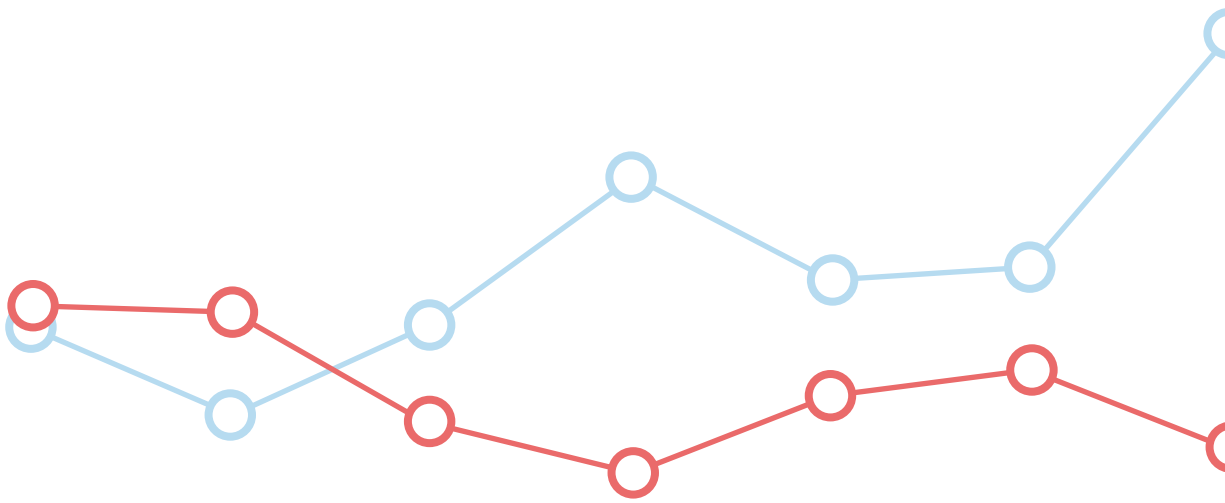
datafocus

像搜索一样简单

首个中文自然语言数据分析系统

2019 Progress Report



Content

04 重新定义BI

4 您的痛点是什么?
7 DataFocus重新定义BI
8 商业智能BI的发展趋势
9 DataFocus与数据中台的融合

10 DataFocus核心优势

12 搜索式分析
12 自带大数据仓库
12 强大的可视化能力
13 快速实时计算，无须CUBE
13 完善的多端应用
14 行列级权限控制
14 笛卡尔透视
14 高斯之瞳

15 DataFocus设计架构及配置

16 DataFocus架构设计
17 各版本软硬件配置

18 DataFocus功能概述

19 数据处理
19 数据源支持
23 多表关联
25 新增公式
26 行列转换
28 筛选过滤
30 中间表应用
31 智能洞察
32 数据支持
33 搜索分析
34 多维搜索分析
35 时间关键词搜索
35 日期关键词搜索
36 排序关键词搜索
37 字符关键词搜索
38 复合语句搜索
39 数据可视化
37 丰富的图表类型
38 个性化的图表配置
41 多样的图表交互效果
43 个性化布局配置
44 自定义组件
45 数据钻取
45 大屏联动
47 企业级管控
47 用户管理
48 数据权限
49 多端展示
49 移动端展示
49 大屏应用

50 产品性能测评报告

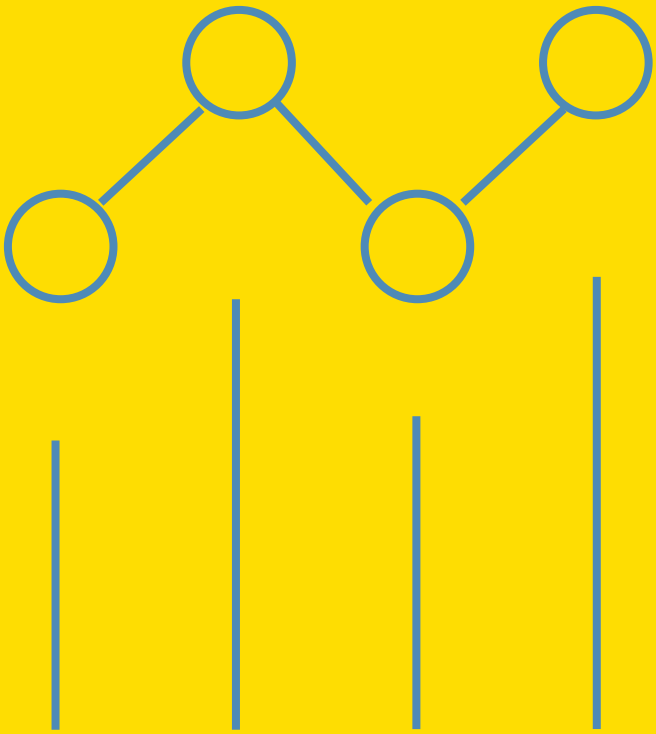
51 测试环境
52 数据准备
53 测试过程

61 关于DataFocus

01

重新定义BI

Redefine BI



您的痛点是什么?

“成功的管理都是相似的，有问题的管理各有各的问题。”

作为一名企业管理者

您是否时常感到蒙在鼓里，或者总感觉财务、
销售等部门提供的数据与自己的感觉不符?
明明公司已经有了ERP、CRM、HR等等各种业务系统，为什么它们的数据时常打架
，或者通过它们，您仍然对公司的业务只能雾里看花。

作为公司的信息部门负责人

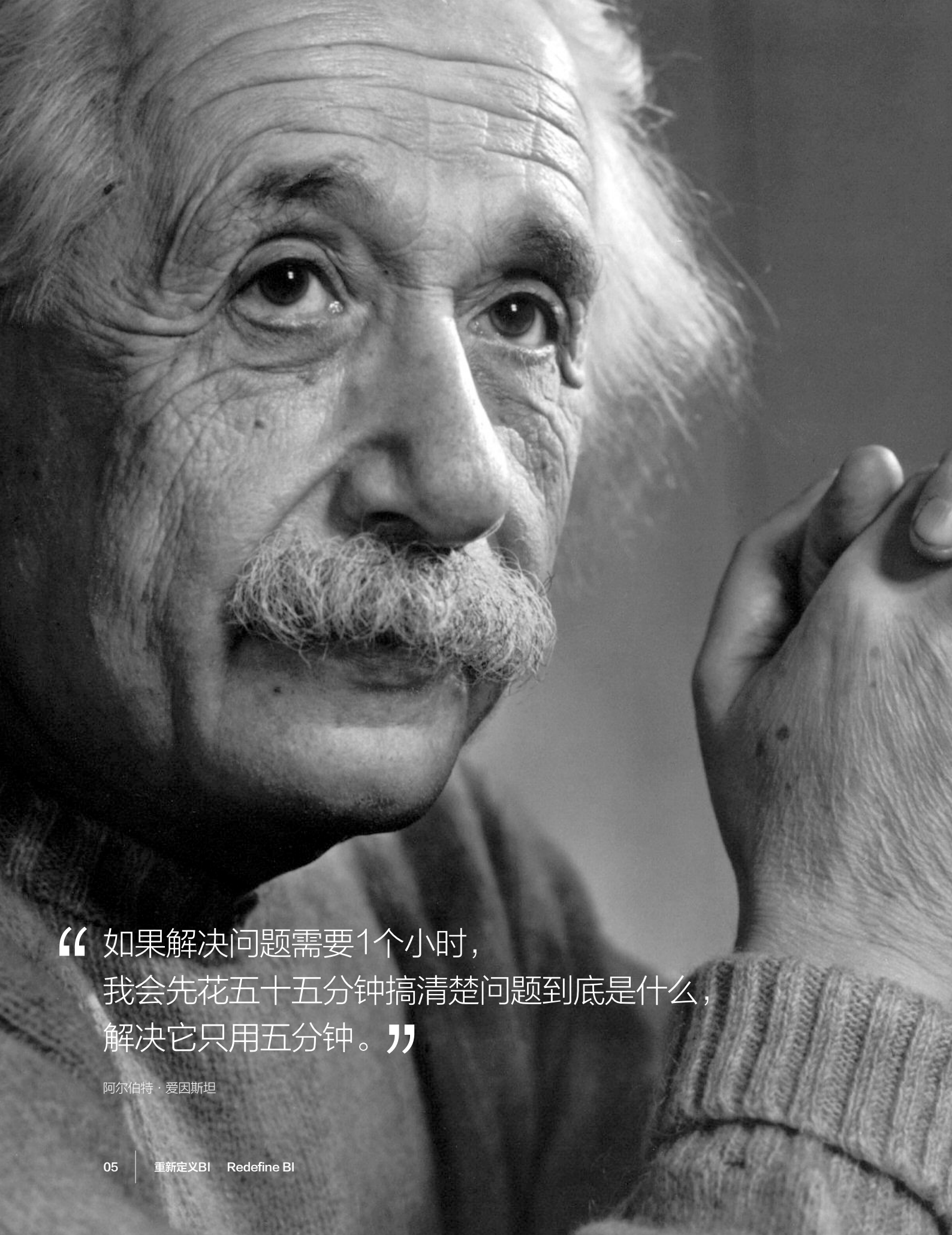
您是否已经被无穷无尽的报表开发需求淹没?
或者已经很努力可仍然被指责不能及时解决业务部门的数据问题?
明明已经购买了各种报表软件，为什么工作量反而越来越大，效率反而越来越低?

作为公司业务部门如销售、财务生产等部门负责人

您是否觉得IT部门总是无法正确的理解您的需求?
或者他们总是在再三催促后才会交出您想要的数据报表?
CRM、财务软件、ERP系统提供的数据分析还是几年前系统上线时候的样子，
已经远远不能满足需求；哦，看到信息部门同事的黑眼圈，
其实您们都已经不愿意再反复要求他们帮忙了。

作为公司负责人

每一次IT系统的投入，都会形成巨大的连带成本，
这本账您最清楚。花钱购买软件往往只是一道开胃菜而已，后续的维护成本，
人才流失成本会不会成为无底洞? 以前购买的报表软件已经是一个教训了，
刚培训出来的数据库工程师、数据分析师不满意工资待遇离职了，
还需要继续投入吗?



“如果解决问题需要1个小时，
我会先花五十五分钟搞清楚问题到底是什么，
解决它只用五分钟。”

阿尔伯特 · 爱因斯坦

问题的关键不在于
您需要再购买一套报表系统，
或者升级到商业智能BI，
真正的问题是，
您需要一种随时随地
分析数据的能力！

The point is not that you need to buy another reporting system or upgrade to BI . The real problem is that you need the ability to analyze data anytime, anywhere!

DataFocus重新定义BI

传统商业智能BI，一方面由于架构设计落后，无法高效处理大容量、复杂业务系统大数据分析问题；另一方面其使用门槛高，需要用户花费大量时间学习才能掌握相应的分析技巧，给企业带来了使用成本高、数据生产和消费分离、分析周期长等一系列问题，从而造成企业、组织数字化治理的严重障碍。

7分钟
即可开发完整大屏

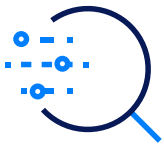
30分钟
培训熟练掌握

100倍
效率提高100倍以上

DataFocus通过引入人工智能技术，完全革新了商业智能BI的交互方式。

通过链接企业各种数据源，DataFocus为您整合所有的数据资产，在此基础上，您只需要通过搜索问答的方式向它提问。AI驱动的数据分析能力，比传统的报表软件效率提高100倍以上，比传统的拖拽式BI工具更智能；更有价值的是，掌握这种搜索式分析技能，不需要熟练的数据库工程师或数据分析师。精通业务的非工程技术人员，经过半小时的培训，即可熟练掌握，7分钟即可开发一个完整的数据可视化大屏。

商业智能BI的发展趋势



到2019年自助分析和BI用户创造的分析结果将超过数据科学家 — Gartner

自助式数据分析是大势所趋

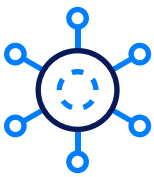
传统的数据分析模式，都是通过IT信息部门集中管控。需求由业务人员发起，IT工程师根据业务人员的数据分析需求，写代码到各业务系统取出数据，制作报表。这种方式效率低下，费时费力。往往还因为业务部门和IT人员的认知不同，导致沟通偏差，未能准确满足需求。

未来，以业务人员为中心的，IT部门辅助的自助式分析模式，将从根本上解决以上问题。有了DataFocus这类数据分析工具，业务人员从自身需求出发自助进行数据分析成为现实。现在，IT部门只需要事先将数据进行好标准化规范，业务人员通过简单搜索即可完成。

你需要的不仅仅是一种工具，而是一个工具箱

全栈式BI已成为潮流

传统的BI或报表工具，只局限在数据的分析和展现，往往还需要搭配第三方的数据仓库才能有效运行。DataFocus自带大数据仓库，企业数据分析不需要再采购第三方组件，系统部署完成后，从数据抽取、建模到搜索分析，分享、预警和移动端展现，All in one！ DataFocus引领全栈式BI的潮流，将企业从BI部署、实施到上线运行的周期缩短70%以上。



