DataCanvas竞品分析

**一、简介**

产品名称：DataCanvas

体验平台：PC端 Web

Slogan：五分钟拥有大数据能力

面向用户：数据分析师、数据科学家

目前使用用户群体：金融行业

**二、用户特征、市场概要**

目前，DataCanvas产品的客户都是来自于金融行业。该类客户注重数据隐私和安全，数据保密性极高，面向使用用户是算法工程师、数据分析师，具备一定编程和算法能力的人群等。

用户特征：

1. 面向使用用户局限性高，因为使用DataCanvas产品的用户，要有具备一定的编程核数据分析能力，并不是面向广泛的群众。

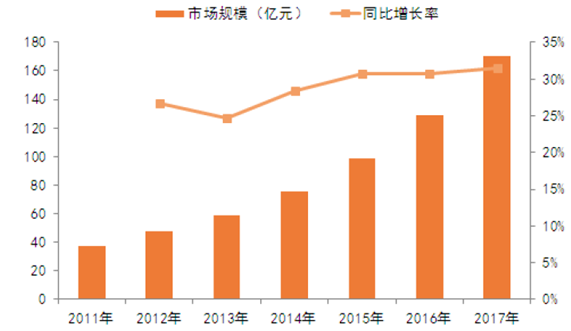
2. 熟悉各种业务，根据不同的业务场景，使用不同的算法模块，每天使用DataCanvas产品的用户，通过对业务的熟知和理解，可以从多维度角度得出不同的结果，从而更精准的分析和决策，创造商业价值。

3.考核KPI导致高频，DataCanvas的算法模块量，目前已被多家客户采取作为衡量这些使用者们的考核KPI，从而间接促进了 高频使用产品。

4.分析结果，面向的人群大部分是boss，或公司决策者。以结果为导向，作为重要参考分析依据，更能准确定位分析、决策，从而创造更多的商业价值，所以这些人更注重分析结果，而不是过程。能方便直接查看分析结果，是这些用户的刚需。

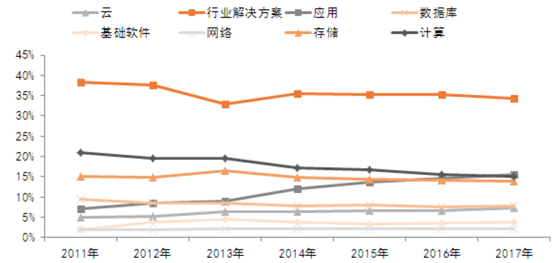
市场概要：

大数据已上升为我国国家战略，2016年3月，“十三五”规划纲要全文（以下简称“全文”）发布，全文第二十七章明确提出“实施国家大数据战略”，彰显国家对于大数据战略的重视。在国家政策的推动下，我国大数据行业市场规模高速增长。根据相关数据，2016年我国大数据市场规模增速将在30%以上。

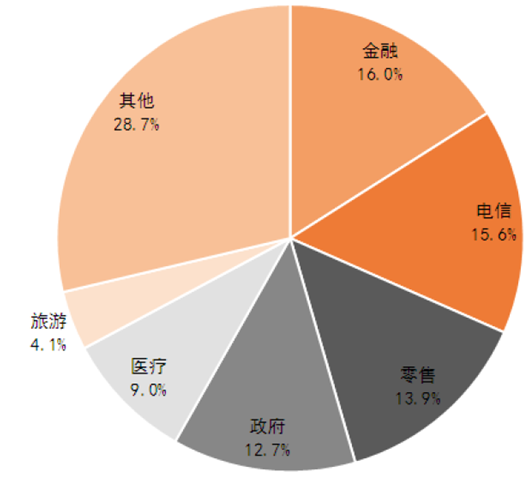


2011-2017 年我国大数据市场规模及增速

未来在我国大数据市场，行业解决方案细分领域将高速发展。根据数据，2011-2017年，全球大数据市场规模中行业解决方案的占比最高。我国目前的大数据市场，据相关研究结论，大数据整体解决方案/大数据行业解决方案产品和服务严重不足。参考全球大数据行业的发展轨迹。未来在我国大数据市场，行业解决方案细分领域将高速发展。



全球大数据市场各细分领域占比



2014年中国大数据市场行业投资结构

注：数据摘自<http://www.chyxx.com/industry/201606/426475.html>

**三、产品定位**

【目前】

DataCanvas是基于超融合架构的大数据平台，为行业客户提供数据基础架构的迁移，数据的导入，存储，快速查询和分析，为企业的业务和应用提供大数据的支撑，更快更好地帮助企业实现数据化决策，产生商业价值。

