DIGITAL DIALOGUE

Data Integration

Data Sheet

Unify Your Data Landscape with Data Integration

In today's data-driven world, organizations are overwhelmed with vast amounts of data scattered across various sources, systems, and platforms. This creates a fragmented landscape where valuable insights remain trapped in the disconnection of data.

Digital Dialogue's Data integration solution acts as a bridge, connecting these disparate sources and enabling seamless data flow to insights generation. It empowers businesses to break down barriers, collecting, importing myriads of data sources, including real-time streaming data, and moving them along their data journey, to be stored, accessed, and analyzed.

With a good data integration solution, organizations can streamline data processes, ensuring data is ready to be analyzed for decision-making in a timely manner, ultimately fostering innovation, driving efficiency, and growth.

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 Data Integration solution is driven by our products under CUBIKA Big Insights family; CUBIKA Big Insights for Data Ingestion and CUBIKA FileGuard. 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

Low-code Integration: Create simple-to-sophisticated data integration projects with a highly intuitive Action Flow function for ingestion workflow and connection creations. Low-code Ingestion Data Stream function offers you "cross-function, cross-cloud" ingestion 10x faster than NiFi.*

Connects All Data with Agility: Easily connect to any databases, cloud data lakes, on-premises and SaaS apps, Web Services, ServiceNow, SAP S/4HANA, SOAP API, REST API, SFTP, ERP applications and any data warehouses and ingest data of any source, any pattern, at any latency at scale in real-time and batch with no data limit.

Full data integrity: Ensuring that your imported data quality is in its top form with Smart Data Assets Scanner function.

Highest Level of Security: "Security-by-Design" approach. All data and workload security are weaved into every stage of the data integration lifecycle with AES encryption for all of data at REST and in-transit by integrating a transport protocol like HTTPS (SSL, TLS) and SFTP to ensure your data is always guarded and secure.

Customizable Granular-level data security: CUBIKA FileGuard accommodates varying organizational security policies requirement through sophisticated file-level control mechanisms. Each file can be secure, utilizing AES-based dynamic encryption, providing customizable security protocols tailored to specific divisional requirements

Enhancing data democratization: CUBIKA Big Insights empowers to diverse users, from business analysts and data scientists to data stewards and citizen integrators. We provide a user-friendly, zero-coding environment, tailored to each role's specific needs, enabling active data usage and fostering collaboration and data-driven culture across your organization.

*Depends on the server capability/specification.

Scale as you grow: Digital Dialogue's Data Integration solution helps you with cost optimization with our dynamic pricing model based on ever-evolving business needs and requirements.

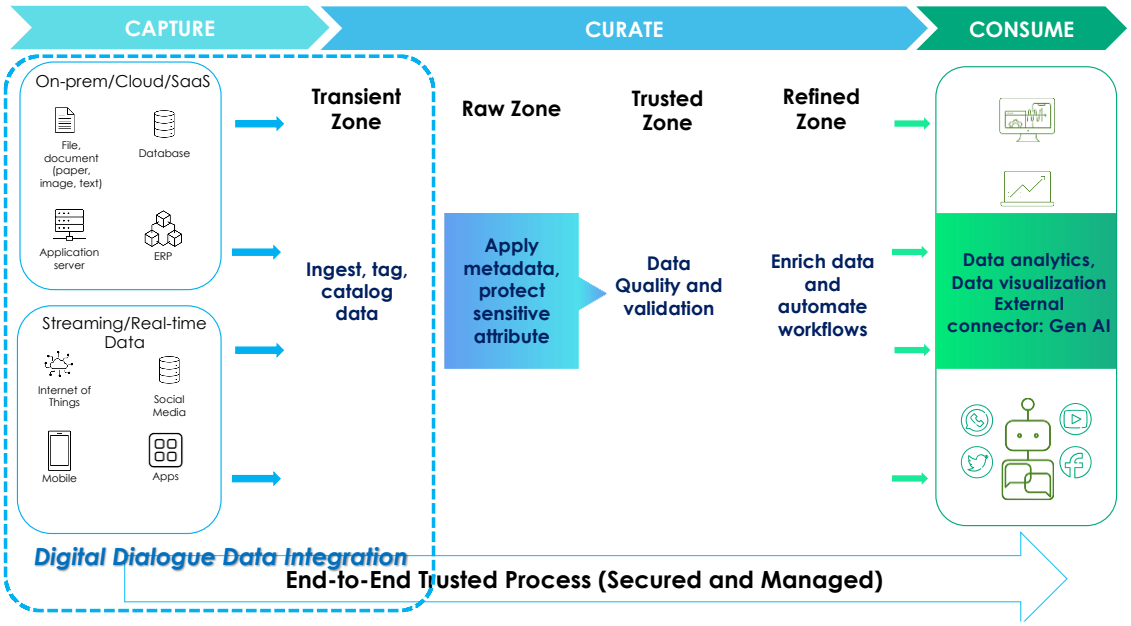


Figure 1: How Digital Dialogue's Data Integration Powers Your 3C Data Journey: Capture, Curate, Consume

Technical capabilities

CUBIKA Big Insights for Data Ingestion is an essential part of CAPTURE phase.

