

# 宽拓资讯数据整合系统 白皮书

## Financial Data Synthesizer

北京宽拓智融科技有限公司

2020

BEIJING QUANTEX FINTECH CO.,LTD.

---

## 目 录

一、FDS 背景和概述 .....	4
1.1 行业背景 .....	4
1.2 FDS 概述 .....	5
二、FDS 结构 .....	5
2.1 FDS 整体数据架构 .....	5
2.2 FDS 内部架构 .....	6
2.3 ODS 层: .....	6
2.4 Einfo 层: .....	6
2.5 Info 层: .....	6
2.6 数据维护工具: .....	7
三、FDS 功能列表 .....	7
3.1 多源对应 .....	8
3.2 统一模型 .....	8
3.3 多源整合 .....	8
3.4 数据校验 .....	8
3.5 事件驱动 .....	9
四、FDS 优势 .....	9
4.1 实现资讯统一管控 .....	10
4.2 全面数据校验 .....	10
4.3 提升资讯数据覆盖面 .....	10

---

4.4 择优选数，提升数据使用效率 .....	10
4.5 数据跑批更新时效提升 .....	10
4.6 帮助提升数据清洗能力 .....	11
五、FDS 部署环境要求.....	12

## 一、FDS 背景和概述

### 1.1 行业背景

系统是骨骼器官，数据是血液。金融行业对于数据使用逐步转变，从数据可有可无模式转到凭数据说话模式，但这个阶段对于数据要求仍是有数可用即可。当前又从有数可用模式转向精选数据使用模式。

目前资管场景下的交易风控环节，通过精选数据，校验数据方式提升数据准确性和可靠性尤为关键。比如，人工录入标的的数据和数据商资讯数据未实现统一管控，会直接导致数据更新问题，数据口径差异；风控限额的把控完全可能因为单个券的发行主体数据错误而流产；完美的分析模型完全可能因为数据不完整而影响模型效果，从而影响投资结果。

