



MININGLAMP
明 略 数 据

Discovery : 明略大数据商务智能平台

产品白皮书

明略数据

2015/10/1

使用条款

明略软件系统有限公司（以下简称“明略”）所有文档，接收者有保密义务。未经明略书面许可，任何人或任何机构不得向第三方披露、泄露有关本文件的任何内容或细节。明略拥有修改、调整、增补本文件的权利。明略™和MiningLamp™及相关延展标识为明略公司在中国和/或其他国家或地区的注册商标或商标。

目录

1 产品概述.....	5
2 产品特点和核心优势.....	6
3 系统架构.....	7
3.1 数据引擎.....	8
3.2 BI Server.....	8
3.3 交互分析模块.....	8
3.4 API.....	9
4 功能介绍.....	10
4.1 数据处理.....	10
4.1.1 数据源.....	10
4.1.2 数据关联.....	11
4.1.3 ETL数据转换处理.....	12
4.1.4 多维数据库Cube.....	12
4.1.5 分布式部署.....	13
4.2 查询引擎.....	13
4.2.1 传统关系型数据库.....	13
4.2.2 分布式（MPP）数据库.....	14
4.2.3 NoSQL数据库.....	14
4.3 OLAP分析.....	15
4.3.1 任意多维度分析.....	15
4.3.2 多层钻取.....	15
4.3.3 排序.....	16
4.4 Dashboard.....	16
4.4.1 图表类型.....	17
4.4.2 图表编辑.....	18
4.4.3 参数和过滤器设置.....	19
4.4.4 Dashboard布局.....	21
4.4.5 保存和发布.....	21
4.5 权限管理.....	22
4.5.1 权限控制的层次.....	23

4.5.2 权限的粒度	23
5 性能与可靠性	24
6 结束语	25

1 产品概述

BI (Business Intelligence , 商业智能) 的概念最早在1996年提出。当时将商业智能定义为一类由数据仓库 (或数据集市) 、 查询报表、数据分析、数据挖掘、数据备份和恢复等部分组成的、以帮助企业决策为目的技术及其应用。目前, 商业智能通常被理解为将企业中现有的数据转化为知识, 帮助企业做出明智的业务经营决策的工具。商务智能系统中的数据来自企业其他业务系统。例如商贸型企业, 其商务智能系统数据包括业务系统的订单、库存、交易账目、客户和供应商信息等, 以及企业所处行业和竞争对手的数据、其他外部环境数据。而这些数据可能来自企业的CRM、SCM等业务系统。

商业智能能够辅助的业务经营决策, 既可以是操作层的, 也可以是战术层和战略层的决策。为了将数据转化为知识, 需要利用数据仓库、联机分析处理 (OLAP) 工具和数据挖掘等技术。因此, 从技术层面上讲, 商业智能不是什么新技术, 它是数据仓库、OLAP和数据挖掘等技术的综合运用。

把商业智能看成一种解决方案应该比较恰当。商业智能的关键是从许多来自不同的企业运作系统的数据中提取出有用的数据并进行清理, 以保证数据的正确性, 然后经过抽取 (Extraction) 、 转换 (Transformation) 和装载 (Load) , 即ETL过程, 合并到一个企业级的数据仓库里, 从而得到企业数据的一个全局视图, 在此基础上利用合适的查询和分析工具、数据挖掘工具、OLAP工具等对其进行分析和处理 (这时信息变为辅助决策的知识) , 最后将知识呈现给管理者, 为管理者的决策过程提供数据支持。

而随着大数据时代的到来, 急剧膨胀的数据规模、精准营销等对细粒度数据探索的需求, 以及对决策时效性的需求愈加旺盛, 传统的BI方案在解决数据规模、分析的时效性、数据粒度方面越发显得困难。明略Discovery产品正是在这样的背景下开发的一款面向大数据商业智能解决方案的产品。

明略Discovery大数据商业智能平台是由明略软件系统研发的基于Hadoop技术框架的大数据商业智能平台。我们以更快、更高效、更便捷为目标, 致力于帮助用户实现数据的价值。该平台旨在帮助用户挖掘大数据环境下的数据价值, 解决传统BI工具在大数据环境下无法使用的困境, 帮助普通用户不经过专业培训就能方便快捷的使用大数据商业智能平台进行大数据分析。

2 产品特点和核心优势

➤ 基于大数据技术的优异性能

Discovery是一款真正支持海量大数据的BI产品。经过测试，Discovery对于分组汇总表的计算，百亿条数据规模，汇总计算均在3s内完成。对于明细表的展示，均在1s内即可完成。该测试数据基于实际实施项目案例。

➤ 数据模型建立简单

Discovery支持直接连接RDBMS或者Hadoop数仓、ROLAP和MOLAP连接Hadoop数仓，通过简单的Drag & Drop即可实现数据模型的建立。

➤ 开放架构方便集成

Discovery可以结合数据仓库使用，也可以直接连接ERP/CRM等业务系统的数据源，扩展现有系统的功能，更好的利用现有IT资产。Discovery采用开放的SOA架构，提供了丰富的API接口，能够方便地融合进其他应用系统，能方便实现与第三方系统的用户和权限整合。

➤ 易于非技术人员使用

传统BI的繁杂之处主要体现在两个方面：技术人员需要花费大量时间准备数据以及业务人员基于数据偶得的一些分析需求实现过程复杂。

Discovery基于明略MDP产品、结合BI的需求，提供完整的ETL工具协助用户自行进行数据准备，大大降低了非IT人员使用产品的门槛。

➤ 实施周期短，维护简单

由于Discovery特有的数据准备模块，除去原始需求沟通时间，技术人员在实施时无需花费过多的时间来进行数据处理，而Discovery特有的建模流程能够极大的减少需求多变导致的数据准备和沟通交流时间，真正做到了实施周期短，后期也无需投入过多的人力和时间来维护系统，解决需求上的变动。

3 系统架构

Discovery总体上属于客户端-服务器（B/S）架构。

它基于大数据技术构建，可同时支持大数据平台和传统的RDBMS；提供文本文件、JDBC、Cube、Search等多类型的数据源支持；支持ROLAP和MOLAP两类OLAP分析模型；支持细粒度的访问权限控制。

