DTEC
Advanced Digital Technology
先进数通

SharkData

数据智能研发管理平台

产品白皮书

北京先进数通信息技术股份公司

目　　录

目　　录 i
1. 什么是SharkData 1
2. SharkData一体三面产品理念 1
3. SharkData数据视角解决之道 2
4. SharkData功能列表 3
5. SharkData典型案例 6
5.1. XX城商联盟 6
5.2. XX银行 6

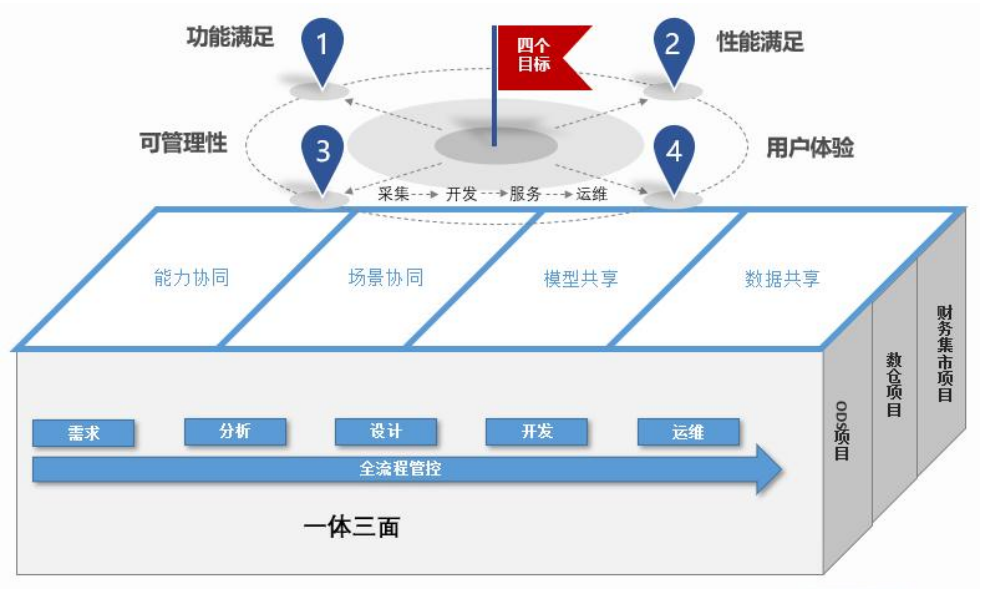
1. 什么是SharkData

在大数据时代，数据中心云化的情况下，企业在目标保证数据线建设的**快捷、有序、规范**建设的同时，面临着基础数据技术平台的高频更替，混合的存储计算引擎，混搭的云环境，跨平台统一调度需求，以及管理部门要求对数据治理工作不断深入等情况。在考虑到对数据资源相关最大限度的整合使用，企业逐渐从零散的使用工具平台到建设一体化平台进行支撑。

企业智能数据研发管理平台（SharkData）是北京先进数通信息技术股份公司自主研发的一款用于支持企业构建统一的元数据和数据资产管理，并在统一的元数据和数据资产管理的基础上支持企业以全程面向数据的方式进行内外数据统一采集，数据集成开发，统一数据服务管理，数据管控工作，同时提供统一的调度运行服务及整体运行管理功能的软件。

SharkData支持企业进行全面的数据采集、数据研发、数据服务、数据管控、数据运维能力的构建，支持企业以SharkData为支撑平台构建数据中台，支持企业快速开展如数据交换、数据仓库和应用集市等信息应用的建设。

2. SharkData一体三面产品理念



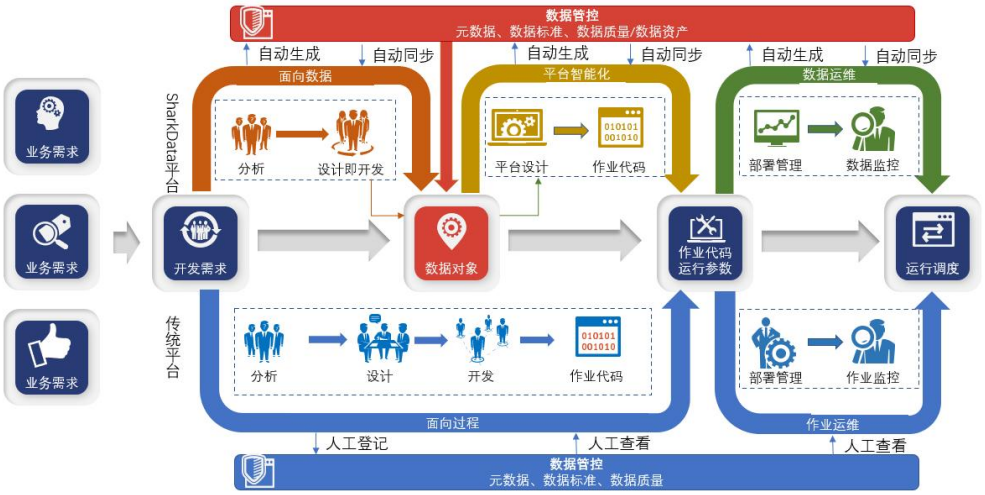
SharkData是在先进数通多个数据线产品的基础上，通过对产品概念、应用模式及界面风格进行一致性设计，深度整合多个数据线产品构建为支撑数据全生