【未来】

逐渐发展成Data Science Platform 平台。

**四、产品核心竞争力**

1.分布式架构存储

将企业原有架构迁移到基于X86的大数据分布式计算架构之上，确保海量数据存储的高扩展性和运算的高效性。

2.实时数据处理

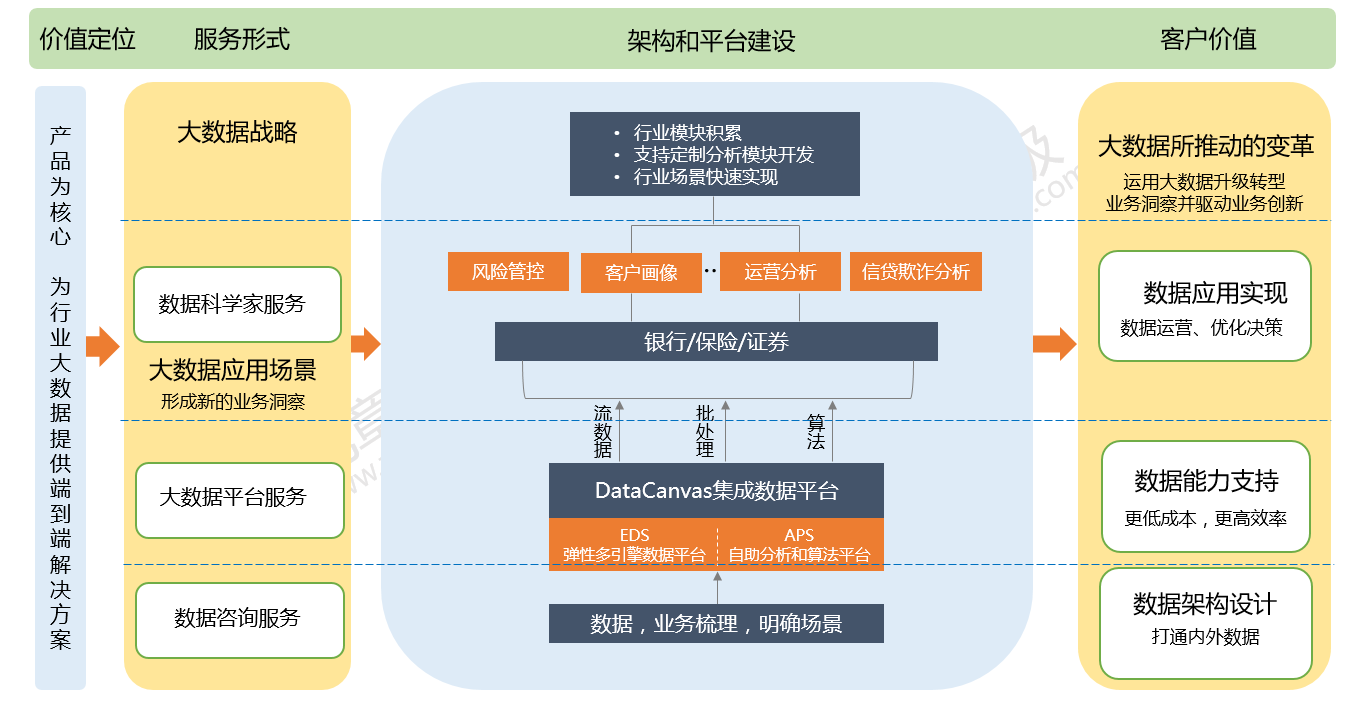
支持数据的实时入库、实时计算、实时查询、让实时数据处理变更更加现实和可靠。

3.全新自助式分析

基于机器学习模型的自助式大数据分析平台，提高数据分析效率，使数据分析周期缩短80%。

4.可视化应用

数据分析结果可视化呈现，直观揭示数据内在联系，帮助企业洞察商机，快速知道商业决策。



产品概况及愿景

**五、竞品分析**

**EDS方面的竞争对手**

竞争对手：星环

产品：Transwarp Data Hub （TDH）

技术优势：

1. 极致的性能与可扩展性。

TDH的批处理速度是开源 Hadoop的10-100倍,是MPP的5-10倍,可以对从GB 到PB级的数据量实现复杂的查询和分析。

2. 图形化的大数据开发工具套件。

Transwarp Studio 是TDH中的大数据开发工具集,包括元数据管理 Governor、工作流Workflow、数据整合工具Transporter, Cube设计工具Rubik以及报表工具Pilot。用户可以使用这些图形化工具来提高大数据的开发效率,降低技术门槛。