Discovery 各个部分的主要功能如下表所示：

模块	组件	主要功能
数据源	ERP/CRM	支持对接客户已有系统。
	CDH/Hortonworks	支持对接其他商用发行版。
	RDBMS	支持对接常见的关系型数据库。
	Excel/CSV	支持常用的文本文件格式。
数据预处理	MOLAP Cube	配置多维OLAP Cube的维度、指标、数据更新方式等。
	Search index	配置针对非结构化数据(也可以是结构化数据)的索引。
数据引擎	OLAP Engine	支持ROLAP、MOLAP两种OLAP分析，为上层交互提供Roll up、Drill down、Slice & Dice等支撑。
	Ad-hoc Engine	针对关系型数据库提供的查询引擎，支持常见的RDBMS、Hive、Impala等。
	Search Engine	针对非结构化数据提供的搜索引擎，支持对文本、图片等非结构化数据的检索分析，同时也支持对结构化数据进行明细粒度的分析。
BI Server	Security	BI Server提供针对访问连接(Https)、身份认证等安全保证的功能，确保数据安全不外泄。
	Administration	针对Dashboard、图表、数据预处理、分享等的权限控制，同时支持细粒度的数据访问权限控制。
	Business logic	提供针对 workflow、配置等的业务流程管理。
	Repository	管理数据库、表、Cube、索引等。
交互分析引擎	Report Engine	通过可视化界面创建到CSV、RDBMS、Cube和Index的连接, Drag & Drop数据源中的维度和指标，选择图表展现样式，生成chart。Discovery基于D3渲染图表，可支持大部分常用的展现形式并支持自定义。
	Dashboard Design	将Report Engine生成的图组装为Dashboard，

	er	并提供自定义样式等的接口。
API	-	Discovery对外提供的Dashboard接口和Report接口，方便第三方App将Dashboard和Report整合入
可视化	-	将API下引擎提供的各类功能封装为基于Html5的网页，支持pc、mobile等展现和下载、分享。

3.1 数据引擎

Discovery的数据引擎包括OLAP、Ad-hoc和Search三部分，分别面向基于Cube的查询、即席查询和明细搜索三种业务场景，并由BI Server进行统一调度。

Discovery的OLAP引擎以Discovery的多维数据库为核心，旨在对原始数据进行抽取，转换和加载，支撑Discovery的在线分析，是Discovery必不可少的一部分。Discovery的OLAP引擎同时支持ROLAP和MOLAP两种模式，并且基于Hadoop生态系统进行了针对性的优化。OLAP引擎支持使用离线cube进行数据存储，支持cube数据定时全量以及增量更新。

Discovery的Ad-hoc引擎基于大数据生态体系的SQL引擎构建，支持在海量数据集上进行分钟级别的即席查询。

Discovery的Search引擎支持对数据进行raw data级别的探索发现，在可观的时间延迟内，支持对原始数据集进行灵活的数据探索分析。

3.2 BI Server

BI Server支持对Discovery平台内的元数据、权限等进行统一授权管理。BI Server可支持到行列级别的数据权限管理。

3.3 交互分析模块

Discovery的分析模块以Dashboard为分析的载体，每个分析中可向Dashboard内加入任意数量的组件和控件。所有分析中的操作均通过拖拽和点击完成，无需写SQL。

支持任意维度和指标的切换，可以对已有的表样字段进行切换，从而进行自由分析。任意维度和指标切换的功能保障了当查看分析的人员在查看分析时，如果针对已有的表样产生额外的分析需求或改变了已有的分析需求，不需要重新制作一次分析，而可以直接通过切换维度和指标实现。

可以选择数据快速创建表格或者图表以使数据可视化、添加过滤条件筛选数据，即时排序，使数据分析更快捷。

3.4 API

Discovery提供灵活的数据接口，对于有一定开发能力或者已有可视化工具的客户，Discovery可以通过标准的数据接口对外输出经过处理和分析的数据，帮助用户快速的生成定制报表。

4 功能介绍

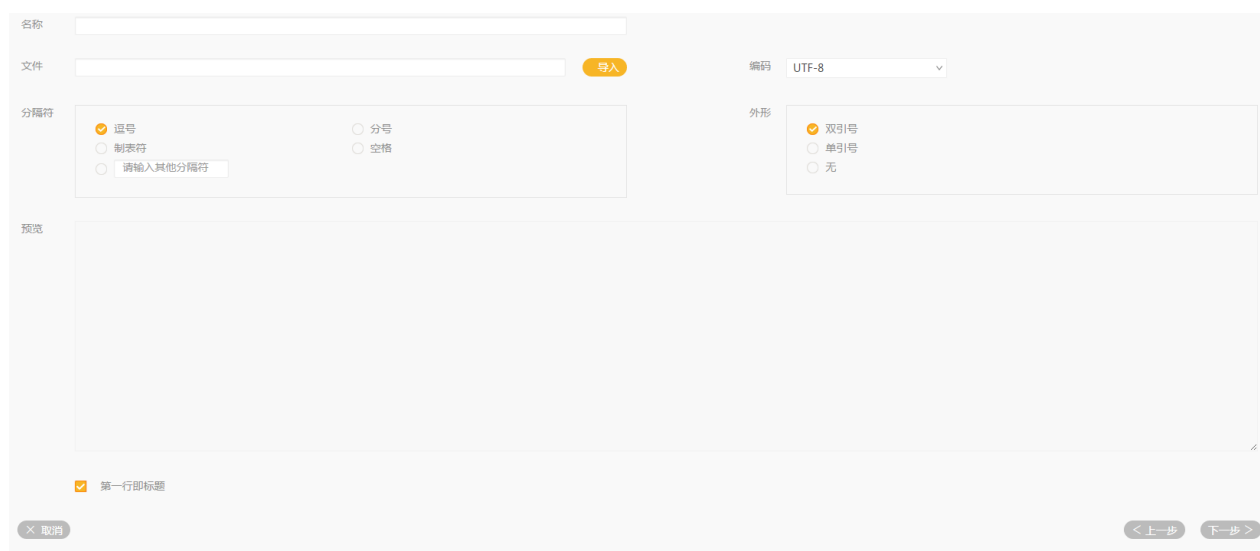
Discovery功能模块可以分为数据处理、查询引擎、OLAP分析、Dashboard和权限管理。下面介绍具体的功能模块以及权限控制。

4.1 数据处理

4.1.1 数据源

4.1.1.1 文本数据源

Discovery支持用户将CSV、Excel等文本类型的文件导入分析。用户可以选择文件的本地路径、编码、分隔符等信息，并且Discovery提供文本内容的预览功能。



配置好文本的信息后，可以对文本的具体字段再进行编辑：

	名称	类型	数据格式	长度	精度
<input checked="" type="checkbox"/>	Row ID	Numeric	#	0	0
<input checked="" type="checkbox"/>	Order Priority	String	#	19	0
<input checked="" type="checkbox"/>	Discount	String	#	31	0
<input checked="" type="checkbox"/>	Unit Price	Numeric	###0.###	0	16
<input checked="" type="checkbox"/>	Shipping Cost	Numeric	###0.###	0	16
<input checked="" type="checkbox"/>	Customer ID	Numeric	#	0	0
<input checked="" type="checkbox"/>	Customer Name	String	#	42	0
<input checked="" type="checkbox"/>	Ship Mode	String	#	21	0
<input checked="" type="checkbox"/>	Customer Segment	String	#	21	0
<input checked="" type="checkbox"/>	Product Category	String	#	22	0

