

Trf Retrieval

Tuesday, August 26, 2025 11:00 PM

- 21个大的view
- **asriclickthru** 是一种机制或功能，用于将结构化财务数据（如 Fundamental 或 FinancialStatementLineItem）与其原始来源（如 SEC 报表、年报）进行链接。
- 内容分析师通过它可以验证每一个数据点的来源，确保数据的准确性和合规性。
- 这对于金融数据产品至关重要，尤其是在审计、监管或高精度分析场景中。
- 对终端用户（如投资者、分析师）而言，点击某个财务指标时能直接跳转到原始报表的具体位置（如 10-K 第 42 页），极大地提升了数据的透明度和信任度。
- 内容分析师的工作就是确保这种跳转是准确的，标签是清晰的，语义是正确的。
- **asriclickthru** 的存在使得内容分析师可以更容易地发现数据异常、标注错误或来源不一致的问题。
- 在数据标注和机器学习训练中，它也可以作为“监督信号”，帮助模型理解财务语境和结构。

Trf category

Tuesday, August 26, 2025 11:05 PM

TRF Retrieval (数据检索) 21个大的view

10 billion 1.4TB 150kb per data

核心功能：用于精确获取某个公司、某个时间点、某个字段的财务数据或文档。

内容分析意义：

- 点对点查询：适合内容分析师在处理特定公司或特定报表时，快速定位和验证数据。
- 数据溯源：结合 FinancialSourceBookMark，可以直接跳转到原始报表位置，验证数据准确性。
- 适合审计和验证流程：Retrieval 是内容审核和数据质量控制的关键工具。

示例用途：

- 获取某公司 2023 年 Q2 的净利润原始值。
- 下载某个 10-K 报表的 PDF 文件。
- 验证某个 Fundamental 数据是否与原始报表一致。

TRF Screener (数据筛选) 18个表

核心功能：用于批量筛选符合某些条件的公司或数据项，支持投资分析、内容分类、数据挖掘。

内容分析意义：

- 批量分析与分类：适合内容分析师进行行业分析、主题研究、数据标注。
- 构建数据集：用于构建训练集、测试集或专题数据集（如 ESG 高评分公司、 $ROE > 15\%$ 的公司）。
- 支持自动化标注：结合 NLP 或机器学习模型，筛选结果可用于自动化内容分类。

示例用途：

- 筛选所有 $ROE > 20\%$ 、市值 $> 10B$ 的科技公司。
- 找出所有在过去 3 年中 EPS 连续增长的公司。
- 构建一个包含财务稳健企业的分析样本。

功能	TRF Retrieval	TRF Screener
目标	精确获取单点数据	批量筛选符合条件的数据
用途	数据验证、溯源	数据挖掘、分析、分类
内容分析价值	高度准确性、审计支持	高效构建数据集、支持自动化

接口类型

[REST API / 文档下载](#)

[API / 筛选器表达式](#)

FundamentalAnalyticItemCurnKeysplit 的意义

1. 支持筛选器表达式的构建

- 将指标定义（Analytic）与其时间序列（Series）和计算逻辑（Relationship）结合起来，使内容分析师可以构建复杂的筛选条件。
- 例如：筛选 $ROE > 15\%$ 且 EPS 连续增长”的公司。

2. 实现指标的语义分割与映射

- CurnKeysplit 表示对指标进行“货币单位”、“周期”、“行业”等维度的拆分，使得筛选器可以精确控制筛选范围。
- 例如：只筛选以美元计价的市盈率或科技行业的资产回报率”。

3. 支持内容分类与自动化标注

- 内容分析师可以利用该表将公司或数据项自动归类到某个分析标签下（如“高成长”、“低估值”）。
- 这对于构建智能内容推荐或专题分析非常关键。

应用场景示例

- 构建一个 Screener 表达式： $ROE > 15\% \text{ AND } MarketCap > 10B$
- ROE 来自 FundamentalAnalytic
- 15% 是筛选条件，映射到 FundamentalSeriesV2
- $MarketCap$ 与 ROE 的关系由 Relationship 定义
- CurnKeysplit 控制单位和周期（如 FY2023）

在 TRF Screener 的数据结构中，表 FundamentalAnalyticItemCurnKeysplit 是一个用于支持筛选器表达式和分析逻辑的关键表，它通过以下三个核心表进行关联构建：