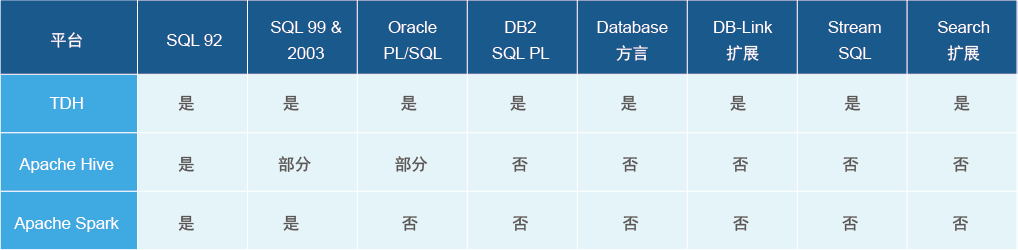
3. 低延迟的流处理

Transwarp Slipstream是同时支持事件驱动和微批处理的流处理引擎,计算延迟最低可至 5ms。它提供标准的SQL编程接口,还支持高可用性 (HA)和Exactly-Once的语义,从而支持7x24小时的生产业务。

4. 大数据上的全文搜索

Transwarp Search支持通过 SQL实现大数据上的秒级全文搜索,它利用层次化存 储、堆外内存管理等创新性技术,极大的提高了系统 的可用性。此外,Search还可以结合Inceptor提供较 强的数据分析能力。

5.平台支持



核心产品特点：

产品1： Transwarp Slipstream

特点： 提供实时计算

突出优势：

1. 基于事件驱动的计算引擎可将延迟时间缩减到5毫秒，是spark streaming引擎延时的1/100。

2. 支持复杂事件处理能力（CEP），因此用户可以基于Slipstream用SQL语言开发比较复杂的在线流计算业务。

3. 提供完善的高可用性（HA）和Exactly-Once语义。

产品2：Transwarp Discover

特点： 分布式机器学习平台

突出优势：

1. 包含丰富的分布式算法库，内置了多个行业应用模块。

2. 提供R语言、Python和SQL接口，以帮助数据科学家开发自己的数据挖掘算法。

3. 内置NoteBook工具Zeppelin,可以非常灵活的支持数据工程师和科学家之间的团队协作。

星环产品定位：

要做一家专注于大数据基础软件平台一站式产品的大数据公司，而不是一站式服务的大数据公司，也不是大数据服务提供商。

（摘自：<https://baijia.baidu.com/s?old_id=425768>）

星环目前客户群：

目标客户群是拥有极大数据量的公司，主要集中在金融、电信、交通、电力、政府等领域。

目前服务的客户有三大运营商、国家电网、民生银行等公司。

（摘自：<http://www.sohu.com/a/109292346_460376>）

TDH与DataCanvas EDS比较分析：

1.产品侧重点：

DataCanvas 更侧重平台提供服务，未来逐渐倾向数据分析，服务于直接面向用户是数据科学家、数据分析师等。

TDH 更侧重 海量数据的存储、处理，专注基础平台研发，以自主研发技术支撑产品，以此提供平台服务。

2.技术层面：

TDH 是源于Hadoop，并超越Hadoop，基于Hadoop，自主研发多项技术，体现在分析型数据库、流处理技术、大规模分析搜索引擎以及图形化的大数据开发套件方面等。

DataCanvas EDS 虽然没有自主研发技术，但是采用了很多成熟的技术，如Hadoop、Apache Ambari、Apache Hawq、Apache Hive、Apache Flume、Spark等。