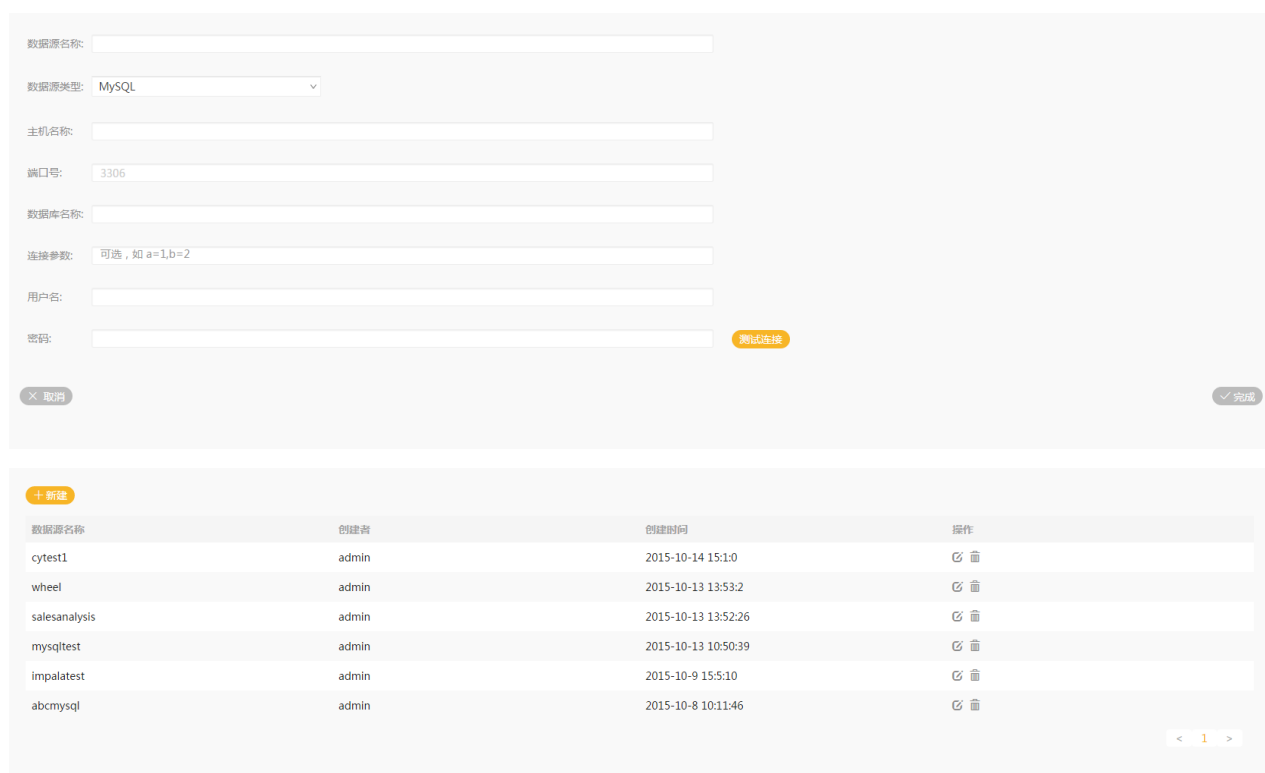
< 1 2 3 >
取消 上一步 完成

4.1.1.2 RDBMS数据源

Discovery支持用户配置直接到关系型数据库的JDBC连接,进而在生成图表和Dashboard的过程中引用。通过集成各种商业关系型数据库的JDBC驱动程序,Discovery支持从Oracle、DB2、SQL Server、Sybase、SQL Server Analysis Services、Essbase等数据库系统中读写数据。用户可以在Discovery的数据源配置界面配置好相关数据库连接,之后可以在编辑图表、数据探索和分析过程中直接引用该数据源。

除去常见的MySQL, Oracle等关系型数据库外,Discovery还支持Hive、Impala等大数据环境的关系型数据库引擎。

Discovery提供了统一的新建、编辑、删除数据库连接的功能。



数据源名称	创建者	创建时间	操作
cytest1	admin	2015-10-14 15:10	✕ 编辑
wheel	admin	2015-10-13 13:53:2	✕ 编辑
salesanalysis	admin	2015-10-13 13:52:26	✕ 编辑
mysqltest	admin	2015-10-13 10:50:39	✕ 编辑
impalatest	admin	2015-10-9 15:5:10	✕ 编辑
abcmysql	admin	2015-10-8 10:11:46	✕ 编辑

无论是传统的关系型数据库还是大数据环境下的关系型数据库,在通过Discovery连接上数据源之后,客户端可以通过Discovery统一的数据查询接口向服务器端发送ANSI SQL查询语句,由Discovery REST API负责从数据库系统获取数据返回给客户端。

4.1.1.3 OLAP数据源

Discovery集成了强大的Mondrian ROLAP引擎,可以在关联多个数据表之后生成ROLAP Cube数据结构,并可接受来自客户端的MDX多维数据查询语句返回,利用这一特性,Discovery能够对多维度数据进行自由灵活的上卷下钻、钻透等操作。

4.1.2 数据关联

数据间的关联提供了数据库表之间字段值的对应规则，是连接各个数据库表间的桥梁，有了数据库表间的关联，各个表之间的数据才能在一个分析中得到体现。此外，数据关联控制了外键表数据的有效性，并在数据库的层面体现了业务层的逻辑关系。所以，数据关联也是Discovery多维数据库自动建立数据模型的基本依据。

数据之间的关联是通过给当前数据库表建立添加主键表的方式来建立的，多个表间可以直接继承数据库已设的外键关系，也可以手动建立表间关系。

数据关联：1：N关系、N：1关系和1：1关系都可以通过数据关联建立起关联，系统会自动判断主外键关系。

4.1.3 ETL数据转换处理

Discovery OLAP提供数据ETL的功能，方便用户把散落在各个数据源的数据进行汇总和联接后进行数据立方体的创建。整个数据ETL的过程和传统的ETL无异，主要包含数据抽取、转换、装载这三个过程，使得从数据源抽取所需的数据，经过数据清洗，最终按照预先定义好的OLAP数据模型，将数据加载到数据仓库(比如Hive，传统关系型数据库等)中去。

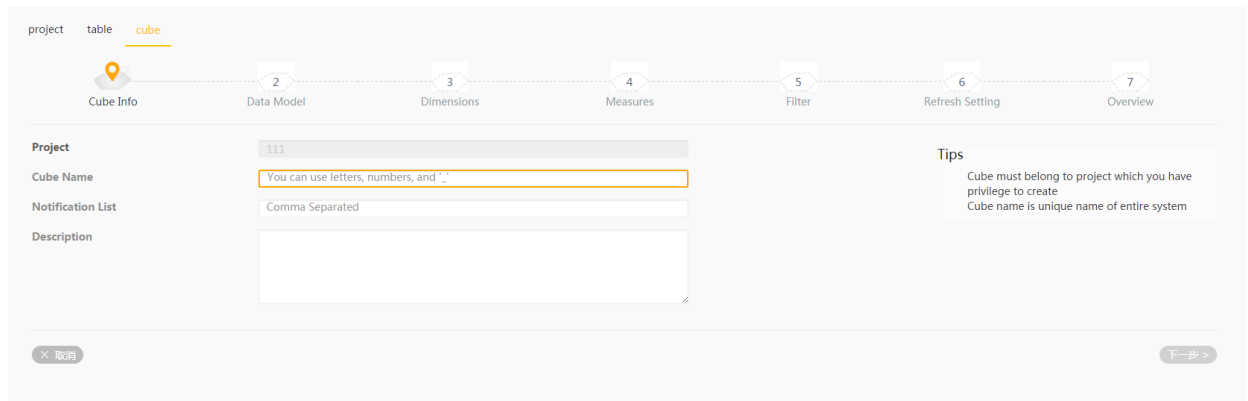
同时Discovery OLAP也支持实时ETL，它能把业务数据源中产生的数据实时的抽取加载到数据仓库，该实时特性确保了业务系统的数据发生变化后，以最少的延误，保持数据的更新。

