

Delta Lake Platform

Data Sheet

Key Benefits:

- **Scalability:** Seamlessly accommodate growing data sets without compromising performance.
- **Flexibility:** Store and process structured, semi-structured, and unstructured data from various sources.
- **Effective:** Monitoring your big data ecosystem performance in a centralized control center.
- **Robust Operation:** Enhance your data security procedures with AI-detected PII (Personal Identifiable Information) or sensitive info and automatically masked.

Build Secure and High-performance Data Foundation, Improve decision-making

Welcome to the age of digitalization where massive amounts of data are generated every single day. Every click, every search, every purchase generates a wave of information. It is essential for organizations to ensure the consistency and availability of their big data management system, ready to take on the waves of data.

Digital Dialogue's Delta Lake Platform solution empowers enterprises to effortlessly manage and store massive datasets. The solution provides a secure platform to manage your data (Distributed and Non-Relational), allowing you to store structured, semi-structured, and unstructured data within HDFS (Hadoop Distributed File System).

Query your data with the familiar ease of RDBMS or table formats, ensuring seamless integration with your existing workflows. With Digital Dialogue's Delta Lake Solution platform, you can manage a reliable data lakehouse system that drives better decision-making and fuels valuable business insights via web-based GUI.

CUBIKA Big Insight's flexible and scalable data platform solution is ready to enhance your data journey and helps enterprise accelerates their shift to become a data-driven enterprise, providing a decoupled architecture and ensuring high scalability through both horizontal and vertical scaling approaches.

Our Delta Lake Platform solution is driven by our products under CUBIKA Big Insights family; CUBIKA Big Insights for Data Access and Storage, CUBIKA Big Insights Cockpit and CUBIKA Big Insights for Data Security. All software platforms under CUBIKA Big Insights family are designed for seamless integration and unified operation across the entire platform ecosystem.

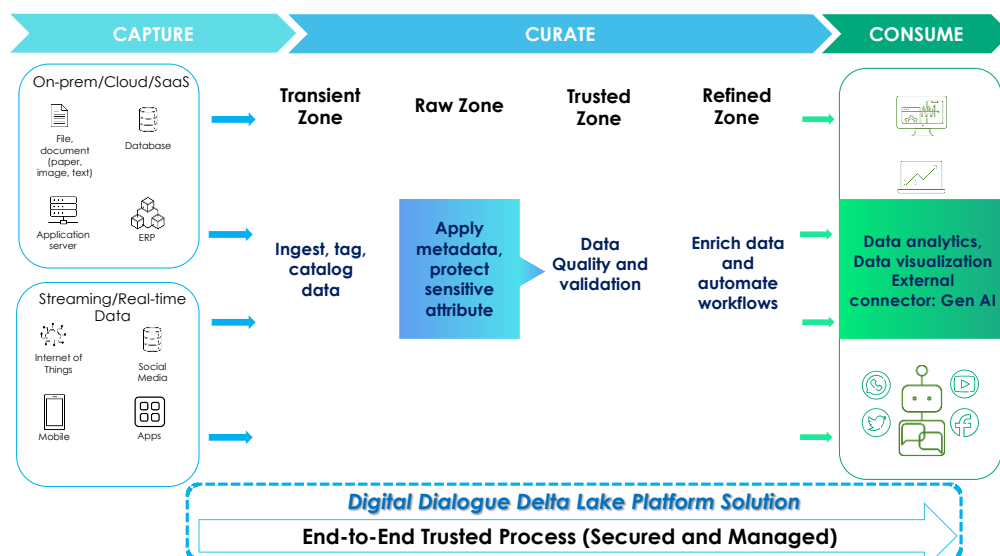


Figure 1: Digital Dialogue Delta Lake Platform Solution Within 3C Data Journey: Capture, Curate, Consume

Key Capabilities and Features

CUBIKA Big Insights for Data Access and Storage

Data Format and SQL Capabilities

CUBIKA Big Insights Storage utilized the Apache Iceberg table format to provide robust data access and manipulation through standard ANSI SQL. It ensures full ACID compliance, delivering Atomicity, Consistency, Isolation, and Durability for enterprise-grade data operations.

