

# 国双Zeta企业级大数据平台 产品白皮书

在大数据时代，企业都遵循着数据指导业务的模型，期望通过数据来深挖新的价值创造点，为企业带来商业模式的创新和变革。

然而在企业转型过程中，通常面临着业务系统数据不连通造成的信息壁垒问题；数据来源渠道责任不明确，数据填报缺失导致的数据质量问题；企业内大数据技术人员有限，造成大数据准入门槛问题；指标零散不成体系，业务系统数据结果不一致导致的内部管理等问题。

国双 Zeta 企业级大数据平台是以 Hadoop 生态圈开源软件技术栈为基础，自研产品为核心，面向企业级用户提供海量数据分布式存储、计算和分析挖掘的高可靠、高性能的大数据平台。本文将梳理 Zeta 平台的架构和能力，深入了解建设 Zeta 平台的意义和价值，以及 Zeta 平台的特点。

## 第一部分：Zeta平台是什么

数字化转型与大数据平台

Zeta平台定义

## 第二部分：Zeta平台功能

管理套件

- 大数据管理
- 安全中心

开发套件

- 数据集成
- 任务调度
- 自助分析
- 实时流计算

数据套件

- 多维分析
- 数据管理

## 第三部分：为什么选择Zeta平台

高性能的大数据平台

行业标准制定的参与者

全流程的数据安全管控

支持丰富的数据采集场景

一站式数据管理能力

图形化的数据开发界面

# 1.Zeta平台是什么

## ● 1.1 数字化转型与大数据平台

企业数字化转型不仅是借助创建新技术迭代产品与业务,更核心的是经过搭建大数据平台实现经营、管理、运营、产品、服务等各维度由内到外的数字化;发挥大数据的驱动作用,充分挖掘企业数据的核心价值。

尽管企业已意识到数据在数字化转型过程中的重要性,但在转型过程中几乎都会遇到如下问题:

### ○ 数据孤岛问题

企业内部各业务系统已独立运行多年,存在着数据管理分散,数据彼此不相通等问题。

### ○ 数据质量问题

企业内数据来源渠道多种多样,数据填写录入不规范,不完整,从而导致数据质量差。

### ○ 指标维度体系问题

业务部门提出需求,IT部门提供数据,在这个过程中由于指标零散,口径不一致,影响数据价值的转化和输出。

### ○ 使用门槛问题

大数据的采集、开发、计算、分析需要大量大数据技术人才,传统企业人才储备不足。

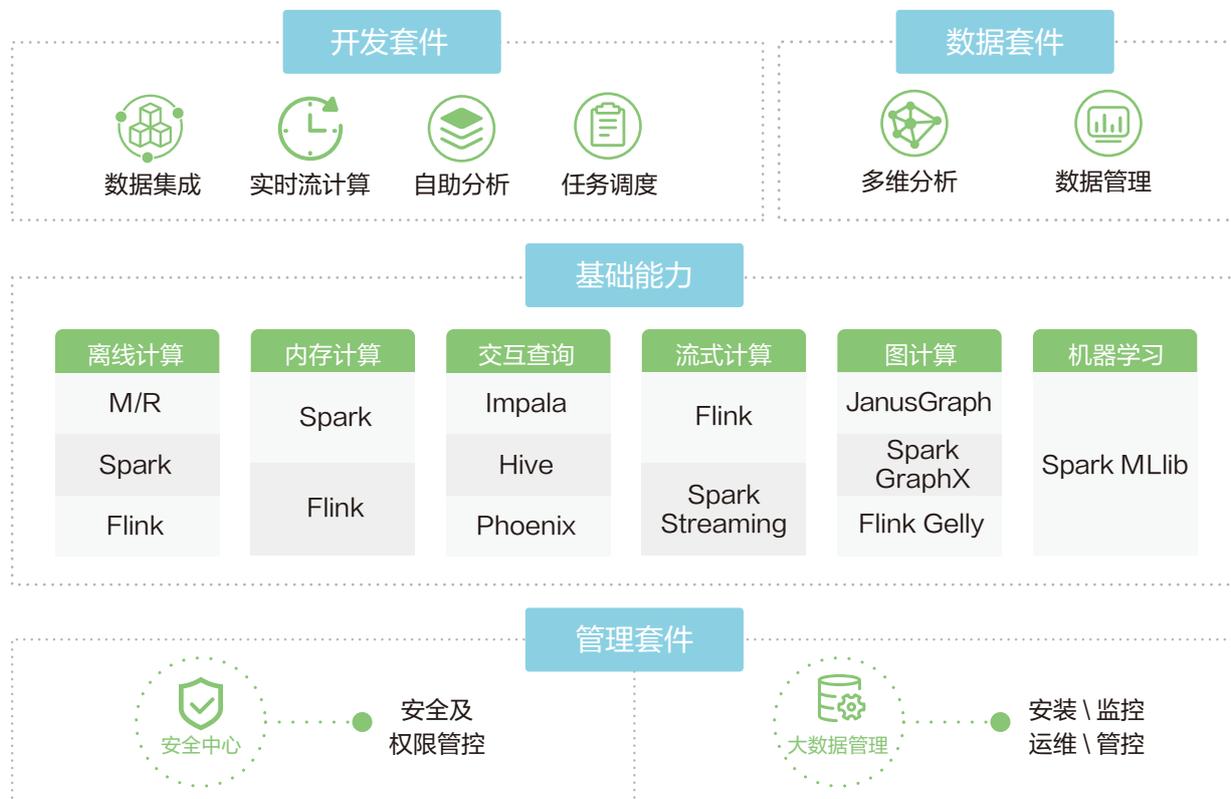
随着企业数据量逐年递增,传统的数据仓库已无法满足和处理大批量数据的查询、计算和分析等业务场景,大数据平台应运而生。通过大数据平台可构建开放、灵活、可扩展的企业级数据仓库,打破数据系统界限,有效解决数据孤岛问题;通过平台数据管理能力,可帮助企业搭建统一规范的数据标准,对数据进行全生命周期及数据质量监控,有效解决数据质量问题;通过平台的指标分类管理能力,可提供给企业一个统一标准的完善指标体系;可视化配置界面能够大大减少代码的写入量,降低开发人员要求,降低准入门槛。

## ● 1.2 Zeta平台定义

国双 Zeta 企业级大数据平台,是以 Hadoop 生态圈开源软件技术栈为基础,自研产品为核心,结合国双多年业务和技术实践沉淀下来的一套针对海量数据进行分布式存储、计算和分析挖掘的稳定可靠、高性能的平台。

Zeta 平台通过安全管控、数据集成、任务调度、多维分析、数据管理、可视化操作等能力为企业用户提供一站式服务,从数据采集到数据存储,从数据计算到数据开发、从数据分析到数据展示及全生命周期的数据管理。

## 2.Zeta平台功能



Zeta 平台产品架构

**基础能力:** 通过集成 Hadoop 生态圈开源技术栈, 支持 Zeta 平台所需的离线计算、内存计算、交互查询、流式计算、图计算、机器学习等基础能力。

**管理套件:** 包含大数据管理、安全中心两个产品, 是面向平台管理人员, 提供一站式、高效智能的多租户、权限、资源管理。

**开发套件:** 包含数据集成、自助分析、任务调度、实时流计算四个产品, 是面向大数据开发者群体, 提供友好

易用的开发体验, 让开发者可充分享用大数据平台的海量异构数据存储能力和并行计算能力, 实现数据资产的开发、加工与整合。

**数据套件:** 包含多维分析、数据管理两个产品, 提供数据分析师面向业务的多维度指标建模, 提升数据查询效率; 提供数据管理员数据全生命周期的管理工具, 为提升数据质量、实现企业数据资产盘点提供支持和服务。



## 2.Zeta平台功能

### • 2.2 开发套件

#### 2.2.1 数据集成

数据集成是一款高可靠、高性能的可视化数据采集工具。支持从 RDBM 到 Zeta 平台的离线和实时数据采集，并为后续的数据 ETL、数据开发、数据管理、数据分析等服务。

数据集成工具支持 20+ 种结构化、半结构化、非结构化数据源，能够满足市面上大部分数据采集需求场景；同时支持多种常用的数据转换操作，如数据过滤、类型转换、日期转换、混合运算、字段合并和拆分等。除全量采