除外，在具体实现ETL时，Discovery OLAP在以下几个方面加强了功能：

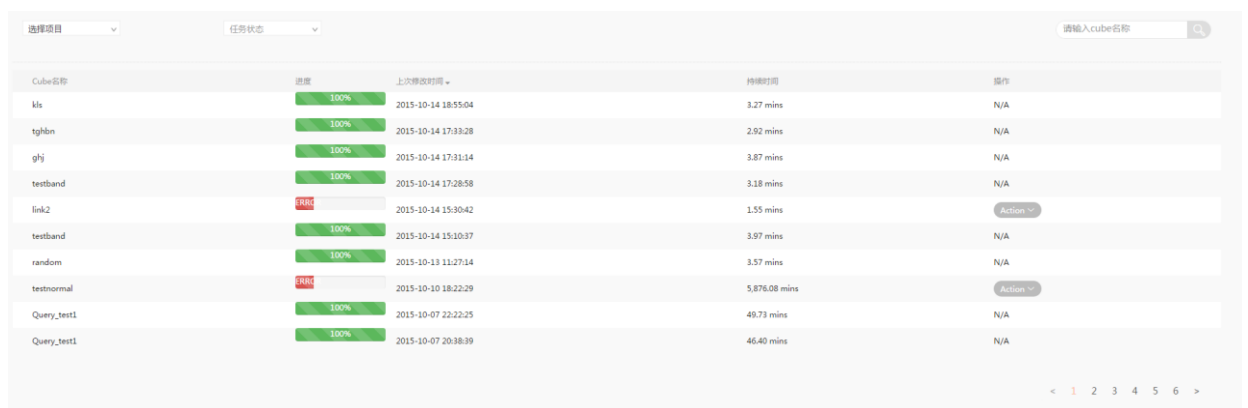
类别	支持
支持数据源	CSV，数据库表，流式数据，日志文件等
空值处理	可捕获字段空值，进行加载或替换为其他含义数据，并可根据字段空值实现分流加载到不同目标库
规范化数据格式	可实现字段格式约束定义，对于数据源中时间、数值、字符等数据，可自定义加载格式
拆分数据	依据业务需求对字段可进行分解。例如可把电话号码按照区域码和分机号进行分解
验证数据正确性	可利用哈希表及拆分功能进行数据验证
数据替换	对于因业务因素，可实现无效数据、缺失数据的替换
建立主外键约束	对无依赖性的非法数据，可替换或导出到错误数据文件中，保证主键唯一记录的加载

4.1.4 多维数据库Cube

Discovery根据业务场景，进行Cube的设计、生成与管理。Cube采用MOLAP的形式，在处理大数据量的问题上具有优良的支撑。



Discovery可以对进行build的Cube任务进行实时的监控和管理。在任务主页上，可以看到各个任务的名称、对应cube的名称、当前build进度等信息。



Cube名称	进度	上次修改时间	持续时间	操作
kls	100%	2015-10-14 18:55:04	3.27 mins	N/A
tqhbn	100%	2015-10-14 17:33:28	2.92 mins	N/A
ghj	100%	2015-10-14 17:31:14	3.87 mins	N/A
testband	100%	2015-10-14 17:28:58	3.18 mins	N/A
link2	Error	2015-10-14 15:30:42	1.55 mins	Action
testband	100%	2015-10-14 15:10:37	3.97 mins	N/A
random	100%	2015-10-13 11:27:14	3.57 mins	N/A
testnormal	Error	2015-10-10 18:22:29	5.876.08 mins	Action
Query_test1	100%	2015-10-07 22:22:25	49.73 mins	N/A
Query_test1	100%	2015-10-07 20:38:39	46.40 mins	N/A

对于处于build中的cube，可通过操作一栏下拉框进行任务中止（Discard），对于build失败的cube，可根据log信息在平台或者数据等错误排除后，选择恢复（resume）进行任务的恢复。

4.1.5 分布式部署

Discovery分布式集成即可提升Cube的生成速度和即时分析取数的查询速度，简单来说，就是在服务器端配备多台服务器，其中一台为主服务器，其他为分布式服务器，Discovery工程发布在主服务器上，主服务器主要工作就是均衡调配，负责调配将Cube生成在分布式服务器上，每台分布式服务器上均会生成一个Cube，所有分布式服务器的Cube组合起来，则为整个数据库的数据，即时分析查询取数时也是主服务器收到取数查询请求，根据查询的表负责将取数操作分配到各分布式服务器上，最后汇总结果，返回给客户端。

当现有的服务器硬件不能满足数据更新和在线分析的性能需求时，分布式部署可以在多台服务器中实现并行运算以提高性能，充分利用了现有的服务器资源。

4.2 查询引擎

4.2.1 传统关系型数据库

关系数据库是建立在关系模型基础上的数据库，借助于集合代数等数学概念和方法来处理数据库中的数据。现实世界中的各种实体以及实体之间的各种联系均用关系模型来表示。标准数据查询语言SQL就是一种基于关系数据库的语言，这种语言执行对关系数据库中数据的检索和操作。

由于Discovery OLAP组件良好的分层设计，并且技术层级之间是通过JDBC通用接口来进行数据交互的，理论上只要能提供JDBC驱动的关系型数据库都可以作为Discovery OLAP组件的查询引擎，目前已经测试并验证过的关系型数据库有：Oracle、DB2、MySQL、Teradata、Sybase、SQL Server等各类主流的关系型数据库。

4.2.2 分布式（MPP）数据库

分布式MPP（Massive Parallel Processing）数据库架构，它由多个SMP服务器通过一定的节点互联网络进行连接，协同工作，完成相同的任务，从用户的角度来看是一个服务器系统。其基本特征是由多个SMP服务器（每个SMP服务器称节点）通过节点互联网络连接而成。

基于分布式数据库查询引擎，Discovery对Ad-hoc式的实时查询也有业界领先的性能表现。

4.2.3 NoSQL数据库

传统的关系型数据库如Oracle、MySQL等均是基于关系型表格建立，没有采用该模式的数据库被泛称为NoSQL数据库，其底层数据存储结构有键-值模式、图模式、文档模式等。由于其易于设计，对于集群可横向扩展，并且更方便控制操作，越来越多的大数据产品与实时查询的web应用都选用NoSQL数据库作为其数据存储介质。NoSQL数据库为了追求可用性、可扩展性和速度牺牲了存储的连续性，查询语言的通用性等关系型数据库的优点。

作为一种NoSQL数据库，Hbase拥有上述的特点，是Apache的Hadoop 项目的子项目，是一个适合于非结构化数据存储的数据库。Hbase也是目前应用最广泛的NoSQL数据库之一。

