Data Management and Enterprise Data Service

Data Sheet

Manage Securely, Define Understandably, Discover Simply, Refine Intelligently, Lead with Insights Successfully

With the wave of digitalization, data has emerged as the vital of organizations, fueling decision-making, innovation, and growth through data-driven insights. However, the sheer volume, velocity, and variety of gargantuan amounts of data generated today present significant challenges.

To truly harness the power of data and lead with insights, organizations need a robust and reliable big data management system. Ensuring the consistency, quality, availability of your data and data ecosystem is not just an option, but a strategic imperative for success in the digital age.

At Digital Dialogue, we believe in data democratization where everyone is equipped with the tools to derive insights to further drive the insightful "dialogue".

Hence, our Data Management and Enterprise Data Service solution's cornerstone is to help organizations see the complete picture of their data with clarity and to gain those winning insights while maintain full security, compliancy, and integrity of data.

From cataloging data, managing metadata, securing data, cleansing data all the way through deriving insights. The possibilities are here.

CUBIKA Big Insights' flexible and scalable data platform solution is designed to accelerate your data journey and empower enterprises in their transition to a data-driven organization, offering a decoupled architecture with both horizontal and vertical scalability. The platform provides a user-friendly graphical user interface (GUI) accessible via modern web browsers—including Google Chrome, Firefox, and Safari and is deployable on both Windows Server and Linux environments.

Data Management and Service solution is driven by our software platforms under CUBIKA Big Insights family; CUBIKA Big Insights for Data Catalog, CUBIKA Big Insights Data Security and CUBIKA Big Insights for Data Science Tool Packs. All software platforms under CUBIKA Big Insights family are designed for seamless integration and unified operation across the entire platform ecosystem.

Key features and benefits

CUBIKA Big Insights for Data Catalog is your intelligent solution for the CURATE phase, transforming data chaos to data order. It empowers organizations to **build trusted data assets, refine, understand, govern, and unlock the value** of their data with confidence.

The all-in-one platform for intelligent data asset management that automatically organizes, catalogs, and secures enterprise, enable your team to find what they need. With AI-powered metadata management and semantic search, you can discover the right datasets in seconds while maintaining data integrity, accuracy with complete security and compliance control.

Data Scanning	**AI-powered Metadata Management**	**Data Catalogue**	**Data Preview**	**Data Secure Insights**	**Data Standard**
Identify and classify data	Organize, refine and maintain	Searchable inventories of your data	Snippet of data health	Locate where PII is	cataloging and complying to data policy and security of the organization

Data Scanning: Identify and classify data

- **Connect Everything** - structured, semi-structured, unstructured data
- **Smart Classification** - Automatically identify sensitive data like PII (Personal Identifiable Information), financial records, and customer information
- **Map Relationships** - Automatically detect data structures, schemas, and connections between datasets without manual work
- **Enable Governance** - Build the foundation for data compliance and security policies

Datasource Browser: locating registered data sources, zones, schemas, locations metadata throughout your database within a few clicks.

- Datasource browser helps you discover quality and reliability of your data by both "technical rules" and "business rules", which in turn, helps businesses in maintaining data integrity. A quick preview to data quality health-check. Compatible with database technology such as SQL, MySQL, MSSQL Server, Hive, MongoDB, Huawei OBS, DB2, Vertica, Oracle, PostgreSQL.
- Automatically capture comprehensive comments including both technical and non-technical such as schema, tables, fields, data types, Primary Key, Foreign Key, constraints, and also comments from source system (data asset's comments/descriptions).

AI-powered Metadata Management: Organize, refine, maintain

A robust metadata management capability that enables organizations to organize, refine, and maintain various types of metadata across your organization at scale—automatically and securely.