It is a a unified platform for managing and transforming data from various sources from structured, unstructured to stream data, with a focus on automation, low-code interface, and AI-driven functionalities. It supports both real-time and batch processing, offering flexible tools for technical and business users to build scalable data pipelines across databases, APIs, and files and manage them with high efficiency. Minimizing integration and data management complexity with all-in-one platform.

Capability Overview

CUBIKA Big Insights for Data Ingestion is engineered for maximum efficiency with minimal operational complexity. High technicalities, complex tasks can be complete with a few clicks or drag-and-drops through 4 key modules – Action Flow, Data Flow, Action Director, and SQL Lab.

Action Flow - The heart of CUBIKA Big Insights for Data Ingestion - a powerful yet simple tool that lets you build complete data pipelines without complex coding. Whether you're moving data between systems, transforming it for analysis, or running machine learning models, Action Flow adapts to your specific business needs. It combines AI-powered data mapping, SQL-friendly UI for technical aspects, drag-and-drop, low-code data pipeline management and external app integration in one intuitive platform, so you can focus on insights instead of infrastructure.

Search index: A full-text indexing capability for ingested data to support search functionality.

- Indexing both structured and semi-structured content
- Enabling rapid lookup, filtering, and retrieval of records
- Suitable for large-scale document search and data base
- Designed to work seamlessly with CUBIKA Search

Data Stream: An intuitive ETL interface for streaming or transferring data between different platforms or data sources.

- Supports database-to-database, file-to-database, and API-to-database flows and vice-versa
- Compatible with formats like CSV, JSON, Parquet, and common RDBMS
- Real-time and batch scheduling with manual trigger (RUN NOW)
- Displays execution history and flow health (Last Run, Last Update)

SQL Stream: SQL-based ETL tool that allows users to manipulate data directly using SQL commands through secure API and web-based GUI.

- Source and target must be database systems only such as MSSQL, MySQL, Hive)
- Supports SQL commands such DDL and DML, for example, CREATE TABLE, INSERT INTO, SELECT.
- AI-assisted suggestion and edition with syntax-highlighted editor and execution logs
- Ideal for advanced users who require precise query-level control

Model Flow: Orchestrates the execution of machine learning models for the data ingestion process.

- Supports integration of pre-trained ML models (classification, prediction, etc.)
- Designed for inference, scoring, and model-driven data enrichment

- Future support for model registry integration (e.g., MLFlow, JupyterHub)
- Secure and isolated execution environment for model runs

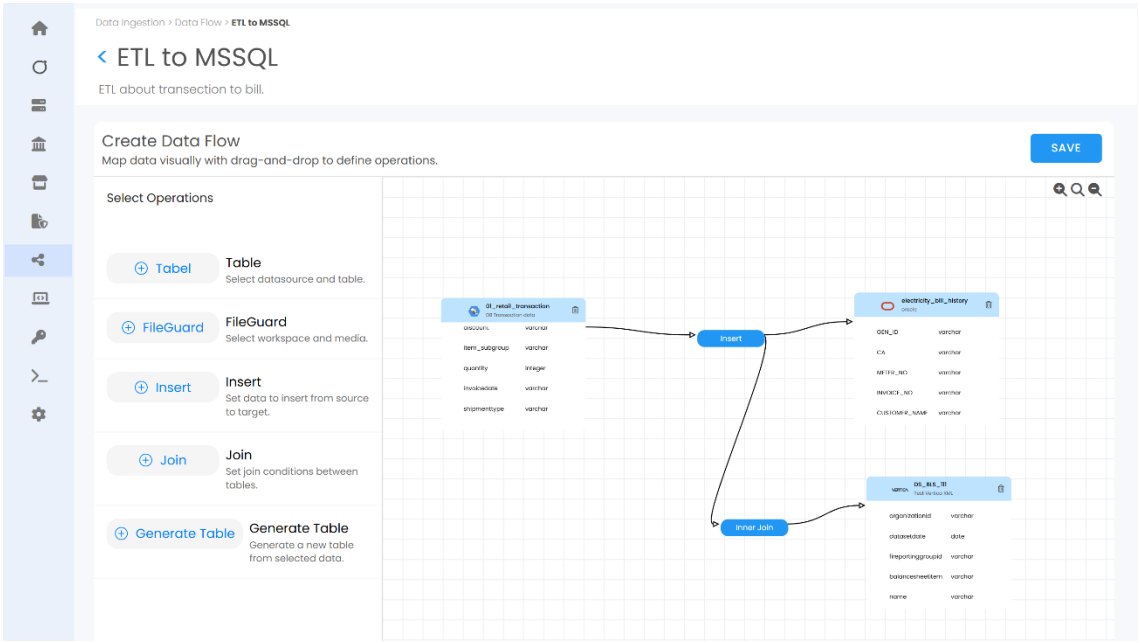
App Runner: A tool to execute external applications as part of the ingestion workflow.