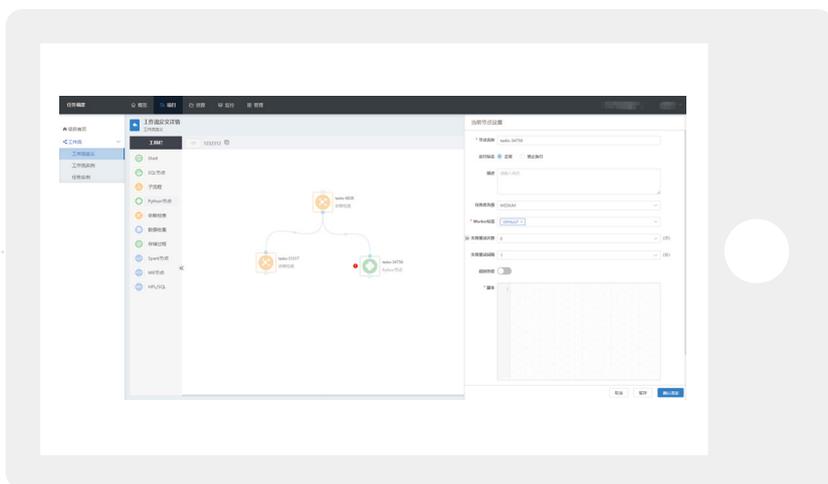
集外，数据集成工具还可在特定条件下支持增量采集，以满足用户不同业务场景下的采集需求。

数据集成作业极易受外部因素的干扰，诸如网络闪断、数据源不稳定等因素很容易让同步到一半的作业报错停止。Zeta 平台数据集成工具可以做到线程级别、作业级别多层次局部 / 全局的重试，保证采集作业稳定运行。针对采集结果，工具提供了采集报告，便于开发人员对任务结果进行校验和纠错。

#### 2.2.2 任务调度

任务调度是一款图形化的工作流设计、调度和分析的数据开发工具。用户可针对不同业务类型的任务自定义工作流，将一组任务以相互依赖的方式配置成 DAG（有向无环图），并进行调度和监控工作流执行状况。支持包括 Shell 脚本、SQL 节点、Python 节点、Spark 节点、Flink 节点、MR 节点、数据集成节点等多种任务类型。

任务调度提供可视化编辑功能，用户通过拖拽的方式，可快速构建出符合场景的任务流，方便开发人员快速完成任务节点的在线开发。任务调度支持以项目的方式隔离工作流，既能满足同组用户协同办公，又可做到项目间工作流互不干扰。



## 2.Zeta平台功能

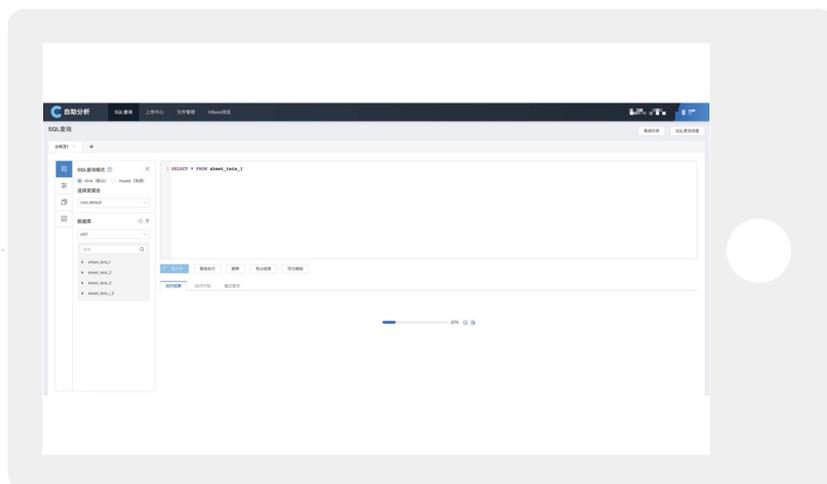
### 2.2.3 自助分析

自助分析是一款协助技术人员进行数据开发、SQL 调试、自助分析等能力的可视化数据分析工具。它能够有效降低开发人员的大数据技术门槛，帮助企业快速落地大数据需求。

自助分析提供了 Web 图形界面在 Hive 和 Impala 引擎下通过 SQL 执行数据查询、计算和分析，并提供语

法检测，协助开发者进行自助式的数据分析，提升开发人员工作效率。

除此之外，可视化的 HDFS 文件系统和 HBase 查询界面，使得对 HDFS 中的数据的操作和 HBase 数据的查询完全能够通过 UI 界面完成，降低开发人员的使用门槛。



