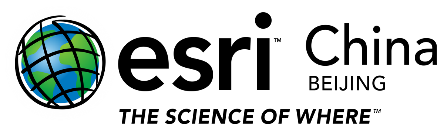
Esri大数据平台搭建指南概述



易智瑞中国信息技术有限公司

2017 年 3 月

**——制定及修订记录——**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **版本** | **完成日期** | **编写/修订纪要** | **编写者** | **备注** |
| **V1.0** | **2017-3-1** | **初稿** | **胡源** |  |
| **V1.1** | **2017-3-19** | **修订** | **胡源** | **内容完善** |
|  |  |  |  |  |
|  |  |  |  |  |

目录

[Esri大数据平台搭建指南概述 1](#_Toc479175649)

[1. 为什么空间大数据越来越重要？ 4](#_Toc479175650)

[2. 空间大数据分析面临哪些问题？ 4](#_Toc479175651)

[3. Esri提供的空间大数据解决方案 4](#_Toc479175652)

[4. 这种解决方案是怎样的？ 5](#_Toc479175653)

[5. 大数据文件共享三种方式区别与联系 6](#_Toc479175654)

1. **为什么空间大数据越来越重要？**

当今世界已进入大数据时代，大数据的竞争越演越烈，而产生的所有数据中，高达80%都包含地理空间要素。无论信息技术变得多么成熟，任何东西都取代不了时间和位置这两种独特数据的结合体。因此，地理空间数据才会拥有今时今日的地位。

随着网络和现实世界变得更加你中有我、我中有你，地理空间数据将被越来越多地加以利用，让我们能够追踪人和物体的位置，提供基于位置的新服务。如果政府部门掌握地理空间大数据的分析与处理技术，就能更正确并及时地做出决策；如果企业可以掌握地理空间大数据并与面向用户的应用结合起来，就将在可预见的未来获得竞争优势。大数据分析能力已经成为政府部门和企业的一大资本。

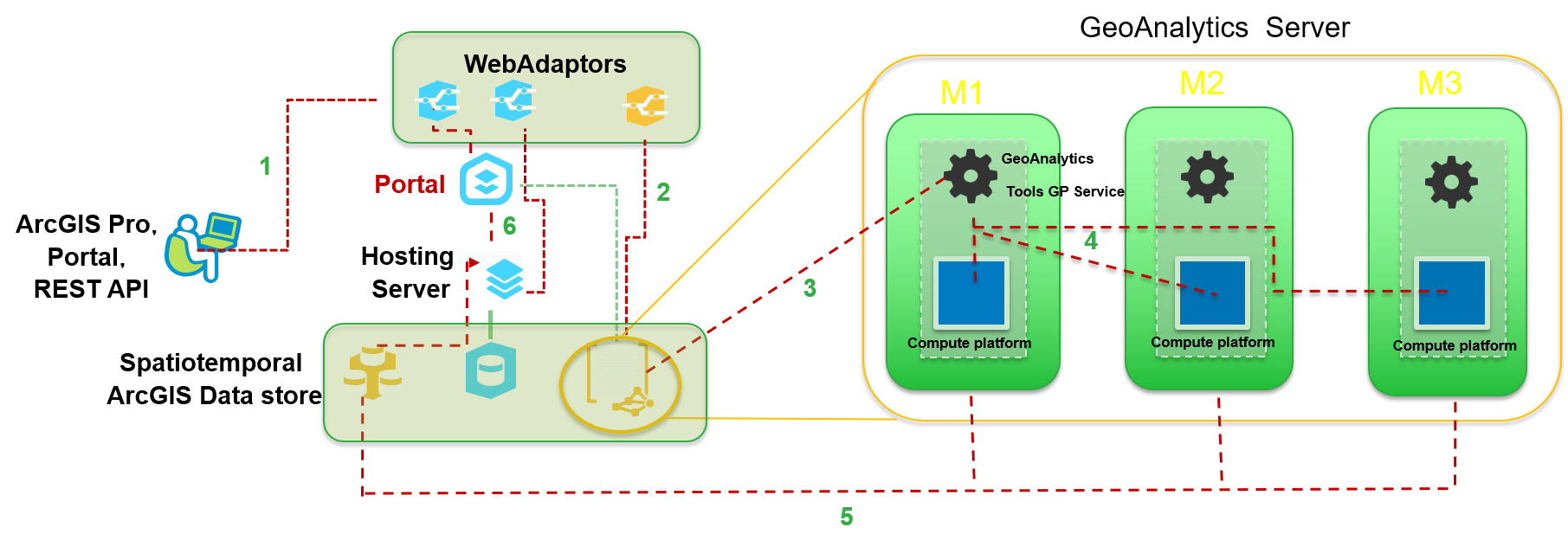
1. **空间大数据分析面临哪些问题？**

大数据有4V的说法，4V即Volume(数量)，Velocity(速度)，Variety(种类)，Value(价值)。地理空间大数据有着极其丰富的应用与价值，同时也带来了一系列独特的挑战。

1. 这么大的数据量如何存储？如何保障数据的存储安全？
2. 这么大的数据量如何计算？如何进行快速计算？

1. **Esri提供的空间大数据解决方案**

在最新的ArcGIS10.5产品体系中，加入了空间大数据高级分析软件产品GeoAnalytics Server，它是ArcGIS Enterprise中ArcGIS Server的一个角色，支持大数据文件共享的形式读取数据，并且内置了Spark这种RDD（弹性分布式数据集）的分布式计算框架，来处理带有时间和空间属性的矢量或者表格数据。其系统架构如下图所示：



可以通过ArcGIS Pro或者Portal for ArcGIS的Web站点执行大数据分析，也可以通过REST方式进行远程过程调用。从中可以看到，Portal将原本复杂的数据分析功能移植到了Web端，用户在任何时间、任何地点都可以通过浏览器快速、方便地进行海量数据的分析与挖掘。

1. **这种解决方案是怎样的？**

GeoAnalytics Server目前提供了十一种空间大数据分析工具与一种大数据管理工具：

* **汇总数据：聚合点、连接要素、重新构建追踪、汇总属性、范围内汇总**
* **查找位置：查找相似位置**
* **分析模式：计算密度、查找热点、时空立方体(ArcGIS Pro)**
* **临近分析：创建缓冲区**
* **管理数据：复制到数据存储**

GeoAnalytics Server目前支持三种方式的大数据文件共享：文件共享、HDFS、Hive。在后面的文档中，我将记录三种方式的搭建过程。

1. **大数据文件共享三种方式区别与联系**

传统文件共享与HDFS区别：

|  |  |  |  |
| --- | --- | --- | --- |
| 存储方式 | 分布式存储 | 多副本容错 | 数据大小上限 |
| 文件共享 | 否 | 否 | 有 |
| HDFS | 是 | 支持 | 无 |

HDFS与Hive联系：Hive可以将HDFS中结构化的数据文件映射为一张数据库表。