

第11章 时空大数据平台介绍

书中已经系统地介绍了时空大数据系统的核心技术，期间提到过一些在不同的方面比较突出的时空数据平台。本章会更详细地介绍目前在时空大数据方面比较有代表性的产品。

11.1 时空大数据平台分类

时空大数据平台大致可以分为3大类：传统GIS公司的大数据产品、具有海量时空数据管理能力的PaaS云平台和具有时空大数据管理与分析能力的SaaS化产品。

传统GIS公司是最早接触时空数据的企业。就像第1章讲到的那样，它们从传统地理空间数据的管理开始，逐渐接入并管理时空数据。随着数据量的增加，开始出现瓶颈，从而将大数据、分布式、云计算技术引入系统，最终发展出了自己的一套时空大数据的解决方案。这类平台基本上都是以打包产品的形式出售，私有化部署在用户自己的环境内。

与传统GIS公司不同，云厂商从一开始就是面向大数据和云计算技术的，以解决通用数据场景为主。随着业务的不断扩展，客户的需求不断丰富，时空数据成为了绕不开的课题。我们前面曾提过，城市产生的数据有80%都与地理位置有关，且数据量巨大。因此，主流的云厂商近几年都在大力发展时空数据的管理能力。

近几年，随着SaaS技术的发展，时空数据软件开始越来越多地以云平台的形式提供服务。很多时空数据厂商跨越了桌面时代、单机服务时代，基于云原生技术，从设计开始就拥抱云计算，并融合先进的DevOps迭代模式，开发出完全云原生的时空SaaS产品。

11.2 传统GIS公司的大数据产品

11.2.1 ESRI

ESRI公司最核心产品是ArcGIS。它是一套覆盖了时空数据几乎所有场景的系统组合。本节要介绍的是ArcGIS最经典的两款产品：ArcGIS Pro和ArcGIS Enterprise。大多数老用户可能更熟悉它们对应的前身——ArcMap和ArcGIS Server。

1. ArcGIS Pro

ArcGIS Pro是ESRI的新一代桌面端产品，2015年1月发布了1.0版本，目前已经迭代到3.1版本。为什么要推出这款产品呢？最主要的原因是，上一代产品ArcMap已经完成它的历史使命，无论在产品架构还是功能丰富度以及分析和显示效率上，ArcMap已经明显更不上时代的脚步。ArcMap是初代产品ArcInfo的升级版，从1978年至今已经服役近半个世纪。它的历史地位不必多言。从事地理信息相关行业的人中，相信很少有人没听过它的大名。ArcMap太优秀了，一度成为了行业标准。只要有新的界面、新的功能发布，很多GIS厂商都会竞相模仿。简单来讲，ArcMap提供了一套简单高效的桌面端地理分析、处理和可视化工作台。用户可以非常方便地将自己的数据与它相连接，最终输出精美的可视化分析结果。但是，随着时空数据体量的增加以及业务场景越来越复杂，ArcMap陈旧的架构体系无法满足需求。ArcGIS Pro的出现就是为了解决这些新时代的新问题。下面列举它的一些优势。

□ 更加稳定

ArcMap是32位程序，无法充分利用现在系统的丰富计算资源。在处理数据量较大的图层时，即使电脑的配置并不低，它也会经常出现崩溃的情况。ArcGIS Pro是原生64位的程序，因此它能够使用更多的资源进行数据分析，所以相对于ArcMap来说更加稳定，很少出现程序崩溃的情况。

□ 效率更高

ArcGIS Pro中的大部分工具都支持多线程处理，对多核CPU的算力利用更加充分。虽然以前ArcMap通过加装工具、后期配置等方法也可以实现多线程处理，但是原生的支持会更加稳定。

除了并行处理算法外，ArcGIS Pro还另外提供了GeoAnalytics Desktop大数据工具箱。这个工具箱里的工具都是基于ArcGIS Pro内置的Spark内存计算框架进行数据处理的，通过Spark框架，充分使用服务器的内存资源来运行，从而能够极大地提高数据运算的效率。这些工具在处理数据量较大的图层时，性能优势尤为明显。ArcGIS Pro2.7软件中的GeoAnalytics Desktop工具箱包含的20多个工具如图11-1所示，不同版本的ArcGIS Pro中的GeoAnalytics Desktop工具箱的工具可能有些差别。

□ 整合了Jupyter Notebook

ArcGIS Pro内置了Jupyter Notebook，可以用来执行分析并在地图窗口中即时查看分析结果。ArcGIS Pro中的所有Python功能都可以通过ArcGIS Notebooks来获得，包括核心Python功能、Python标准库、ArcPy、ArcGIS API for Python以及ArcGIS Pro附带的众多第三方库，例如NumPy和pandas。ArcGIS Pro可以使用ArcGIS Pro Package Manager通过开源库进行扩展。

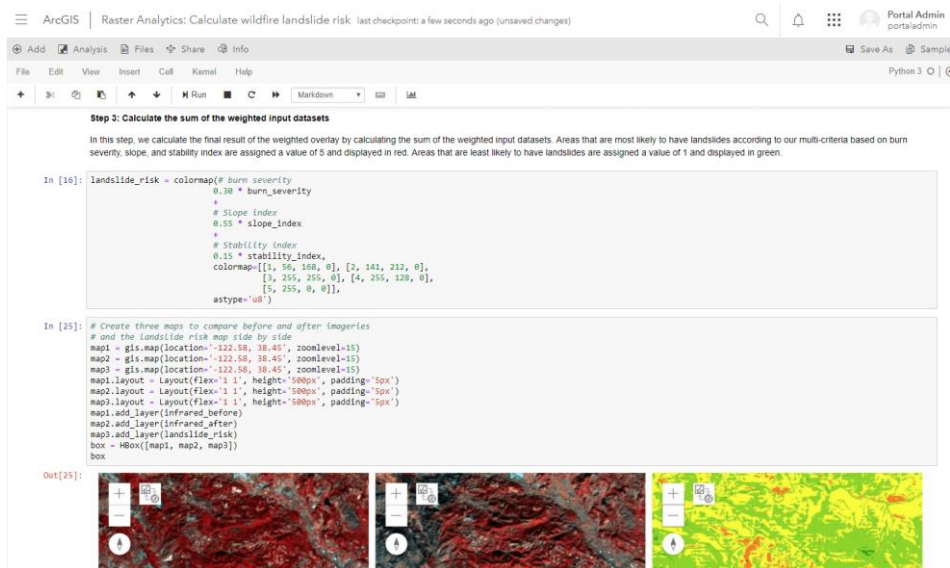


图11-1 ArcGIS Pro Jupyter Notebook

□ 与云环境协同

ArcGIS Pro支持连接到ArcGIS云环境ArcGIS Online，可以访问云环境上的数据和功能，从而可以便捷地加载和使用大量的在线地图数据。

□ 二三维一体化

ArcGIS Pro可以在同一个工程中加载和显示2D和3D数据，实现2D和3D数据的浏览、编辑、制图可视化，并支持2D和3D视图的联动，极大地提高了信息获取的效率（见图11-2）。ArcGIS Pro中整合了以前的ArcMap、ArcCatalog、ArcGlobe、ArcScene、CityEngine等软件的功能，使用起来非常方便，不需要在多个软件间来回切换。

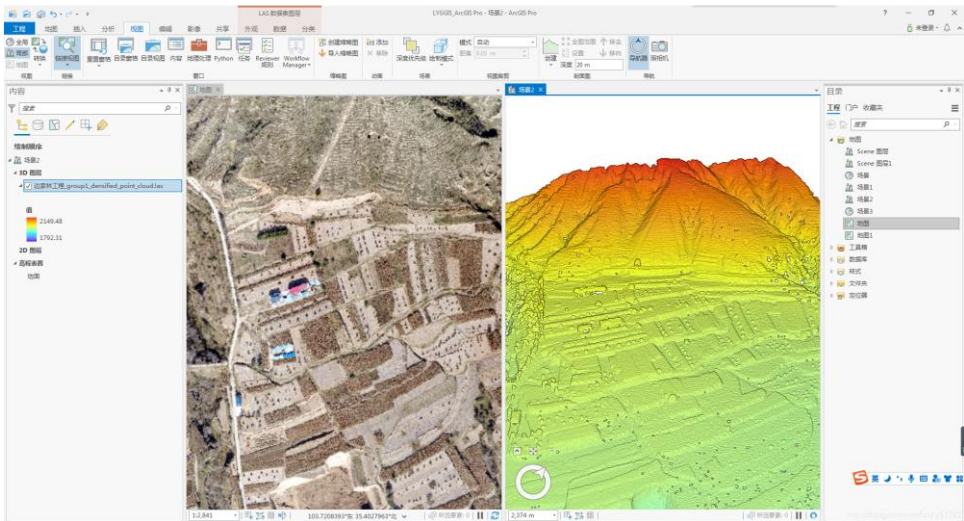


图11-3 ArcGIS Pro 二三维一体化

□ 大数据与AI

ArcGIS Pro实现了AI与地理空间的结合。ArcGIS Pro集成了主流的机器学习框架，内置先进的机器学习方法，可为空间环境系统提供强有力的支持，可以更准确地洞悉、分析和预测周围环境。比如，能够监测水域、农田、森林等环境系统的变化，从而帮助相关研究人员、自然保护主义者和政策制定者采取有效行动来保护我们的家园。