Apache Iceberg enables seamless integration across multiple compute engines including Apache Spark, while supporting various storage backends such as Amazon S3, Google Cloud Storage, Azure Blob Storage, and HDFS without vendor or technology lock-in. Enabling seamless integration to your current workflow and tools with ease.

Real-Time Streaming Integration

To handle streaming data workloads, CUBIKA Big Insights for Data Access and Storage leverages an open-source message broker - Apache Kafka. This integration creates a real-time streaming architecture that delivers high-throughput messaging capabilities while providing key operational benefits including decoupling between systems, horizontal scalability for processing demands, high availability, and zero-downtime operations.

Data Lakehouse Architecture

Supporting the modern approach of storing data – Data Lakehouse. Data Lakehouse evolves from the data warehouse concept. While both architectures serve as repositories for massive data volumes, the Data Lakehouse model provides significantly enhanced flexibility and operational capabilities.

As the name suggests, it is the combination of data warehouse and data lake. By bringing the best of both worlds, the reliability and performance of data warehouses with the flexibility and scalability of data lakes, makes it a top choice for organizations that need to handle various sources of data types and perform various analytical workloads.

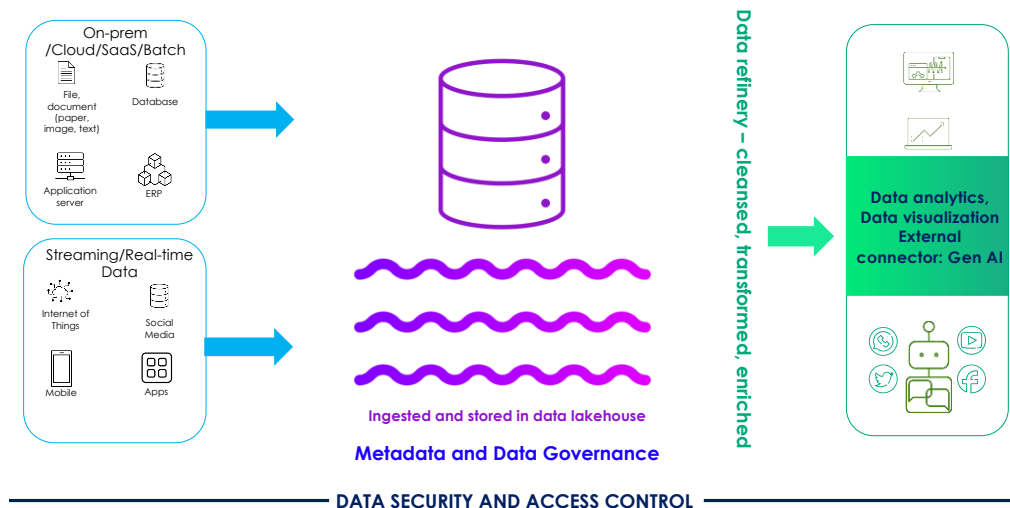


Figure 2: A quick View of Data Lakehouse concept

With CUBIKA Data Access and Storage, organizations can harness HDFS's massive scalability to support thousands of nodes and over 100 petabytes of storage capacity in a single cluster (with the appropriate hardware).

The platform is also equipped with a built-in data snapshot functionality to ensure rapid data recovery and protection against loss, corruption, user errors, unintended modifications, or application failures.

With the future expansion in mind, the cluster can be enlarged either by increasing the sizing of the existing machine in the cluster (Vertical Scaling) or adding more nodes to the existing cluster (Horizontal Scaling), all without disrupting operations.