未来已来，将至已至
既往不恋，纵情向前

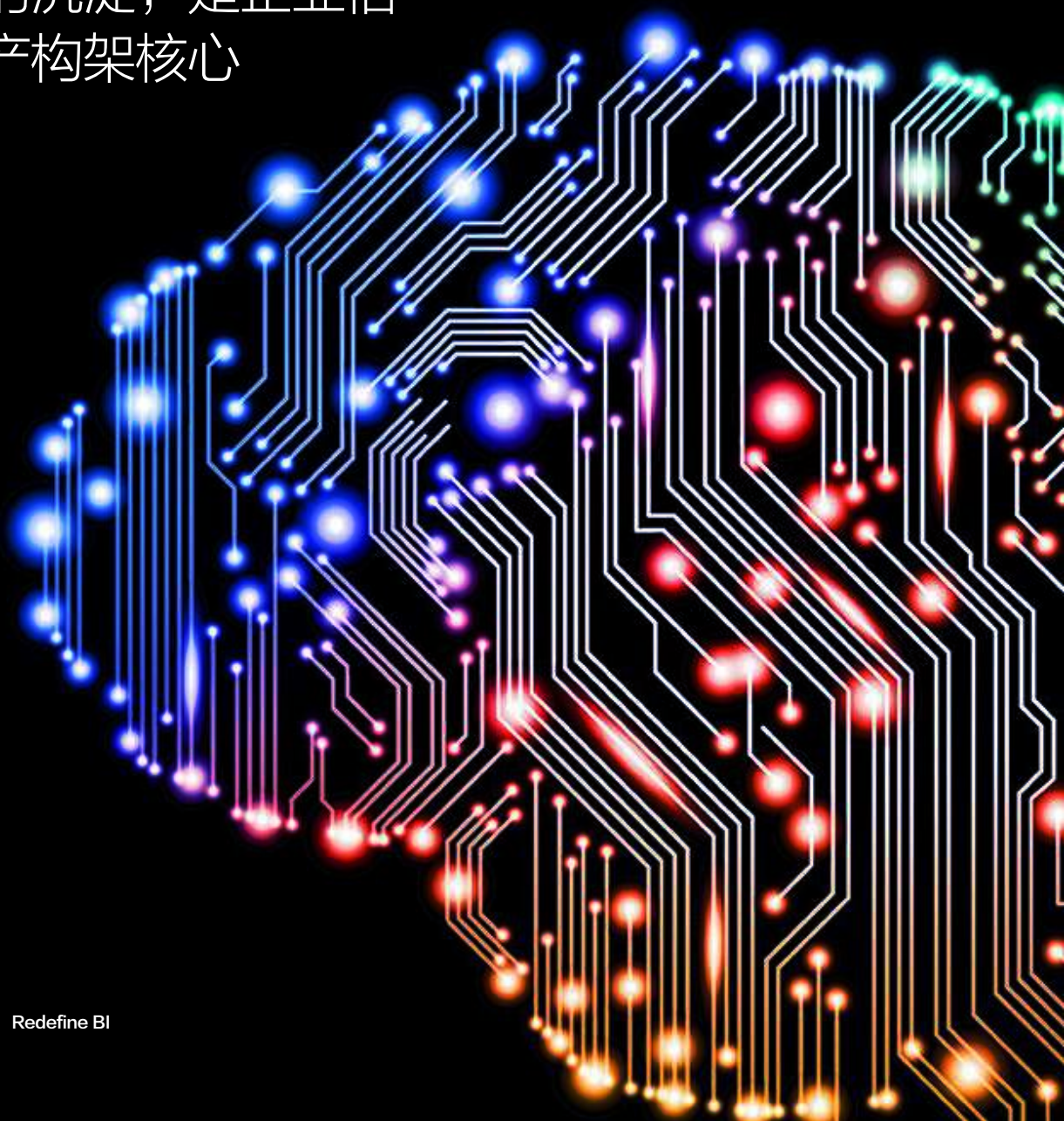
自然语言搜索式分析将成主流

DataFocus与数据中台的融合

大型企业往往IT系统繁多，只有通过建立大数据平台，将所有数据进行标准化规范，通过统一的规格、可封装的技术对外提供，数据才会从固定资产变为流动资产，最大化的释放价值。

DataFocus在设计之初就考虑到提供全栈式功能部署，运用基于自然语言的数据库搜索引擎技术，同时解决大数据的存储、规范和可复用的输出三大问题，轻松建立企业的智能化（AI-Driven）数据中台。企业还可以通过DataFocus的企业数据总线API模块，自主封装、开发各种灵活的前端运用，也可以通过DataFocus专家版的机器学习模块，根据各自的业务场景，训练可以解决特定问题的AI模型，解决更高级别的问题。

数据中台是企业“业务+数据”的沉淀，是企业信息化资产构架核心



02

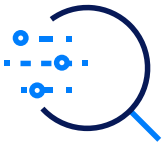
核心优势

Core advantages

DataFocus是一款
运用最新理念设计的数据分析系统，
它包含了商业智能BI的各项功能，
也可以作为商业级的
智能数据中台部署和使用。
企业可以围绕DataFocus打造全系
列的数据分析和AI解决方案，
而无须依赖更多第三方组件。

DataFocus is a A data analysis system designed with the latest concepts, which includes the functions of BI.Can also be used as a commercial grade Intelligent data center deployment and use. Companies can build a full range of data analytics and AI solutions around DataFocus.There is no need to rely on more third–party components.

搜索式分析



“DataFocus既是国际上第一家正式投入商用的，同时支持中英双文搜索的大数据分析产品，也是国内第一家开创搜索分析交互方式的产品。”

DataFocus既是国际上第一家正式投入商用的，同时支持中英双文搜索的大数据分析产品，也是国内第一家开创搜索分析交互方式的产品。搜索式数据分析不仅仅是一种技术变革。它有效的降低了数据分析的门槛，使得非技术出身，但是精通业务的人员也可以快速上手进行数据分析，这有效的打破部门墙，让自助式数据分析真正成为现实。

搜索式分析，还使得企业培训数据分析师变得容易。从而打破了商业智能工具买得起用不起的尴尬。让大数据分析像搜索一样简单，这是DataFocus一直以来的目标。

自带大数据仓库

大型企业集团多年的IT信息化建设，积累了海量的数据，这些数据分散在多达数十种的业务系统数据库中，形成了一个个数据孤岛。这一方面导致了数据标准不统一、无法准确的进行分析；另一方面，传统关系型数据库的查询性能也越来越低，导致分析效率的低下。

DataFocus标准版以上版本自带的大数据仓库，能够轻松对接各类传统关系型数据库抽取数据或直连数据。独有的雪花模型，支持复杂的数据库结构；通过多数据源整合，支持跨系统分析，彻底打破数据孤岛。

强大的可视化能力

优秀的数据展现形式，既可以提高信息获取效率，又能很好的吸引注意力。DataFocus的图表分析可以自适应多达40种可视化图形，告别密密麻麻的表格数据。在此基础上，提供了大量图形通用配置功能，以供追求完美。

DataFocus的数据可视化大屏开发不仅可以做到高效，更可以做到媲美专业的数字大屏开发效果。增强后的数据看板功能模块，可以进行高度、灵活的自定义。比如设定复杂的大屏背景、组件风格、界面配置、全局摆放，动态效果等等，开发出震撼的可视化大屏甚至无须设计师参与。

快速实时计算，无须CUBE

海量数据进行关联分析，往往有效率瓶颈。传统的OLAP分析方式会借助第三方工具进行CUBE预先计算的方式，以空间换时间。但是数据分析的工作是连贯的，这种定时跑任务的分析方法，只适用于传统的固化报表作业模式。

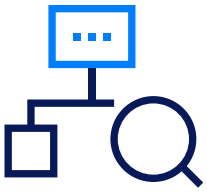
信息瞬息万变，决策毫秒之间。DataFocus采用列式数据存储的方式，通过自带的内存计算引擎，无须预先建立CUBE，数据分析实时交互，完全满足管理决策中经常遇到的临时性分析、多变的业务需求和频繁的结果刷新。IT部门将从此告别延时报表分析，亿级数据秒级响应。

数据立等可取

完善的多端应用

数据分析的结果往往需要与更多人分享或讨论，有时候还需要及时发送预警邮件。DataFocus同时支持企业微信和钉钉，通过简单的分享操作，即可快速实现手机查看。优化过的移动端展示效果，使得数据看板和图表的展现更加人性化，便于您旅途中也可以进行数据分析。

行列级权限控制



精确到行、列的细粒度数据权限控制，分析报告自动根据用户权限现实对应的数据，

DataFocus精确到行、列的细粒度数据权限控制，分析报告自动根据用户权限现实对应的数据，这可以轻松满足大型企业集团精细化的数据控制需求。固化的分析报表或数据看板只需要制作一份，再根据权限配置，即可实现千人千面的数据展示效果。此外，DataFocus还支持LDAP服务器集成，实现多业务系统的单点登陆。

笛卡尔透视

尽管有了搜索式分析的便利，企业在进行浩繁的数据分析探索时，仍然是低效的。举例来说，零售类企业有数以万计的SKU，数以千计的连锁门店和可分析时间段，数以百计的业务人员和促销渠道。分析师需要穷尽数亿级的组合才有可能找到一些典型的特征或趋势，比如某个单品，在特定时间段或特定门店会有快速增长的销量。这种工作理应交给机器。

DataFocus专业版的智能洞察模块，可以自动进行数据探索，从数亿的可能组合中，找出数据趋势和特征，并快速生成数据分析报告。

“从数万个类别中发现隐藏规律，需要分析数百万中可能，人类分析师需要100年！”

高斯之瞳

数据分析可以找到规律、发现问题，企业更加需要AI解决问题。DataFocus提供了各种武器帮助企业进行数据探索，企业在获得确定见解后，需要采取行动。有时候需要AI驱动的工具帮忙。比如，持续观察到某种商品库存占比过高，借助DataFocus专家版的算法模块，训练出库存预测AI，指导企业的采购进行调整。企业的数据分析，从描述性统计，到预测型分析，不需要额外购买更多软件，DataFocus帮你一站式搞定。

“用DataFocus建立算法模型，优化您的业务！”

03

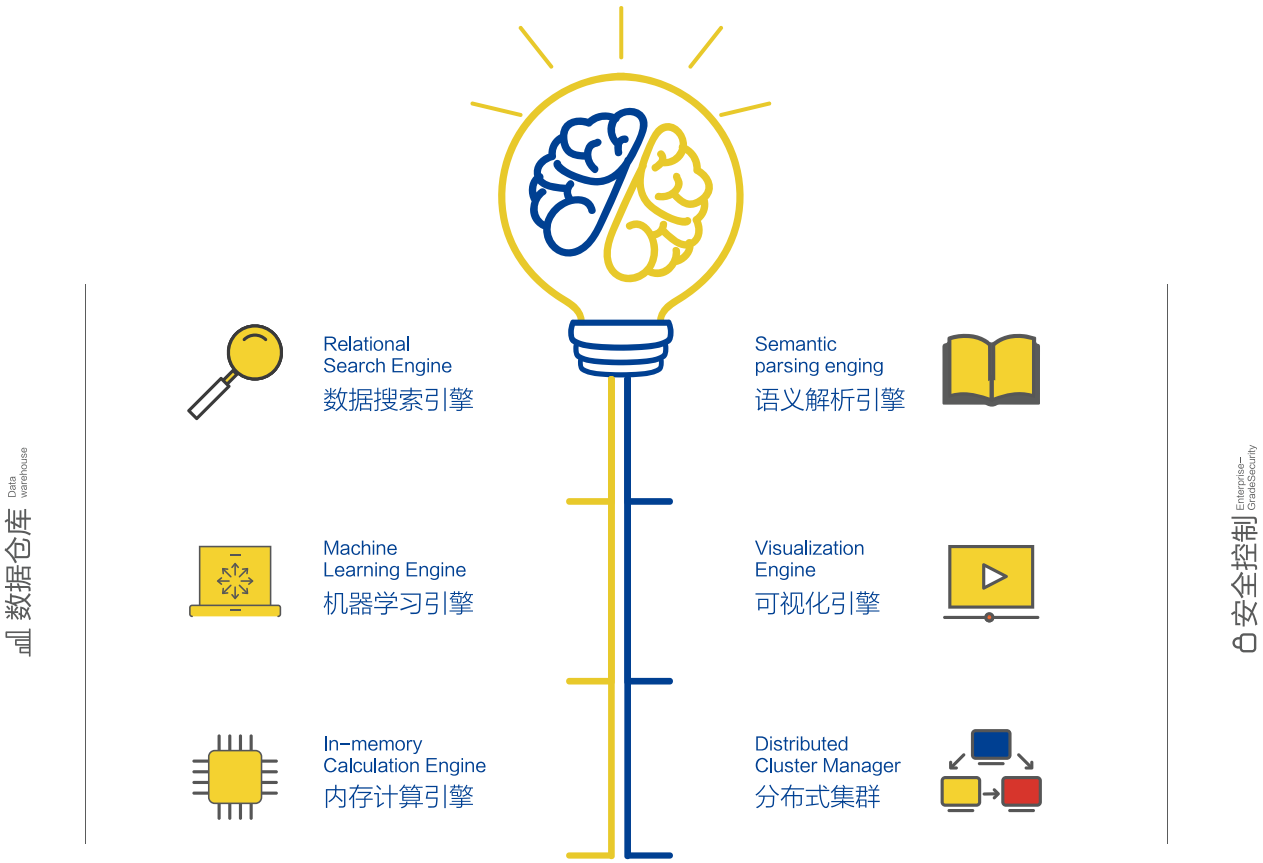
设计架构及配置

Design architecture and configuration

DataFocus架构设计

DataFocus是一个可以实现特定业务场景下人机对话的AI系统，该系统尝试理解人类关于数据的问题，并以丰富的图表和数据回答相关问题，与传统的数据分析方式相比，交互更加智能，效率提高100倍以上。用户不必花费大量心力去学习和了解难懂的计算机语言，而是用自己最习惯的语言如中文或英文，来使用计算机，就跟谷歌搜索一样简单。

整个系统架构包括八大模块：数据仓库、数据搜索引擎、语义解析引擎、机器学习引擎、可视化引擎、内存计算引擎、分布式集群、安全控制。其中数据搜索引擎和语义解析引擎，是组成自然语言搜索的关键部分，大大地简化了数据分析系统的操作难度，其中语义解析引擎会主动猜想用户的下一次输入，将用户搜索输入效率提高40%以上，并根据用户使用习惯优化推荐排序，促使系统愈用愈智能 。



为了满足大数据计算的场景，系统内置的内存计算引擎采用MPP架构的列式数据存储技术，一次可缓存数亿行数据，内存实时运算，千万级数据常规查询秒级响应。DataFocus专家版还提供了机器学习集成平台，用户可以根据自身的业务需求场景，训练自己的AI模型，用于解决销售、库存预测，故障预警或产品推荐、用户画像等经典机器学习问题。

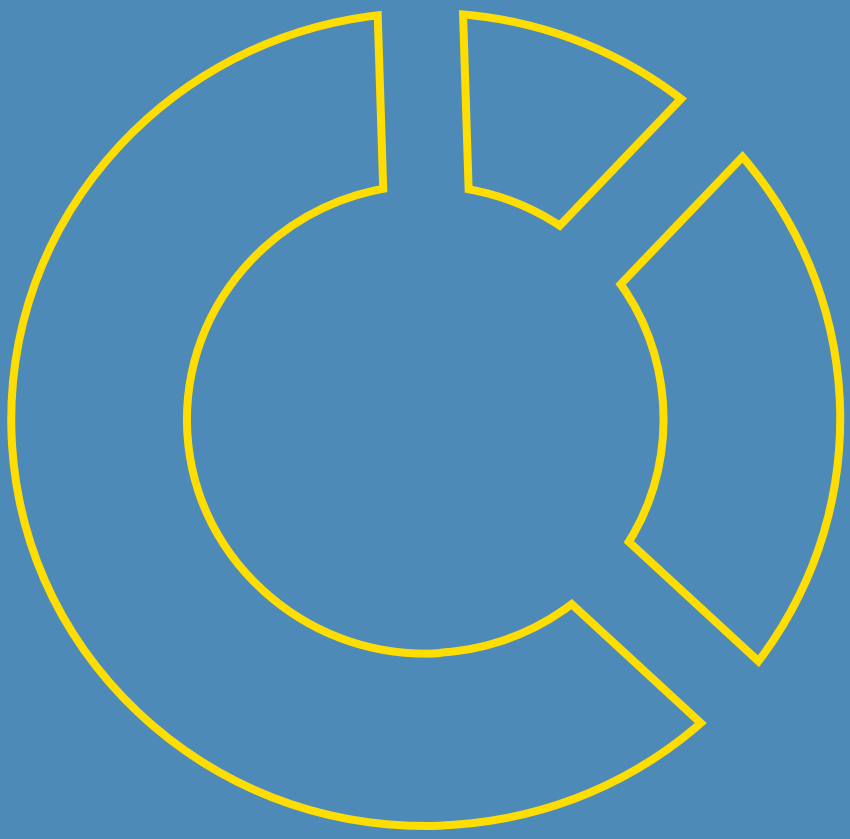
各版本软硬件配置

产品名称	操作系统支持	内存	硬盘	CPU	数据容量	备注
DataFocus Mini	Win7/8/10 64位系统	4G	20G可用		10万行以内数据	内存需2G可用
	CentOS 7.2以上	4G	/home 下30G可用 /root 下20G可用			内存需2G可用
DataFocus Standard	CentOS 7.2以上	20G可用	/home 下50G可用 /root 下20G可用	双核	不限	数据容量受 内存大小制约
DataFocus Professional	CentOS 7.2以上	32G可用	/home 下50G可用 /root 下20G可用	双核	不限	数据容量受 内存大小制约