2. ArcGIS Enterprise

ArcGIS Enterprise是ESRI的新一代服务器端产品，是给用户提供的整套完整的“GIS平台”产品，具有强大的地图制作、空间数据管理、空间分析与数据科学、空间信息整合、发布与共享的能力。ArcGIS Enterprise的部署既支持本地环境也支持私有云环境。其内部整合并升级了ArcGIS上一代的服务器产品，比如ArcGIS Server、ArcSDE等。它最核心的组件如图11-4所示，其中实线框表示必选组件，虚线框为可选组件。

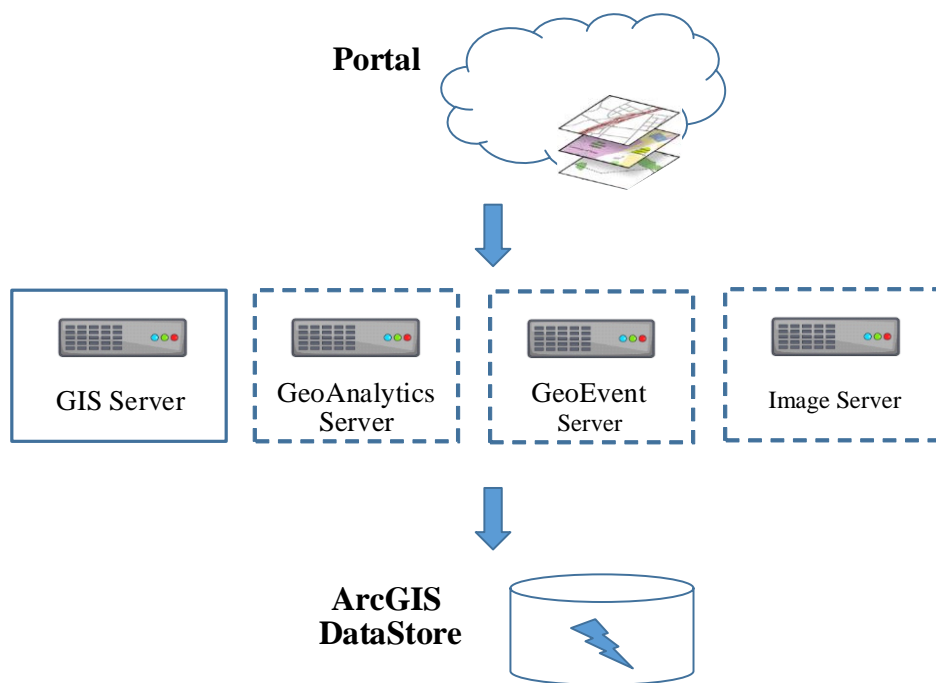


图11-4 ArcGIS Enterprise主要组件架构图

□ Portal for ArcGIS

Portal for ArcGIS是整个平台的门户，也是连接用户与GIS服务器资源和工具的一个Web界面。通过该门户，用户可以便捷地发现和使用GIS资源。Portal for ArcGIS可实现多维内容的管理和整合，将多元业务数据通过以地图为中心的方式进行管理。同时，它还提供了强大的制图和应用创建功能，通过内置的智能制图技术和多种制图模板，可以对多元业务数据快速上图。

□ GIS Server

GIS Server是一款可以独立部署的服务器产品，可以将地理资源转化为在线服务，包括地图、地理编码、地理处理、3D地理数据、要素编辑、网络分析、OGC支持、数据访问、移动数据提取等。这些在线资源可以供用户通过桌面、Web、移动等多客户端使用。ArcGIS GIS Server是ArcGIS Enterprise的核心组件，作为Web GIS平台的托管服务器，是ArcGIS Enterprise系统必须安装的组件，其他Server也需要GIS Server进行管理。

□ GeoAnalytics Server

GeoAnalytics Server提供矢量大数据的处理和分析能力。它可以直接从.csv文件、.txt文件、shapefile和大数据源（例如云存储、HDFS和Hive）读取信息。它内置的分布式计算框架可以为进行大数据分析的一系列工具提供支持。通过Server内部提供的聚合、回归、检测、聚类分析，可以更深入地探索海量数据，洞察可能隐藏在数据中的模式、趋势和异常等。

□ GeoEvent Server

GeoEvent Server是一个流式数据处理组件，它支持将基于事件的实时数据流集成为企业级GIS中的数据源。用户可对事件数据进行过滤、处理并将其发送到多个目的地。它可以与几乎所有类型的数据流建立连接，并在发生指定情况时自动提醒工作人员。上述所有操作全部实时进行。

□ Image Server

Image Server同样是一个支持分布式计算和存储的组件，可为大量影像、高程数据、栅格和其他遥感数据的分析处理和服务提供支持。它可用于汇集、处理、分析和管理来自不同传感器、源和时间段的大量重叠、多分辨率影像和栅格数据，也可以发布动态影像服务，以将源影像和其他栅格数据按需转换为多个影像产品，而无需预处理数据或存储中间结果，从而节省时间和计算机资源。

□ ArcGIS DataStore

ArcGIS DataStore是新一代数据存储中间件，它可以便捷地设置和配置三种数据存储，关系型、切片缓存和时空大数据存储。关系数据存储，用于存储门户的托管要素图层数据、从空间分析工具的输出结果中创建托管要素图层、从Portal for ArcGIS中上传shapefile等文件并发布托管的要素服务等；切片缓存数据存储，用于存储二、三维图层缓存，打造ArcGIS平台的三维GIS应用；时空大数据存储，用于存储GeoEvent Server获取的实时观测数据，以及存储GeoAnalytics Server大数据分析的结果。

11.2.2 SuperMap（超图）

超图是中国最大的地理信息技术提供商，成立于1997年。其最主要的产品为SuperMap GIS，目前已经更新到11i版本。整个产品有5大模块，如图11-5所示。

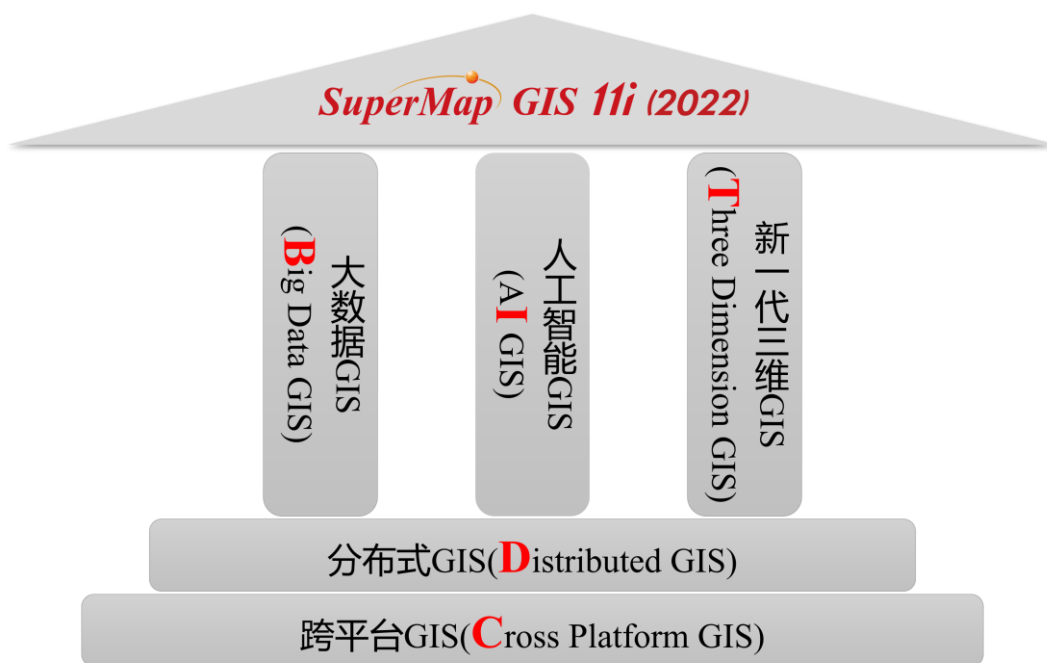


图11-5 Super Map产品架构图

□ 跨平台GIS

跨平台GIS是整个SuperMap的内核，里面包含了所有GIS基础能力。因为全部基于C++编写，并支持多种操作系统和CPU架构，所以目前其内核支持x86、ARM(如鲲鹏、飞腾)、MIPS(如龙芯)、SW-64(如申威)等多种架构的CPU。

□ 分布式GIS

分布式GIS技术体系包括分布式空间数据引擎技术、空间区块链技术、分布式空间分析与处理技术、云原生GIS技术和边缘GIS技术，支撑海量经典空间数据和空间大数据的存储、管理、分析、处理、可视化与发布，实现GIS在高可用、高并发、高弹性、高性能、大容量、高可信等方面的重大突破，构建云边端一体化GIS的分布式协同新模式。

□ 大数据GIS

大数据GIS包括空间大数据存储管理、空间大数据分析、流数据处理与空间大数据可视化等技术，致力于提供全面支持大数据的GIS基础软件与服务，让更多用户能够轻松挖掘空间大数据的内在价值。