Technical capabilities

- All data types can be managed;
 - *Structured data* - Relational Database Management System such as Oracle, MySQL, MSSQL Server, PostgreSQL
 - *Semi-structured data* - For example, Log file, Text file, csv, xlsx, xls, TXT, JSON and XML
 - *Unstructured data* - Various file types and streaming data, MongoDB, IoT, social media
- Supporting both replication and erasure coding setup to ensure data reliability and availability.
- Support both Python and Java on importing/exporting data, data processing, data analyze.

Data Management Console

CUBIKA Big Insights Cockpit is a highly intuitive management console for monitoring, provisioning, and managing big data ecosystem. Moreover, the service can help administrators improving their work efficiency.

The dashboard is powered by web-UI and REST API, enable administrators to monitor health, status or key performances of each, Hadoop cluster, get notification of certain events or system alerting with one centralized dashboard for Hadoop Ecosystem - Hadoop Distributed File System (HDFS), Zookeeper, Yet Another Resource Negotiator (Yarn), Spark, Hive, HBase, Hadoop MapReduce. CUBIKA Big Insights Cockpit, powered by the Alert Framework, can send dispatch notifications to alert user of status changes through JMX or provided email, SMTP. Rest API and Python API are supported to provide historical information of clusters, nodes, applications.

CUBIKA Big Insights Cockpit empowers enterprises via web-based in

1. **Monitoring resource overview and check system performance** - node, resource and components in Hadoop ecosystem within one dashboard, enabling system administrators to get a quick view on Resource Utilization status right away.
2. **Simplifying big data ecosystem configuration** - providing system administrator easy, efficient, effortless tool for managing Hadoop ecosystem
3. **Taking proactive measures** – Live Notification is available for system errors and warning on each service through CUBIKA Big Insights Cockpit platform notification system. Moreover, alerts and notifications configuration to automatically notify relevant personnel when system outages or failures occur via email
4. **Data Snapshot** – easily create and roll-backed snapshots to help ensure your data is back-up when needed
5. **Smart Self-Healing** - automatic detection of Data node problems and perform auto-restart to remedy process
6. **Universal Database Support** – compatible with your preferred open-source and commercial databases without technology constraints

Technical Capabilities

Works with operating systems with 64-bit CPU as indicated below:

- CentOS
- Ubuntu Linux
- Red Hat Enterprise Linux

Users can select which module they wish to configure via web-based GUI

- Hive
- HBase
- Hadoop
- Flink
- Kafka
- Spark
- Tez
- YARN + MapReduce
- ZooKeeper

CUBIKA Big Insights for Data Security (CBI Data Secure)

CUBIKA Big Insights for Data Security offers comprehensive data-centric protection, at-rest and in-transit through advanced dynamic masking capabilities, enterprise role-based access control, privacy compliance management frameworks, and detailed audit tracking for complete data governance and security.

Technical capabilities

- Supporting user authentication via Native user authentication, LDAP user authentication, Active Directory (AD, Microsoft Entra ID).

Integration with CUBIKA Big Insights Data Secure to manage user authentication mechanisms based on:

Privileges	An authorization level for managing users and policies for preliminary authorizations at startup.
Role Based	An authorization level is based on the assigned role within an organization of a user.
Permission	An authorization level is based on the access policy for a user: 1. View – user with this access level can access and only view details and data within the system. 2. Modify – user with this access level and access and modify any settings and configurations within the system.

- Enabled data warehouse security.
- Providing capability for Role-based Access Control (RBAC) working in tandem with HDFS Access Controls (ACLs)
- Dynamic data anonymization (data obfuscation) is available for data pseudonymization, enhancing data security.
 - Cryptographic Hashing
 - Partial Masking/masking
 - Data Redacting
 - Dynamic encryption with AES-256 algorithm based on Regular Expression
- Supporting Data-in-Transit security with TLS 1.2, 1.3 encryption
- Supporting Data-At-Rest security with AES based encryption

- **Adaptive Data Safeguard** - supporting dynamic trickling-down policy enforcement, PDPA & regulatory compliance with automatic policy enforcement from the top to bottom level where top-tier policies are mandatory enforced and cannot be revised by the lower-tier. The lower-tier policies can be added, if needed. The policy flow setup process can be fully configured.