CUBIKA Big Insights for Data Catalog leverages AI to help organizes in handling vast amount of data, extracting from connected data sources, and displaying:

Technical metadata	- Database schemas, table structures, fields - Data types, constraints - Foreign key, Primary key
Business metadata	- Business definitions and glossaries - Data ownership and stewardship - Usage context and business rules - Data lineage and transformation logic - Quality standards and validation rules
Operational metadata	- Access logs and usage statistics - Data refresh schedules and frequencies - User activity and usage pattern - System performance and monitoring data - Compliance and audit trails
Social metadata	- User ratings - Collaboration - Data source comments, user comments

All metadata is automatically cataloged and indexed in one central location, making it easy to search for data, trace its origins, and control who has access. This ensures your organization stays compliant with data policies and audit-ready at all times.

Data Catalogue: Searchable inventories of your data

- **Catalog Everything -** Maintain complete metadata coverage for all datasets
- **Trace Data Journeys** - Track lineage, ownership, and usage patterns from source to destination**.**
- **Deploy Anywhere** - Support cloud, on-premise, and hybrid environments with consistent functionality**.**
- **Enable Self-Service** - Create a single source of truth that empowers teams to discover data independently.
- **High Visibility of Data Activity Trails for Auditability**– Clear view of what changes happened and who carried those actions to any metadata through simple web UI with no coding skills required.
  - o Assigned personnel can review any additions, deletions and modifications to metadata, including any configurations to data governance policies and process

Governance Center: a single and comprehensive snapshot view of all registered and scanned data within the data catalog. The real-time alert via email can be sent to notify a system administrator on data completeness, consistency and accuracy.

Business Glossary - A central hub for your business's business glossary, a data dictionary of that dataset where users can view the related metadata. helping everyone in an organization understand the meaning and business context of the information they are working with.

Data Asset Synopsis – A "read-me" capability to help users in getting the quick, essential information about the data asset in an article-like, easy to digest format.

Assets Discovery: organizations must thoroughly understand their data and its underlying context to get the most value from it. The powerful semantic search capability helps discover the most relevant and trusted data assets through a well-maintained organization-wide:

- Registered data within the database by a table, a column, data pipeline, lineage. Dataset preview is available.
- Data Lineage allows users to trace data flows, understand dependencies, and assess the potential impact of changes.
- Relevancy score – ensuring high-relevancy search results.

To unlock real value from data, organizations need to understand not just what they have, but what it means. CUBIKA Big Insights for Data Catalog's intelligent semantic and indexed search, no-code, user-friendly interface, makes this effortless—simply type what you're looking for in plain language, and discover the most relevant, trusted datasets, at the tip of your fingers.

Whether you're the sales manager, looking for a sales record or a data scientist, looking for your data to build analytics model, the data assets you're looking for is just one search away.

Data Assets Access & Permission Control: balancing security governance with user productivity, maintaining data controls without creating access bottlenecks through:

- **Role-Based Access Management**: Configure data asset permissions based on user roles and responsibilities, ensuring appropriate access levels.
- **Self-Service Access Requests**: Users can request access to restricted data assets through a web-based interface with configurable approval workflows for streamlined authorization.

Information Services Portal: provides robust data services that support seamless data integration and exchange both within and across organizations. The platform enables authorized users to access, query, and export data in standard formats such as CSV and Excel, based on role-based access control. This ensures secure, governed data sharing while maintaining flexibility for downstream use and reporting. In addition, the platform offers a user-friendly graphical user interface (GUI) that is accessible via modern web browsers—including Google Chrome, Firefox, and Safari.