□ 人工智能GIS

地理信息人工智能 GIS（AI GIS）是人工智能与 GIS 相互融合的技术统称，包括融合 AI 的空间分析算法（GeoAI）与相关的流程工具，基于 AI 技术增强 GIS 软件的功能和交互端用户体验，提高 GIS 软件的智能化，基于 GIS 对 AI 算法输出结果进行管理、分析与可视化。

□ 三维GIS

以二三维一体化 GIS 技术为基础框架，拓展了三维空间数据模型及其分析计算能力，更全面地融合了实景三维模型、地理实体数据、BIM、激光点云、三维场、地质体等多源异构数据。基于分布式的数据处理自动化工具，实现手工建模数据、BIM、实景三维模型、激光点云、地形等三维数据的高效全流程管理。集成 WebGL、高保真渲染、虚拟现实(VR)、增强现实(AR)、混合现实(MR)、3D 打印等 IT 新技术，推动构建室外室内一体化、宏观微观一体化、空天/地表/地下一体化的数字孪生空间，赋能全空间的新一代三维 GIS 应用。

11.3 具有海量时空数据管理能力的PaaS云平台

11.3.1 阿里云Ganos

阿里云作为中国云计算市场份额最高的平台厂商，在云计算技术方面有着非常大的优势。因此，在一些新技术的探索和新场景的应用上阿里云相比其他厂商相对超前。由阿里李飞飞博士团队研发的空天数据库引擎Ganos，是新一代时空数据库中间件。Ganos与阿里云多个存储引擎相结合，使这些通用存储组件具备时空数据管理的能力，并提供了适合时空场景下的多种接口API，方便用户调用。

1. Lindorm Ganos

Lindorm是面向物联网、互联网、车联网等设计和优化的云原生多模超融合NoSQL数据库，支持宽表、时序、文本、对象、流、空间等多种数据的统一访问和融合处理，并兼容SQL、HBase/Cassandra/S3、TSDB、HDFS、Solr、Kafka等多种标准接口，无缝集成三方生态工具，适用于日志、监控、账单、广告、社交、出行、风控等场景。

Lindorm Ganos时空服务是Lindorm宽表引擎的原生扩展服务，在宽表基础上，为各类空间/时空数据提供云上的高效存储、查询和分析计算能力。在功能上，Lindorm Ganos时空服务与PostgreSQL中的PostGIS插件、MongoDB的地理范围查询相似。Lindorm Ganos时空服务支持标准的空间数据类型（Geometry），提供时空索引能力，能够帮助用户查询各类时空数据。Lindorm Ganos提供了两个

独立的版本，分别是使用SQL接口的版本和兼容GeoMesa接口的版本。这两个版本主要特征如表8-1所示。

表8-1 Lindorm Ganos不同接口对比

类别	SQL 接口	兼容 GeoMesa 接口
接口	标准 SQL 接口	GeoTools API/ECQL 接口
功能	覆盖 GeoMesa、MongoDB、Redis 在时空方面的功能。 提供了大量兼容 PostGIS 的函数。 支持轨迹计算相关的函数	继承开源 GeoMesa 功能
性能	通过查询改写、并行查询、shard、优化空间逼近算法等技术，性能优于开源版本的 GeoMesa	优于开源版本的 GeoMesa 和 HBase
定位	适用于原有使用 SQL 接口（如 PostGIS）和 GeoMesa、MongoDB、Redis 时空应用的迁移	适用于所有开源 GeoMesa 用户

2. PolarDB Ganos

PolarDB是阿里巴巴自研的新一代云原生关系型数据库，在存储计算分离架构下，利用了软硬件结合的优势，为用户提供具备极致弹性、高性能、海量存储、安全可靠的数据库服务。PolarDB 100%兼容PostgreSQL 11、PostgreSQL 14，高度兼容Oracle。

PolarDB Ganos时空引擎提供了一系列的数据类型、函数和存储过程，用于在云原生关系型分布式数据库PolarDB中对空间/时空数据进行高效的存储、索引、查询和分析计算。它内置了7大类时空模型，包括几何模型、栅格模型、路径模型、点云模型、轨迹模型、格网模型和矢量金字塔模型。针对这些模型，它分别配置了6类SQL函数：Geometry SQL、Raster SQL、SpatialRef SQL、PointCloud SQL、Trajectory SQL、GeomGrid SQL。除此之外，还可以搭配GeoServer通过配置PostGIS数据源进行数据接入，对外提供REST API。

3. RDS for PostgreSQL Ganos

阿里云关系型数据库RDS（Relational Database Service）是一种稳定可靠、可弹性伸缩的在线数据库服务。基于阿里云分布式文件系统和SSD盘高性能存储，RDS支持MySQL、SQL Server、PostgreSQL和MariaDB引擎，并且提供了容灾、备份、恢复、监控、迁移等方面的全套解决方案。RDS PostgreSQL完整实现了SQL规范并支持丰富多样的数据类型，包括JSON数据、IP数据和几何数据等。除了完美支持事务、子查询、多版本控制（MVCC）、数据完整性检查等特性外，RDS PostgreSQL还集成了高可用和备份恢复等重要功能。

因为同样是关系型数据库，因此和PolarDB Ganos非常相似，RDS for PostgreSQL Ganos时空引擎提供了相同的数据类型、函数和存储过程，并且SQL语法也完全一致。

11.3.2 华为云 CloudTable

作为国内另一家云厂商巨头，华为云的CloudTable存储引擎也支持时空数据管理能力。表格存储服务（CloudTable Service，简称CloudTable）是基于Apache HBase提供的全托管NoSQL服务，提供毫秒级随机读写能力，适用于海量(半)结构化的时空、时序数据存储，可被广泛应用于物联网、车联网、金融、智慧城市、气象等行业。

CloudTable内置了分布式时空处理中间件GeoMesa。通过官网看并没有做额外的优化和扩展，因此用户只需要将其作为原生GeoMesa HBase来使用。整个产品架构图如图11-6所示。

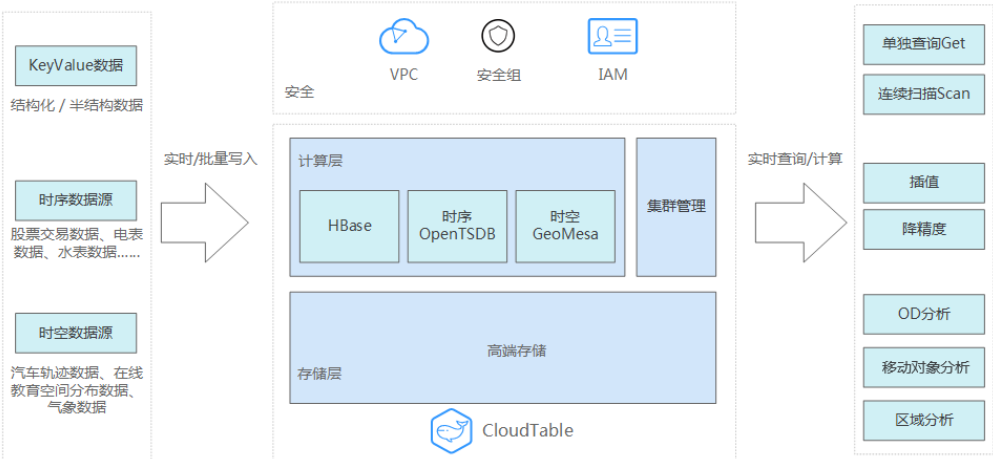


图11-6 CloudTable系统架构图

11.4 具有时空大数据管理与分析能力的SaaS化产品

11.4.1 ESRI ArcGIS Online

ArcGIS Online第一个版本上线于2012年，是一款在公有云上的SaaS产品。相比ArcGIS传统的桌面端产品，这款产品更方便用户使用和共享。ArcGIS Online是一个基于云技术的制图和分析解决方案。将其用于制作地图、分析数据以及共享与协作。用户可以在任何地点访问工作流特定的应用程序、地图和数据，以及移动的外业工具。用户先注册ArcGIS Online账号，然后就拥有了自己组织的工作台。用户最开始可以编辑组织的主页，发布后可以作为所有时空资源的入口，这里可以配置一些组织的概况及资源描述，如图11-7所示。在工作台内部ArcGIS Online为用户提供了5大块核心功能。



图11-7 ArcGIS Online主页编辑

1. 创建地图、场景及应用

ArcGIS Online支持创建2D web地图、创建3D web场景以及创建web应用程序。通过Map Viewer、Map Viewer经典版和3D Scene Viewer，用户可以访问底图和智能样式的图库，也可以上传自己的数据资源以浏览和可视化数据（见图11-7）。