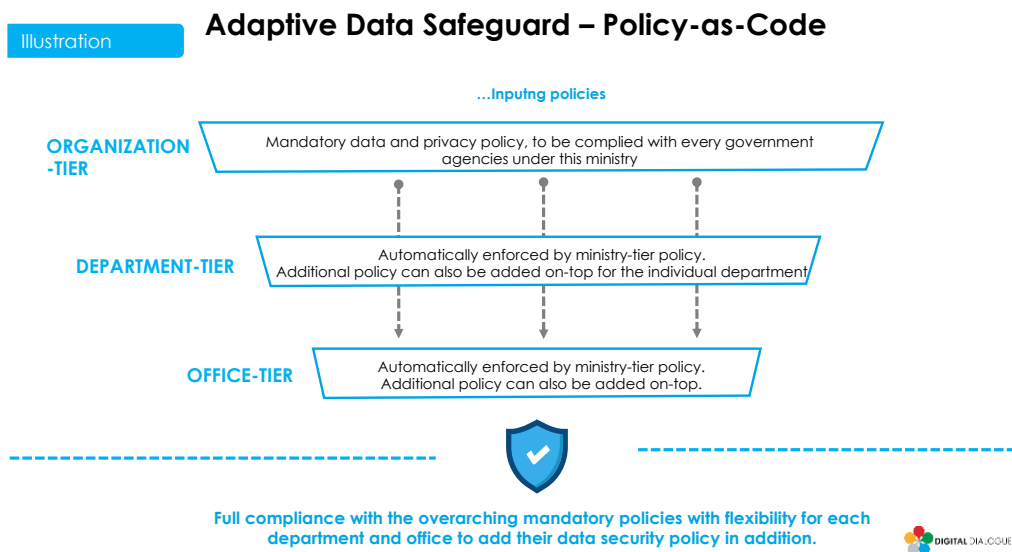


Figure 3: Illustration of How Adaptive Data Safeguard - Dynamic Trickling-down Policy works through CUBIKA Big Insights Data Secure

- **Data access audit** - every previous event and activity in Apache Hadoop can be audited from securely stored activity logs which can be accessed and searched via GUI by administrators and auditors. The activity log will keep all task below:
 - Hadoop service access
 - User log-in
 - Policy admin access
 - Addition logs as below:

HDFS Audit Logs	2 different types of audit logs for HDFS 1) Audit logs of user activity 2) Audit logs of service activity (Hadoop Service)
------------------------	--

MapReduce Audit Logs	2 different types of audit logs for MapReduce 1) Audit logs of user activity 2) Audit logs of service activity (Hadoop Service)
YARN Audit Logs	Audit log events for YARN are logged in daemon log files.
Hive Audit Logs	Using Hive Metastore for logging Hive audit events. Auditors can audit username and IP and Hive service log event by filtering logs with <i>HiveMetaStore, audit</i>
HBase Audit Logs	HBase audit log events and services containing information and event in Column Family, Column and Table.

- **Secure Data Zone Setup** – Create business-focused and encrypted data zones based on your organization's data management strategy. With CBI Data Secure, data zones can be easily configured with full support for Access Control Lists (ACLs) and Role-Based Access Control (RBAC) for role and authorization management. The platform also enables fine-grained access control at the object or file level.

CBI Data Security helps your enterprise in architect and organize data zone with data organization strategy as a backbone, support heavy curation of data for various business use.