04

功能概述

Functional Overview



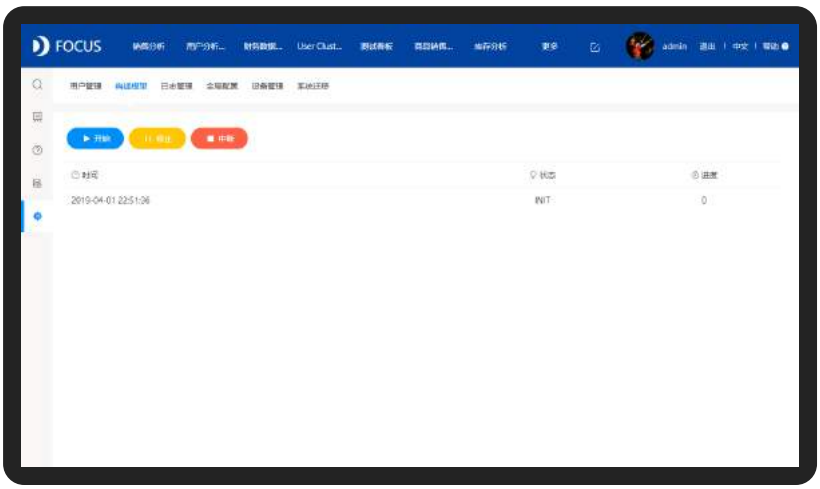
数据处理

数据源支持

DataFocus可支持多种不同的数据库类型，也支持不同的数据连接方式。

企业对于数据时效性的要求，有需要实时更新数据的。例如“双十一”时期，企业需要实时掌握销售情况，这时候就要求销售大屏数据实时更新。也有对数据时效性要求没有这么高的，或者固定周期更新的，例如财务数据，人事数据都有固定的更新周期，以周报形式或者月报形式为主。另外还有部分数据可能是外来数据，比如同行业的对比数据等，是来自于外部的数据，此时就需要有可以上传本地数据的途径。

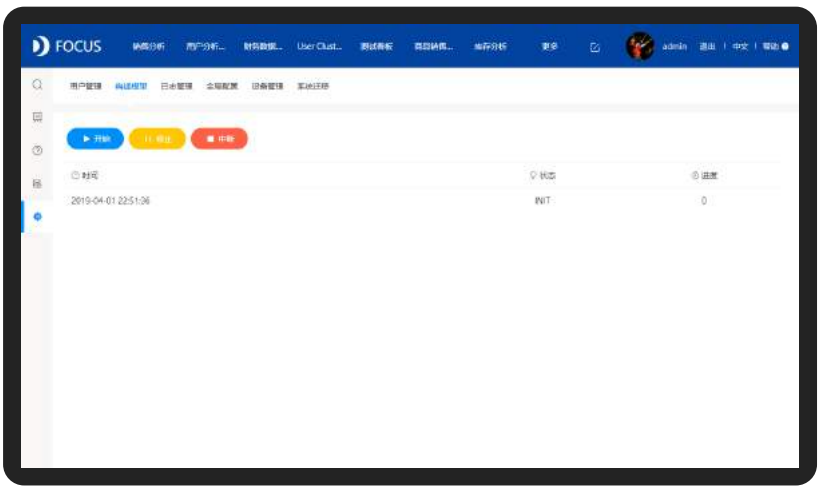
考虑实际业务需求，DataFocus开发了导入数据、直连数据、本地文件上传三种方式。



导入数据

即把DataFocus作为一个数据仓库，将各个业务系统中的数据统一导入到DataFocus中，导入数据形式支持定时更新，支持自由选择更新方式。像是财务数据、人力资源数据、客户数据等都可选择此种连接方式。

导入数据支持的数据源类型主要有：PostgreSQL、Oracle、SqlServer、MySql、Sybase ASE、Sybase IQ、SAP。



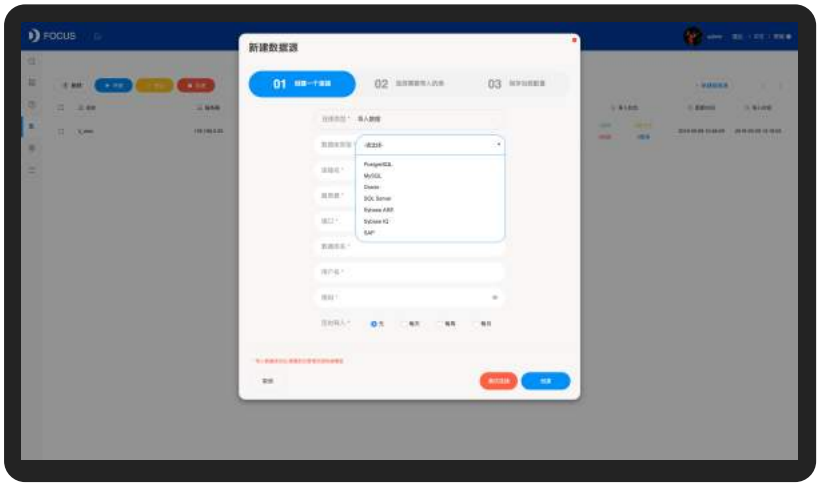
更新频率：

DataFocus支持用户以每天、每周、每月的频率在指定时间对数据源开始更新，用户可以根据自身需要，随心设置数据更新的频率。只需要在创建数据源或之后编辑数据源时配置定时导入选项，方便系统自动化的更新数据表，方便操作的同时，避免了遗漏情况的出现。

直连数据

即DataFocus直接连接企业数据库，不将数据库中的数据导入DataFocus，像是销售数据等对时效性有较高要求的，可以选择直连数据库的方式。直连的数据表其查询速度以及效率等，主要与企业的数据库性能有关，与DataFocus性能关系不大。

直连数据支持的数据源类型主要有：MySql。

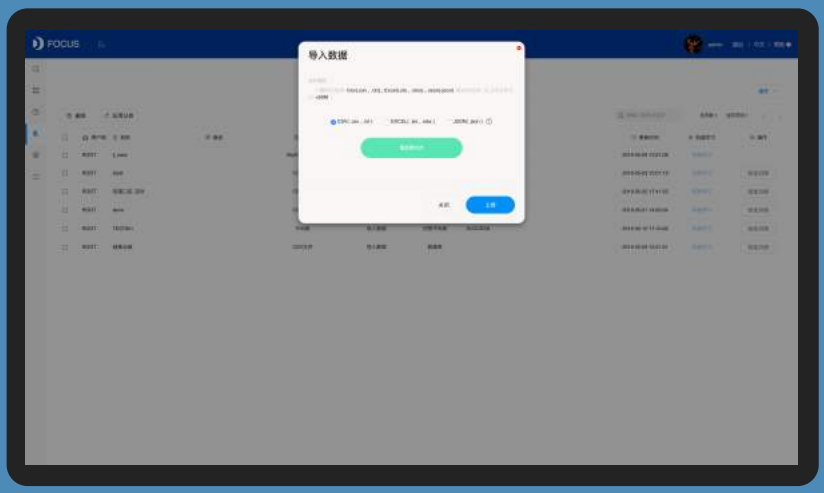


更新频率：

直连数据库支持实时更新，即数据库中数据有变化，DataFocus中相应的数据以及图表都会随之更新。

本地文件上传

即导入本地的文件，一般导入的是业务系统中没有的数据，来自于外部的数据，比如天气数据、同行业数据等。



本地上传支持的文件类型有：

csv 、txt、xls、xlsx、json

多表关联

DataFocus提出了多表联合查询的概念。在企业实际应用中，
需要分析的数据往往来自于多个不同的数据表，例如分析企业全面的销售情况，
则需要对销售订单数据表、销售交货数据表、
销售发货数据表等一系列销售相关的数据表进行联合查询分析，
不能只单独查看分析某一个数据表，因为部分的销售情况并不能代表企业的全部面貌，
在做数据分析的时候，最忌讳的就是以偏概全。如果只分析了销售订单数据，
那么有可能会夸大了企业实际的销售情况，因为在后续销售过程中，不乏有很多退货的情况。
故我们需要综合所有的销售数据，以更全面的角度来分析。

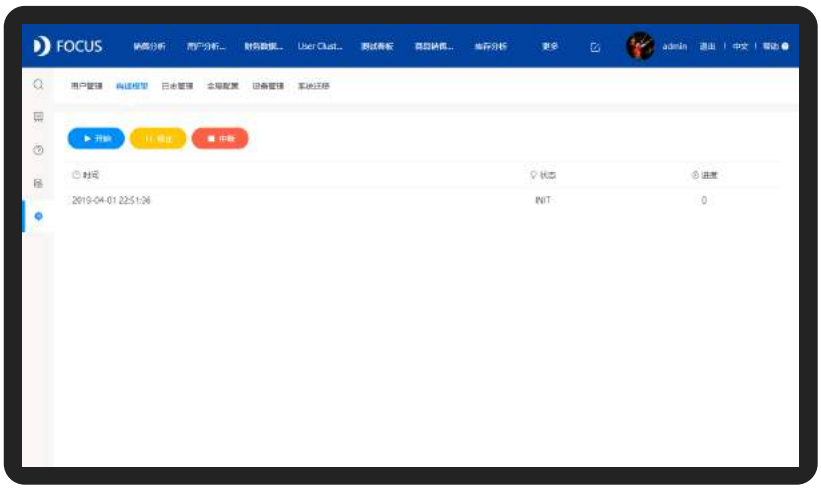
DataFocus中支持多表联合查询，但是在多表进行联合查询之前，需要先创建关联关系，
并且构建模型。什么是创建关联关系？通俗地解释，就是例如excel中vlookup的功能，
即通过多个表之间的某一列或者某几列进行匹配，然后将这几个表联立起来，
例如上述各个销售数据表中采用的匹配字段是销售订单号。

例如上述各个销售数据表中采用的匹配字段是销售订单号

多表关联

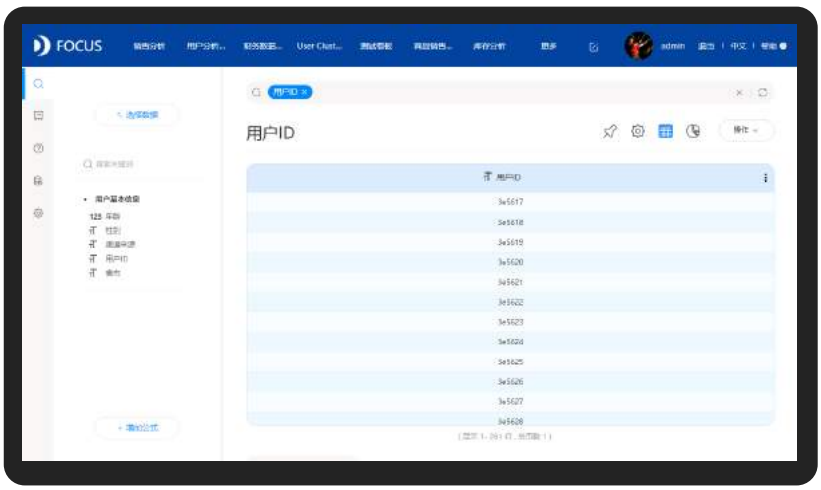
构建模型

创建完关联关系之后，并不能直接在DataFocus中起效，需要先构建模型，构建模型的具体过程无需用户操作，用户只需要打开构建模型的开关即可。



事实表与维度表进行联合

上述为了更全面的分析数据是多表查询的其中一个目的。还有另外一个目的，就是将事实表与维度表进行联合。什么意思？意思就是比如现在有一份销售订单数据，订单数据表中有物料ID、公司代码ID、客户ID等，但是这些字段都不是具体的值，而是一个个的字符串代码，这时候，就需要我们用关联关系来将销售订单数据表（事实表）与物料主数据（维度表）、公司代码主数据（维度表）、客户主数据（维度表）进行关联，从而得到例如销售订单数据中的物料AS001代表的是什么货物，公司代码A001代表的是哪家分公司，客户ID 3e5617代表哪个客户。



新增公式

DataFocus中公式可以解决大部分数据分析问题，可以计算各项关键性指标，可以嵌套使用。

在实际业务中，不管是什么分析，公式都是最常用的分析手段。比如时常需要计算的各项指标、各项比率等，都是通过公式计算所得。

DataFocus支持添加公式列，其所拥有的公式基本满足企业的需要，覆盖了大部分excel的公式函数等。例如在计算销售完成率、库存周转率、存销比等都需要利用原始数据进行简单的计算，从而生成一列新值。

对于DataFocus中支持的公式类型，大致可分为以下几种：

聚合公式：average、count、sum、max、min、sum_if、group_max等。

转换公式：to_date、to_string、to_double、to_integer、to_bool等。

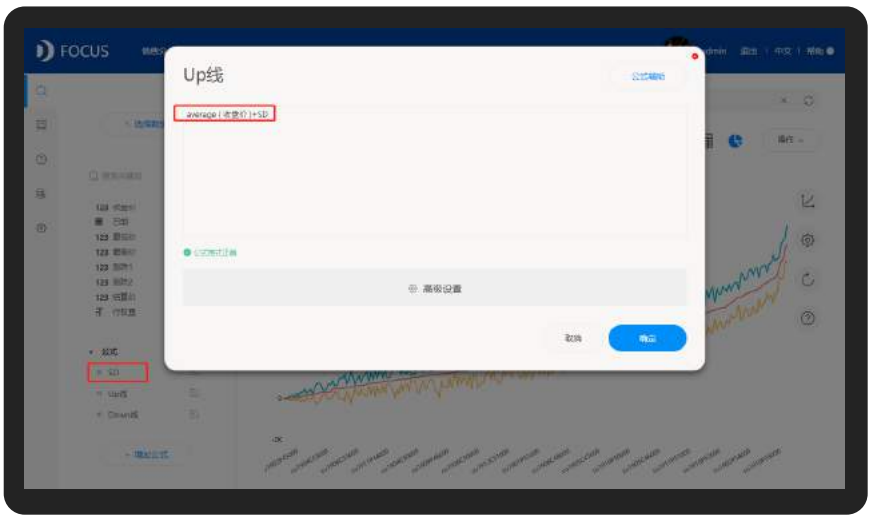
日期时间公式：data、day、day_number_of_week、diff_days、month、year、time等。