整体来看，金融机构在资讯数据使用时，会有一些共同的痛点，从行业调研的资讯数据的使用问题来看，大致可归为以下几类：

- 缺少资讯数据管控
- 数据覆盖面不足
- 数据使用效率低
- 数据错误被动发现
- 数据质量提升投入成本高

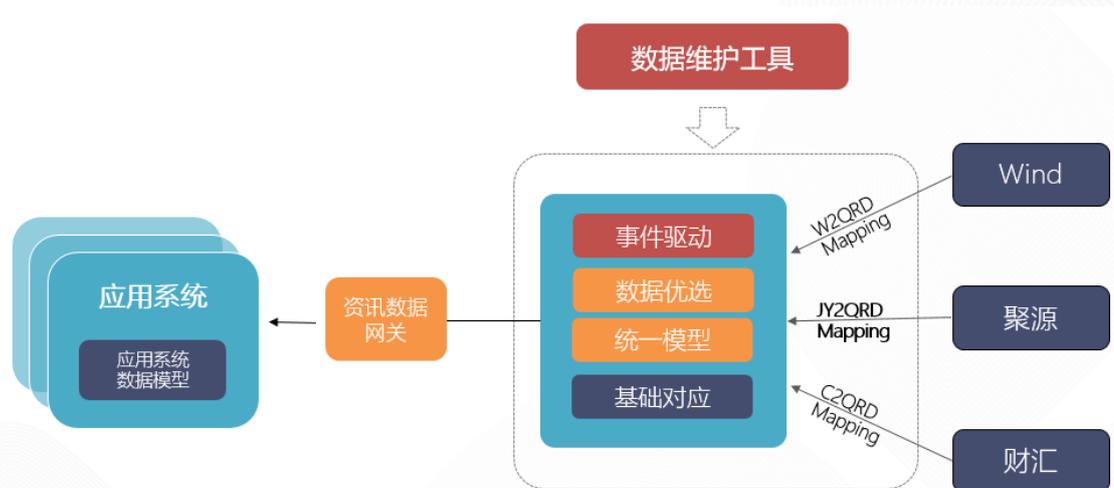
## 1.2 FDS 概述

**FDS (Financial Data Synthesizer)**是宽拓科技开发的主资讯数据整合系统，能够帮助资产管理机构实现多套资讯数据源整合、转换、校验等管理工作，并增补宽拓另类数据，最终输出一套高质量的、标准化的资讯主数据，解决资管机构多套资讯数据不一致的问题，全面提升金融资讯数据的质量、使用效率以及数据覆盖面，为核心业务提供基础数据支撑能力。FDS 的定位为：打造统一资讯中心，实现的主要途径为：资讯数据统一管控，数据整合和提升数据质量。

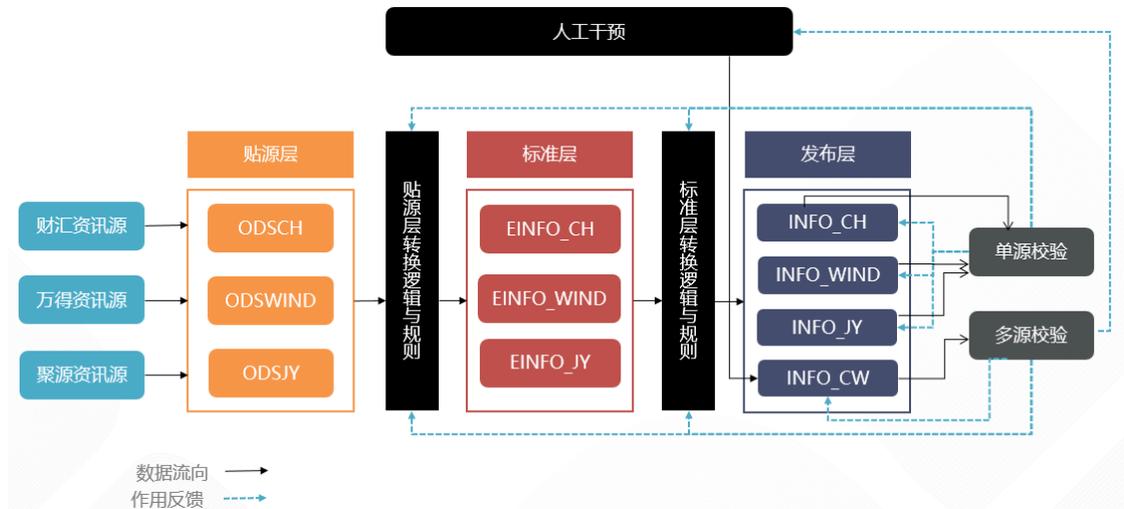
## 二、FDS 结构

FDS 采用面向对象的动态脚本语言 Python、开源 ETL 工具 Kettle、PL/SQL 结构化语言实现数据抽取、处理、加载。整个内部架构分三层，分别为贴源层 (ODS 层)、标准层 (Einfo)、发布层 (Info)。