Discovery的OLAP引擎试用的Cube数据存储于Hbase，提供Hadoop之上的SQL查询接口及多维分析能力以支持超大规模数据。用户可以采用前端UI的方式或者命令行的方式来进行Cube的创建、Build与管理。Discovery具有以下特性：

特性	描述
快	基于Hadoop和HBase技术，专为百亿级条目、10TB以上数据规模设计；亚秒级查询响应，远高于传统OLAP引擎在RDBMS上的表现。
高并发	基于HBase技术，并实现了加速查询的HBase Coprocessor，可承压千级别以上的查询并发。
易扩展	基于Hadoop和HBase，性能和并发量可以根据业务需求随集群规模线性扩展。
SQL接口	支持大部分常用的ANSI SQL标准，无学习成本。
JDBC	提供JDBC包，上层业务可无缝整合。

4.3 OLAP分析

多维OLAP分析是BI工具分析功能的集中体现，其应用特性主要体现在两方面：一是即时查询到效果（Online），这要求后台数据的计算速度和前台浏览器的展示速度都要很快；二是多维度自定义分析，这要求BI工具的多维数据库应该具有较大的灵活性，可以随用户的要求组合任意的指标和维度。只有同时满足这两个特性的交互分析过程，才是多维OLAP分析，才能保障用户即时看到其分析需求对应的数据统计结果，以及通过切换维度和改变条件等方式，满足根据上一步的结果即时产生的新的分析需求。

Discovery提供了各种常见的OLAP分析操作，除基本的分组汇总外，还可以进行任意多维度的分析，钻取分析、排序、过滤等等分析功能。

4.3.1 任意多维度分析

Discovery提供任意维度的数据分析，针对要分析的数据，我们可以任意添加需要分析的维度。图表设置过程类似，需要注意的是绝大多数的图表无需刻意添加分析。

支持任意维度和指标的切换，可以对已有的表样字段进行切换，从而实现自由分析。任意维度和指标切换的功能保障了当查看分析的人员在查看分析时，如果针对已有的表样产生额外的分析需求或改变了已有的分析需求，不需要重新制作一次分析，而可以直接通过切换维度和指标实现。

4.3.2 多层钻取

Discovery支持基于Mondrian的ROLAP数据分析操作。前端界面可以将多维数据源的维度层级关系以树形的结构显示出来，通过折叠、展开不同层级的维度，实现上卷、下钻的功能。

Discovery还支持对某个单元格中的数据进行钻透（Drill Through）操作，用户可以任意选择钻透时要显示的维度和指标数据，Discovery支持对钻透的结果进行分页显示。

数据连接方式
 mdx
 多维数据源
 salesanalysis
 维度
 version
 Product
 Time
 (All)
 Years
 Quarters
 Months
 指标 +
 Quantity
 Sales

行 Years Quarters Months
 列 Line
 指标 Quantity
 过滤

Years	Quarters	Months	Classic Cars Quantity	Motorcycles Quantity	Planes Quantity	Ships Quantity	Trains Quantity	Trucks and Buses Quantity	Vintage Cars Quantity
2003	QTR1	Jan	334				81	370	572
		Feb	120	229	460	301	38		301
		Mar	929	170				93	563
	QTR2	Apr	465	212	401	323	69	279	244
		May	934	251				248	584
		Jun	338	23	416	319	116	160	313
	QTR3	Jul	961	408				257	519
		Aug	339	440	371	318	106	136	264
		Sep	1,139	47		234	93	410	587
	QTR4	Oct	2,032	661	879	428	132	456	1,143
		Nov	3,879	1,163	1,081	921	320	1,238	2,260
		Dec	1,292	427	225		45	409	563

保存 取消

4.3.3 排序

基于查询出来的结果的排序，根据维度自身进行排序，根据汇总指标的大小对维度进行排序展示，根据公式值进行排序。可以进行升序、降序和自定义排序。选择了排序方式，数据会根据所选排序方式自动排序。使用各种排序和TopN功能的组合可以简单直观的分析重点的局部数据。

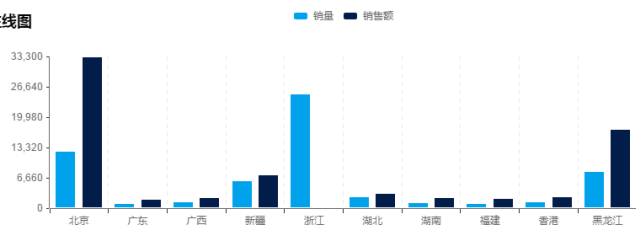
4.4 Dashboard

Dashboard作为Discovery的主要前端展现形式，具有丰富的组件展示模型以及控件类型。其中组件的模型包含有表格以及多种图表类型。

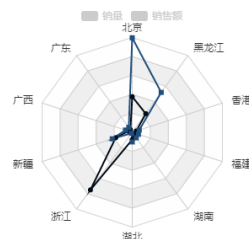
数据分析者可以使用Discovery的即时分析模块制作满足自己业务需求的Dashboard的结果，可以在编辑、查看。

Dashboard具有极强的交互性，企业中高层对于分析人员制作好的Dashboard结果，可以进行查看、交互。

柱线图



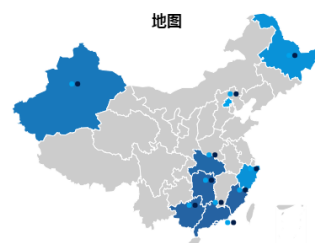
雷达图



仪表盘



地图

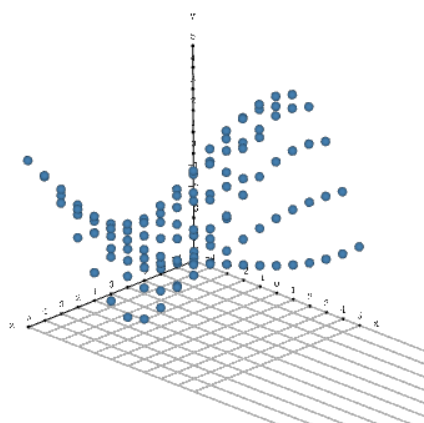
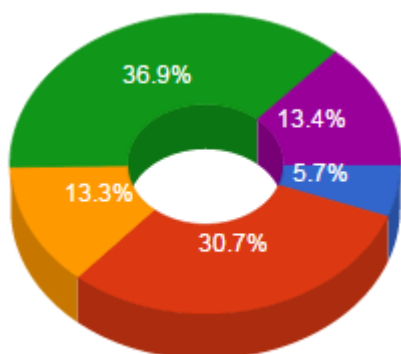


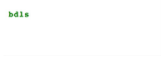
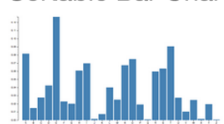
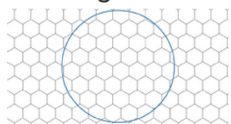