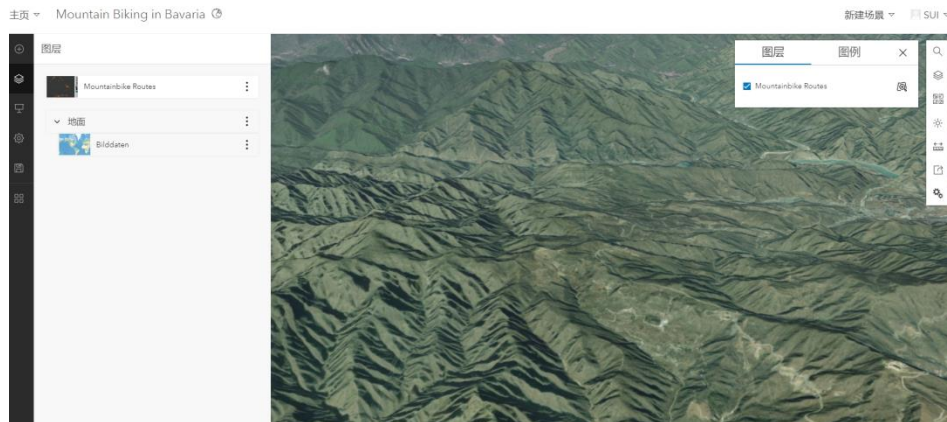


图11-8 ArcGIS Online 3D 设计界面

2. 共享与协作

ArcGIS Online提供了丰富的权限管理机制，用户可建立仅通过邀请加入的私有群组，或者对所有人开放的公共群组，也能与组织以外的其他用户共享内容，还可以与其他组织建立协作，以共享和使用彼此的内容。共享地图和其他内容的方法还有：将其嵌入Web页面、博客、Web应用程序及社交媒体；使用专用应用程序，在外业、办公室或社区与同事进行协作。

3. 浏览和分析数据

ArcGIS Online包括交互式地图和3D场景，方便用户在此之上了解、探索和测量地理数据。用户可以资源社区“ArcGIS Living Atlas of the World”，该社区里面包括地图、3D场景、数据图层、影像、分析工具以及应用程序。用户还可以使用Map Viewer经典版中所含的分析工具来探寻新模式、

寻找适宜地点、丰富我们的数据、了解附近信息并汇总数据。用户还可以通过ArcGIS Notebooks利用ArcGIS API for Python来运行数据科学脚本，进而深入研究数据。

4. 管理数据

ArcGIS Online可以方便地添加、管理和共享自己的数据，可以将数据作为web图层发布到ArcGIS Online。这些web图层都会托管为ESRI的云端资源。这些资源会有一些的限额，用户可以按需升级。用户还可以将图层添加至地图，并允许其他用户使用这些图层。

5. 外业作业

ArcGIS Online支持外业活动。利用内置工具和应用程序来收集数据、导航、协调和监控工程。用户可以选择一个区域来创建用于离线使用的地图。然后设置同步，以便离线编辑者可以获取地图的最新版本。

11.4.2 Carto

Carto是一款在线的SaaS位置智能平台。它于2012年由Javier de la Torre 和 Sergio Álvarez 在马德里创立。目前平台在不断壮大，现在拥有数十万的用户，让数百万人通过时空的视角来洞察事物和解决问题。下面介绍Carto主要的能力模块。

1. 空间资源云连接

在Carto内，用户可以连接到一些主流的云端数据库（见图11-8），然后注册已有的数据资源。目前支持的数据库有Google BigQuery、Snowflake、Amazon Redshift和Databricks。所有已注册的数据，都可以用于后续的分析与可视化。

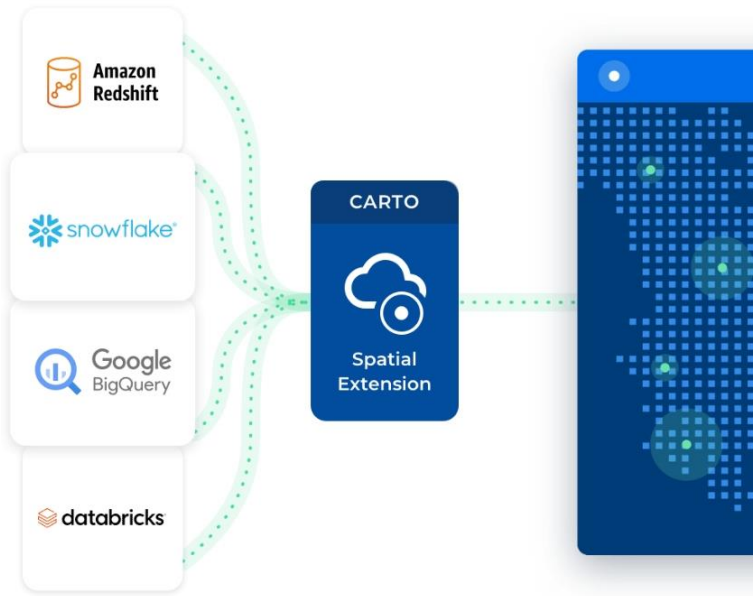


图11-9 Carto云端资源连接

2. 数据集市

除了连接云端的数据资源之外，用户还可以使用Carto数据集市中的数据（见图11-9）。这些数据来源非常广泛，涉及人口、兴趣点、环保、交通、住房等10多个行业的数据，包括12000份开源/外采的地理空间数据集。整体数据集的可靠性比较高，数据提供商都是垂直领域比较权威的机构，包括MasterCard、Vodafone、美国审计署等。用户进入数据集后，可以查看数据来源和描述。还可以通过地图或者表格对部分数据进行预览。如果决定使用当前数据，可以进行订阅或者接入部分免费的样本。

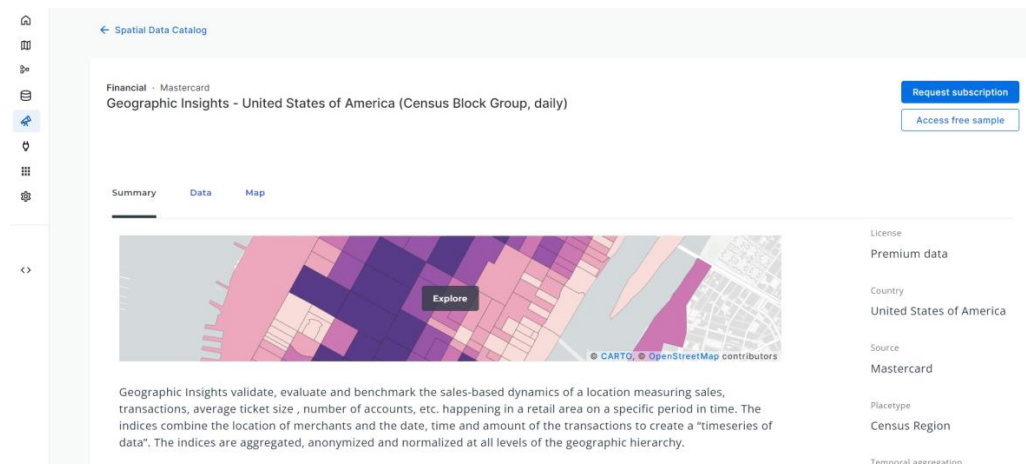


图11-10 Carto数据集市

3. 数据浏览器

数据浏览器用于管理已有的数据资源，包括前面提到的云资源以及数据集市数据。浏览器模块以目录树的方式展示所有数据表。用户可以预览每个表格的数据，并支持在地图展示。除此之外，这里还支持基于当前数据创建地图、创建瓦片等操作。

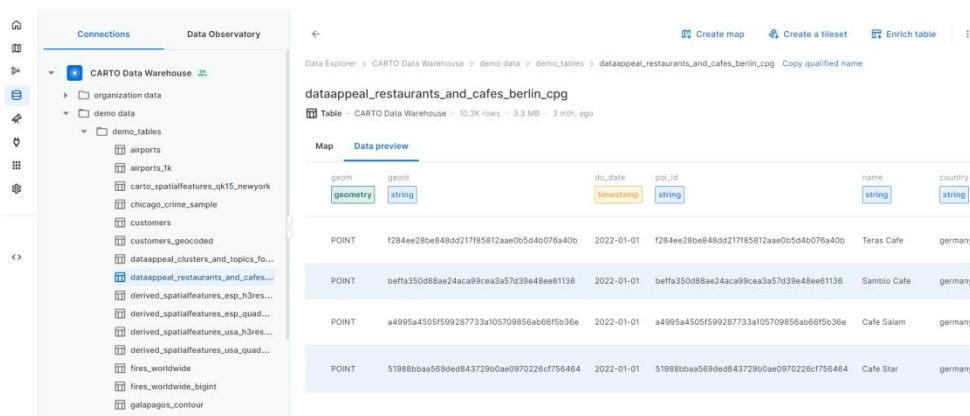


图11-11 Carto数据浏览器

4. 地图可视化分析

基于地图的可视化分析是Carto的核心能力。前面介绍的数据接入和管理，都是为此模块服务的。

此模块有几个重要的功能，可以帮助用户完成一幅分析地图。

添加数据源

Carto支持三种数据资源的加入（见图11-11）。第一种直接加入数据浏览器管理的资源。第二种是对已有资源进行SQL查询，将结果表作为数据源。第三种支持用户上传本地的文件，支持csv、geojson以及shapefile。