### 2.2.4 实时流计算

实时流计算是一款实时流计算工具，通过工作流描述构建分布式流计算应用程序。支持 Flink、Spark Streaming 等主流计算引擎，为用户提供了不同场景（高吞吐和低延迟）适配的需求。在 TB 级别的数据上，能保持高吞吐、低延迟的性能。

实时流计算通过 Stream SQL 支持实时 ETL 数据处

理能力以及实时流计算，达到处理 1800 万条消息 / 秒（每条消息 100 字节），4 节点集群上的延迟为 8 毫秒；同时支持大规模分布式集群，集群弹性伸缩，按作业使用的资源扩容和缩容集群，最大化节省成本。

## 2.Zeta平台功能

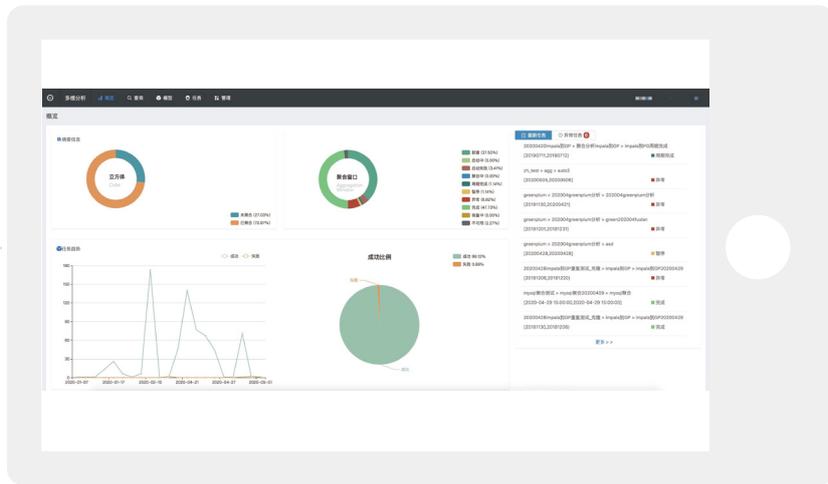
### 2.3 数据套件

#### 2.3.1 多维分析

多维分析是面向业务的维度指标建模与分析工具。基于此工具引擎，数据分析师可以快速构建一个具备多维度剖析功能的数据分析系统。

多维分析是国双基于过往的业务经验，自主开发并拥有多项自主知识产权的组件。工具基于 OLAP( 联机分析

处理 ) 多维分析技术，采用多维视图 Cube 来描述数据集的结构，提供了灵活的模型搭建能力。工具按照业务需求以维度和指标搭建数据模型 (Cube)，可根据需求可任意构建聚合组、聚合窗口。工具提供预聚合功能，可有效提高查询效率，大幅度降低查询时间。



#### 2.3.2 数据管理

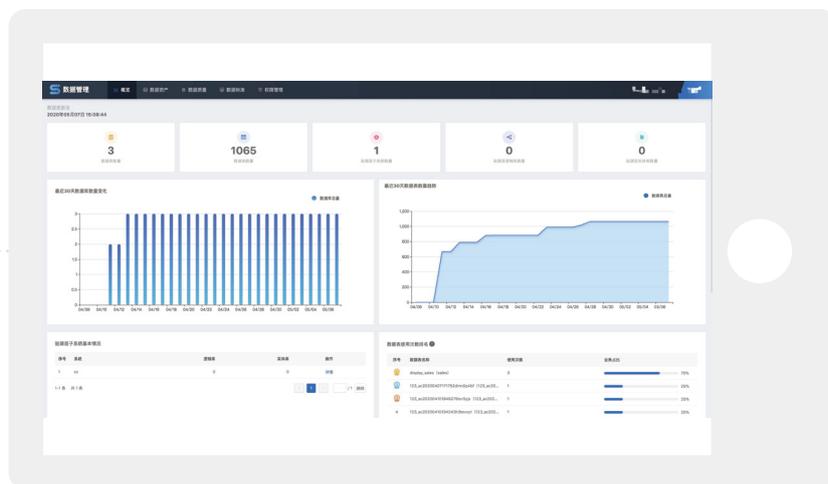
数据管理是大数据平台内数据的全生命周期管理工具，提供了包括元数据管理、数据标准、数据质量、数据服务、数据资产统计等功能，协助企业完成数据资产管理工作。