表结构组成与内容分析意义

表名	内容分析师视角下的作用
FundamentalAnalytic	定义了分析维度和指标的语义，例如“市盈率”、“资产回报率”等，是筛选器的核心指标库。
Relationship	建立指标之间的逻辑关系，例如“ ROE 是净利润与股东权益的比值”，用于支持表达式计算和筛选逻辑。
FundamentalSeriesV2	提供指标的时间序列数据，支持跨期筛选（如“过去三年 EPS 连续增长”），是 Screener 的数据基础。

Eikon and data plantform

Wednesday, August 27, 2025 3:47 AM

核心区别：Eikon vs Refinitiv Data Platform

特性	Refinitiv Eikon	Refinitiv Data Platform (RDP)
定位	桌面终端，面向金融专业人士	云平台，面向开发者和企业系统
访问方式	GUI界面 + Eikon Data API	REST API / WebSocket / Python SDK
数据覆盖	实时市场数据、新闻、分析工具	实时数据、历史数据、ESG、AI优化数据等
使用者	交易员、分析师、投资经理	数据工程师、开发者、企业系统
部署方式	本地桌面应用	云端或本地集成
未来发展方向	逐步整合进 LSEG Workspace	战略重点平台，持续扩展功能

Refinitiv Eikon 的功能与使用场景

- 实时市场数据：股票、债券、外汇、商品等
- 新闻与研究：整合路透新闻和分析报告
- 图表与分析工具：高级技术分析、回测、场景分析
- 可定制仪表盘：用户可根据需求配置界面
- 集成交易系统：与交易平台无缝连接
- 适合人群：金融分析师、投资经理、交易员

④ Refinitiv Data Platform 的功能与使用场景

- 统一数据访问接口：REST API、WebSocket、Python SDK（如 refinitiv-data, lseg-data）
- 数据类型丰富：
- 实时数据
- 历史数据（Tick History）
- ESG数据
- 公司基本面
- 宏观经济指标
- 文本分析（如财报、电话会议）
- AI与自动化支持：
- 用于训练模型、自动化交易、风险分析
- 与 LangGraph、BigQuery 等工具集成
- 适合人群：数据科学家、开发者、企业数据团队

VectorWise

Wednesday, August 27, 2025 3:53 AM

VectorWise 是一个专为高性能分析任务设计的列式关系型数据库系统，现由 Actian 公司维护，产品名为 Actian Vector

1

。它源自荷兰 CWI 的 X100 项目，强调 向量化查询执行 和 压缩列存储，在 TPC-H 基准测试中表现优异。

核心技术特点

1. 向量化查询执行 (Vectorized Execution) :

- 数据以 “向量” 形式处理，每次处理多个值，充分利用 CPU 缓存和 SIMD 指令集
- 2
◦
- 比传统 “逐行处理” 方式更高效，尤其适合 OLAP 查询。

2. 列式存储 (Columnar Storage) :

- 数据按列存储，便于压缩和按需读取。
- 支持块级过滤 (min-max 索引)，减少 I/O

3

◦

3. 压缩与延迟解压 (Late Decompression) :

- 每列可使用不同压缩算法 (如 LZ4)，提高压缩率和查询效率
- 4
◦

4. Positional Delta Trees (PDTs):

- 一种专利结构，用于支持小规模事务更新，避免频繁重写主存储
- 1
◦

5. 混合存储模型:

- 支持传统 Ingres 行存储与 VectorWise 列存储并存，适应不同工作负载

性能与使用场景

- 数据仓库与 BI 分析：
- 适合处理大规模数据集，支持复杂联接、聚合、Top-N 查询等

6

◦

- 报表与数据挖掘：
- 在 DBT-3 基准测试中表现优异，适合电商、金融、供应链等场景。
- 替代传统数据库：
- 在某些场景下比 SQL Server、Redshift、Snowflake 更快