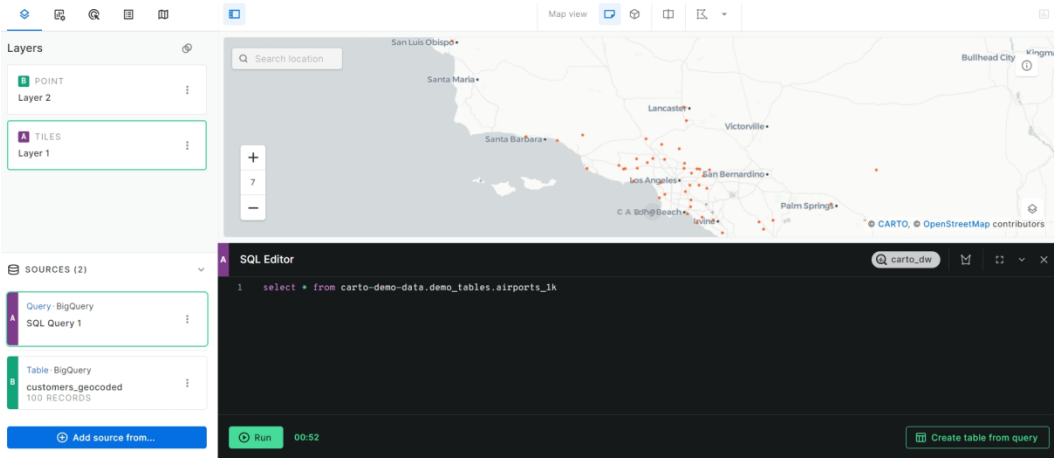


图11-12 Carto配置数据源

配置地图

配置好数据源后，系统会自动添加相应的图层。用户可以对图层进行样式配置、调整顺序等地图常规操作。除此之外，Carto还支持在地图上添加一些统计控件（见图11-12）。每个控件都会绑定一个数据源，用户可以为统计空间配置不同的的方法及字段。目前支持公式、直方图、分类等控件。

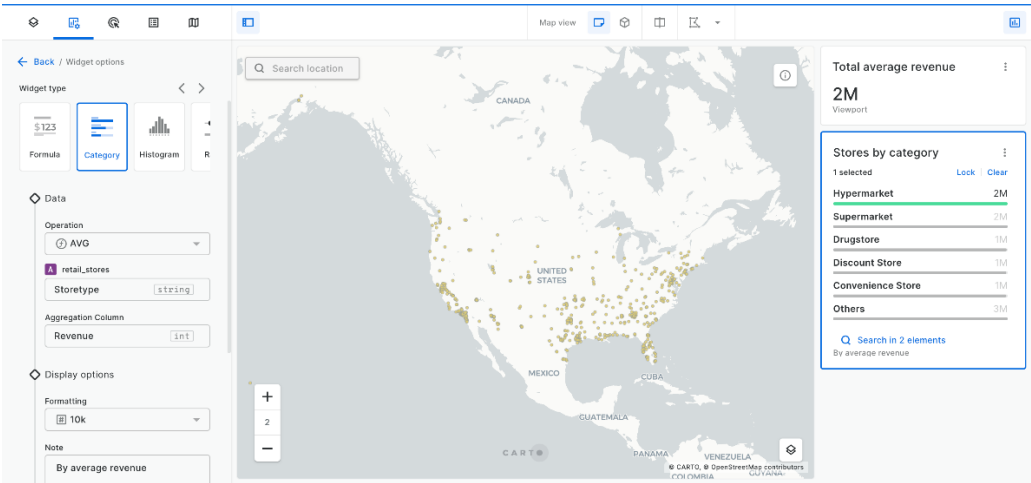


图11-13 加入公式控件和分类控件

设计地图

配置好地图控件后，Carto为图层提供了丰富的样式效果。用户可以对不同几何类型的数据继续进行配置。比如，点数据可以配置聚类、热力图有效果，还可以将点聚合到动态网格内。除此之外，还可以对图层的显隐、颜色、符号进行配置。

5. 进行空间SQL分析

除了将数据源直接进行样式展示，Carto还提供了丰富的SQL分析功能，支持将分析后的结果进行展示。不过，根据数据源类型的不同，支持的分析函数也有区别。表8-2展示了数据源和分析能力的对应关系。

表8-2 不同数据源与分析能力的对应表（*表示需要使用分析工具箱）

分析\数据源	Carto数据集市	BigQuery	Redshift	Snowflake	PostgreSQL
相交聚合	✓	✓	✓		✓
缓冲区分析	✓	✓			✓
添加字段	✓	✓	✓	✓	✓
筛选字段值	✓	✓	✓	✓	✓
计算质心	✓	✓	✓	✓	✓
KMeans聚类	✓	✓ *			✓
等值线生成	✓	✓ *	✓ *	✓ *	

6. 发布地图

最终搭建好的地图效果可以分享给其他用户，Carto支持组织内分享和公共分享。通过一个链接URL即可访问地图。

7. 分析工具箱

CARTO工具箱是一套功能和程序（见图11-13），扩展了云数据仓库中可用的地理空间功能。它目前可用于 Google BigQuery、Snowflake、Redshift、Databricks和PostgreSQL。该工具箱目前包含100多个高级空间函数，归属于不同的模块。

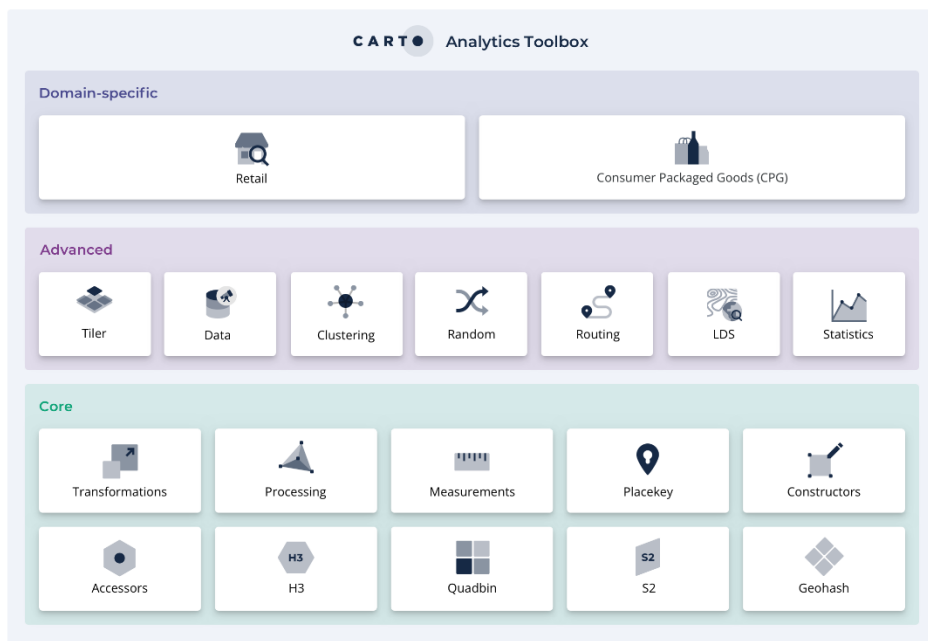


图11-14 Carto分析工具箱功能架构图

11.4.3 京东JUST

京东JUST（**J**D **U**rban **S**patio-**T**emporal Data Engine）是由京东科技集团研发的一款云原生的时空大数据引擎。系统主要面向需要在城市场景内对时空数据进行管理、分析、可视化的客户。JUST从构建初期开始就基于云计算和大数据技术，天然面向海量时空大数据场景，支持高效的存储、索引和可视化渲染。系统内置了大量的分析算子，可以快速支持各种时空数据需求。JUST一直在时空数据管理的基础理论、前沿技术和项目实践三个方面不断探索，从2019年建设以来，发表了20多篇国际顶刊论文，先后申请专利50多项，获得了公安、民政等权威机构的认证。JUST在几年间服务了近20个大中型智慧城市项目，涉及市域治理、监测预警、智慧园区、智慧公安等场景。

JUST既有公有云产品供大众进行时空数据在线管理和分析，也有私有云版本供政企客户使用。整个系统对应时空大数据系统的体系结构，共有5层6个模块（见图11-14和图11-15）。