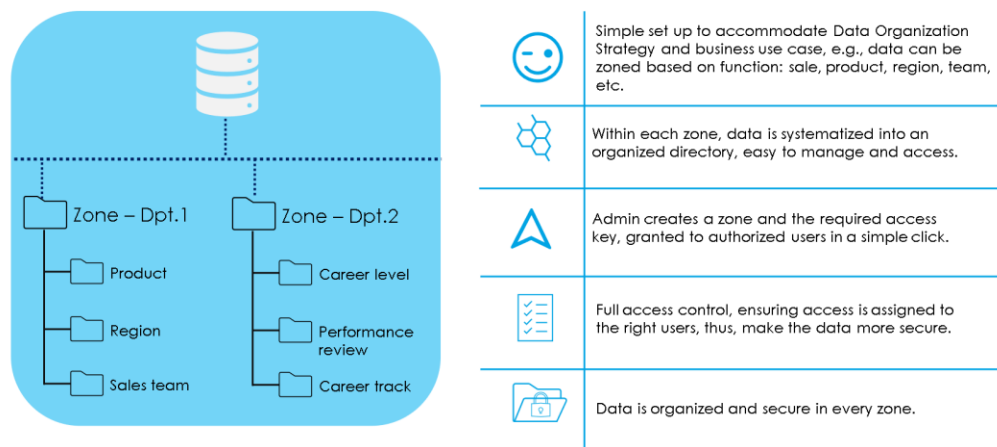


Figure 4: Illustration of Secure Data Zone setup and its benefits

About CUBIKA Big Insights



Low Code, Big Insights

CUBIKA Big Insights products applies machine learning, analytics, and Digital Dialogue's Thai Natural language processing in automating tasks, categorizing, and standardizing not only English but Thai data across enterprise's big data environment. CUBIKA Big Insights help everyone from business users, data engineers, analysts to IT in achieving their task, gaining understanding of data, and turning them into actionable insights.

CUBIKA Big Insights enables organizations to capture, curate and consume data with the speed of business, embedded with 5Vs of Big Data (Volume, Variety, Veracity, Velocity and Value) as a cornerstone in order to transform data into meaningful insights for business. CUBIKA Big Insights handles mountains of data (Volume) in various shapes and forms; unstructured, structured and semi-structured data from various sources through API integration (Variety and Velocity), process them through data management, including profiling and deduplicating, leveraging open-source technologies in Hadoop ecosystem, ensuring data's integrity, lineage and accuracy with no anomaly (Veracity). Data is now ready to be analyzed by intelligent analytics model, powered by machine learning and Digital Dialogue's owns NLP engine (TH/ENG) which serve as a brain to understand the context of data and help with data de-duplication efforts to make sure the data is clean even before entering the storage (in-transit context analysis). Users can harness and unlock powerful, actionable insights in data usage stage via industrialized tool such as Power BI (Value) or Tableau or KNIME. With our principle in democratizing data, CUBIKA Big Insights empowers users in accessing data with simple and user-friendly tools and user interface in real-time and on-demand.

CUBIKA Big Insights, part of CUBIKA - Digital Dialogue's suite of intelligent product, helps organizations in democratizing data and analytics, enabling the organizations to harness meaningful and valuable business insights which transforms into opportunities for business growth, powered by Digital Dialogue's proprietary Thai NLP engine and framework.

About Digital Dialogue

Digital Dialogue is a professional intelligence solution platform and software developer company. Powered by Digital Dialogue's market leading Thai NLP, we help enterprises in numerous industries harnessing powerful AI, machine learning, data analytics, CRM and intelligence technologies for data management, data governance, underpinned by our extensive expertise in the industry, depths and breadths of capabilities, partner network and deep commitment in helping clients accelerating their business growth and realize tangible business value.

CUBIKA, Digital Dialogue's suite of intelligence products aims to empower an enterprise to be a data-driven and lead with insights, ranging from AI-powered chatbot, power apps (low code and no code) to data analytics, voice analytics and data management platform, all infused with Thai NLP capability.

PSP-CBI-DPF-101-V1.2025



© Copyright 2025 Digital Dialogue. This document is provided for information purposes only and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. Digital Dialogue shall not be liable for technical or editorial errors or omissions contained herein. CUBIKA is a registered trademark of Digital Dialogue in Thailand. All third-party trademarks are property of their respective owners.

Contact our specialist to get started with CUBIKA Big Insights and harness valuable business



contact@ddlghq.com



02-088-0795