混合公式：!=、>=、<=、<、>、=、least等。

数值公式：+、-、*、/、^、abs、cbrt、cos、ln、log2、sq、sqrt等。

判断公式：and、if…then…else、ifnull、isnull、not、or等。

字符公式：concat、contains、strlen、strpos、substr等。

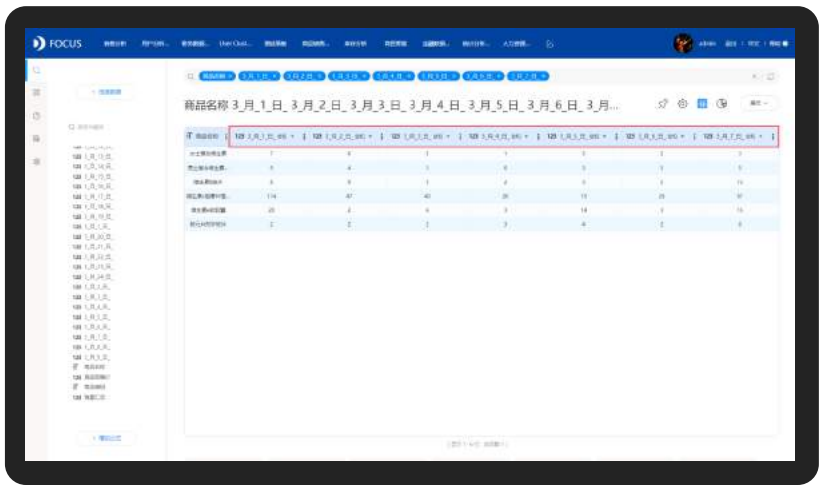


公式嵌套

DataFocus中公式的利用，不仅可以通过添加公式创建新列，还可以再次利用已创建的公式进行二次创建，即公式嵌套。例如在计算金融指标保利加通道指标项时，就需要先通过公式计算出收盘价标准差的两倍，再利用这个公式计算出它的Up线，也就是收盘价平均值加上刚刚计算出的收盘价标准差的两倍，以及它的Down线，即收盘价平均值减去刚刚计算出的收盘价标准差的两倍，这里相当于公式嵌套了一次。

行列转换

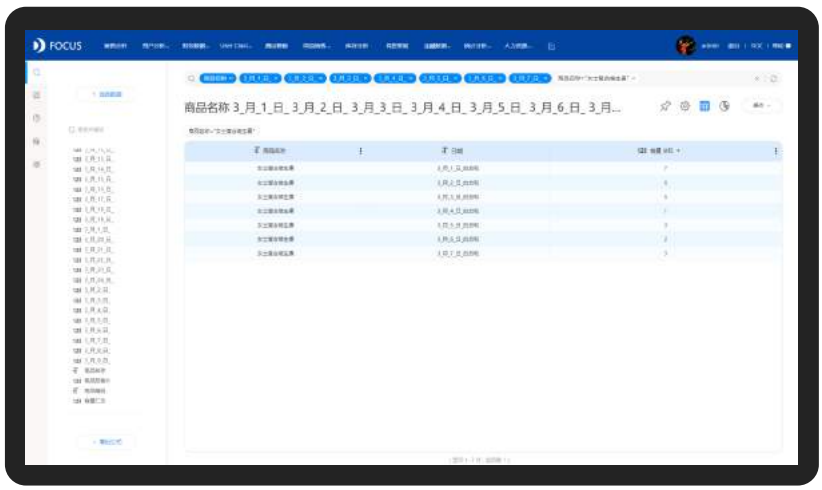
在实际的业务中，我们经常会遇到如下的场景，比如日期不是作为单独的日期列中的值，而是把日期的值作为一个个单独的列，如下图，每一日作为一个列。



但是有时候并不想要上述的数据格式，想要的是日期作为一列，销量作为一列，以便更好的看出其销量的趋势变化情况，这时候就需要用到行列转换的功能。

考虑到实际业务情况，DataFocus开发了行列转换功能，不仅能将上述的列转换成行，也可以将行转换成列。此功能解决了大多数实际业务场景中难以解决的问题。

如下图，即上述列转行的结果。DataFocus中，在上述搜索结果的基础上直接选择操作——行列转换，选择需要转换的行或者列，即可实时的看到转换后的效果。



列转行之后，即可将日期作为X轴，销量作为Y轴，数据显示成柱状图、折线图等常用图形，可以直观地观察销量的数据变化情况以及变化趋势。

数据填报

DataFocus专注于数据分析。尽管如此，考虑到存在较为广泛的场景，企业需要从外部收集数据，比如收集调查收据，或者由连锁门店店员填报数据以进行统计和汇总。因此，最新版本的DataFocus支持了轻量级在线数据填报，虽然不具备业务流程管理功能，但管理员可以为用户配置填报、审批权限，可自定义编辑填报模板，填报审核后生成对应的工作表，用于后续的搜索分析，制作历史问答、数据看板。

配置填报、审批权限，可自定义编辑填报

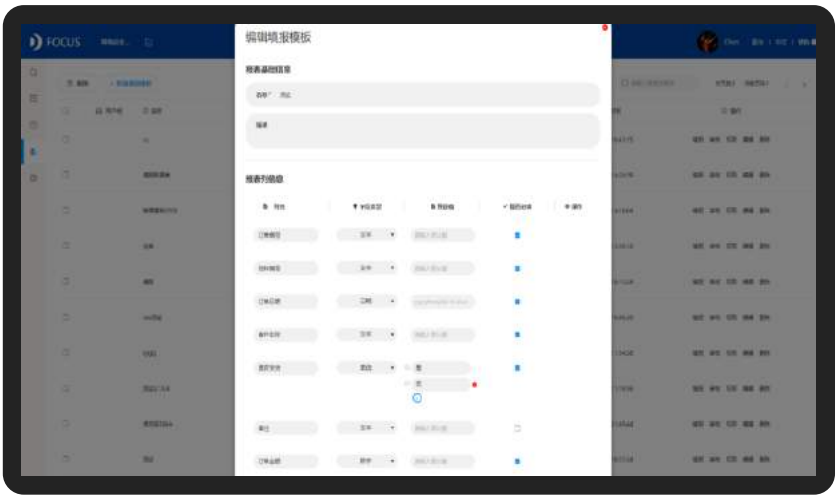


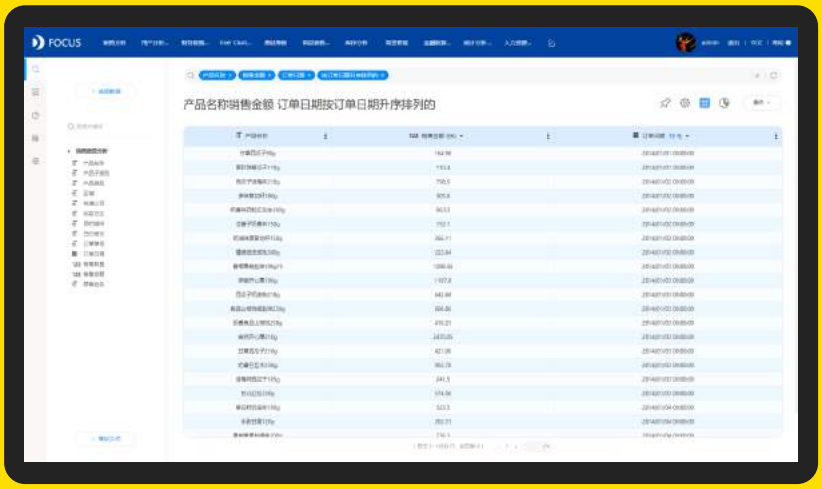
图 4　1-4 填报数据模板属性编辑

筛选过滤

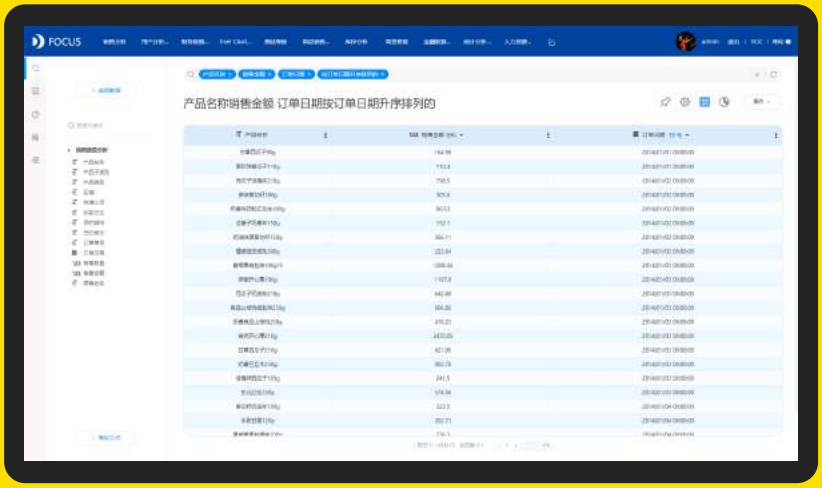
筛选过滤，是我们数据分析中最常用的手段之一。比如在对店铺进行销售分析时，经常需要单独查看某一家店铺的具体销售情况，这时候就需要对店铺进行筛选过滤，或者在进行时间序列的数据分析时，需要具体分析某一时期，这时候就需要对时间序列进行筛选过滤，又或者在分析销售数据时，需要知道哪些是没有达到销售目标的，故需要对销售金额或者销售量进行过滤，以上三种应用场景都是最常见的。

DataFocus中针对不同列类型的筛选过滤有不同的操作。其列类型有3种，分别为属性列、数值列、时间日期列，对这三种不同类型的列都可以直接进行筛选过滤。例如以下示例，是对不同列的筛选过滤，DataFocus中可以单独进行筛选，也可以进行多重筛选。

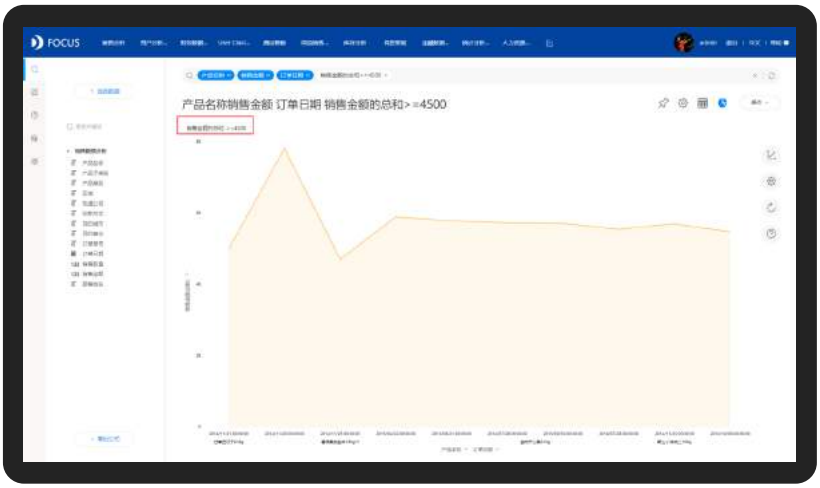
原始值：



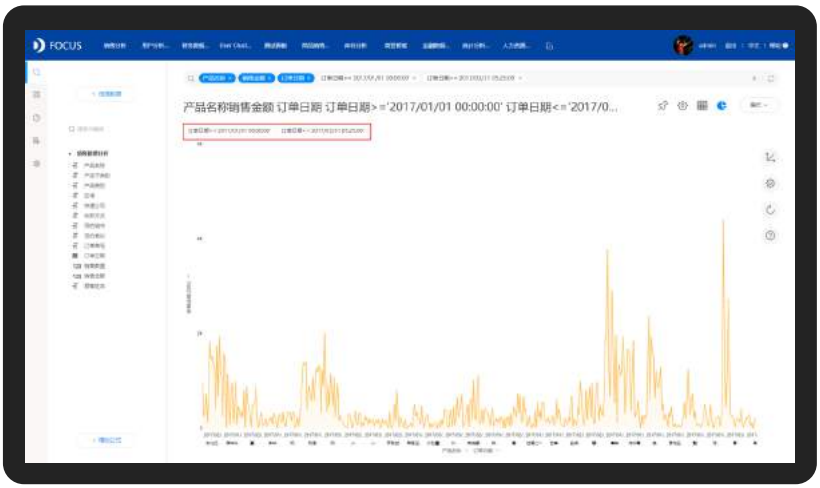
属性列筛选过滤：



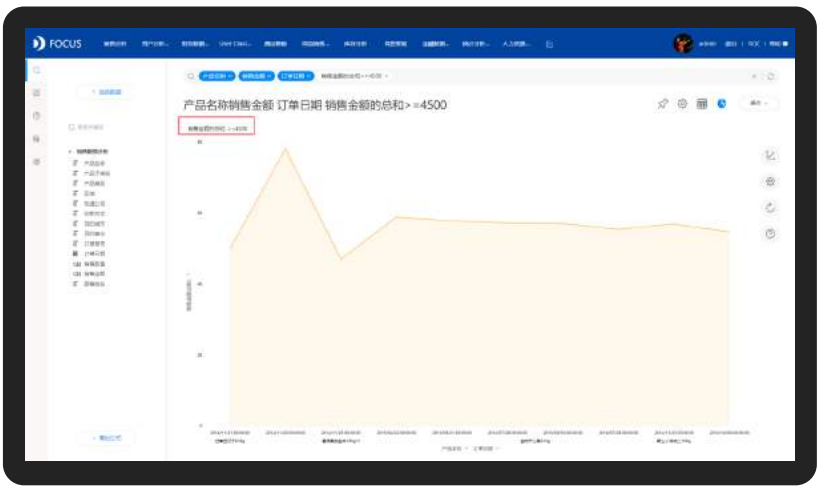
数值列筛选过滤：



时间日期列筛选过滤：



多重筛选：



中间表应用

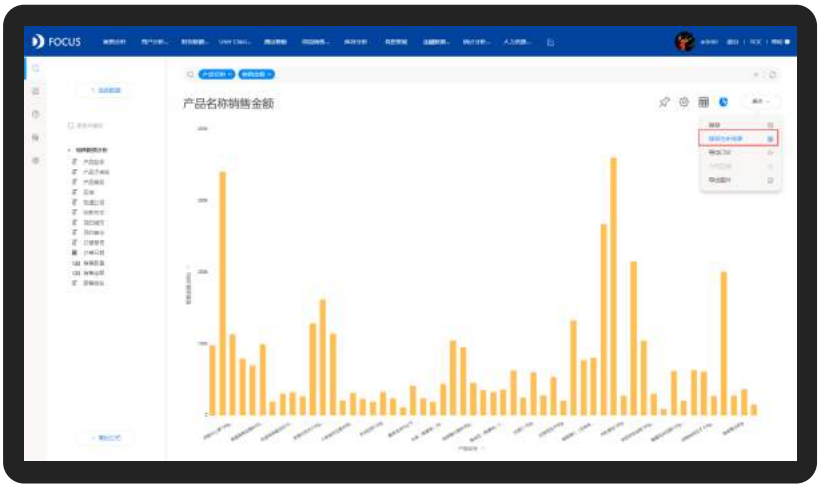
在DataFocus中，提出了一个新的概念，就是中间表的概念。中间表是什么，有什么用？这是很多人在第一次听到中间表之后的瞬间反应。DataFocus的中间表，就是用来处理数据的一种手段，比如在分析一个列非常多的数据表，但是只需要分析利用其中的某几列，这时候，可以先创建一个中间表，中间表中可以只包含需要分析的列，再用此中间表进行分析，这样不仅会提高系统分析的速度效率，也会节省用户的操作时间。