Data Preview: Snippet of your data health

- **Secure Data Preview** - Preview sample data subsets without full dataset exposure and access
- **Data Profiling Insights** - View data profiling metrics including null values (count, percentage), outliers, frequencies, distinct count, data distributions such as min.,max.,avg. values, standard deviation, and statistical distributions in one dashboard
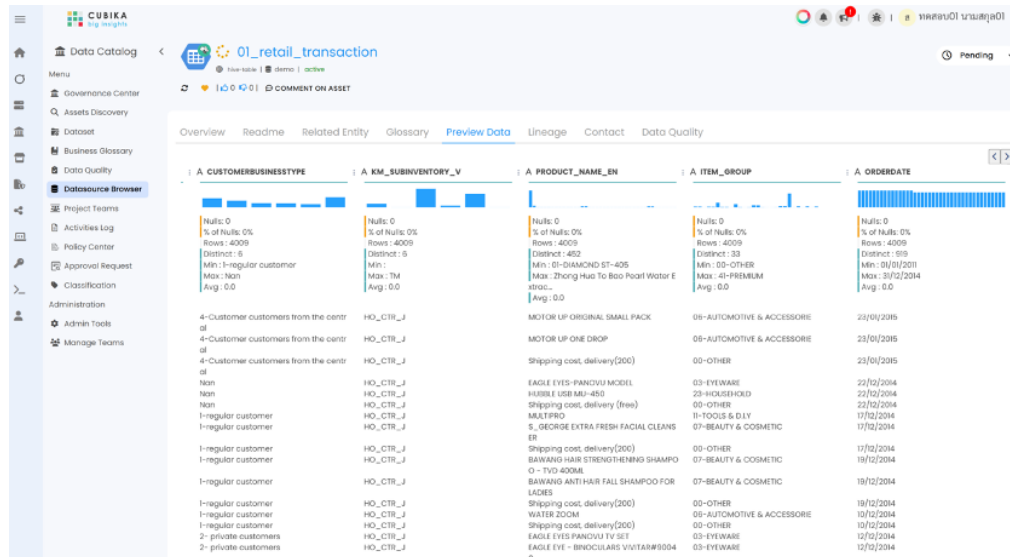

Illustration of Data Profiling

- **Anomaly Detection** - Identify data quality problems and anomalies before they impact processing or productions
- **Accelerated Data Selection** - Enable analysts and data scientists to quickly evaluate and select optimal datasets

Data Secure Insights: Locate PII and sensitive data

- **Automated PII Discovery** - Automatically scan and identify personally identifiable information across all data sources
- **Risk-Based Classification** - AI-detected, auto-tagged and data anonymization applied (hashing, masking, redacting) for PII (Personal Identifiable Information) such as social security number, ID card number, name or based on the organization's data security and privacy policy through CUBIKA Big Insights for Data Security
- **Policy-Driven Access Control** - Seamlessly integrate with existing data access policies and user role management systems through user logging, role-based access control (RBAC) through CUBIKA Big Insights for Data Security
- **Access-level Governance** – At data usage level, it controls who can access which data, how they can see it, use it and under what conditions through CUBIKA Big Insights for Data Security
- **Regulatory Compliance Support** - Ensure adherence to PDPA, GDPR, and industry-specific data protection standards

Data Standard Compliancy: Catalog while staying compliant

- **Enterprise Standards Alignment** - Ensure all metadata consistently follows organizational data standards and naming conventions
- **Automated Compliance Validation** - Verify schema naming, classification rules, and documentation practices against established guidelines
- **Empowered Data Leadership** - Provide Data Stewards and Data Protection Officers with comprehensive tools to govern data assets with confidence
- **Organization-Wide Data Trust** - Build transparency and reliability that gives all data users confidence in one single source of truth

CUBIKA Big Insights for Data Science Tool Packs
Data Refinery and Visualization, Ready for Business Usage

Low-code data exploration and visualization: explore crooks and crannies of your data to uncover hidden insights. Creating a data-driven visualized dashboard, where your imagination takes you, based on your organization's data. Be the Data Picasso, paint that insightful picture.