- Supports Python 3.6 or higher, any third-party or custom scripts
- Enables external logic execution for data validation, transformation, or enrichment
- Suitable for integrating external APIs or invoking microservices
- Includes manual execution, output logging, and runtime monitoring

Data Flow – An interactive and intuitive, drag-and-drop data pipeline builder that enables users to design, configure, and execute data transformation logic easily but efficiently. Simplifying complex ETL tasks like building with modular components, you simply drag and drop data sources, connect transformation blocks, and assemble complete pipelines—no manual coding required.

It is intended for data engineers, analysts, and developers who prefer an intuitive, interactive workspace to model data relationships and generate new tables or views for data pipeline management.

Data Flow provides complete transparency into your data ingestion flow, making it easy to understand, modify, and troubleshoot pipelines at any stage.



Data Flow

Key Capabilities:

- **Drag-and-Drop Interface**
  Users can visually select operations such as Table selection, Join conditions, and Table generation using blocks and connectors. This reduces coding overhead and enhances accessibility for non-technical users.

- **Supporting comprehensive data operation to enhance and ensure data integrity and standards**
  - **Data transformation**: Convert and restructure data formats, schemas, and values to maximize compatibility with any applications through aggregation, column generation, enrichment, filtering, merging, pivoting, sorting operations
  - **Data cleansing:** Correct data quality issues through deduplication, inconsistencies and discrepancies removal, replacing with correct values, filling missing values, and error-free, high consistency data formatting

- **Table Selector Component**
  Allows users to select data sources and specific tables from connected databases. Metadata such as column names and data types are automatically displayed in the block.

- **Join Block**
  Supports various join types (e.g., Inner Join, Left Join) between tables. Join keys can be defined interactively by connecting compatible columns across tables. Each join condition is validated to ensure referential integrity.

- **Generate Table Operation**
  Once the transformation logic is defined (e.g., source tables + joins), users can create a new target table with a single block. The schema is inferred from upstream inputs, and users can modify column definitions if needed.

- **Preview and Execution Pipeline**
  Users can simulate and preview join results before executing. Final pipelines can be saved and executed, with outputs stored in the designated database.

Action Director – The built-in orchestration and scheduling engine behind data ingestion process.

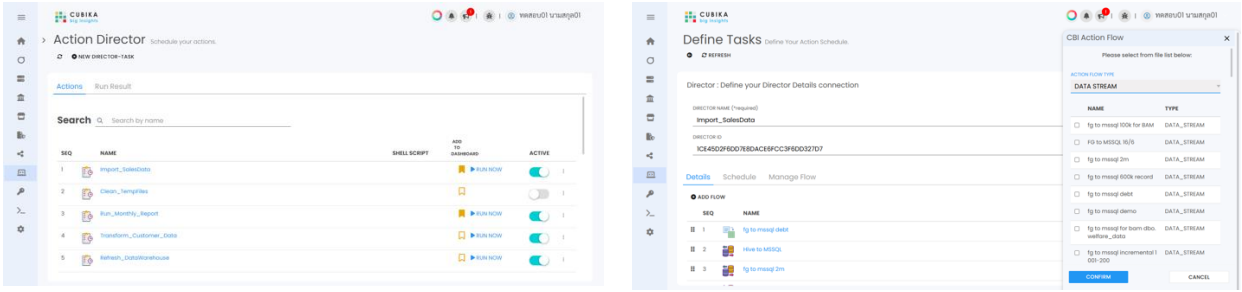
It lets you chain together multiple data processes, set flexible schedules, and monitor everything from one place—whether you're a data engineer building complex pipelines or an operations team managing and monitoring daily ingestion task executions.

Think of it as your workflow conductor: you define the sequence, schedule the timing, and Action Director handles the execution automatically. When issues arise, you get instant visibility and control to pause, restart, or troubleshoot without losing your work.

Action Director supports manual runs, event-based execution, and flexible scheduling while offering centralized monitoring, error handling, and task lifecycle management.