DataFocus中创建中间表有两种方式，一种是通过question的方式创建，即通过搜索的方式，将搜索结果保存为中间表，此种方式保存的中间表为聚合后的数据。

通过prejoin的方式

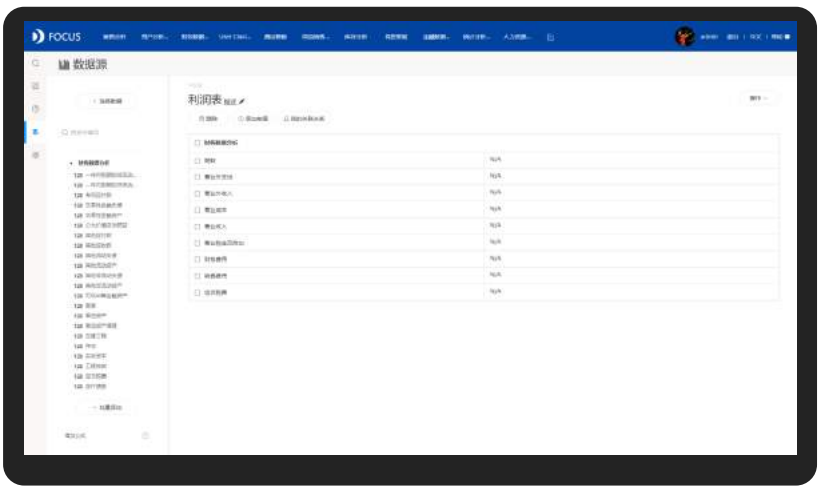
另一种是通过prejoin的方式，即按照原始数据创建中间表，表中数据是没有经过聚合的。

上述两种不同的创建方式，其入口也不同，前者是在搜索之后直接进行保存，后者是在数据管理页面中，选择创建中间表，进入到如下的创建中间表页面。



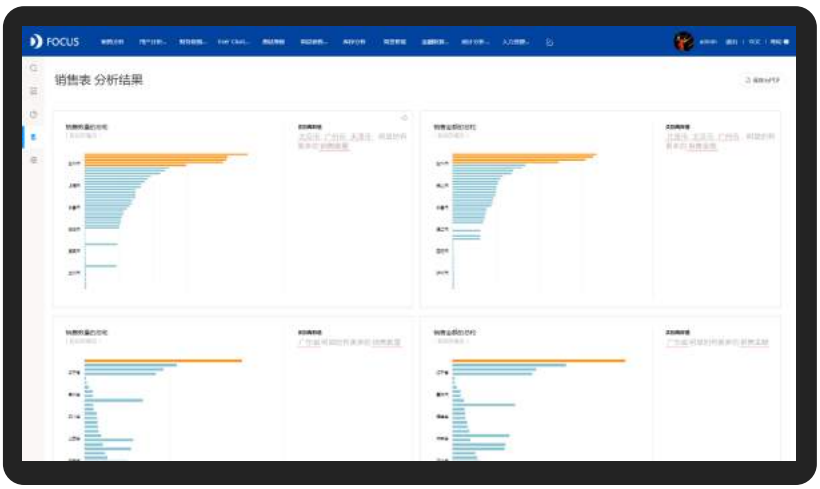
化繁为简

有时候，在创建完关联关系后，会将多个表中需要分析的列选出，保存到同一个中间表中，即化繁为简，后续分析相关的数据，就不需要再重复选择多个数据表，只需要选择一个中间表即可，此方法同样也加快了系统的分析效率。



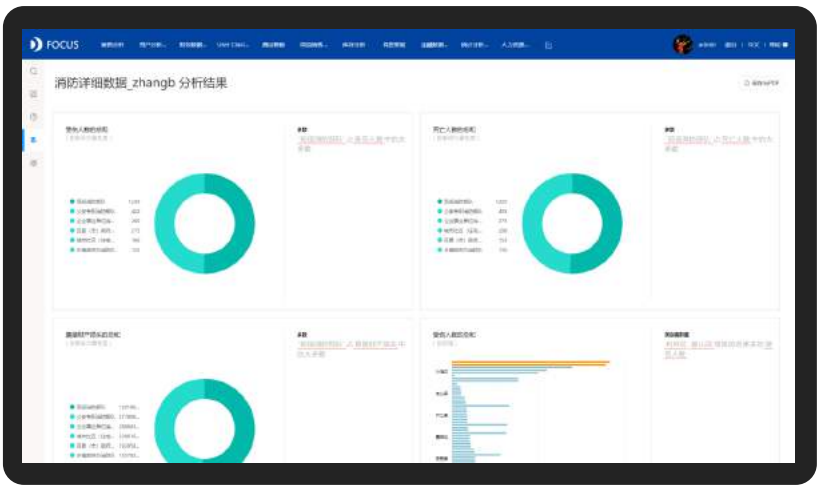
智能洞察

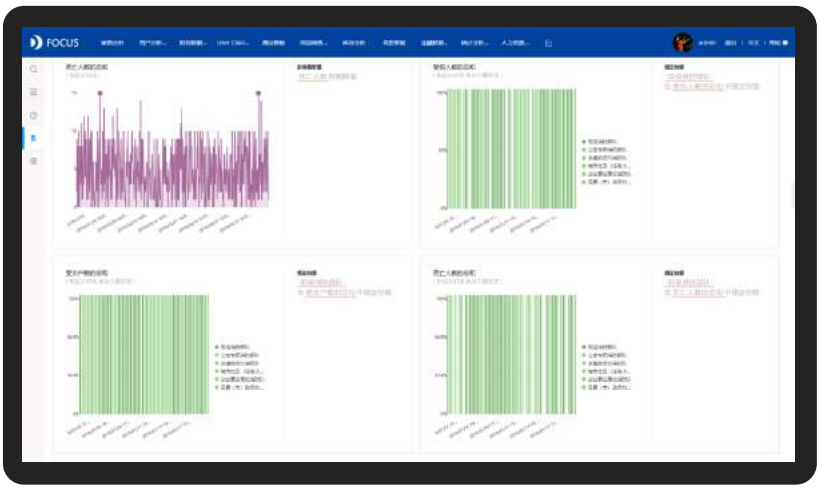
智能洞察是DataFocus非常具有竞争力的一个功能。其集合了人工智能技术，只要一键，就能从数千万的组合中，发现数据之间的各种规律，此功能大大减轻了用户的工作量。主要适用于分析商品种类众多的销售数据、库存数据，因为商品种类众多，在分析不同类别对销售的影响时，工作量会非常大，每个种类都需要进行分析，这时候，如果采用智能洞察一键分析，让系统来做这些工作，系统会从各个维度全面的分析所有数据，寻找数据之间的关系以及规律，形成一份报告，从这份分析报告中，我们可以清楚地知道哪些数据之间有线性关系，哪些数据中有异常值。如下图，即智能洞察分析出部分数据的类别离群值、时序离群值以及部分数据占多数情况、稳定份额情况等。



一键智能洞察

数据管理页面，一键智能洞察，从数千万自由组合中发现数据异常及规律，分析报告还可进一步导出成PDF文件，以供后续辅助进行详细分析。





数据支持

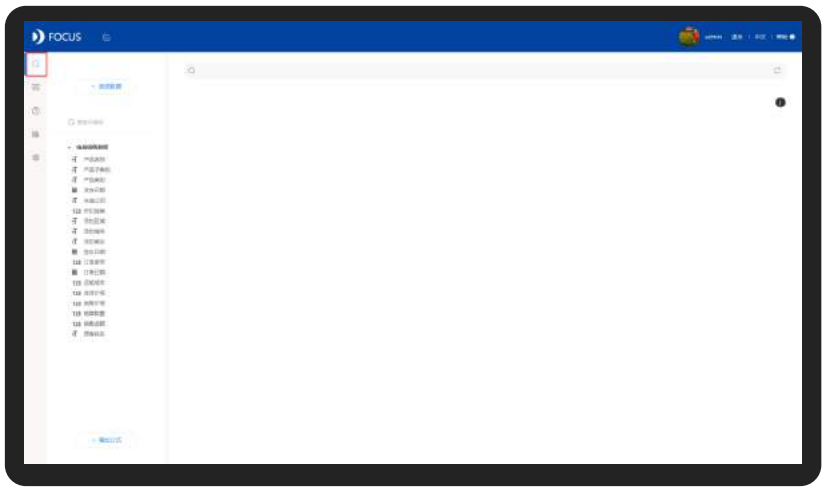
DataFocus连接数据的方式很全面，就大类来说，支持导入本地数据文件和连接到数据源两种。无论是零散的本地报表或数据文件，还是企业ERP、OA等系统中的数据库都可以覆盖支持，而如媒体、互联网、铁路管理等等对即时分析要求较高的一些行业，DataFocus还提供了直连数据库的功能，真正做到可以随取数据随时展现。

数据接入方式	支持格式
本地数据上传	excel文件、csv文件、json
数据源导入	Oracle、PostgreSQL、SqlServer、MySQL、Sybase、SAP
数据源直连	MySQL、Oracle、SqlServer

搜索分析

搜索模块是DataFocus系统中最核心的功能模块之一，也是将业务数据进行可视化的主要模块。从下图可以看出，DataFocus系统中特设了一个搜索模块辅助进行数据分析。点击页面左侧的导航栏，就可以切换对应的模块。

点击页面左侧的导航栏，就可以切换对应的模块。



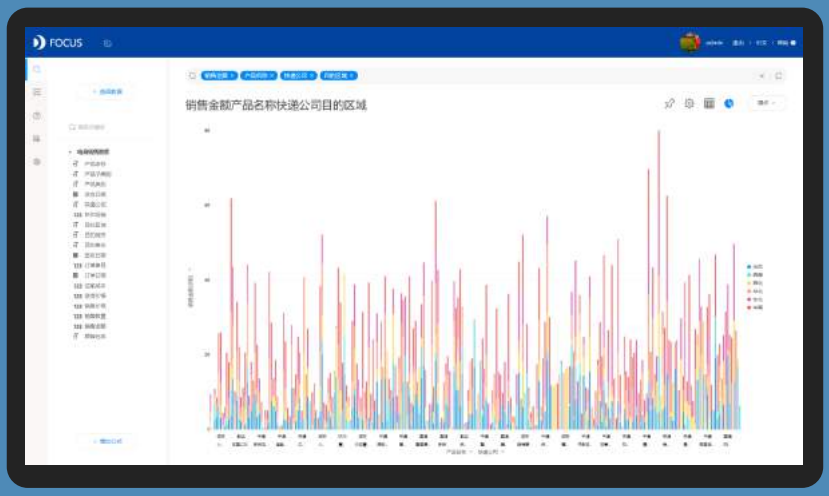
本章节介绍的内容主要包括两大部分，首先是介绍简单的多维度搜索分析，其次就几个较有代表性的关键词，进行分类讲解，介绍一下DataFocus首创的中文自然语言搜索的功能，也就是关键词搜索的内容。

先讲述常用的中文关键词，后介绍英文环境下的关键词。关键语句可以对其划分系列，主要包括时间日期相关、排序相关、字符相关以及较复杂的复合语句搜索。

维度搜索分析

在DataFocus中进行数据分析，唯一的交互方式就是搜索。最常见，也是使用最为频繁的数据分析模式，就是对数据进行维度分析。无论是单维度，还是多维度，在DataFocus系统中用鼠标双击需要分析的字段，或者在搜索框中输入需要分析的字段名，系统就会自动进行计算，并将分析结果以最合适的图形展现出来。相比于拖拽式分析，维度搜索不需要先配置图表，也不需要进行多次拖拽配置。

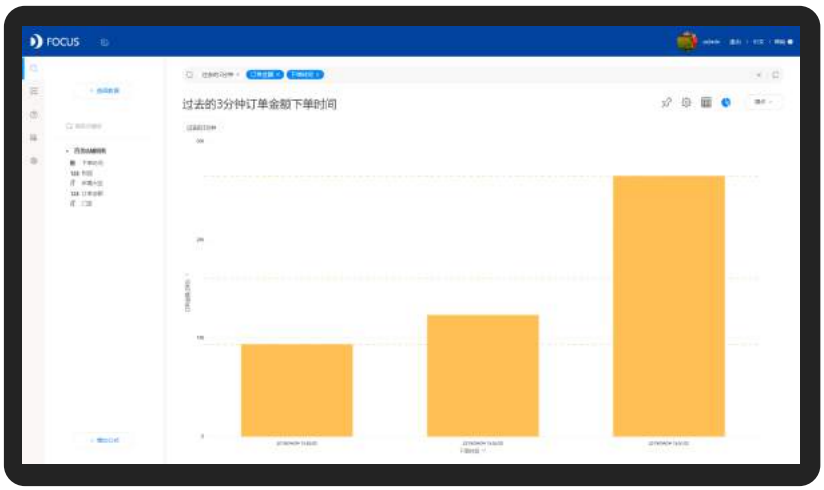
搜索框中输入需要分析的字段名，系统就会自动进行计算



通过关键词搜索分析

与拖拽式分析工具相比，你没有必须要去做复杂的处理，甚至也不需要记忆全部的关键词

简单的维度分析，并不能充分展示搜索式分析的优势。如果需要进行较为复杂的分析，则需要辅以对应的关键词进行搜索。比如，当我想知道最近3分钟内的订单情况，在搜索栏输入“过去的3分钟”、“订单金额”、“下单时间”，即可查看过去3分钟下单的订单金额之和，非常智能化。与拖拽式分析工具相比，你没有必须要去做复杂的处理，甚至也不需要记忆全部的关键词，你只需要在搜索框中输入“过去”，系统将自动提示下一步的输入内容，您只需要根据提示补齐对应的内容即可。



利用关键词搜索

同样，如果我想查看早前X天的数据情况，可以利用关键词“最近/过去/前X天”。在搜索框键入“前X天”，然后双击选择“产品”、“实际完成量”，就可以得出这些产品设备以当天为起始的前3天的实际完成量。

通过关键词搜索分析

数据分析中经常用的同比和环比，又是如何通过搜索实现的呢？输入关键语句“按订单日期计算的销售金额的总和的月增长率”，（再次，您不需要进行完整的输入，只需要在搜索框中键入“按”，系统将自动提示您完成输入。）可以对销售金额数据的总和就订单日期进行月环比增长率的计算，直接得出环比增长率变化情况，浅蓝色代表下跌，深蓝色代表上升。