命周期应用建设，集数据采集、数据研发、数据服务、数据管控、数据运维为一体的，一站式平台。

SharkData提供企业用户对数据开发过程，即数据需求、设计、开发、部署、运维全流程多角色工作的支撑，形成对数据需求处理闭环的支撑；SharkData支撑企业用户进行端到端的数据应用的建设，形成对数据从源端产生到末端消费的闭环支撑；同时SharkData通过提供能力和场景的协同，数据资源和开发资源的共享，支撑多项目协同开发管理，形成对整体集成开发管理的闭环支撑，最终形成一体三面的企业智能数据研发管理平台。

3. SharkData数据视角解决之道

SharkData的核心是面向数据进行数据开发、部署、运维和服务，与传统的面向过程，需求从提出到设计后都以作业的视角进行开发、部署和运维的方式不同，整个过程企业用户都是以数据的视角进行需求、设计、开发、部署、运维和服务，结合统一数据资产管理和数据管控，保障数据线建设各角色、各环节的数据资产信息一致，消除了业务人员与技术人员理解上的差异；同时结合数据开发的特性，平台提供对应的脚本生成工具，满足用户实现设计即开发，保障设计、开发到部署的各环节的信息不脱节，提高效率的同时，为后续运维提供了便利的基础。



4. SharkData功能列表

数据采集 SharkStart	提供对多种存储计算引擎的数据进行离线数据批量采集，包括传统关系型数据库：Mysql，Oracle，DB2，Informix等，MPP数据库：GaussDB，Greenplum，Teradata，GBase等，大数据平台：FusionInsight，星环Transwarp Data Hub等，文件：本地结构化文件和非结构化文件，远程ftp/sftp上的结构化文件和非结构化文件。
数据开发 SharkBuilder	提供对开发使用的规范进行管理，包括基于模式的文件交换规范，数据命名规范和调度对象命名规范等；提供对组件和模板的管理功能； 提供对数据需求的登记，并支持对需求处理情况的监控； 提供体系化、系统化、规范化的建模能力，将建模理论以工具化、半自动化的方式实现；提供图形化的、更符合建模习惯的在线数据模型设计方式；同时对接数据标准，可通过依据标准的逻辑模型设计，自动化的生成物理模型；对自定义设计编辑的数据模型提供标准比对的功能； 提供多项目协同开发，通过项目进行资源隔离，人员、开发资源和开发任务的分配，并根据项目提供版本管理和发版的导出导入部署包的功能。 提供场景化的开发模式，内置数据接入、数据加工和数据供应三种模式应对绝大多数数据开发场景；其中数据接入提供多种异构存储计算引擎之间的读取和写入，并在过程中完成对数据的标准化，清洗检核等操作；数据加工提供多种存储计算引擎上的数据加工计算，提供场景化的模板，支持用户快速的完成数据标准化，贴源数据加工等；数据供应提供向多种异构提供存储计算引擎同步数据和交换数据的功能；提供参数化、配置化的开发方式，并提供批量开发方式，通过导入批量开发模板，批量完成数据开发；提供组件引用、模板配置和自定义配置多种自动化程度的开发方式；提供面向数据对象的开发配置方式进行开发，开发过程即是对数据元数据信息的不断配置，最终