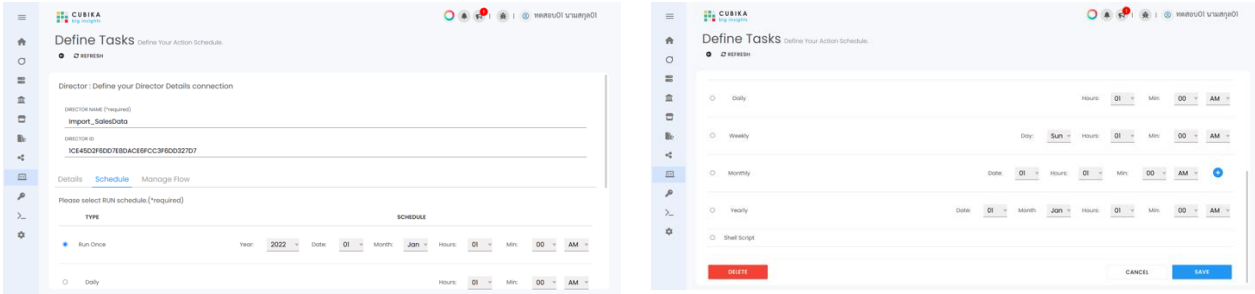
Key Capabilities:

- **Task configuration and deployment (Flow Orchestration)** - Connect multiple flow components, processes, applications and custom scripts such as Data Flow, SQL Stream, App Runner, Shell Scripts) into complete end-to-end pipelines that run exactly when and how you need them.



Example of task configuration and Deployment – Action Director Main View and New Data Pipeline Configuration

- **Flexible Scheduling Options and Deployment** - Supporting multiple schedule types:
  - **Run Once**
  - **Daily / Weekly / Monthly / Yearly**
  - Specific time or interval-based execution
  - **Shell Script Integration** for custom logic

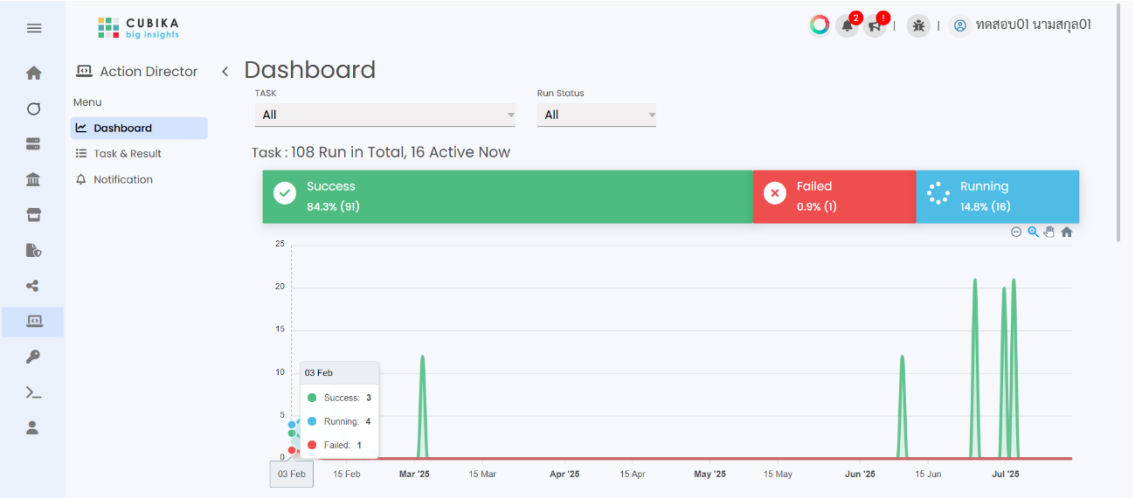


Example of task scheduling and deployment through an intuitive Action Director

- **Flow Lifecycle Control -** Pause or resume any task without rebuilding your workflow.
    - Activate or deactivate each task independently without deleting or recreating the flow
    - The option to **Stop** or **Continue** flows based on given conditions or the system administrator's decision
    - Supports versioning and updates without affecting runtime history

- **Manual Execution & Trigger Controls**
    - Tasks can be triggered immediately via **RUN NOW**
    - Useful for on-demand verification, backfill operations, or user-initiated jobs

- **Monitoring Everything**
  See the status of all your workflows at a glance through the comprehensive status dashboard, in-app notification or alerts via email when something needs attention, and drill down into details when troubleshooting.



Action Director's Center Monitoring Dashboard

Action Director is equipped with a **real-time monitoring panel** for all scheduled or manually triggered tasks. Rather than simply showing task configurations, this view is focused on the **outcomes of each execution**, providing operational transparency and actionable insights for system administrators, DevOps teams, and data engineers.