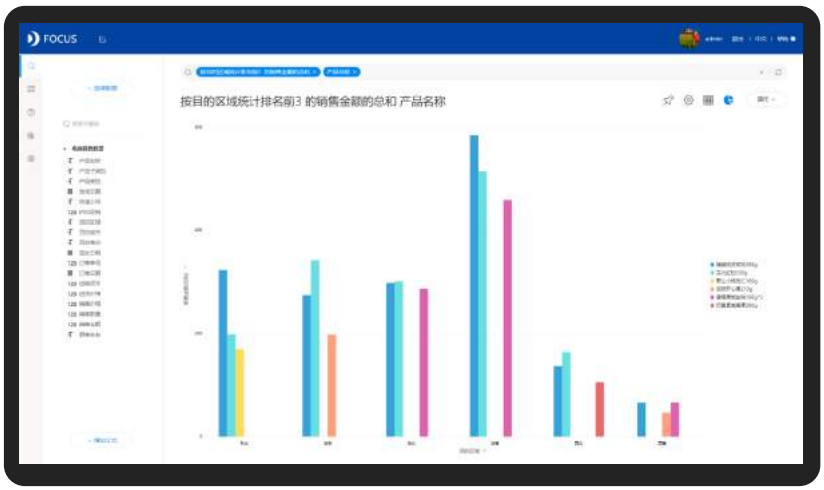
计算同比情况则是在环比的基础上添加一句“与往年同期相比”即可。

依次类推，所有您在数据分析时需要进行的复杂分析，都可以通过简单的关键词搜索得到结果。比如，“排名前x的产品销售额”，“姓名开头是‘王’的销售员上个月销售金额”，等等。不用担心关键词的记忆，系统会非常高效的提示您的输入。

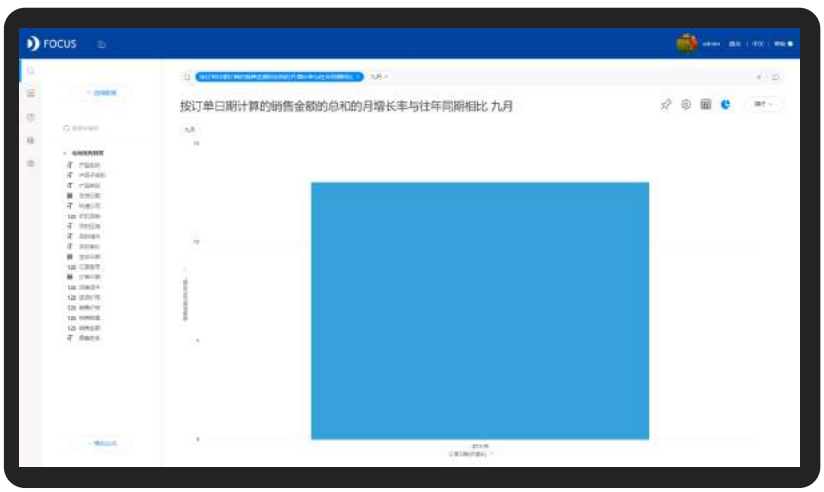
高级搜索分析

有人会问，这种搜索的方式能否覆盖所有数据分析的可能场景呢？换句话说，DataFocus没有提供SQL查询接口，以便用户在搜索无能为力的时候，通过写代码来实现吗？答案是否定的。DataFocus在设计之初的目标，就是要报证用户在不写一行代码的前提下，完成所有的分析，这样说可能会让一些工程师产生反感。但是，请不要武断的生气，想想看，SQL语言也不过是一门语言而已，一般情况来说，自然语言的表达比程序语言丰富得多，哪怕是有规则的自然语言，如DataFocus的关键词语言体系，也是如此。

我们可以通过组合各种关键词，进行相对复杂的搜索，以完成各种高级分析场景。举例来说，如果想对电商销售数据中的各目的区域中销售额排名前3的数据进行统计，查看不同区域销售的产品之间的差异，可以在搜索栏键入“按目的区域统计排名前3 的销售金额的总和”、“产品名称”。



还可以多重叠加关键语句，比如在搜索栏键入“按订单日期计算的销售金额的总和的月增长率与往年同期相比”、“九月”，可以单独提取9月份的同比情况。



高级搜索分析

还有一种对数值进行分组分析的场景，如在搜索栏键入“按销售金额 分7组统计的”、“销售金额的数量”，可以得出销售金额平均分为7组后，各组所占的销售金额的数量。

数据可视化

单个的图表可视化

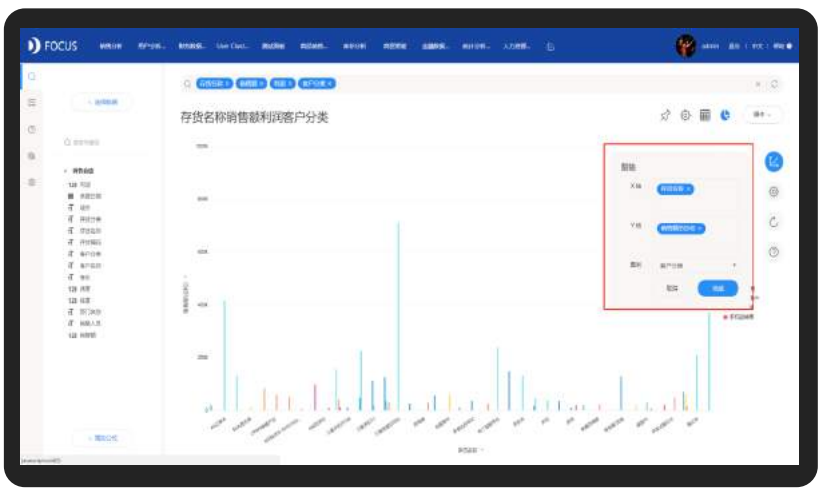
DataFocus中包含丰富的图表，具有超强的个性化设置，支持移动端操作系统。另外，用户可根据需求，自行导出图形或图形对应数据信息。目前，系统支持的图表类型高达36种，分为基础图表类型和高级图表类型，系统会根据用户当前分析的字段个数和种类，自动推荐合适的图标类型。

个性化的图表配置

DataFocus支持图表样式个性化配置，用户可以修改坐标轴、图例等属性，图表布局和风格设置，使图表更加美观。

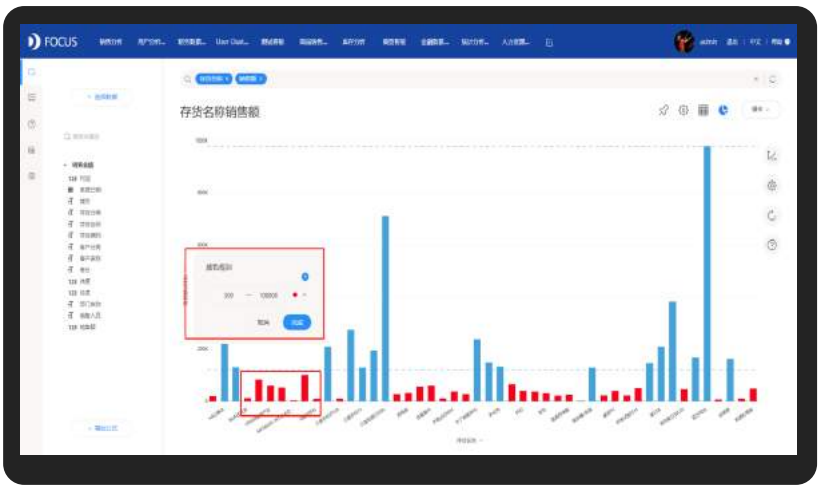
图表颜色：
提供默认主题颜色，还支持自定义修改图形颜色。

配置图：
支持多个x轴，当图形中包含多个x轴变量时配置框中会显示所有的x轴变量，用户可以对其进行更改。



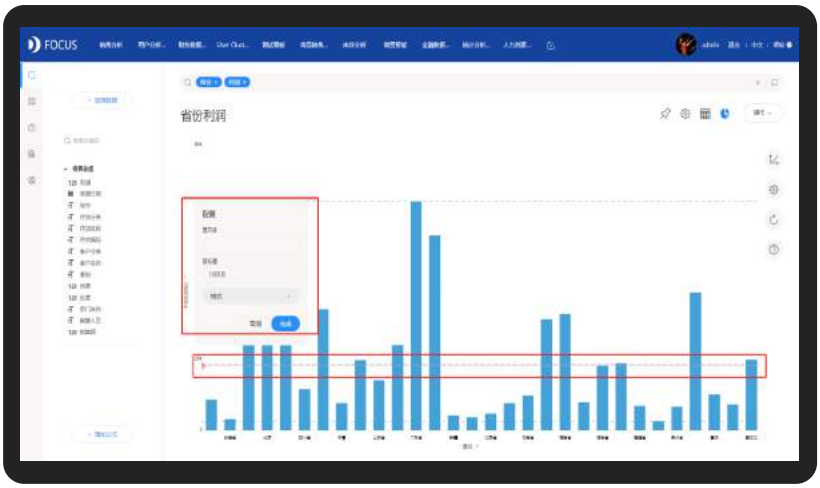
配置颜色规则：

可设置将某一个自定义范围内的值在图表中显示一个颜色。



可配置标准线：

设置一个标准目标值在图中显示。



配置悬浮文本：

可设置数据标签文本，添加额外的文本信息。

表格属性：

提供字体大小、文本颜色、财务型负数极其颜色，以及热图模式的主题颜色配置，根据用户的审美和自身需要进行调整。

图形属性：

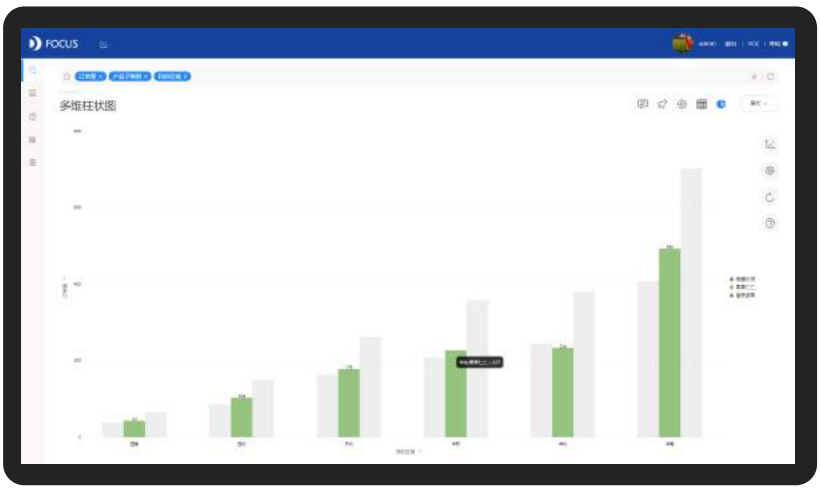
提供主题颜色、字体大小、图例显示位置、线条粗细、半径起点等微小属性设置，根据用户的审美和自身需要进行调整。

多样的图表交互效果

DataFocus具有丰富多样的图表交互方式，帮助用户更好的查看数据，极大的提升用户体验。DataFocus支持的交互效果：

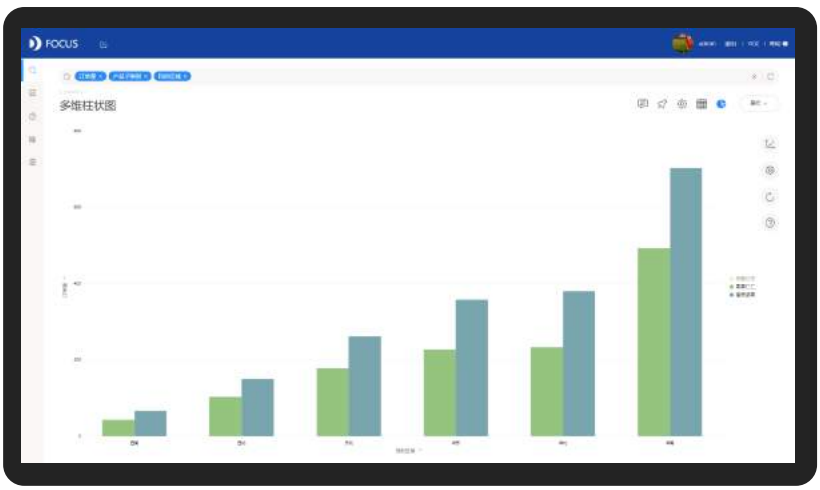
数据高亮显示：

当鼠标移动到图表的某一数据点时，会显示该数据点所在的坐标值和相关信息，并且该图表线条或区域将高亮凸出显示；当鼠标移动到某一图例时，该图例对应的线条或区域将高亮凸出显示。