数据管理依托于国双在大数据应用领域多年的实践经验，通过元数据管理帮助管理员快速定位问题数据，从而实现数据的全流程管理。数据标准协助管理人员梳理企业数据标准，完成企业统一规范的数据标准建设；并提供了

30+ 种数据质量规则模板，支持数据管理员自定义数据质量检核任务，实现对数据质量的闭环管理。

数据目录可帮助平台用户梳理业务信息，通过业务对象将技术信息和业务信息结合成主题式数据卡片的形式对外提供服务，打破了技术人员与业务人员的隔阂，降低沟通成本，帮助企业全面盘点数据资产、实现数据互通互联提供支持。

## 2.Zeta平台功能



## 3.为什么选择Zeta平台

### 3.1 高性能的大数据平台

国双大数据平台在 2018 年一次性通过 7 项 DCA 大数据能力评测，是当时该评测有史以来一次性通过评测数量最多的企业。并行数据仓库系统是当时唯一一个通过分布式分析型数据库性能评测的产品。

在 Hadoop 性能测试中，同等测试环境下，国双 SQL 性能测试勇夺第一名，在 TB 量级下部分 SQL 执行时间可低于 10 秒钟，海量数据检索速度遥遥领先。

### 3.2 强大的数据管理能力

国双 Zeta 平台在 2019 年作为众多厂商中唯一一个满足所有评测项的产品通过了由中国信息通信研究院组织

的数据管理平台基础能力评测，在数据管理能力上远超同期厂商。

## 3.为什么选择Zeta平台

---

### ● 3.3 行业标准制定的参与者

国双科技有限公司作为通信标准化协会列席会员，深度参与 TC601 “数据管理平台标准 2.0” 的制定。

### ● 3.4 全流程的数据安全管控

Zeta 平台采用 Kerberos 和 LDAP 的方式进行用户级别的安全认证，满足细粒度到列级别的数据赋权，通过支持多租户、租户隔离、保证租户间资源互不可见，互不影响，通过审批机制，层层把关，保障数据安全。

### ● 3.5 支持丰富的数据采集场景

Zeta 平台提供了 20+ 种以上的结构化、半结构化、非结构化数据源作为数据采集端和目的地以及 10 多种转换器可以在添加到收集器中；在采集过程中，Zeta 平台提供了多种数据转化工具，能达到在传输过程中的数据脱敏、补全、过滤等目的，支持多种数据采集业务场景。

### ● 3.6 一站式数据管理能力

Zeta 平台提供了从数据采集、数据存储、数据计算、数据分析以及数据共享的全流程一站式管理能力，并通过视图的方式展示数据给用户，减少由于技术背景不同所带来的使用瓶颈，帮助用户高效地获取数据价值。整个过程受到安全管控和权限限制，最大程度保证数据安全。

### ● 3.7 图形化的数据开发界面

Zeta 平台提供图形化操作配置界面，通过拖拽方式配置任务节点依赖关系和调度，有效降低大数据平台对开发人员的能力要求；同时减少开发任务的代码输入量，提高开发效率。



## 关于国双

国双 (NASDAQ:GSUM) 是中国领先的企业级大数据和人工智能解决方案提供商。基于国双大数据平台独有的分布式数据架构和先进的实时、多维度关联性分析技术，同时利用自然语言处理、知识图谱等人工智能技术，国双的解决方案能够使客户充分洞悉数据间的复杂关系，获得全新的商业洞察，帮助企业和政府客户作出更好的业务决策，有效驱动产业智能化和数字化转型。

## 服务领域



工业互联网



智慧能源



智慧司法



新零售



航空及旅游



汽车



运营商

## 合作伙伴



## 服务客户



### 北京总部

地址：北京市海淀区北四环中路229号国双大厦  
 电话：(86-10) 8261 9988  
 传真：(86-10) 8261 9993



国双官方微信

# 国双产业人工智能平台