7

Cat Content sets

Wednesday, August 27, 2025 3:55 AM

数据集名	内容类型	覆盖范围	主要用途
dealsmn a	并购交易 (M&A)	全球, 自1970年起	并购分析、交易结构、顾问排名 1
dealsni	新发行 (New Issues)	全球, 自2000年起	IPO、债券发行、承销分析 2
dealspe	私募股权 (Private Equity)	全球, 自1970年起	PE/VC投资、基金、退出、募资分析 3
dealsmu ni	市政债发行 (Municipal New Issues)	北美, 自1960年起	市政债券发行、承销商、评级分析 4
dealspjf	项目融资 (Project Finance)	通常涵盖基础设施、能源等	项目融资结构、贷款方、期限等 (未找到详细页面)
dealsrep	股票回购 (Repurchase)	企业回购交易记录	回购行为分析、市场影响 (未找到详细页面)
dealsloa ns	贷款市场 (Syndicated Loans)	全球	联贷、杠杆贷款、贷款条款分析 5
esg	环境、社会、治理评分	12,000+公司, 76国, 自2002年起	可持续投资、ESG筛选、争议分析 6
trf	标准化财务数据 (TR.F.*)	全球上市公司	财务建模、因子分析、估值模型 7

- **dealscompany** — 交易参与方信息
 - 内容：记录参与交易的公司信息，包括买方、卖方、目标公司、顾问等。
- **dealsauxiliary** — 辅助交易属性信息
 - 内容：提供交易的补充属性信息，如交易状态、交易结构、是否涉及交叉持股、是否为杠杆收购等。

dealsdir — 交易目录与分类信息

- 内容：为交易数据提供目录结构和分类标签，便于数据组织和检索。

More than 30 content set

Star schema

Wednesday, August 27, 2025 4:00 AM

星型模型由一个**事实表（Fact Table）和多个维度表（Dimension Tables）**组成

- Content data sets come from the manual / auto information inputs by content analysts, and are transformed into the formats for feeds and Eikon, including join, filter, picker, normalization, and so on.
- There are many challenges to implement such transformation:
- Every data set needs a lot of development and test works. The normal time to market for each set is 6 months or even longer
- Many similar works are duplicated among the data sets with different implementations
- It is very difficult to make the transform scalable to support TB level data sets
- Many content data sets are hierarchical, such as JSON and XML, transformed from unstructured data such as PDF reports, news, deals, and so on.
- Hierarchical data has many advantages than flat data (such as table), including flexibility, extendibility, self-contained, storage efficiency, and so on.

- However, it is much harder to transform between hierarchical data than flat data
- For batch processing, SQL can handle the transformation of flat data well. But the solutions for hierarchical data are not mature enough yet, including JSONiq, SQL++, ...
- For streaming updates, even SQL does not fully support it on flat data, such as Flink and Calcite. It is much more difficult to support if on hierarchical data.

• **Binary Store**

Purpose

- Convert structured, semi-structured
-  CAT Framework...
- data into structured data
- Generate meta data from accumulated raw data
- Store data into HBase for easy random access
- Generate strong typed data by annotation

Dsa conent set

Wednesday, August 27, 2025 4:26 AM

GreenRevenue

来自 FTSE Russell 的 [Green Revenues Data Model](#)

1

:

- 用于评估公司从绿色产品和服务中获得的收入。
- 覆盖全球 19,000 多家上市公司，提供微观行业分类（10个绿色行业、64个子行业、133个微行业）。
- 支持 EU Taxonomy 对绿色经济活动的分类和合规性评估。
- 在 EDS 架构中，GreenRevenue 数据可能作为一种 内容源，通过 CAT Framework 进行结构化处理并进入 Binary Store 或 Derived Store。

EU Taxonomy

来自欧盟的 [EU Taxonomy Regulation](#)

2

:

- 是欧盟可持续金融框架的核心，用于定义哪些经济活动是“环境可持续”的。
- 包括技术筛选标准（Technical Screening Criteria）和 “Do No Significant Harm” (DNSH) 原则。
- 在 CAT Framework 中，EU Taxonomy 可作为 数据标准或合规规则，用于数据筛选、标签化和报告生成。

StockReportPlus

来自 LSEG 的 [Stock Reports Plus](#)

3

:

- 提供股票、行业、投资组合和市场层面的量化分析。
- 包括盈利趋势、估值、价格动量、风险、内部交易等评分。
- 支持 50 个市场、覆盖 37,000 家公司。
- 在 CAT Framework 中，StockReportPlus 可能作为 内容消费者，从 Derived Store 中获取结构化数据用于生成报告或分析视图。

