube.ai and mcube.data, delivering AI capabilities and data management seamlessly through unified platform services.

www.tcgdigital.com/tcg-mcube

TCG DIGITAL SOLUTIONS
265 Davidson Ave, Suite 220 Somerset,
New Jersey 08873

✉ contact@tcgdigital.com
🌐 www.tcgdigital.com