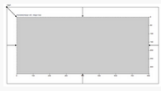

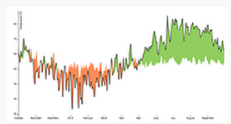


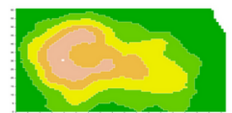







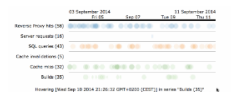
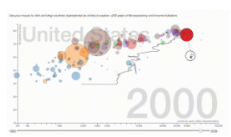
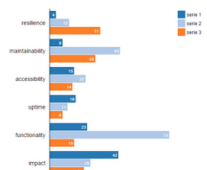
4.4.1 图表类型

Discovery中的图表使用echarts+D3组合方案实现。ECharts基于html5 Canvas，是一个开源的纯JavaScript图表库，有比较完备的基础图表类型，因此我们选择ECharts作为Discovery基础图表类型的图表库。D3是基于JavaScript技术实现的图表的函数库，故D3也常被叫做D3.js。D3.js提供了大量原生的可视化效果，非常适宜根据实际可视化的需求进行图表展现形式的定制，为可视化展现提供了非常丰富灵活的选择。

Discovery支持多种图表且类型可以随意切换，支持的图表类型有：柱形图、柱形堆积图、折线图、堆积面积图、组合图、条形图、堆积条形图、饼图、仪表盘、地图。

除了基本图表类型，Discovery还可支持多种复杂的图表类型(包括以下三维图表，其他未全部列出)：



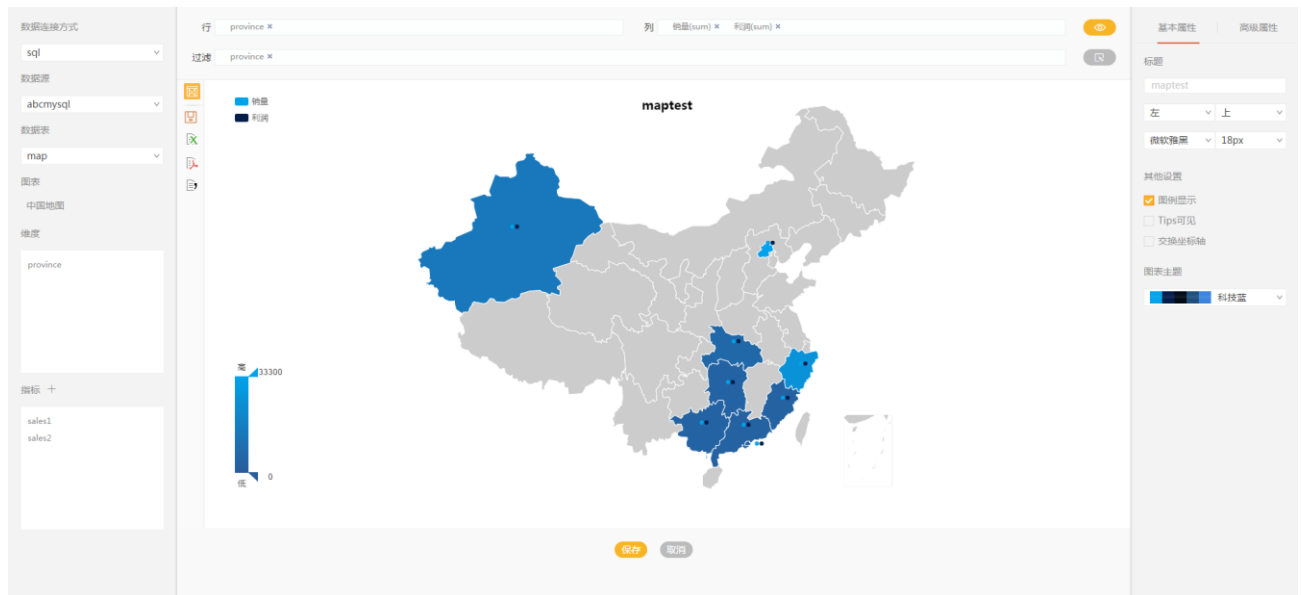
General Update Pattern 	Sortable Bar Chart 	van Wijk Smooth Zooming 	Progress Events 
Margin Convention 	Focus+Context via Brushing 	Difference Chart 	Pie Chart Update 
Hexagonal Binning 	Contour Plot 	Build Your Own Graph 	Modifying a Force Layout 
Spline Interpolation 	Dispatching Events 	Better force layout selection 	v45 web theme using SVG 
Interactive force layout 	EventDrops: Zoomable time series 	Draggable scatterplot with motion trails 	Horizontally grouped bar chart 

对于D3原生不支持的图表类型，可以基于D3的代码和框架，通过定制Discovery的图表来支持。

4.4.2 图表编辑

在用户进入报表编辑页面之后，Discovery会将用户之前配置好的数据源列举出来，用户从中选择了一个数据源，Discovery会根据该数据源中字段的类型自动划分为维度和指标两个类别，使得业务人员更直观地区分不同字段的作用。

通过拖拽任意多个维度和指标到中心区域，由Discovery自动编辑成查询请求发送给服务器端，默认情况下Discovery会自动触发作图操作，使得用户在第一时间看到数据的展现效果。Discovery拖拽式的操作体验易于上手，不需要进行特殊的培训。