ICW (Indices, Constituents and Weightings) 是 LSEG (伦敦证券交易所集团) 提供的一项关键数据服务，专注于指数的构成和权重信息

ICW 的核心内容

1. Index Composition (指数构成)

- 提供全球主要指数的成分股列表。
- 覆盖主流指数提供商，如 MSCI、FTSE Russell、S&P Dow Jones 等。

2. Weightings (权重信息)

- 每个成分股在指数中的权重。
- 支持实时、日终、历史数据（可追溯至 1950 年）。

3. 数据格式与交付方式

- 支持 CSV、JSON、XML、SQL 等格式。
- 可通过 API、FTP、SFTP、桌面工具等方式访问。
- 可部署在本地、云端（Azure、Snowflake）或虚拟机环境中。

dsa

Wednesday, August 27, 2025 4:28 AM

New version of cat

8 content set

lcw

Mcap

Mifid,

Starmine

Stockreportplus

Greenrevenue

ICW-IndexAnalytics-Analytics 是 LSEG (伦敦证券交易所集团) 旗下的一一个数据产品模块，属于 Indices, Constituents and Weightings (ICW) 系列，专注于指数分析和构成数据的深度挖掘。它在 CAT Framework 或 EDS 架构中通常作为一个分析层模块，用于支持投资研究、筛选器、报告生成等功能。

ICW IndexAnalytics 的核心功能

1. 指数构成分析
 - 提供全球主要指数的成分股列表。
 - 支持主流指数提供商（如 MSCI、FTSE Russell、S&P Dow Jones）的数据。
2. 权重分析
 - 每个成分股在指数中的权重，支持实时和历史数据。
 - 可用于估值、回测、绩效分析等。
3. 时间序列分析
 - 提供约 50 万个指数实体的日终数据，历史可追溯至 1950 年。
 - 支持趋势分析、基准比较、图表生成等。
4. 数据格式与交付方式
 - 支持 CSV、JSON、XML、SQL、Python 等格式。
 - 可通过 API、FTP、SFTP、桌面工具、云平台（Azure、Snowflake）访问。

Wednesday, August 27, 2025 4:39 AM

Dealsni join

Wednesday, August 27, 2025 4:55 AM

你提到的字段 `ni_1m_japan_shelf_filings_amt` 是一个指标字段，通常用于表示某一时间段内日本市场的 Shelf Filings（货架注册）金额。根据你的说明，它是通过对以下两个数据表进行 Join 操作 得到的：

◆ 数据来源表

1. NewIssuesBondTrancheInfo

- 包含债券发行的分期信息（Tranche），如金额、发行日期、利率、期限等。
- 每个 Tranche 是一个债券发行包中的子部分。

2. NewIssuesBondPackageInfo

- 包含债券发行包的整体信息，如发行人、注册类型（Shelf/Non-Shelf）、市场区域（如日本）、发行状态等。
 - 一个 Package 可以包含多个 Tranche。
-
- `ni_1m_japan_shelf_filings_amt` 表示过去一个月内，在日本市场通过 Shelf Registration 方式发行的债券总金额。
 - 该指标常用于：
 - 市场活跃度分析
 - 监管报告
 - 投资者情绪评估
 - 与 EU Taxonomy 或 GreenRevenue 数据结合，评估绿色债券发行情况

Dealsni picker

Wednesday, August 27, 2025 5:00 AM

dealsnicouponpayment

- 数据表: NewIssuesBondTrancheInfo
- 字段提取:
- **DealPackageID**: 标识债券发行包，便于分组和归类。
- **SdcTrancheID**: 标识具体的债券分期，用于精确定位。
- **CouponPaymentDateMonthDay**: 表示利息支付的具体日期（按月日格式），用于时间分析。

分析应用场景

1. 现金流预测
 - 识别未来某日的利息支付总额。
 - 构建时间序列图或热力图。
2. 债券筛选器
 - 筛选具有特定支付周期的债券（如每年 3 月 15 日支付）。
3. 报告生成
 - 在 StockReportPlus 中展示债券支付日信息。
4. 合规分析
 - 判断支付结构是否符合绿色金融或 EU Taxonomy 要求。