Key Features:

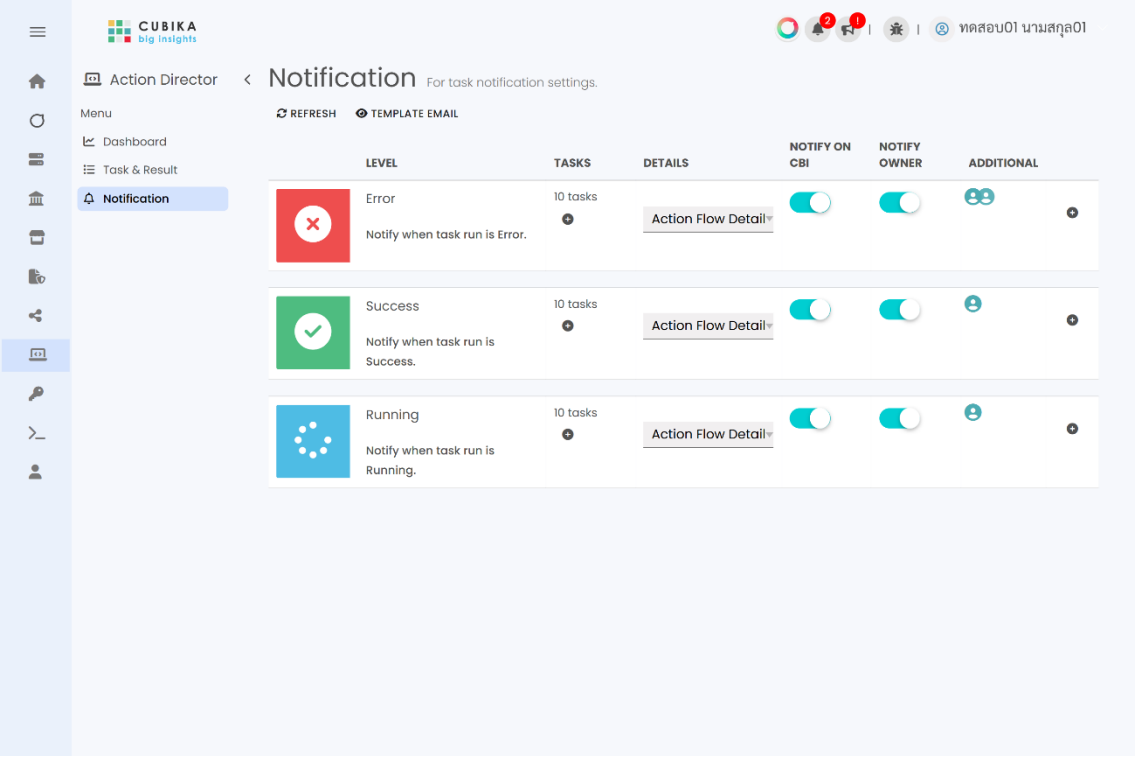
- **Visual Run Summary**
  Displays aggregated statistics on task outcomes:
    - **Success** – Total number and percentage of successfully completed runs

- **Failed** – Failed executions flagged for review
- **Running** – Ongoing processes that are currently executing

- **Interactive Run Trend Graph**
  Time-series visualization of execution trends, categorized by result status (Success, Running, Failed). Users can hover over data points for daily breakdowns.
- **Version Control for data pipeline**
  Business can configure and set the version controller for any data pipelines
- **Execution Count & Status Filtering**
  Drop-down filters allow users to:

  - View tasks by name or group
  - Filter results based on run status (All, Success, Failed, Running)

- **Error Tracking**
  Shows the **Latest 5 Error Tasks** to highlight failed executions. Includes task name, run date, and error context to support rapid troubleshooting.



Action Director's Notification Control Hub

Action Director's Notification Control Hub keeps your team instantly informed when executed actions succeed, fail, or need attention. The smart alerts configuration can be set for each tasks to deliver the right notification to the right people through in-platform notifications or email.

This enables proactive approach to prevent small issues from becoming big problems, ensuring your data operations run smoothly around the clock without constant manual monitoring.

Key Features:

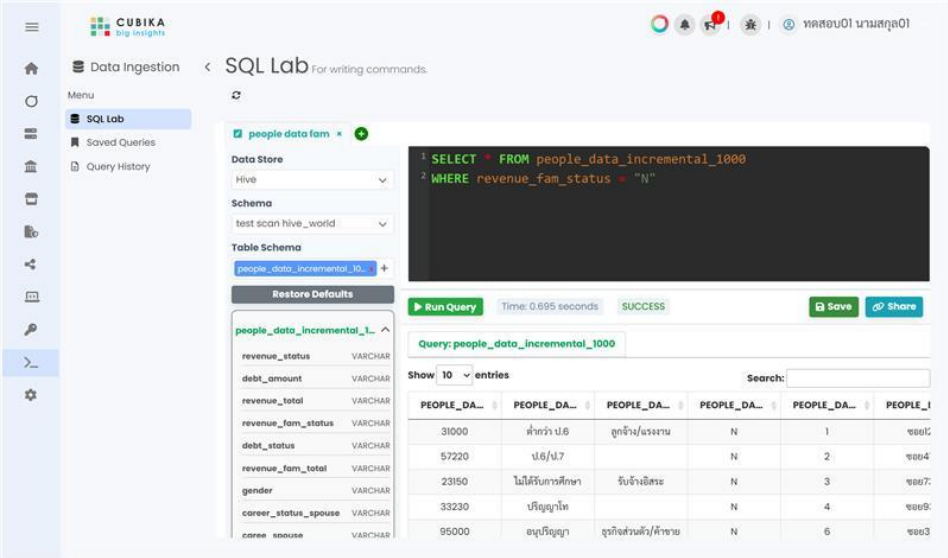
- **Notification by Status Type**
  Alerts can be set for:
  - **Success** – Notifies upon successful task completion
  - **Failed** – Notifies when a task fails or encounters execution errors
  - **Running** – Notifies when a task starts executing

- **Multi-Channel Delivery Options**
  - **In-platform** – Sends notifications to the in-platform notification center
  - **To Owner** – Sends alerts to the task creator or responsible user
  - **Additional Recipients** – Allows users to add additional email recipients or team members for broader alert coverage

- **Task-Level Notification Assignment**
  - Notification rule can be set for **specific tasks** (e.g., 10 tasks per status group)
  - Drop-down menu to define **details or context** included in the alert (e.g., Action Flow Details)

- **Email Notification Templates**
  - Customizable email templates can be configured for each status
  - Supports structured format by automatically populate task name, timestamp, error logs, and system environment

SQL Lab – Your testing ground for SQL queries—a built-in editor where you can write, test, and perfect SQL commands before putting them into production. SQL Lab helps you explore your databases, validate your logic, and see results instantly without affecting your live workflows.



SQL Lab

Key Features:

- **Interactive SQL Query Editor through web GUI**
  Developing and executing SQL commands with syntax highlighting, query formatting, real-time results, auto-completion and smart suggestions that speed up query development.
  - View historical SQL queries

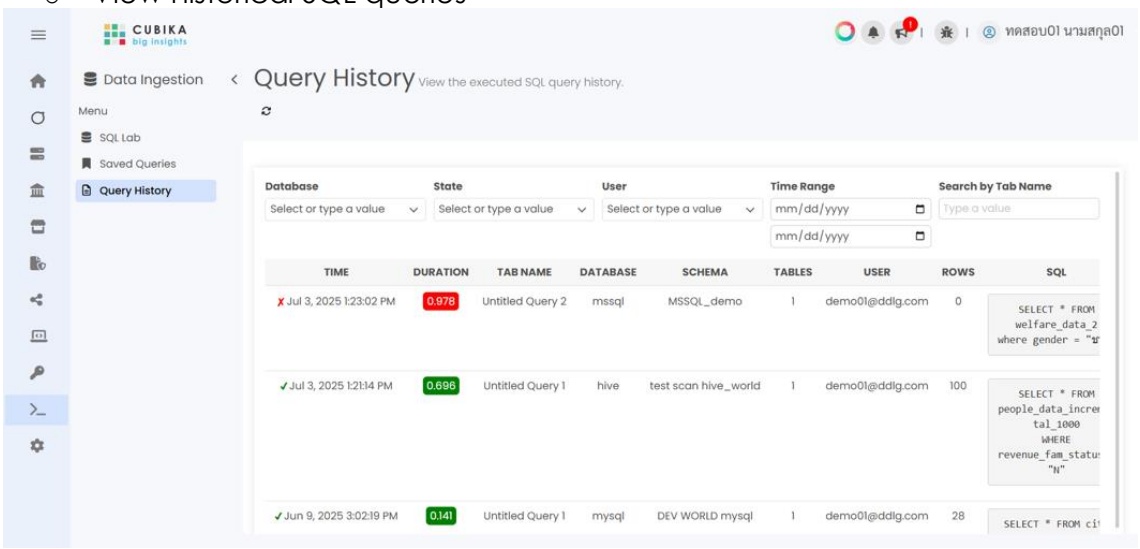


Illustration of Query History UI

  - Configure SQL queries version control
  - Simple grouping function for organizing SQL queries through an intuitive Visual Folder capability

- Time-saving - users can **save frequently used queries** for future use under the "Saved Queries" menu

- **Smart Data Store and Schema Selection**
  - Users can select a connected **data store** (e.g., MySQL, PostgreSQL, Hive) simply from a drop-down list
  - Automatically loading available **schemas and tables** for the selected data source
  - Displaying column metadata (name and type) for quick reference

- **Intelligent Autocomplete & Query Assistance**
  - Real-time suggestions for **table names, column fields, and SQL syntax**
  - Reducing error rates and accelerates query development for all user at any skill level

- **Instant Query Execution and Result Preview**
  Displaying query outputs immediately in a paginated, searchable table format, including execution time and success/failure messages to quickly validate your logic.