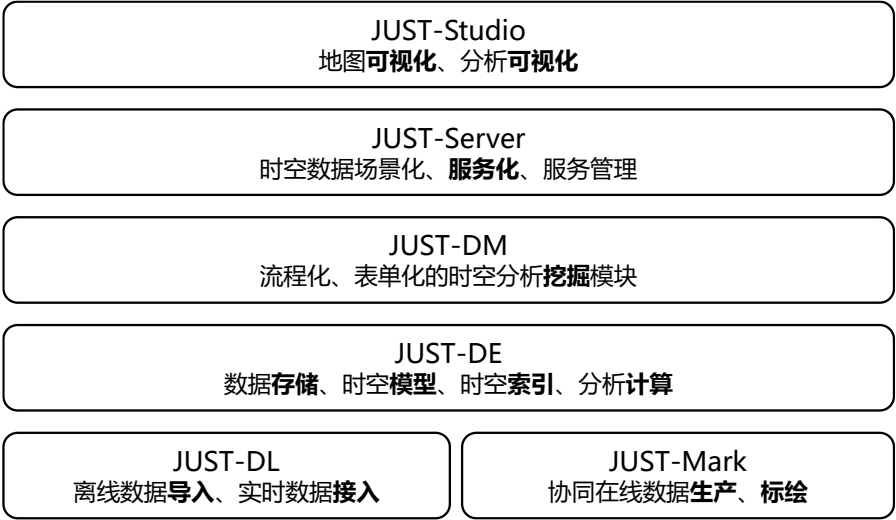


图11-15 JUST模块架构图

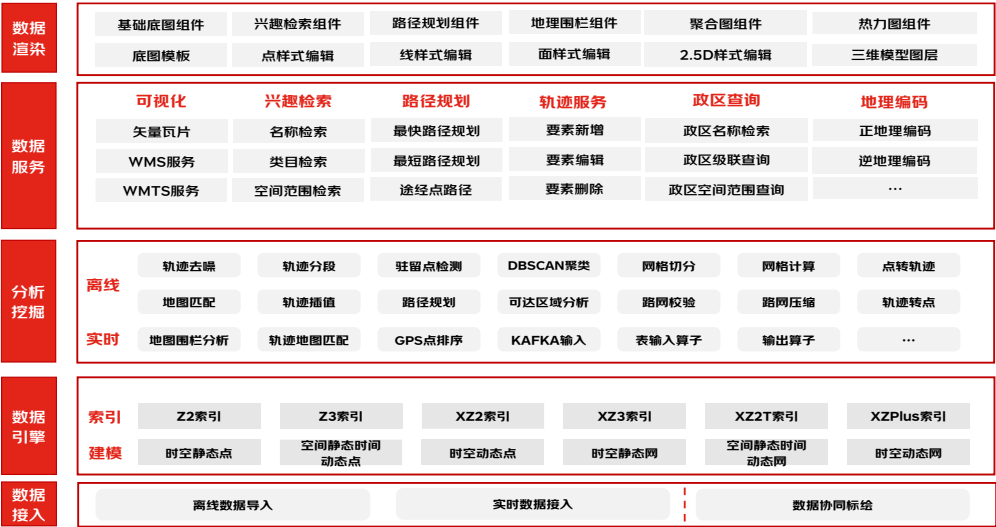


图11-16 JUST系统功能架构图

1. JUST-DL（JUST数据接入模块）

针对多源异构时空数据，JUST提供了一个单独的数据接入模块，支持离线和实时两种接入模式(见图11-16和图11-17)。其中，离线模式支持用户自己上传的本地文件，包括csv、txt以及压缩后的Shapefile文件，同时还支持分布式HDFS文件的导入。用户配置好原始数据后系统会自动判断文件大小，如果超过阈值会自动转为分布式导入模式，加快导入的速度。实时接入的方式，JUST扩展了Flink的能力，一方面增加了JUST类型的Sink，使外部数据源可以直接导入存储引擎，另一方面在Flink SQL内加入了大量的时空处理函数，让用户在接入数据的过程中实时处理、实时过滤。

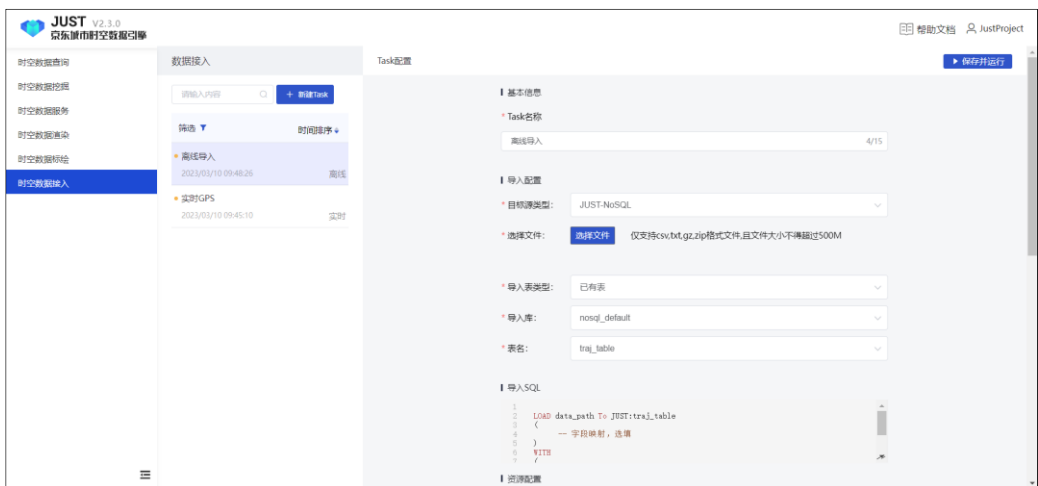


图11-17 离线导入模式

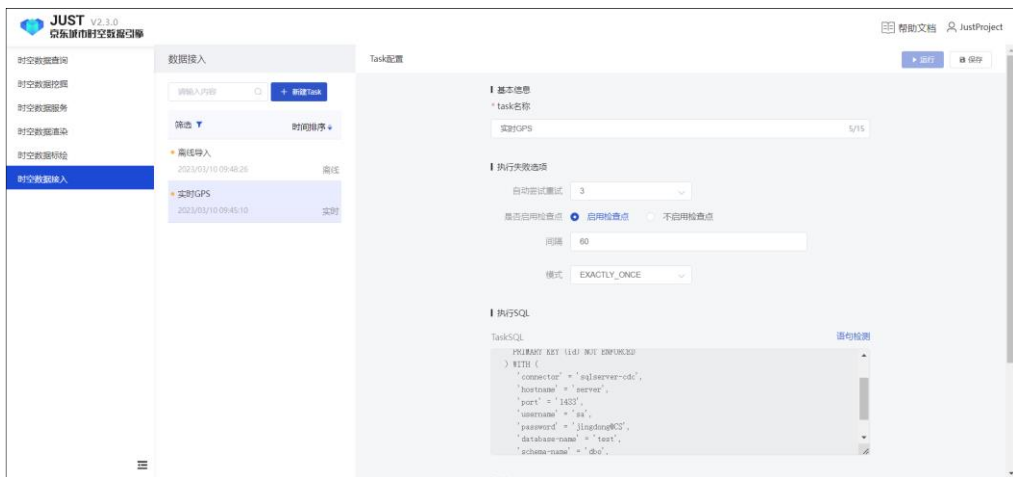


图11-18 实时接入模式

2. JUST-Mark

在一些项目中，很多业务数据用户并不掌握，无法通过数据接入的方式获取数据。比如，交通监管部门，经常要对某些车辆划分一些禁行区域。为了对这些区域进行数字化管理，需要生成区域的空间范围。除此之外，一些物流配送区域、基层辖区、重要地点等都需要一个灵活的数据生成工具。JUST-Mark就是这样一个支持多人在线协同标绘的工具（见图11-18）。

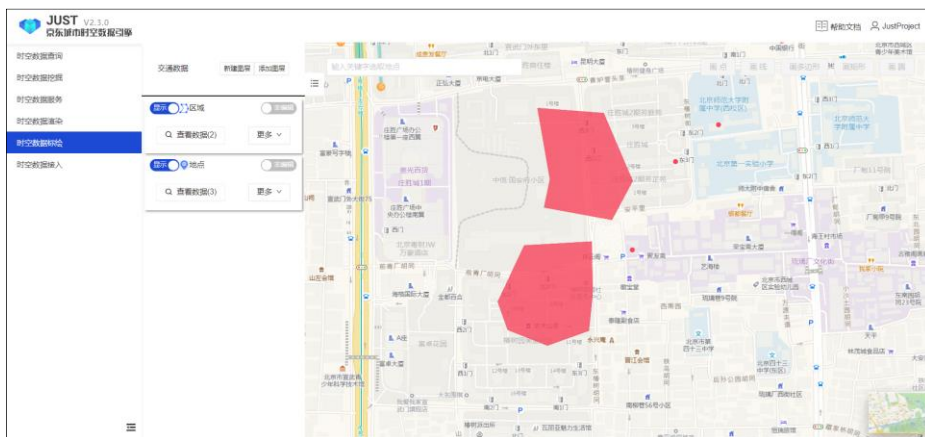


图11-19 JUST-Mark 数据编辑界面

除了在线标绘工具外，JUST目前正在内测一款语义地理实体生成工具，其构建了城市基础地理数据的关系图谱，结合NLP技术对用户输入的地理文本进行分析、拆解、推断、修正，最终生成最可能的地理实体范围。和一般的地址解析不同，这个工具不是简单的查询匹配，而是实体的生成，类似于时空数据的AIGC。这个工具可以大大提升数据生产的效率。JUST已经在一些项目中做过小范围的应用。比如在某市基层治理项目中，政府缺乏社区辖区的地理范围数据，只有对于辖区的描述，并且由于城市快速发展，这些辖区会经常变化，需要一款快速生成范围的工具。在此项目中，工作人员基于JUST地理实体生成工具，只用2天时间就生成了全市9000+个社区，大大降低了人工标注的成本。图11-169是工具的算法流程示意图。