3.资源背景：

TDH 研发团队大多来自知名外企，超过85%的员工为研发工程师，并以博士、硕士为主。

目前，星环已获得C轮融资2.35亿，腾讯领投。

DataCanvas 研发团队大多数来自于985、211学校，其中也有不少不乏来自知名大学，产品研发团队人员相对较少。

（摘自：<https://baijia.baidu.com/s?old_id=845874>）

**APS方面的竞争对手**

竞争对手：dataiku

产品：dataiku DSS

产品特点：

Connectivity 连接性 -- connect to more than 25 data storage

Data Wrangling 数据整理

Machine Learning 机器学习

Data Minings 数据挖掘

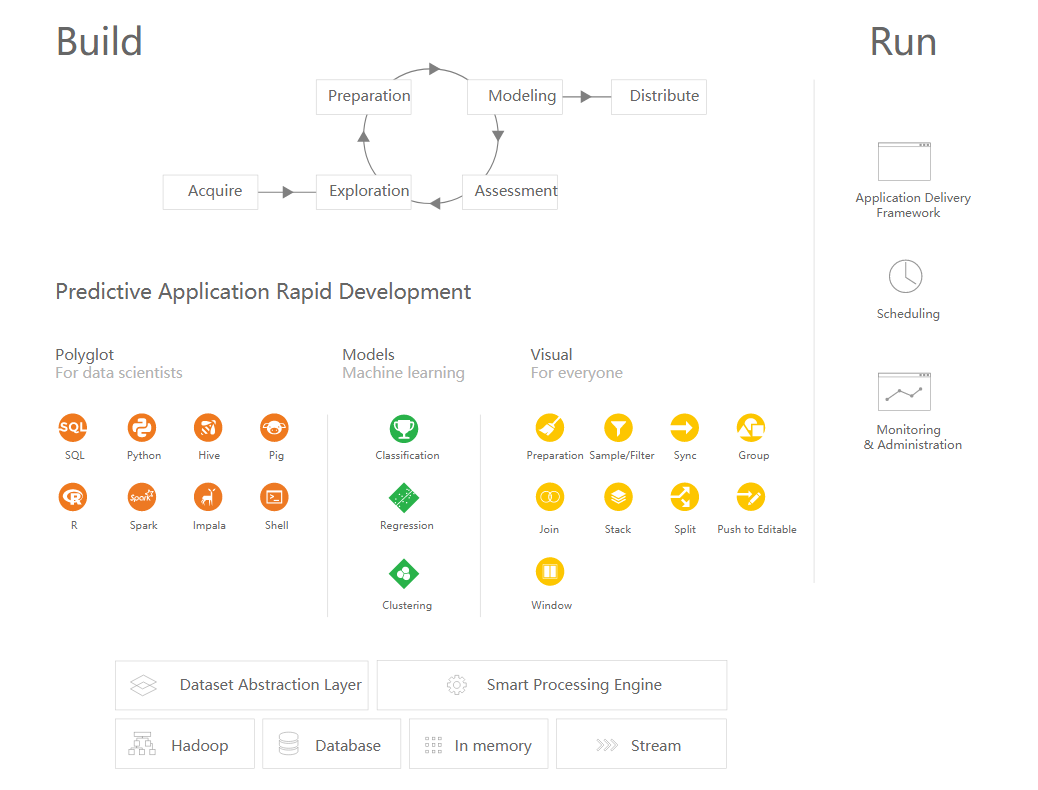
Data Visualization 数据可视化

Data Workflow 数据工作流

Realtime Scoring 实时监控

Collaboration 集成

Technology snapshot：



Dataiku产品定位;

为企业客户提供了一个基于云技术的数据服务平台，让数据专家和没有工程师背景的员工轻松访问、收集海量数据。让企业从难以理解的大数据世界里脱身，更简单、直观地获得数据分析结果。

Dataiku DSS与DataCanvas APS比较分析：

1.功能方面：

Dataiku DSS：功能较为齐全，是一个非常完善的data science platform 。

DataCanvas APS ：功能不够齐全，暂且是一站式数据存储和数据分析平台。

2.界面：

Dataiku DSS 界面从数据挖掘、抽取、 清洗这一系列过程都可以采用拖拽的方式快速形成flow，并且可以通过切换view，能从多维度视角查看flow，简便快捷，界面简洁，一目了然。

DataCanvas APS 界面 需先建立好数据模块、分析模块，通过拖拽模块和连线模块的方式形成一个树状图，操作稍有复杂，而且拖拽前无显示模块输入输出参数。

DataIku DSS 可视化界面做的很全面，各类统计图表也较为完善，用户可以从多维度分析。

DataCanvas 暂且还没有可视化界面操作。

3.资源背景：

Dataiku法国公司 获得1400万美元 A轮融资，领投方是纽约知名风投FirstMark Capital。

六、总结

目前大数据行业形式分析，随着大数据产业政策的不断出台，我国大数据发展风生水起、成绩斐然。

据易观国际统计，2015年我国大数据市场规模达102亿元，2017年有望达到170亿元。

过去10年，大数据的应用逐步从理论走向实践，被越来越多的企业视做未来的核心竞争力，并成为一种常态趋势。这期间，新方法、新技术层出不穷，如Hadoop、HBase、Hive 、Storm和Spark等。

根据行业痛点分析，企业需要的是一个广泛和稳定的集成数据平台来支撑数据存储、分析和应用。

DataCanvas集成数据平台针对国内大数据应用的现状和痛点，以企业需求为导向，帮助企业进行数据基础架构的迁移，并同时提供全能的数据分析平台，一站式解决数据存储和分析的难题，让数据存储、查询、分析和可视化形成闭环并快速落地，让企业在未来数据竞争的时代夺得先机。

DataCanvas的未来发展趋势也许会往DSP（data science platform）平台方面发展：

1.深入用户使用场景，链接服务，协调工作，dsp平台存在的意义就是为了让数据科学家们更能简单、方便、快捷的制作算法模型以及多维度分析数据，从而为企业带来更精准的决策和商业价值。

2.扩宽产品边界，抽象产品平台特性，将其推广用于各行各业，收集各行各业的算法模型，不仅打造适用各种场景产品平台，更能积累价值多样性的算法模型，供其用户参考学习。

（注：数据摘自 <http://www.chinabgao.com/k/dashuju/26612.html>）