- **Collaboration Capabilities**
  - Pre-built and pre-approved SQL Queries can be **shared with teammates. Users can invite anyone to be part of collaborators and comment** to strengthen collaboration through simplified and user-friendly web interface UI.

Highlight capabilities

- Drag and Drop application to upload file via web-based GUI.
- Develop data pipeline process - ETL/ELT/Reverse ETL/Push Down with Hadoop Distribution. This enables data workflow management via a simple web-based interface that is easy to use and configure but provides powerful results.
- Automating data type validation between the data source and the target
- The system can save data type validation results for auditing.
- Create Action Flow by schedule job (Submit Job)
  - Support one-time, weekly, monthly, yearly job schedule
  - Support sequential creation
- Compatible with HDFS (Hadoop Distributed File System) in the following:
  - Supporting multiple data processing pipeline
  - Supporting control over data flow direction for big data platform from design, control, feedback to monitor

- List of compatible connectors, databases, business application software and technology providers:
  - Supporting data import and processing from Apache Hadoop with at least ANSI-92

- Supporting ODBC and JDBC database connection
- Supporting RAW, ORC, Avro, Parquet, Delta Lake and Apache Iceberg connection
- Compatible with major cloud service providers such as Amazon Web Services (AWS), Azure, Databricks, Google Cloud Platform (GCP), Microsoft Azure, Huawei Cloud, NT Cloud, Snowflake
- Compatible with database technology such as SQL, MySQL, MSSQL Server, Hive, MongoDB, MariaDB, DB2, Vertica, Oracle, Oracle ADW PostgreSQL.
- Working with cloud-based storage technology such as AWS S3, Blob storage, Azure Data Lake Storage Gen 2 (ADLS), Google Cloud Storage (GCS), Huawei OBS, and MinIO, Object Storage.
- Can be integrated with ERP, CRM business applications such as SAP S/4HANA, ServiceNow
- Supporting web services and API connection through REST API architecture and SOAP
- Supporting database indexing capability for search with Elasticsearch
- Compatible with structured data format such as Text File, CSV, Microsoft Excel, JSON, XML, Parquet, Avro

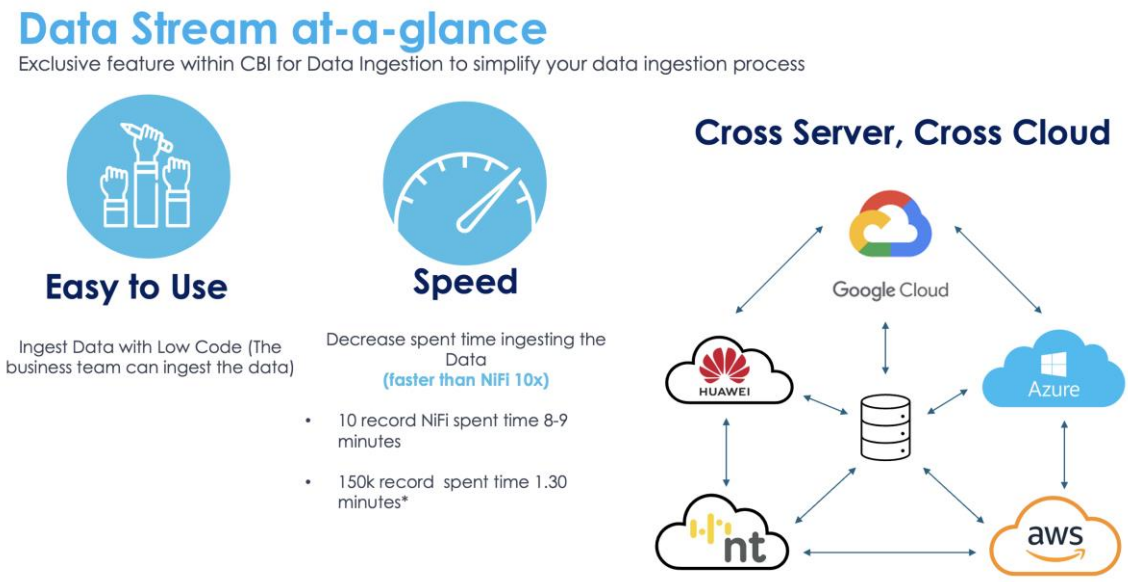
- List of compatible big data tools, technologies, and functions below:

HDFS (Hadoop Distributed File System)	A distributed file system for storing and retrieving structured, semi-structured and unstructured data, providing high throughput access data by providing the data access in parallel. Edit and create data category management and show in Browser Directory (One of File Explorer)
Apache Iceberg	An open-source **high-performance table format** for huge analytic datasets. Iceberg is designed to work on top of data lakes (like S3, HDFS, or Azure Blob) and bring **database-like capabilities** (ACID transactions, schema evolution, partitioning, etc.) to files like Parquet, ORC, and Avro.
Hadoop Management Console	Leveraging Apache Ambari as a tool for provisioning, managing and monitor Apache Hadoop cluster via easy-to-use web-based dashboard.
HBase	A non-relational database management system, supporting structured data in large table
Hive	A data warehouse software for writing, reading, managing large data sets in HDFS or HBase using SQL.
Kafka	A platform for building real-time data pipelines and streaming apps.

MQTT	Real-time & IoT data steaming Messenger
Action Flow	A powerful yet simple tool that lets you build complete data pipelines without complex coding. Whether you're moving data between systems, transforming it for analysis, or running machine learning models
Action Director	The built-in orchestration and scheduling engine behind data ingestion process.
SQL Stream	This allows users to manipulate data directly using SQL commands through secure API and web-based GUI.
Data Stream	An intuitive ETL interface for streaming or transferring data between different platforms or data sources.
App Runner	Run and test your written external command easily. Compatible with Python 3.6 or higher
Spark	A fast and unified analytics engine, capable of processing large sets of data including streaming, combination of batch processing, streaming processing with machine learning.
MapReduce	A YARN-based system for parallel processing of large data set.
SQL Lab	A built-in SQL editor within the CBI Data Ingestion platform, designed to help users test, explore, and validate SQL commands against connected data sources before integrating them into production workflows

- Applicable with Thai and English language by importing/exporting data, data processing, data analyze.
- Collecting joined data points within the single view of truth
- Compatible with any changes in data structure whether on the data structure itself or user requirement.
- Controlled schema version for the correct reporting
- Data Integrity Check in graph type:
    - Total Error and Total Success
    - Error Data Type
    - Start time, end type and duration

- Supporting both low-code and complex data ingestion in Action Flow
- Action Director – to orchestrate automated Action Flow. Action Flow acts as the "navigator", directing the flow of data ingestion process from start which is the dataset form data catalog to storage and access.
- Data processing works with operating systems with 64-bit CPU as indicated below:
    - CentOS
    - Ubuntu Linux
    - Windows Server
    - Red Hat Enterprise Linux
- Supporting real-time processing as follow:
    - Compatible with real-time transformation logic for both input and output data
    - Compatible with Real-time Pipeline and Real-time Streaming
    - Compatible with Distributed Messaging Queue
    - Distributed in cluster data management.
    - Resilient architecture with redundancy
    - High fault tolerance
    - High availability (HA)
    - Horizontal-node scalability
    - Support both Python, Java and Scala in transformation logic
    - Supporting publish and subscribe model
- Compatible with Python and Java Development Kit (JDK)8 64 bit
- Low-code, GUI compatible for drag-and-drop workflow creation. Supporting remote execution and job schedule automation.
- Supports multiple programming languages commonly used in data science, such as Python, R, SQL and Jupyter Notebook, Zeppelin
- Supporting direct read and write on data source from big data system such as Apache Hive and Impala
- Providing accurate performance record (read/write) such as working parameter and assigned KPIs.
- Supporting parallel executive & computing
- Supporting addition scripts
- Supporting additional libraries or plug-ins
- Supporting job queuing for higher efficiency
- Users can provide their own customed data processing pipelines via Java SDK.



Overview of Data Stream Functionalities

CUBIKA FileGuard helps enterprises in securing documents through encryptions from storage to access management that comes with the document automation and search capability, a system built with best-in-class security. With CUBIKA FileGuard, your documents are encrypted and stored with the highest level of protection available. Stored documents are automatically protected based on company's security policy.

- Ensuring protection of document with secure login
- Preventing unauthorized access with file encryption
- File scrambling for de-identification
- Search function, compatible with Adobe Acrobat file format
- Highly secure login process through username and password. Password is securely stored in the database with the One-Way Hash capability. If the information in the database is copied, the system will prevent reverse password and password decryption. Also working with Single Sign-On system such as Azure Entra ID.
- Compatible with various file formats for storing such as Text File, .pdf, .docx, .xlsx, .csv, .json, .xml, .pttx, .jpg, .tiff, .bmp, .mp4, .mpg, .wmv
- Automatic and customizable document properties and metadata record such as document author, who makes the latest revision, creation date, modified date.
- Make sure that your files stay safe with our AES based file encryption for preventing unauthorized access and tampering with dynamic keys.

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.

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-INT-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