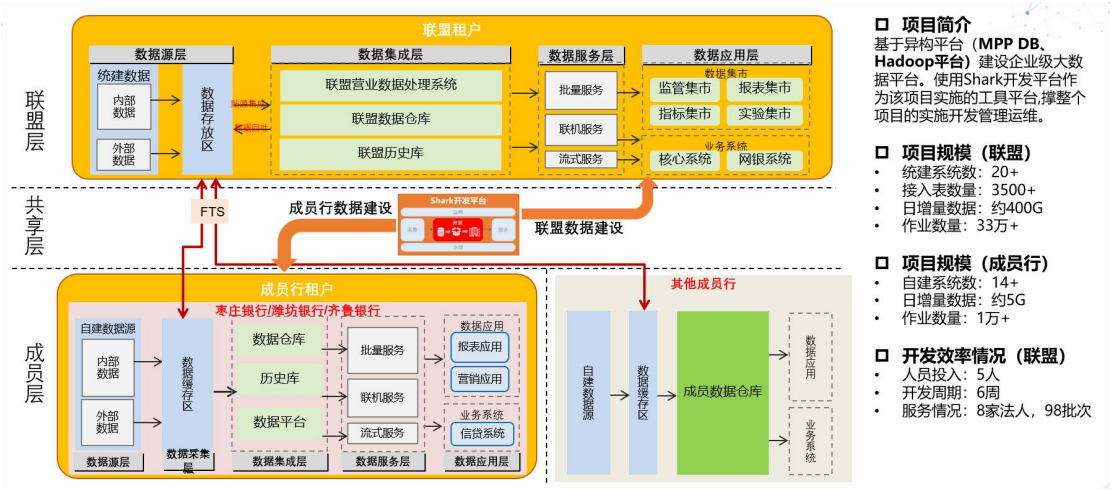
	通过脚本自动生成工具生成脚本代码，实现元数据驱动的开发方式，达到设计及开发的设计理念。
数据资产 SharkAsset	提供全局的数据资产管理，提供对数据对象实例的采集功能，并能根据数据对象或者元数据关联创建对应的业务分析对象，提供了统一的数据资产展现功能及灵活的业务数据分析组织定义查询展示办法。
数据治理 SharkRule	提供统一的元数据管理功能，支持多种类型的元数据的自动采集和模板导入，并以此实现元数据版本管理，差异分析，变更分析，血缘分析和影响分析。 提供统一的数据标准管理功能，支持批量导入及手工录入，支持数据标准的版本管理及落标管理，实现了数据标准及版本的统一发布、撤回等全流程管理，并提供版本对比，落标分析等功能。 提供全局的数据质量管理功能，提供对检核规则的管理，支持关系数据库、hive、Hbase、数据标准、指标、文件多种类型的质量检核，并且根据检核结果创建检核问题以及问题的处理流程的管理，统计分析平台内所有的质量检核数据，提供完整的质量检核报告。
数据运维 SharkOps	提供数据产品线运行、维护、监控平台，数据对象实例管理平台，数据流转全景可视化平台，平台贯穿数据线项目的测试、投产、运行管理阶段。通过定义数据加工流程，通过元数据对数据对象特征进行描述，并以数据目录形式分类管理，形成统一规范的数据视角运维服务。同时兼顾提供传统任务视角的运维监控。提供统一的产品组件调度管理接口。
数据服务 SharkInsight	构建统一的数据服务能力中心，提供将数据发布成服务以开放使用的功能，释放数据资产的价值。提供的服务能力包括批量数据集服务，API在线查询服务，数据展现服务，数据补录服务。支持零代码方式快速将物理表、查询逻辑发布成服务，降低服务开发的成本，以共享+自助的方式支撑用户数据消费，

	同时提供不同粒度数据权限以保证服务的数据安全，提供多种数据治理策略。通过统一服务出口，使数据服务的生命周期可管理、可监控、可量化。

5. SharkData典型案例

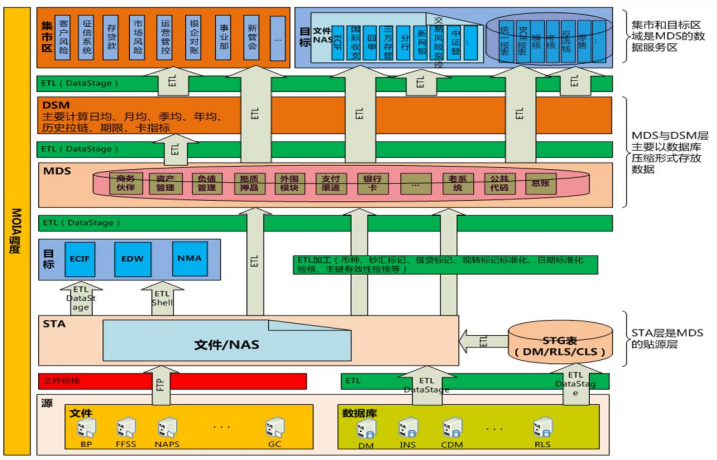
5.1. XX城商联盟

基于异构平台（MPP DB、Hadoop平台）建设企业级大数据平台。使用SharkData平台作为该项目实施的工具平台，撑整个项目的实施开发管理运维。



5.2. XX银行

基于SharkData构建全行企业级数据交换总线，负责采集各源系统数据，并交换至下游系统，是该行数据仓库、数据湖的主要数据来源。其中对涉及 378 个模块，总共 15466 个文件，日均数据总量 2.4T的数据进行采集和相关处理，向 174 个模块，总共 22831 个文件，日均数据总量 5.3T的数据进行下发。



1 简介

基于SharkOps运维平台，建立了全行级批量任务调度运行平台，涉及到全行9个重大数据平台批量任务，包括：MDS、FDM、管会、审计、运营管控、大数据、考核、安全平台、市场风险。

2 数量

58700

3 时效

15:00 - 次日 8：00

4 提升

开发配置效率提升，运维监控处理方便，调度效率提升