用户对拖拽出来的维度或者字段可以进一步修改字段别名、聚合方式、排序方式等属性。修改结果将直接影响查询的输出。



另外，用户还可以通过Discovery提供的高级SQL查询编辑器，根据自己的需求交互式地自定义复杂的查询语句。

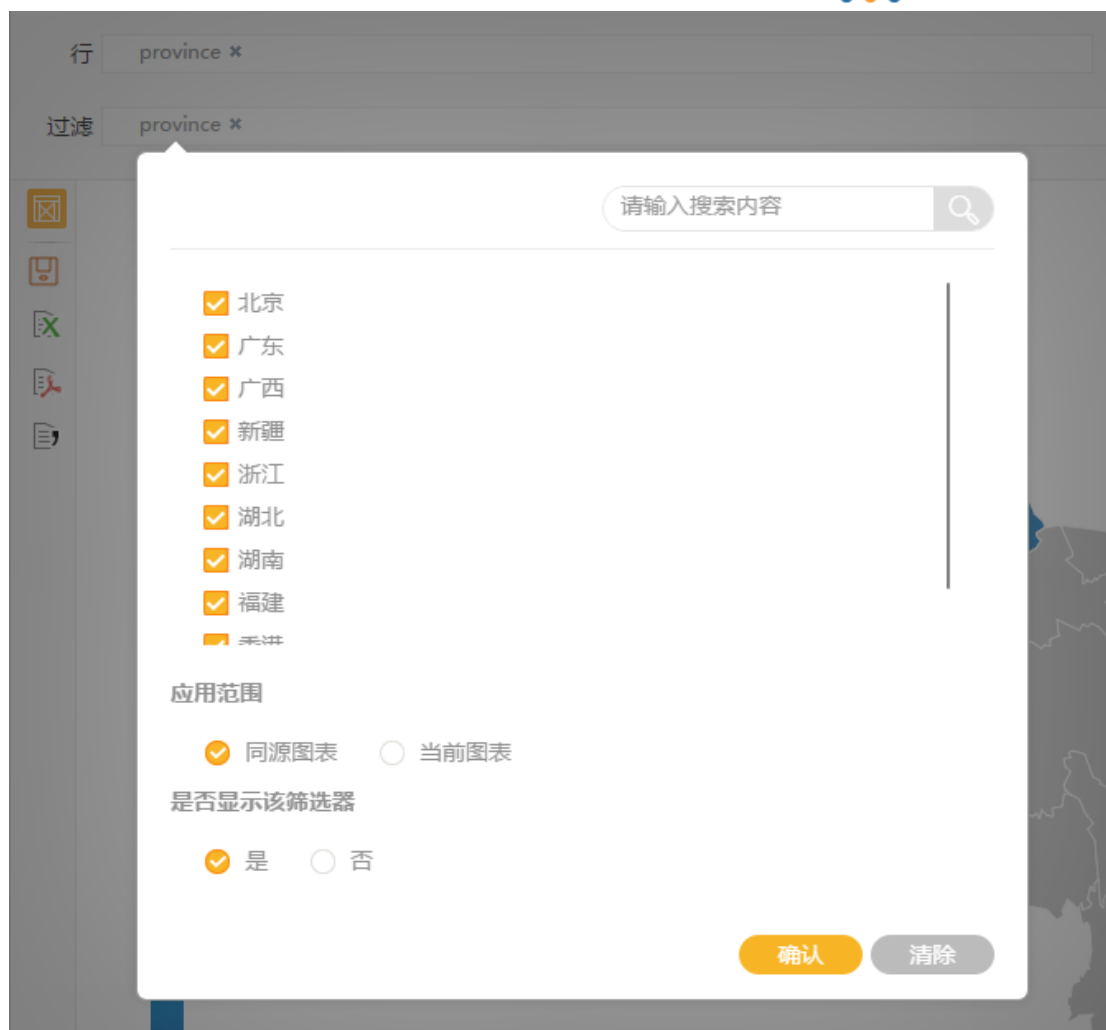
4.4.3 参数和过滤器设置

Discovery的Dashboard编辑器支持参数和过滤器设置。新建一个参数，依然是通过拖拽式的操作，将维度或指标拖拽到中心区域，Discovery将根据维度或指标的类型判断应该使用何种展现形式（比如常见的下拉选择框、日期选择控件等），用户也可以手动切换不同的展现形式，之后保存该参数供后面编辑的图表使用。

在图表编辑页面，从过滤参数栏里拖拽刚刚保存的参数到中心区域过滤器部分，Discovery将该参数和图表绑定在一起，使得用户可以在Dashboard中对参数值进行切换时，实现图表数据的联动。

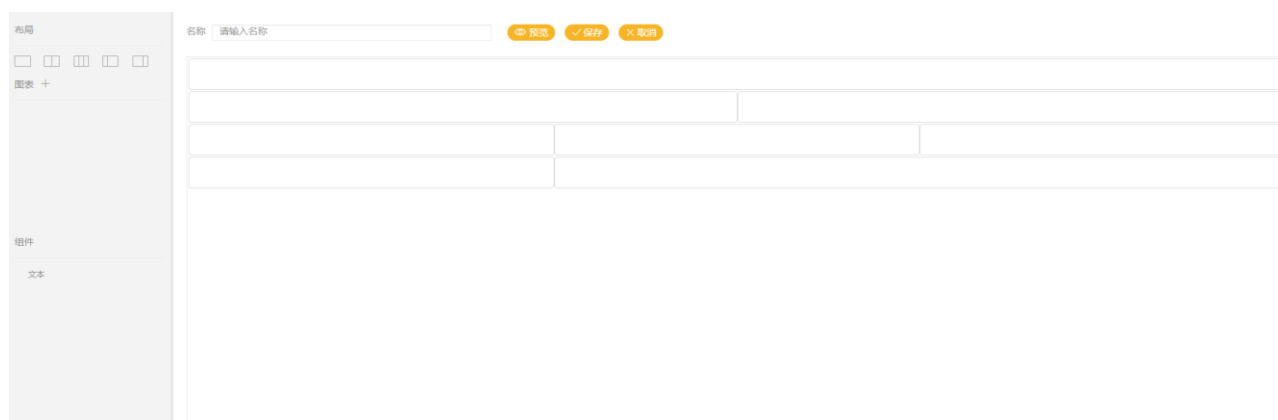
Discovery Dashboard的参数是全局过滤参数，这意味着如果有多个图表绑定上同一个参数，在Dashboard中改变该参数的值后，这多个图表都会同时联动。全局过滤参数支持级联配置。所谓级联配置，意思是一个参数的结果可以依赖于另一个参数的值。在参数配置界面，同样可以拖拽添加好的参数到过滤器区域，将子参数和父参数绑定，此时最终Dashboard上将显示两个参数输入，比如一个下拉选择框列举“国家”列的值，另一个下拉选择框列举“省份”列的值，“国家”列为子参数，“省份”列为父参数，当选择国家的值，省份参数的列表会根据选择的国家参数值而改变。可以将多个参数与多个图表绑定，从而实现Dashboard中丰富的联动效果。

除了全局过滤参数，Discovery还提供局部过滤参数。局部过滤参数的作用是允许用户在编辑某个图表时，加入自定义的条件即时进行数据的过滤。用户通过拖拽维度或指标到过滤器区域，进而可以点击拖拽出来的维度或指标，设置过滤条件，比如选择相等、大于、等于、不等于某个常量，或选择维度、指标值范围内的某个区间作为有效限制范围。