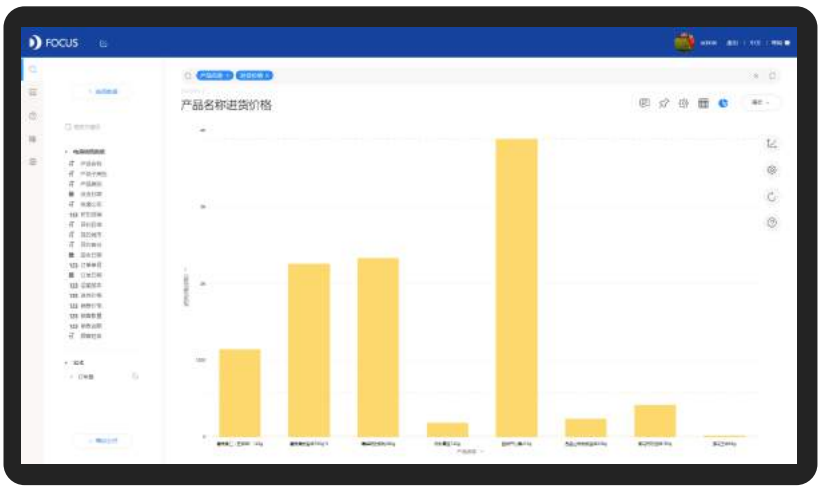
选择性显示：

点击图例可以隐藏该图例对应图形，以方便用户更好的阅读数据。



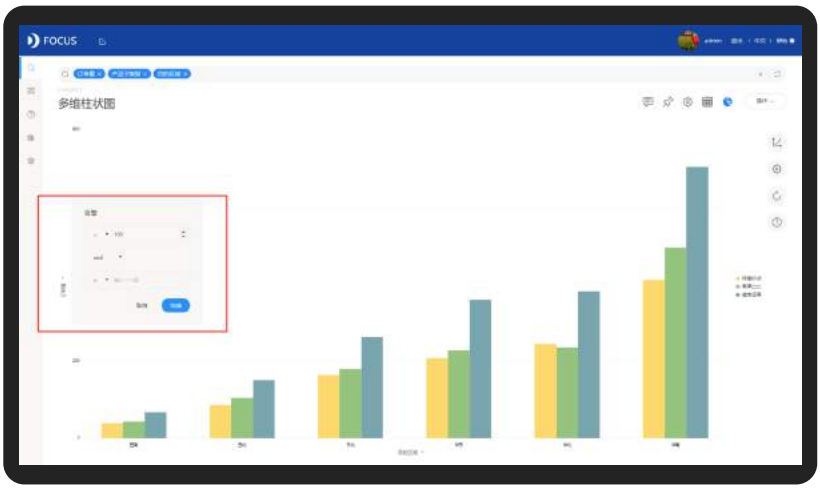
图表切块：

支持拖动鼠标将图表进行切块放大。



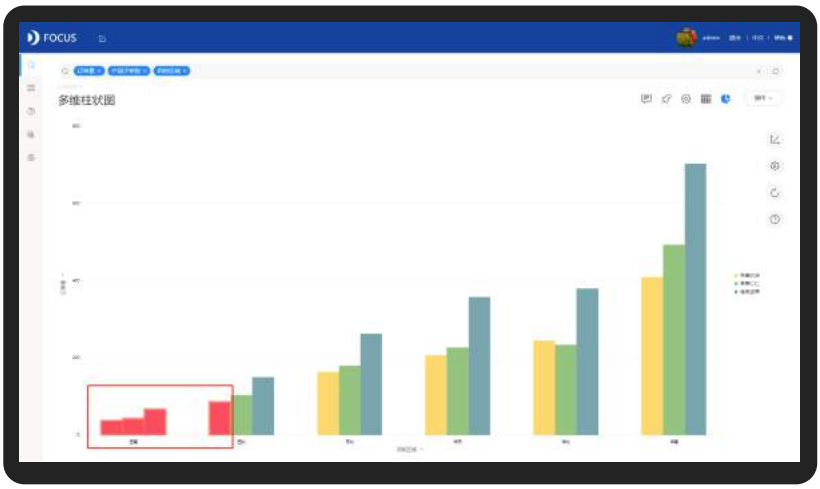
图表翻页：

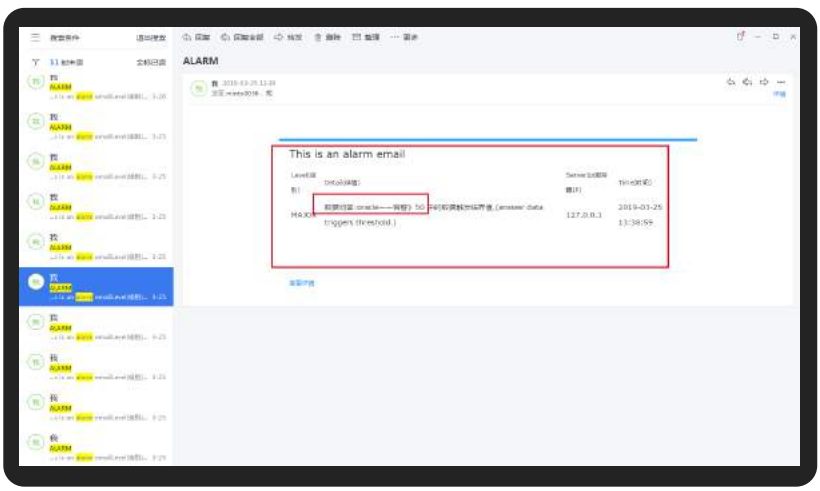
当坐标轴信息列较多时，支持图表翻页查看。



数据预警：

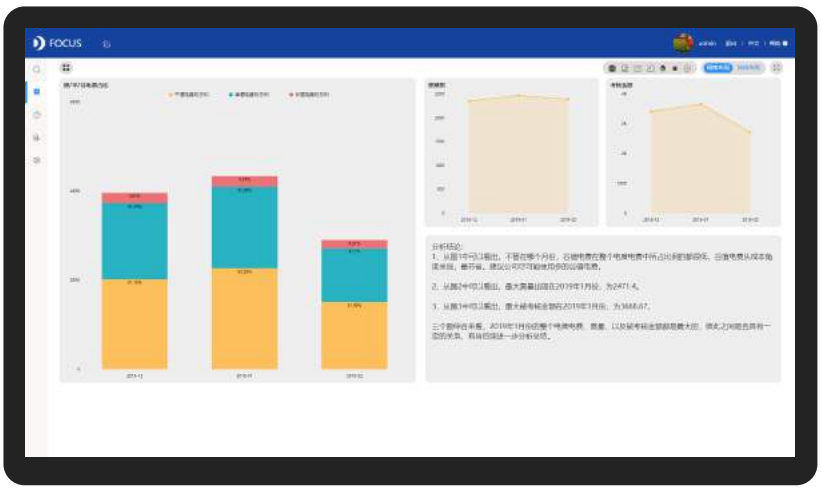
可自行设置数据峰值、区间等预警，当数值复合设定的条件时，将会以高亮闪烁动画的方式突出显示，并发送邮件通知。





数据看板和大屏

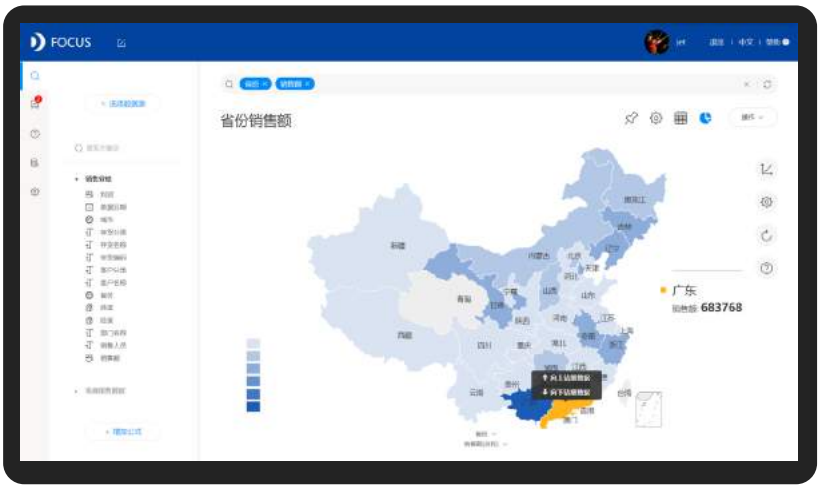
DataFocus种的数据看板（数字化大屏），有网格布局和自由布局两种形式。网格布局可以快速调整图表大小，分为25%、50%和100%三种样式大小，可拖动图表调整图表顺序。网格布局是默认布局形式，用户可以通过向看板中钉入多个问答（图表）的方式，快速开发出专题数据报告。



如果想开发个性化的数据可视化大屏，则可以通过自由布局的方式，通过拖拽图表大小和位置，改变图表或大屏的组件配置参数，如界面背景、组件背景、标题栏背景、组件文字、标题文字、隐藏副标题等美化设计，自由挥洒您的视觉创意。当然，系统也提供了多种预制样式主题，用户可以通过自定义组件的组合，设计出各种图片轮播、动态效果等等，让可视化大屏的可读性和阅读感更强。

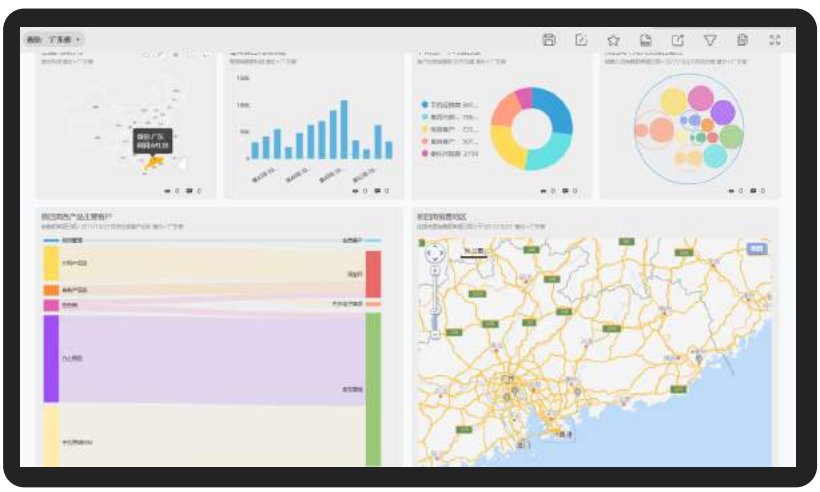
数据钻取

可以在问答（图表）上直接进行数据钻取，也可以在数据看板（数字化大屏）页面，对单个问答（图表）进行上卷（roll up）和下钻（drill down）。DataFocus中进行数据钻取，无须配置钻取路径，也不需要进行额外的其他配置，原则上，只要数据本身具有包含关系，比如省的字段下面包含地市字段，则系统就会自动判断可以钻取数据。例如，将利润最高的省份“广东”向下钻取地级市，查看利润分布情况，只需要右键点击广东地区，就可以看到钻取选项。



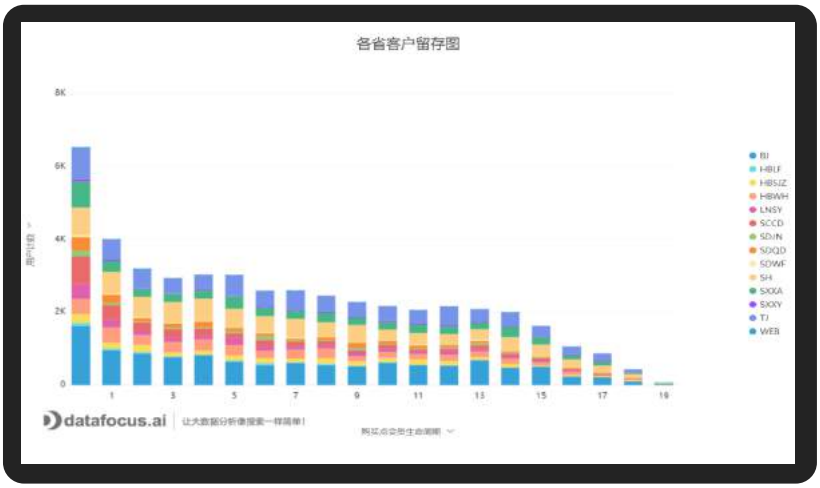
大屏联动

用户在数据看板中可以通过数据联动，实现关联分析，即通过筛选器对数据进行过滤。数据联动是自动进行的，也不需要预先配置，系统会自动判断数据看板中的问答（图表）是否有共用数据，然后将联动筛选条件附加到相应的图表，进行批量筛选，从而实现数据联动的效果。如，在DataFocus看板中点击位置图上的“浙江省”区块，所有字图标联动变化，显示浙江省访问渠道来源比率、浙江省注册用户数变化趋势，浙江省不同年龄段用户差异等。



报告输出

有时候，用户在制作演示报告时需要引用DataFocus系统的数据。系统提供了两种方式：一种是将问答（图表）直接导出为图片格式，另一种是将DataFocus数据看板（可视化大屏）直接打印成pdf文档输出。第一种方式方便用户在文档中引用图表数据，第二种方式，则是用户可以直接利用数据看板制作专题报告。

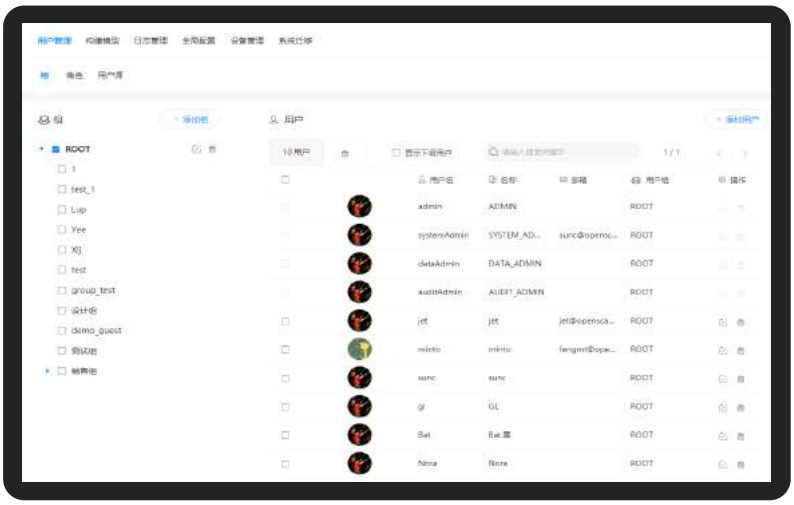


权限管理

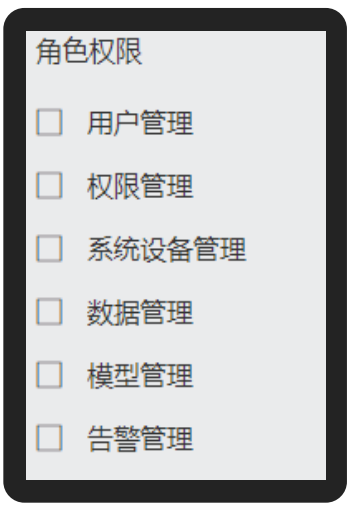
DataFocus的层级式企业系统管控方案，切合主流企业的管理运营模式，集用户管理、组织管理、权限管理、数据资源分发管理于一体，形成一个完整、安全、高效的信息化管理模式，在让系统用户更好地进行数据分析的同时，保障企业资源安全无虞。

用户权限

管理员通过用户管理页面可以对系统内的用户和角色进行统一管理。系统按照组织构架将用户划分进行管理，管理员可以采用手动添加与批量同步用户源数据集两种方式去添加用户。支持平台内置认证和LDAP单点登录，在对接OA、ERP等系统内的员工组织数据库，直接批量导入用户源会十分方便。



系统内置六种角色权限，三种预置角色和四个预置管理员用户账号，拥有用户管理权限的管理员可以对用户进行角色赋予操作，进行系统管理任务的分配。除了默认的几个管理员角色外，有用户管理权限的管理员还可以任意组合这六种权限添加为新的角色，通过将该角色赋予其他用户，让其他用户协同进行系统的集中管理。



预置角色

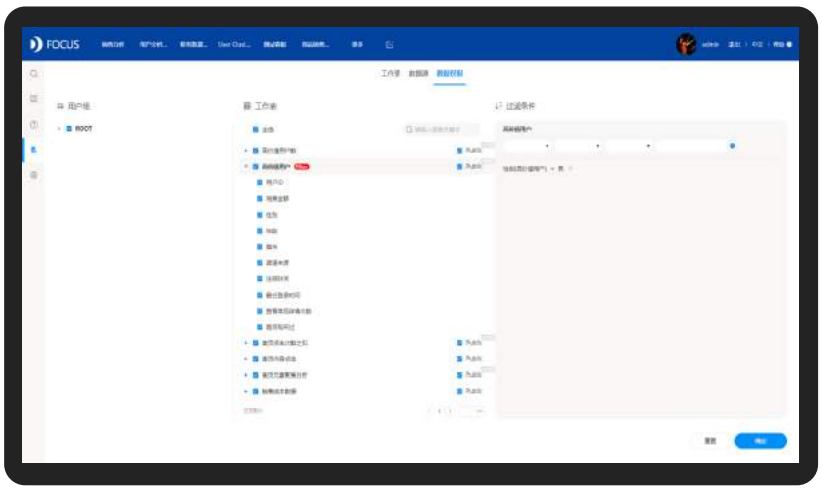
角色名称	拥有权限	备注
系统管理员	用户管理/角色权限管理/设备管理/告警管理	管理员可以配置用户，管理系统设备组件以及查看系统运行时的告警信息
数据管理员	角色权限管理/数据管理/模型管理	管理员可以导入数据配置数据权限以及管理BI模型
审计管理员	角色权限管理/日志管理	管理员可以查看审计日志，监控其他管理员的行为