### 2.1 FDS 整体数据架构



## 2.2 FDS 内部架构



### 2.3 ODS 层:

贴源层负责存储从数据厂商源库抽取过来的数据，数据模型、码值和数据粒度等都和厂商一致，实现源数据汇集的功能。

### 2.4 Einfo 层:

标准层负责将贴源层的数据通过 Mapping 转化成 FDS 数据模型，FDS 的码值，实际实现的是模型统一化和码值统一化，每个数据厂商对应一个标准层节点。

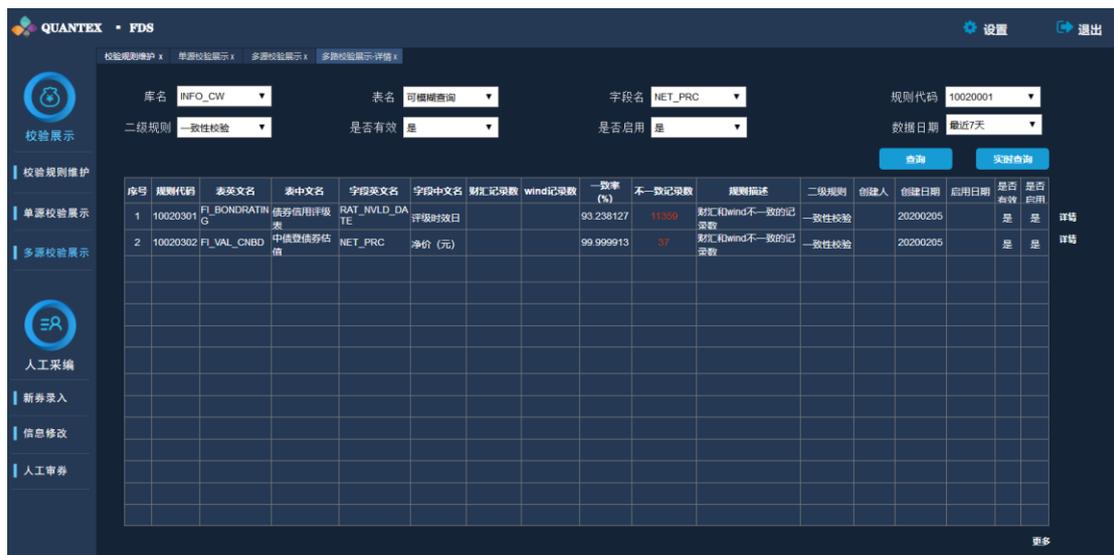
### 2.5 Info 层:

负责将各个标准层的数据进行多源对应，以及通过数据优选的方式实现多源整合，最终发布给下游用户使用，发布层通常只有一个核心节点和多个备用节点，其中每个数据厂商对应一个备用节点。

## 2.6 数据维护工具：

数据维护工具，即 FDS 配套的可视化工具，可实现数据校验规则的维护，校验结果的展示，问题数据的修改，以及人工合券等功能，是数据精细化作业的必备工具。

FDS 数据维护工具图



序号	规则代码	表英文名	表中文名	字段英文名	字段中文名	财经记录数	wind记录数	一致率 (%)	不一致记录数	规则描述	二级规则	创建人	创建日期	启用日期	是否启用	是否启用
1	10020301	FI_BONDRATING	债券信用评级表	RAT_NVL.D_DATE	评级时效日			93.238127	11309	财经和wind不一致的记录数	一致性校验		20200205		是	是
2	10020302	FI_VAL_CNBD	中债壹债估值	NET_PRC	净价 (元)			99.999913	37	财经和wind不一致的记录数	一致性校验		20200205		是	是

## 三、FDS 功能列表

FDS 通过实现多个资讯数据源的内码体系对应，然后构建专业化的资讯业务模型，整合过程中按每个业务模块去选择最优数据，再对整合的数据进行数据校验，事后对可疑数据提供数据预警，并可提供查询、确认、修正错误数据的服务。

### 3.1 多源对应

建立多个数据源证券和机构的内码对应，保证跨数据源选取数据应用时能很好衔接，这是实现数据整合的第一步，主要包含两个细分项

- (1) 多数据源证券内部编码建立对应；
- (2) 多数据源主体（或机构，甚至个人）内部编码的建立对应；

### 3.2 统一模型

提供专业化的，多个源统一的资讯模型。统一模型是数据中台无法规避的核心工作之一，也便于后期的数据校验、数据治理、数据切换等工作；

### 3.3 多源整合

在前面功能基础上，可以实现数据整合工作，按业务块择优选取数据，数据优选狭义的理解是按业务块，或者按表选取优势数据，广义的含义比较复杂，可包含多种优选方式，包括先到先入、交叉互补、指定单源、按券选源等。