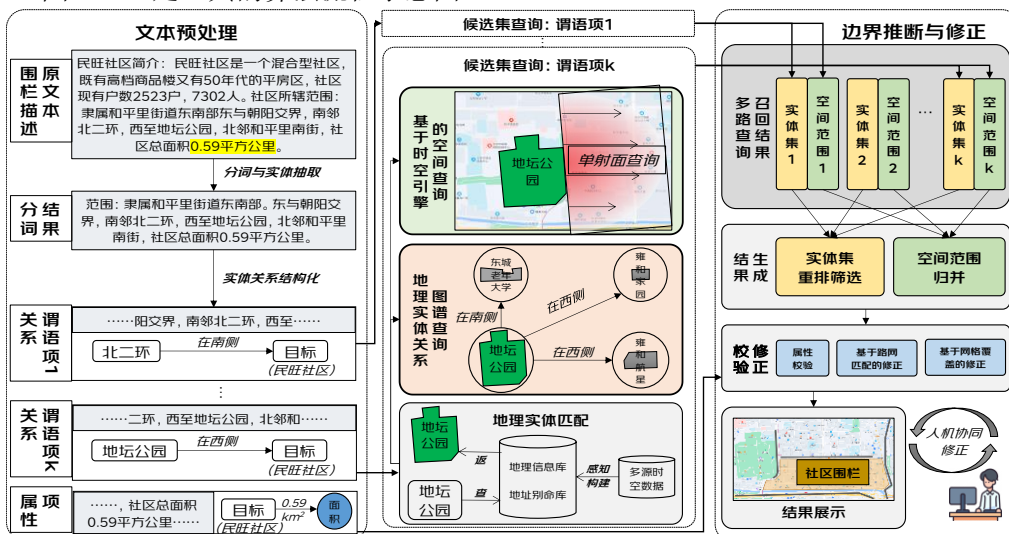


图11-20 JUST语义地理实体生成算法流程示意图

3. JUST-DE

JUST最重要的一项功能是对数据的存储和管理，JUST-DE（JUST-DataEngine）提供了丰富的时空存储模型，并且支持4大类存储引擎，自研了多种时空索引。JUST-DE配套统一的SQL语法让时空数据的管理非常高效。下面介绍该模块的几大特色。

□ 高效索引

JUST-DE本身支持10多种主流时空数据索引，包括四叉树、R树、Z曲线等。除此之外，引擎基于海量数据自主研发了效率更高的Z2T、XZ2T及XZ-Plus索引。Z2T和XZ2T索引如图11-20所示。经过第三方测试机构的评测，在千万量级的数据上进行时空查询，JUST新型索引效率相比PostGIS效率要高近十倍，相比原生GeoMesa索引提高了4倍。

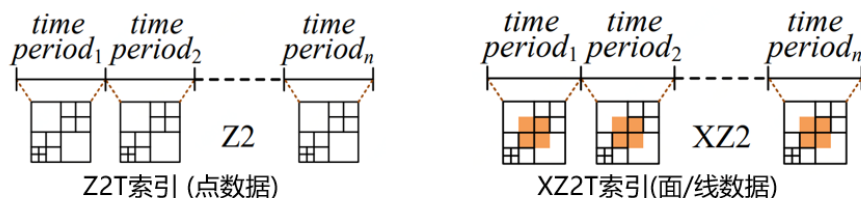


图11-21 JUST自定义索引Z2T以及XZ2T

□ 多种存储模式

时空数据的种类千变万化，应用场景也十分复杂。实践证明，很难利用一种存储方式应对所有场景。因此，JUST-DE扩展了存储组件的类型，目前已支持4种类型：**RDS**时空关系型存储适合千万级及以下数据量OLTP场景、**MDS**时空内存存储适合数据实时快速更新场景、**时空NoSQL**存储适合千万级以上写多读少的场景、**DFS**时空对象存储适合存储时空文件数据。

□ SQL语法支持

JUST为方便用户管理数据，提供了一个基于SQL的数据管理界面（见图11-21）。用户通过JUST-SQL语法，可以对时空数据进行增、删、改、查等操作。针对不同的存储模式，JUST-SQL有80%的语法是相同的，另外20%保留了不同存储模式特色的语法。

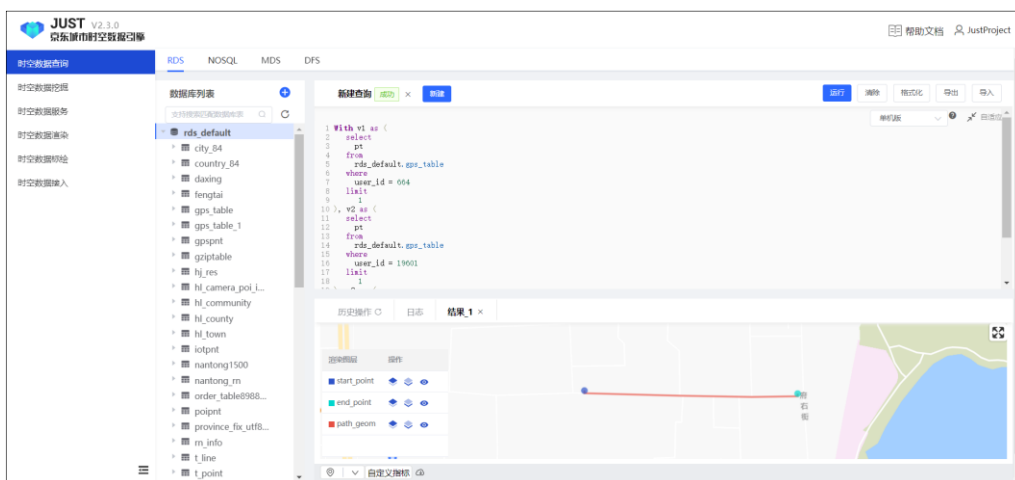


图11-22 JUST 数据管理界面

□ 多种执行模式

针对不同的分析场景，JUST还提供了两种执行模式：单机模式和分布式模式。前者适合小数据量简单查询、计算场景，后者利用Spark引擎可以支持超大体量的计算和分析。系统支持用户自己在界面指定执行模式，同时配有智能模式，即系统通过一些CBO和RBO的策略进行自动判断，如图11-

22所示。

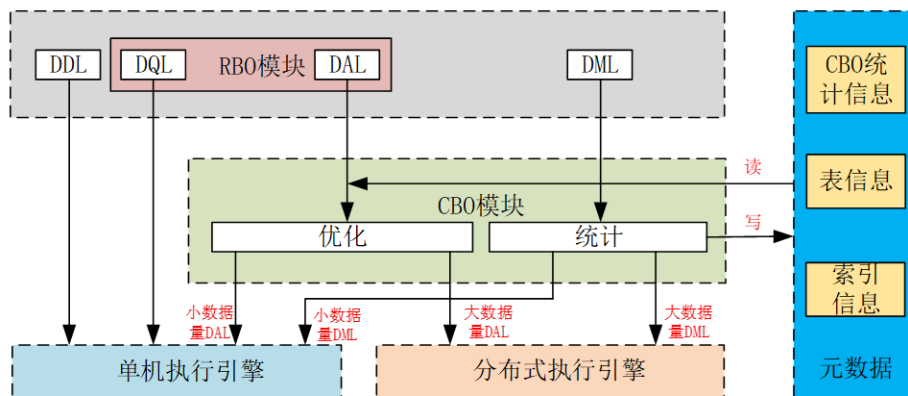


图11-23 JUST-SQL多模式执行流程

4. JUST-DM

JUST-DM是一个时空工作流引擎，适合执行一些周期性的跑批任务或实时分析处理任务。它内部以算子化的形式封装了大量主流的时空分析算法（如图11-23和图11-24）。用户只需要通过托、拉、拽的方式即可快速配置一个处理流程。为方便使用，**JUST-DM**预配置了大量的流程模板，用户只需根据自己的数据，配置好输入和输出算子即可。