预置管理员用户

用户名称	角色	备注
admin	系统管理员/数据管理员/审计管理员	拥有系统最高权限
sysAdmin	系统管理员	系统管理员，运行维护focus系统
dataAdmin	数据管理员	管理系统数据和模型
aduitAdmin	审计管理员	审计监控其他管理员行为

数据权限

DataFocus数据资源管控策略的特殊之处在于数据的权限分配依赖于组织关系而不是用户个人。拥有数据管理权限的管理员可以基于组织关系（即用户组），对数据表的使用权限进行精细到行列的分配。在数据管理-数据权限页面的选择用户组，然后选择需要分配的数据表并进行行列过滤条件的设置。当用户因为组织调动离开当前部门时，所在用户组发生了变动，相同的数据报告也会因为数据权限随组的变动而变化，只能看到当前部门所在组的权限范围内的内容，真正做到了不同角色通过同一系统看到符合自身需求的数据和报表，千人千面。



多端展示

随着科技和移动便携设备的发展，办公设备也越发的多元化，为了切实的满足用户需求和变化，DataFocus数据分析系统也相应的突破了时空的限制，实现了电脑、手机、pad等等的多端信息互通共享，搭建起多屏覆盖的数据分析决策系统，让用户可以随时随地的查看数据报告，即时进行分析决策，更好地掌控业务的发展变化，不愧于智慧管理智能分析的设计初衷。

移动端展示

为了避免浪费用户的移动设备空间，DataFocus与企业微信相结合，直接在企业微信内设置入口，随时查看数据报告。只要您的设备上有企业微信这个软件，无需多余下载，即可纵享数据分析。与PC端数据分析报表相同，DataFocus的移动端报表也做到了完美自适应多种尺寸，并支持多样化的数据展现，无论是钻取联动，还是消息提示，都完美继承了PC端的功能。

大屏应用

现今很多企业都会在会议、活动中应用到数据可视化大屏来展示公司的运营、销售、KPI达成等信息，或者是用来显示一些实时的数据指标来做到对现场、业务的监控。DataFocus的自由/网格双布局模式+全局/组件全面细节设置+多样化图表+丰富配色，可以完美支持各种尺寸要求的数据可视化大屏的制作与展示。

05

产品性能测评报告

Product performance
evaluation repo

测试环境

服务器1	系统	Centos7
	硬件	CPU: 6核 Intel(R) Xeon(R) CPU X5675 @ 3.07GHz 内存32GB 硬盘 固态硬盘 100GB
测试机	硬件	CPU: Inter Xeon L5640 内存 16GB
	系统	Windows7 64位
软件	产品版本	Datafocus-v2.16.1 标准版

数据准备

常规数据：

MySQL数据库新建表 “10000W”， 10000W 数据量；“5000W”， 5000W 数据量；
“500W”， 500W 数据量；

7 个字段。表字段如下，其中 GROUP10 表示 10 个分组，GROUP100 表示 100 个分组以此类推。

名	类型
GROUP10	VARCHAR
GROUP50	VARCHAR
GROUP100	VARCHAR
GROUP1000	VARCHAR
GROUP10000	VARCHAR
GROUP100000	VARCHAR
MEASURE	NUMBER

TPC-DS 数据：

版本： v2.10.1　　数据量 10GB.

详细数据结构：

http://www.tpc.org/tpc_documents_current_versions/pdf/tpc-ds_v2.10.1.pdf

测试过程

1. 常规数据 使用measure列作为ID导入的方式 导入数据所使用时间。测试结果单位（s）：

当数据量在1千万行内，导入时间相对耗时较少，导入三次平均耗时可以控制在15分钟以内，随着数据量的增长。导入数据所消耗的时间呈现指数级增长。当数据量达到1亿行时，导入三次平均耗时为 02:03:00。

根据每5% 导入进度来看：

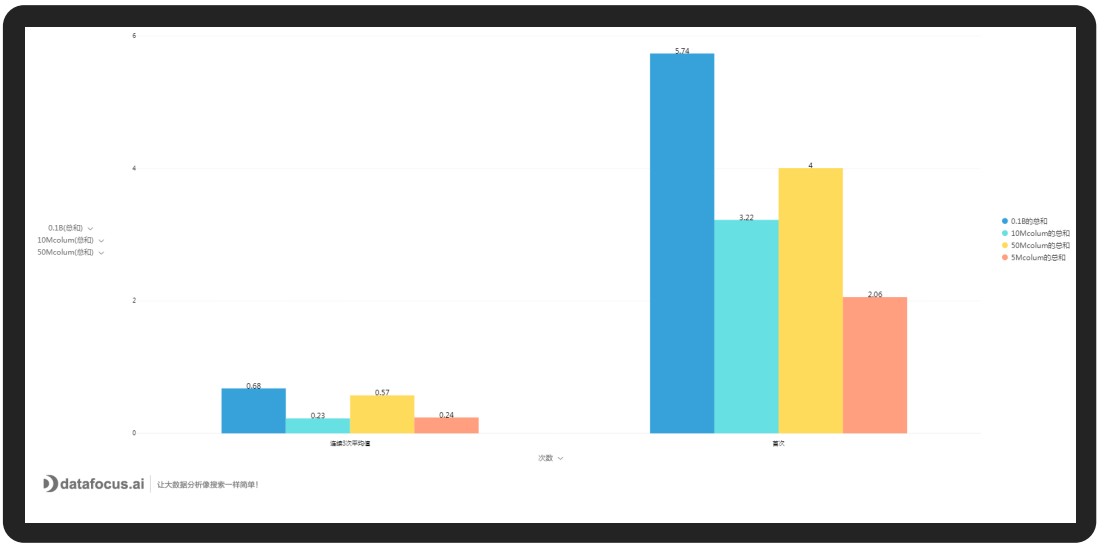
当数据量在1千万内时，每5%进度导入所消耗的时间比较接近

当数据量超过5千万行时，每5%进度导入所消耗的时间随着导入进度的增长而延长

（*）当不使用任何配置导入时耗时会有较大浮动，导入大数据量时（>500万行），建议用户使用ID导入的方式

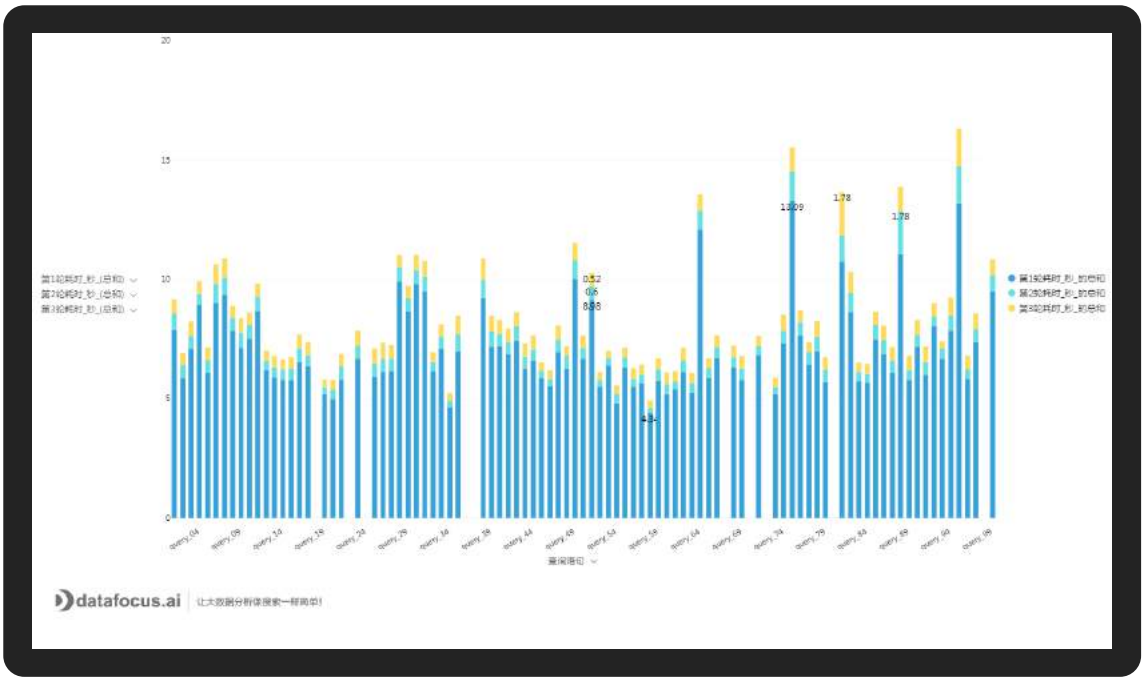
(*) When you do not use any configuration import, it will take a lot of time to float. When importing large data volume (>5 million lines), it is recommended that users use ID import method.

2. 单维度测试：1个attribute，1一个measure指标同时增加关键字搜索（如同比计算的）。



首次搜索时耗时相对较多，当进行多次查询时感知不到等待时间

3. TPC-DS测试结果：



详细数据：

01

测试方法：在Datafocus中依次对执行每个query，共执行3轮，计算每轮中各个query 耗时.

查询语句	第1轮耗时（秒）	第2轮耗时（秒）	第3轮耗时（秒）
query_01	7.771	0.656	0.612
query_02	5.776	0.54	0.499
query_03	6.97	0.55	0.602
query_04	8.814	0.439	0.533
query_05	5.994	0.512	0.541
query_06	8.882	0.77	0.821
query_07	9.212	0.71	0.81
query_08	7.712	0.554	0.498
query_09	6.997	0.661	0.597
query_10	7.407	0.565	0.502
query_11	8.547	0.574	0.557
query_12	6.097	0.401	0.435
query_13	5.793	0.443	0.449
query_14	5.697	0.436	0.409
query_15	5.698	0.487	0.477
query_16	6.438	0.554	0.593
query_17	6.257	0.444	0.561
query_18	0	0	0
query_19	5.124	0.288	0.321
query_20	4.912	0.399	0.398
query_21	5.718	0.55	0.496
query_22	0	0	0
query_23	6.567	0.559	0.603
query_24	0	0	0
query_25	5.825	0.554	0.6123
query_26	6.032	0.553	0.654
query_27	6.062	0.509	0.5877
query_28	9.778	0.565	0.512
query_29	8.525	0.554	0.512
query_30	9.654	0.591	0.6123
query_31	9.356	0.609	0.65
query_32	6.063	0.388	0.4
query_33	6.97	0.521	0.512

详细数据：

02

测试方法：在Datafocus中依次对执行每个query，共执行3轮，计算每轮中各个query 耗时.

查询语句	第1轮耗时（秒）	第2轮耗时（秒）	第3轮耗时（秒）
query_34	4.551	0.298	0.309
query_35	6.891	0.71	0.755
query_36	0	0	0
query_37	0	0	0
query_38	9.077	0.771	0.876
query_39	7.079	0.6	0.677
query_40	7.103	0.501	0.57
query_41	6.763	0.51	0.555
query_42	7.32	0.577	0.601
query_43	6.175	0.494	0.532
query_44	6.491	0.471	0.5784
query_45	5.765	0.31	0.365
query_46	5.447	0.299	0.369
query_47	6.863	0.501	0.577
query_48	6.181	0.509	0.417
query_49	9.881	0.773	0.69
query_50	6.565	0.449	0.52
query_51	8.984	0.599	0.517
query_52	5.415	0.281	0.32
query_53	6.291	0.331	0.299
query_54	4.752	0.381	0.341
query_55	6.228	0.411	0.399
query_56	5.424	0.331	0.435
query_57	5.569	0.371	0.398
query_58	4.336	0.208	0.31
query_59	5.673	0.474	0.447
query_60	5.114	0.411	0.475
query_61	5.314	0.321	0.431
query_62	6.014	0.481	0.522
query_63	5.167	0.399	0.411
query_64	11.912	0.771	0.689
query_65	5.781	0.432	0.401
query_66	6.612	0.439	0.494

详细数据：

03

测试方法：在Datafocus中依次对执行每个query，共执行3轮，计算每轮中各个query 耗时.

查询语句	第1轮耗时（秒）	第2轮耗时（秒）	第3轮耗时（秒）
query_67	0	0	0
query_68	6.221	0.411	0.501
query_69	5.712	0.455	0.512
query_70	0	0	0
query_71	6.714	0.391	0.412
query_72	0	0	0
query_73	5.118	0.289	0.374
query_74	7.198	0.534	0.654
query_75	13.09	1.211	0.999
query_76	7.531	0.55	0.499
query_77	6.343	0.503	0.408
query_78	6.89	0.574	0.685
query_79	5.6211	0.499	0.535
query_80	0	0	0
query_81	10.591	1.09	1.78
query_82	8.49	0.809	0.854
query_83	5.661	0.364	0.412
query_84	5.571	0.389	0.403
query_85	7.382	0.587	0.554
query_86	6.77	0.571	0.601
query_87	5.995	0.48	0.58
query_88	10.891	1.78	1.011
query_89	5.7146	0.4	0.583
query_90	7.076	0.493	0.599
query_91	5.921	0.534	0.636
query_92	7.915	0.408	0.555
query_93	6.581	0.414	0.301
query_94	7.71	0.669	0.712
query_95	12.996	1.527	1.555
query_96	5.751	0.4	0.55
query_97	7.246	0.567	0.631
query_98	0	0	0
query_99	9.361	0.669	0.647

ps:"0"表示未得出计算结果

06

关于DataFocus

About DataFocus

DataFocus不仅给企业和个人带来更高效便捷的使用体验，真正实现像搜索引擎一样与大数据进行交互；更为大数据分析领域带来革新的动力，让停滞多年的商业智能BI领域能够重新开始向更智能的方向进化。

13种不同的行业，服务超过200家企业或组织

DataFocus立足于满足不同行业客户的各个需求，在各个行业领域内建立了良好的口碑，品牌影响力逐年上升，产品关注度持续增长。截止目前，DataFocus已深入到13种不同的行业，累计为超过200家以上企业或组织提供智能化、数字化转型支持，便捷的搜索式分析填平了企业的数据生产者和消费者鸿沟，为企业的全面数字化管理转型奠定了坚实的基础，跟得上思考速度、及时的数据决策为用户带来看得见的价值！

节约成本，高使用效率的人工智能产品

DataFocus诞生于一群充满激情的技术极客之手，为用户打造完美的产品是他们孜孜以求的前进动力。未来他们将继续致力于为广大合作伙伴提供低使用成本、高使用效率的人工智能产品：继续完善和打磨产品的每一个交互细节，提升各种复杂数据环境下的运行性能，在新商业智能领域中和行业同仁共同推动整个行业的发展。

试用期长达1个月

为了让客户能够更好地体验和熟悉这款产品，DataFocus目前采用免费开放试用模式，试用期长达1个月，无论是企业或个人都可免费试用DataFocus强大的搜索式分析，让人人都可成为专业的数据分析师。

“开启您的搜索式数据分析之旅吧”

datafocus

HANGZHOU WISDOM CONVERGE
TECHNOLOGY CO., LTD.