### 3.4 数据校验

包括单一数据源的数据校验，校验细分类型有时效性校验，完整性校验，准确性校验等；还有多路数据比对（即多源校验），该项是数据整合校验中的核心校验方式，是对于多数据源细到字段级的差异化比较，能最大程度判定多

数据源的数据差异，对于高质量要求的资讯数据是不二之选，另外在数据流转过程中，要保证数据逐层同步，DQA 校验是必不可少的。

### 3.5 事件驱动

资讯数据的事件驱动，比较常用的实现方式是：基于血缘分析，实现增量数据相关表的数据更新，目前可实现分钟级的数据跑批更新，能很好地满足资管新规下对于风控的监管要求和对于交易的时效要求。

## 四、FDS 优势

目前市场上有少部分产品也开始或者正在对资讯数据进行整合、时效监控和数据校验的操作，据市场调研情况，上述整合校验的操作通常是局部性的。而 FDS 旨在实现对金融资讯数据做全面业务整合，建立全方位的数据校验机制。同时考虑到各个金融机构数据使用差异性，需要提供丰富的接口，也可以一劳永逸地解决应用端数据迁移问题。下图是 FDS 六大优势：



#### 4.1 实现资讯统一管控

原来多个资讯数据使用源，一般包括：人工源，多个资讯厂商源，FDS 落地后可实现了统一的资讯数据出口，极大地解决了数据口径差异性问题的；

#### 4.2 全面数据校验

目前有成百上千条单源校验和多路比对规则可供选择，用来预警可疑数据和数据错误，同时用 DQA 保证数据流转质量。

#### 4.3 提升资讯数据覆盖面

比如两个源整合后私募债覆盖面上升幅度能大于 9%，同时解决了单一资讯厂商字段很多为空值，数据可能未推送，延迟推送，错误推送等问题；

#### 4.4 择优选数，提升数据使用效率

按业务模块逐一优选数据，提升每一模块数据质量的同时，提升资讯数据的使用效率；同时对于字段为空的值，在验证一致率情形下，可用交叉互补机制提升字段的非空率。

#### 4.5 数据跑批更新时效提升

从常规的数据 T+1 时效或者小时级时效可升级为分钟级（5-10 分钟），很好满足资管新规下对于交易和风控中数据时效要求；

#### 4.6 帮助提升数据清洗能力

FDS 建设过程中，伴随大量数据源的数据问题，随着问题逐一被解决，公司级的数据清洗能力能得到很大提升，可作为后续数据问题处理的经验和模板。

## 五、FDS 部署环境要求

类别	清单名称	要求	推荐硬件	最小硬件
操作系统	Linux	centos 7 及以上	1、 万得财汇整合库  【32 核\128G 内存 \2T 硬盘】	1、 万得财汇整合库  【4 核\16G 内存 \1T 硬盘】
	windows	window7 及以上		
编程环境	python	3.X 版本	2、 调度服务器  &kettle 服务器	2、 调度服务器  &kettle 服务器
	JDK	8		
数据库	oracle 企业版	11G 以上		
ETL 工具	kettle (必须)	7.0	3、 准实时 python 程序服务器  【32 核\128G 内存 \2T 硬盘】	3、 准实时 python 程序服务器  【4 核\16G 内存 \500G 硬盘】
	informatica	需要付费		
	CtrlM	需要付费		
	DophinScheduler (推荐)	1.2.1		
	自有调度工具	需定制开发		
			4、 前端展示服务器	4、 前端展示服务器

# THANK YOU

BEIJING QUANTEX FINTECH CO.,LTD.

宽拓智融科技有限公司

Tel: 010-58341436

Tel: 0755-86329549

BJ Add: 北京市海淀区高梁桥斜街42号融汇国际大厦5层

SZ Add: 深圳市南山区科苑南路3099号中国储能大厦12G

SH Add: 上海市黄浦区河南南路16号中汇大厦2楼氩空间2088

HZ Add: 杭州市滨江区月明路560号正泰大厦1号楼5091室