Credit: https://superset.apache.org/
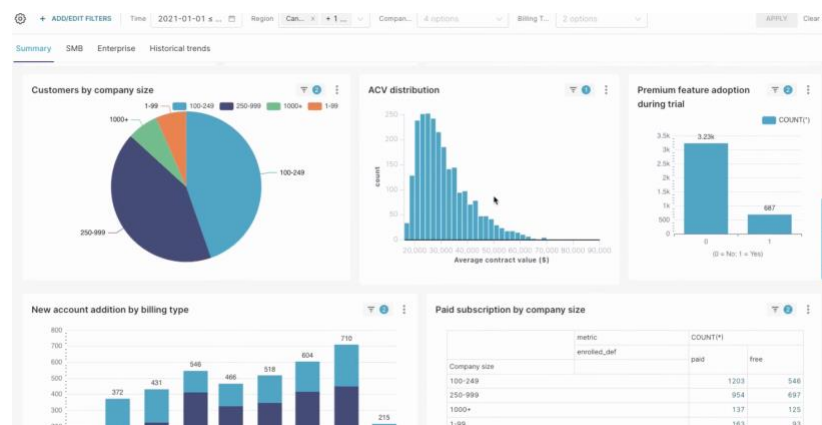
Driving a better outcome with refined data: a holistic and approach to data wrangling and refinery to prepare data before entering insight-generation zone or "consume" zone.

Hypothesis Lab: A "lab" for data science professionals to test their hypothesis and discover new data analytics model.
- Your choice of Web-based Notebook for Python code interactive testing to improve your organization's data science practice with Apache Zeppelin or Jupyter notebook.
- Support Machine learning engine for Classification, Regression, Clustering, Dimension Reduction, Model Selection, Preprocessing though R, Python, and Spark.
- Leverage **CUBIKA Big Insights Access and Storage** to execute queries through Spark-based frameworks (Open Massively Parallel Processing SQL engine) including PySpark, providing

the capability to process large-scale datasets with distributed computing power for enhanced performance and scalability.

Technical capabilities

- Supporting user authentication via Native user authentication, LDAP user authentication, Active Directory (AD), Microsoft Entra ID, and SSL/TLS, including managing user authentication mechanisms as below:

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.

Seamlessly integrated with Cubika Big Insights for Data Secure:

- Enabled data warehouse security.
- Providing capability for Role-based Access Control (RBAC) working in tandem with HDFS Access Controls (ACLs)
- 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.
- Supporting Data-in-Transit security with at least TLS 1.2 encryption
- Supporting Data-At-Rest security with AES-256

CBI Data Refiner

Clean Data   Transform Data   Convert Data Type   Detect redundant data

- Web-based GUI, providing user with Excel-like data view in column
- Powered by Thai NLP (Natural Language Processing), it can automatically detect data duplication and redundancy based on Thai phonetics. For example, the detection in the likelihood of ไกรสอน ไกรสร and ไกรศร as the same entity with different spelling.
- Compatible with column merging and splitting.
- Data error detection system during data cleansing process
- Supporting data transformation as below:
    - Text (string) to integer or number
    - Integer or number to text (string)
    - Text (string) to date format as below:a
        - dd/mm/year
        - year/mm/day
        - dd-mm-year
        - year-mm-dd
        - year/mm/dd
        - year/dd/mm
        - year/dd/mm
        - year-dd-mm
        - Supporting Thai format date such 31 ตุลาคม 2563 and 31 ตค. 2563 as the same entity.
- Supporting changing date to text(string)
- Supporting automatic letter case format, for example, uppercase, lowercase, capitalized word.
- Address format detection and formatting. Embedded with machine learning, the system can automatically detect if the zip code doesn't match the associated district.
- Activity logs are recorded, and redo process is possible.
- Exportable activity logs ins a template format for data cleansing and maintenance
- Data cleansing can be completed via
    - Parsing
    - Data correction
    - Data standardization
    - Data deduplication
- Data visualization supported by Apache Superset
    - Supporting drag-and-drop and SQL queries
    - Customizable charts and dashboards through CSS templates to match your brand's look and feel
    - Creating interactive dashboards faster through dashboard filters and Jinja templating

About CUBIKA Big Insights

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.

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

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-DMS-101-V1.2025

DIGITAL DIALOGUE

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

✉ **contact@ddlghq.com**

📱 **02-088-0795**