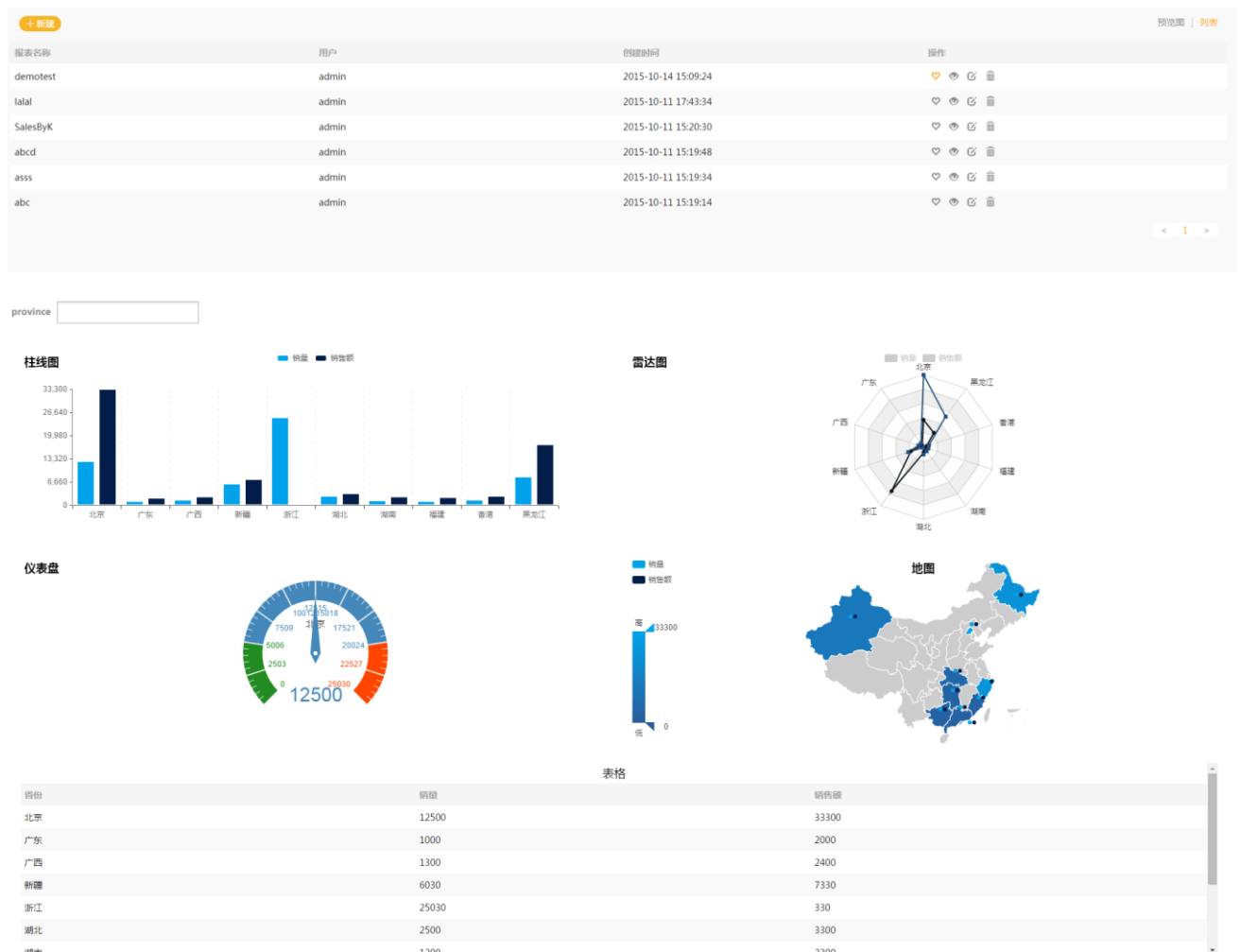
4.4.4 Dashboard布局

编辑好的图表可以通过Discovery的Dashboard布局页面，根据用户的需求实现任意比例的布局。拖拽不同的参数和图表到布局单元格中即可将该元素放置在单元格对应的位置。元素的宽度高度由单元格的相对大小决定。Discovery支持拖拽任意多个布局单元格和任意多个图表到同一个Dashboard结果中。



4.4.5 保存和发布

在布局页面点击保存按钮，选择保存目录和Dashboard文件名，即可将文件保存到Discovery后台服务器。在Dashboard列表页点击预览图标，可以在弹出的对话框中查看当前Dashboard的内容，点击“复制链接”并在浏览器新的窗口中打开即可看到发布形式的Dashboard内容。



除了以URL方式发布动态报告之外，Discovery还支持将图表以及图表中的数据导出到XLS, PDF, HTML,PPT等多种静态格式供用户离线访问。

结合Discovery中的邮件服务API，用户可以通过界面操作订阅某些报表结果，并定义报表数据刷新的周期频率以及将报表推送到用户邮箱的时间安排。

4.5 权限管理

现今，为了保密和业务的需要，各种应用系统都有权限方面的控制，Discovery也不例外，其具体实现过程如下：

- 1) 配置各个用户相应的BI访问权限；

2) 用户第一次访问受保护的资源 (BI中的某个分析模板) 时, 会发出访问请求, 服务器接收到请求后会验证用户权限, 如果没有通过验证则返回登录页面;

3) 如果通过验证用户获取一个唯一标识自身的令牌 (ticket), 并将携带这个令牌进行BI的访问操作。

4.5.1 权限控制的层次

主要分为三个层次:

- 1) 业务包的权限: 支持不同用户对于数据业务包的数据权限仅限于自己权限范围内;
- 2) Dashboard的访问权限: 支持不同的用户可以访问权限范围内的报表;
- 3) 数据的明细权限: 支持不同的用户对于同一张报表, 只能够访问权限范围内的数据。

4.5.2 权限的粒度

很多系统的权限认证只是限制模块的使用, 使得合法用户能够行使自己的权利。Discovery在满足这种整体权限认证的同时, 权限的控制力度可以达到同一个Dashboard的内容在不同权限下展示的效果不一样, 这样就免除了制作大量的Dashboard来实现同样效果, 尤其是在企业内部业务繁杂, 审批麻烦时, 一个Dashboard就可以解决所有问题。

Discovery通过对业务包的权限控制, 从数据包层面控制了不同的用户对于数据的权限, 以达到数据的细粒度控制。Discovery基于ACL和角色控制的权限设置, 实现了对大数据环境下的数据库系统进行行、列级别的权限控制。

5 性能与可靠性

由于获取数据源元数据以及查询数据可能是个耗时较长的过程（取决于网络状况及数据库性能），为了保证用户体验，Discovery针对从数据库获取到的元数据以及查询到的数据做了缓存，初次选择数据源会将数据源的表、字段等按照唯一的ID存放在内存中，初次查询某个语句也会使Discovery将该查询结果存放在内存中，后续访问同一个数据源，执行同一个查询后台服务器将直接返回内存中的结果，避免了每次对数据库进行访问，优化了性能也降低了数据库的负载。

为了避免缓存持续增大，Discovery通过可配置的选项控制缓存的最大值，超过最大值后，Discovery将根据缓存项的最后一次访问时间找出最老的部分进行清理释放部分内存。同时，Discovery也提供了手动清理缓存的API以及定期清理缓存的API供用户调用。

通过使用Nginx或者F5搭建负载均衡服务器，可以将外部访问压力平均分配到多个Discovery服务器上，每个Discovery服务器提供同样的服务，后台数据存储在一个统一的MySQL集群上，保证了后台数据存储的一致性以及读写性能，并且也保证了Discovery系统的可扩展性。

6 结束语

产品设计都是用来满足客户需求的，因此最重要的是其对于客户的使用价值，Discovery的产品价值主要体现在大大缩短项目周期，降低了项目成本，同时还降低了使用人员的技术门槛。

传统BI在数据建模过程中往往需要对数据进行繁杂的处理，例如复杂的查询语句规范、专门的数据仓库构建工具，这些往往是需要专门的技术人员花费大量时间去熟悉掌握的。Discovery提供了基于界面的数据操作模块，只需简单的设置即可满足绝大多数的数据处理需求，而这些过程是不需要任何编码过程的。而针对小部分根据特殊业务需求需要特别处理过程的数据，我们有专门的实施方案来解决此问题。

与传统BI工具相比，明略Discover主要有以下优点：

1. 基于大数据技术的优异性能，真正支持海量大数据的BI引擎；
2. 友好的数据模型方式，仅需要拖拽就能够建立；
3. 开放的系统架构，可以非常方便的对接各种传统业务系统，比如CRM/ERP等；
4. 方便非技术人员，比如业务人员使用，无需复杂的培训，图形化界面，拖拽式操作，大大降低了使用的门槛还降低了项目实施的成本；
5. 实施周期短，维护简单。