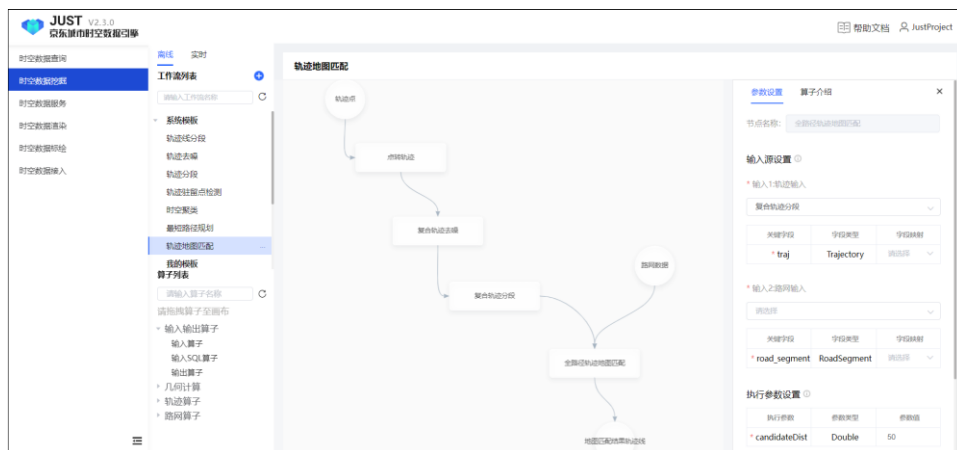


图11-24 JUST-DM离线算法流程设计

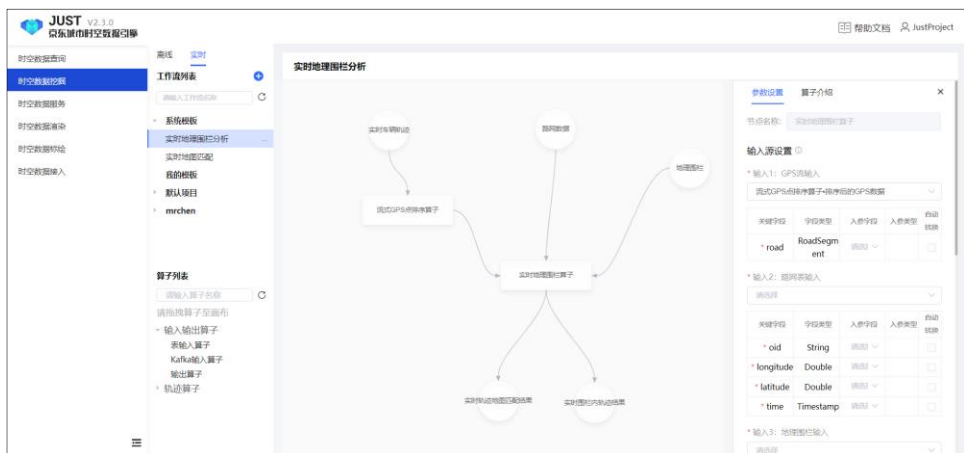


图11-25 JUST-DM实时算法流程设计

5. JUST-Server

我们在第7章讲了时空数据服务的作用和重要性. 时空数据服务是数据共享的重要手段。前面讲到的JUST模块大都服务于数据管理和分析，但在很多场景内，用户需要将已有数据共享给第三方，可能是其他的合作伙伴，也可能是系统前端。因此，共享的标准非常重要，JUST-Server所有服务都符合OGC标准，或者在其基础上进行了扩展，可以快速接入各种第三方系统。JUST-Server针对不同场景设计了两个服务模块。

□ 场景化服务

所谓场景化，就是将所有时空服务进行场景类别划分，用户针对每种类别只需一次数据配置即可获得该场景下所有服务能力。以图11-26所示的几何兴趣检索场景为例，场景内包括多种时空检索的服务。用户只需在页面下方指定对应的数据表名称，映射关键字段即可一次发布所有服务（也支持选择发布部分服务）。每个发布好的服务都有自己的URL和参数。第三方使用者只需根据服务文档即可调用服务。针对每个服务，用户还可以配置超时时间、流量控制、上线下状态等。



图11-26 JUST-Server场景服务构建

□ 自定义SQL服务

虽然系统内置了大量场景服务，但有些项目还会有一些业务定制化需求。为此，系统开放了自定义的SQL服务（见图11-26）。用户可以利用JUST-SQL配置查询逻辑，发布之后，服务会按照指定SQL执行，非常灵活方便。



图11-27 JUST-Server自定义SQL服务

6. JUST-Studio

JUST-Studio是一款时空数据可视化产品，它既可以设计精美的地图样式，也可以进行可视化分析。通过简单的托拉拽，它还能配置简单的地图应用。JUST-Studio内部包含三个模块。

□ 地图设计器

基于JUST内部实现的高效矢量瓦片技术，地图设计器可以配置所见即所得的地图（见图11-27）。通过地图设计器，除了加载JUST内部地图服务，用户还可以接入外部的地图瓦片服务。地图设计器目前支持标准的mvt-zxy矢量瓦片、WMS和WMTS。设计好的地图可以直接生成分享链接。包括两种链接，一种是直接分享地图，其他用户打开链接可以直接访问设计好的地图；另一种是样式文件链接，这个要结合JUST-Map SDK进行加载，并可以在已有的地图基础上进行二次开发。

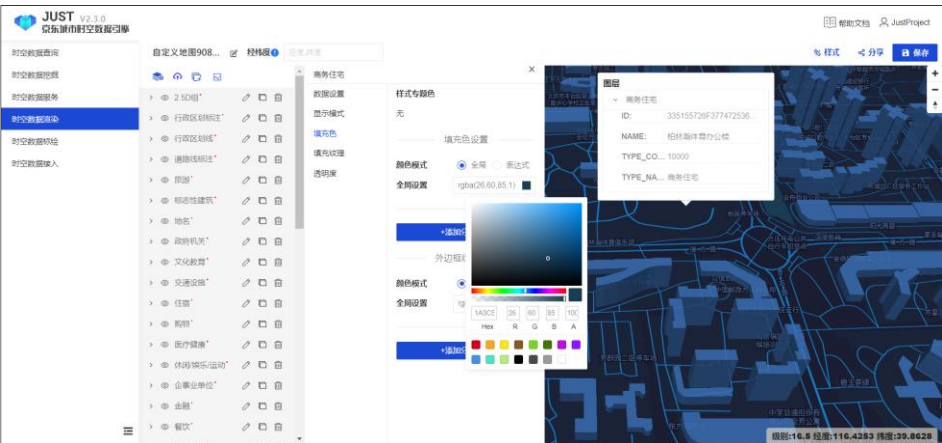


图11-28 JUST-Studio地图设计器

□ 场景编辑器

除了支持地图样式的配置，JUST-Studio还提供了更高级的场景配置工具。用户选择一个配置好的地图作为底图，进入场景编辑器后，可以添加各种控件。目前系统支持四种控件，地图控件用来控制地图视窗，可以改变地图的数据源；图层控件内置了多种时空专题可视化效果，比如热力图、三维柱状图、聚合图等，JUST还有一些独特的时空动态效果，比如实时动态轨迹、监测站点度数等，配置好数据源后可以将效果叠加到当前地图上；分析控件包含了大量地图应用常见的分析能力，如图11-29中POI查询控件、路径规划控件等，用户配置好右侧的数据源，即可使用这些分析能力；通用控件包含地图上一些标准的信息组件，如图11-28中的位置信息组件、比例尺组件等。

配置好场景内的控件后，用户也可以分享给其他用户，同样有两种分享链接。一种可以直接打开作为一个时空Web应用，另一种通过JUST-Map SDK加载后进行二次开发。

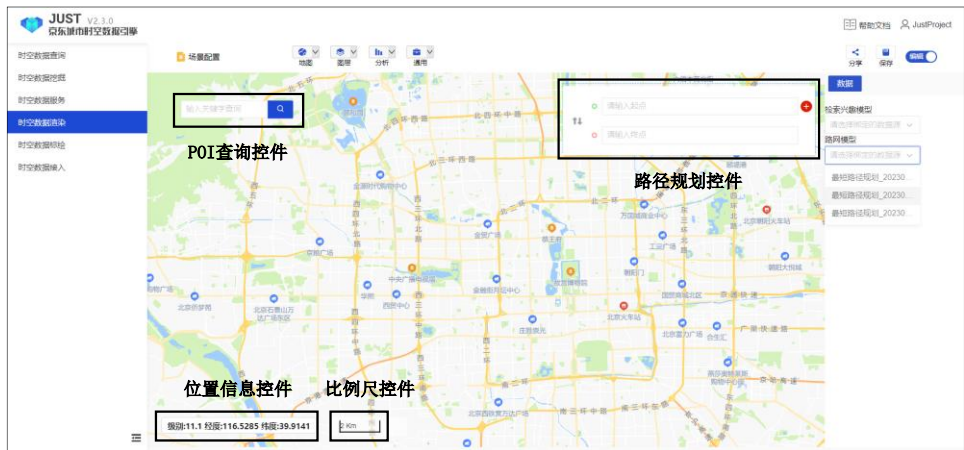


图11-29 JUST-Studio场景编辑器

7. JUST-Map SDK

JUST-Map SDK是一个地图应用的前端开发包，基于它可以从0开始构建一个完整的地图页面。JUST-Map SDK里面开放了大量的可视化接口和分析接口。前面介绍的两个模块就是基于此SDK完成搭建的。更好的实践是基于地图设计器配置好地图，然后利用SDK接入到自己的页面中，根据业务的需求控制地图内数据的展示效果；或者基于场景编辑器，配置好整个页面，利用SDK接入到自己的应用。图11-30是JUST-Map SDK的在线使用文档。该图中的示例展示了如何添加一个轨迹回放组件。

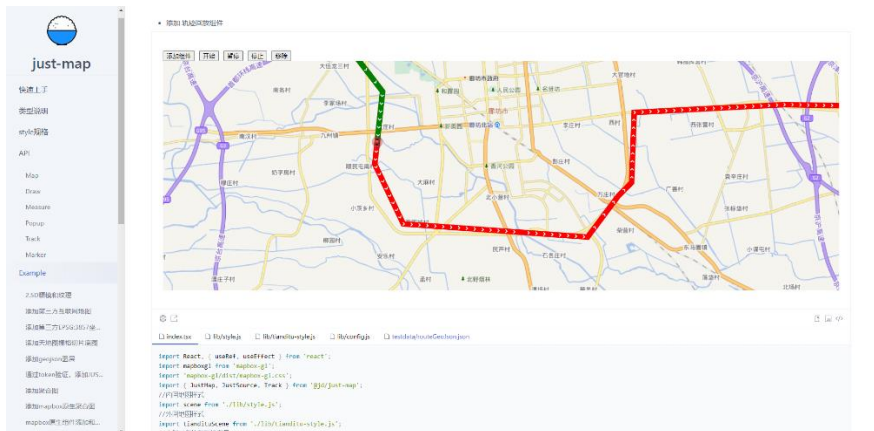


图11-30 JUST-Map SDK的在线使用文档