我和 FME 的故事

摘要：

刚接触 FME 是看到这个比赛才知道的，看到 FME 在空间数据的连接、整合、自动化等方面有很大的帮助。很感兴趣，想多了解 FME，就做了很多功课、调查、学习，操作让自己学习到更多关于 FME 故事。这篇文章是根据自己的理解、操作得到的一些有趣的事跟大家分享一下。谢谢这个比赛让我了解到 FME 这个好玩，有趣的家伙。

FME 的介绍：

FME 是读写、储存、转化空间数据。通过提供在转换过程中重构数据的功能,实现了超过 250 种不同空间数据格式(模型)之间的转换,为进行快速、高质量、多需求的数据转换应用提供了高效、可靠的手段。

FME 处理数据的过程是流线型处理过程。通过在 FME 数据输入的阶段执行多个任务,提供了一套可以以不同方式进行组合的构建模块。

FME 的数据整合。在单个数据转换过程中合并多个数据集,所输入的数据可以是一个或者多个不同的输入格式。

FME 语义转化技术及流程：

FME 语义转换是在数据格式转换的范畴里,但有区别于传统的数据格式转换.语义转换是通过在转换过程中重新构造数据,利用语义文件映射提取源数据的不同层面和内容,来满足最终用

户或系统的需求。FME 语义转换利用可重新定义输入或输出数据的引擎(基于强大的数据模型),来保持内部一致性和可扩展性。输入、输出格式或系统的数据模型映射到该引擎不再需要满足最低公用标准的要求,改变了传统格式转换应用程序对各自系统的种种限制要求。

FME空间数据转换采用语义映射文件来控制整个转换的流程,基于 FME 平台的语义转换流程,如下图 1。图中的语义映射文件数据库实际上是空间数据宿主软件的要素编码对照表,而语义映射文件,是数据转换需遵循的一系列规则,在数据转换过程中,由这些规则驱动所有的 FME 模数操作.FME 语义映射文件拥有自己的宏变量和函数.通过编制语义映射文件,用户可以灵活地控制诸如几何数据变换和坐标系统变换等复杂的数据转换操作。

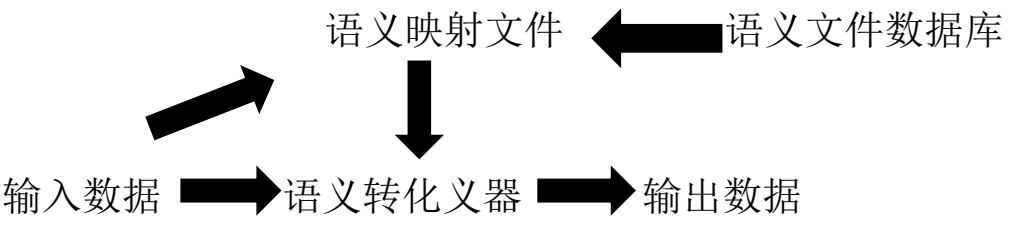


图 1

语义映射文件可以重复使用,并可以进行批量处理,从而为大批量的数据转换供了诸多方便语义映射文件的编制是基于FME平台进行空间数据语义转换的关键,以下为语义映射文件的典型结构图,如图 2。

读写模式配置：

要素厂 1

要素厂 2

.

.

.

⬇

要素厂 n

转化模式配置：

要素定义声明

要素关联规则

图 2FME 映射文件定义结构

Excel 与 FME 之间有趣事：

如果只用到一个或者几个工作表，使用 Excel 就足够了，没有使用 FME 处理的必要了。但是如果需要大量的数据需要处理，还需要跨多个表格，FME 就非常有价值了。手动复制和粘贴和手工输入都会增加风险。FME 的自动化工程能执行 Excel 一样的功能，例如过滤、排序、条件、测试等，能够保护数据的完整性和节省时间。

在 FME 工作中，阅读其文档时，等式关系如下:

Feature Type=Worksheet or Named Range

要素类=工作表格或者命名

Attribute=Column

属性=列

Feature=Row

要素=行

Dataset=Excel

数据集=Excel 文件

打开 Excel 文件

用 FME 中“Add Reader”工具打开 Excel 文件，在 Add Reader 对话框中，选择“Microsoft Excel”作为格式并找到所需要的文件。并用表格来导入和检查数据。

在 sheets to Read 选择要导入的工作表，并在 Attribute 中设置数据类型，设置读模板来识别和保存公式和超链接。

单击“Formatting”下相应的空格，为列中的字体和单元格设置样式元素。

创建 Excel 文件

FME 可以把 Workbench 创建的工作表或数据写到现有的 Excel 模板，不过要先用“Add Writer” 工具添加一个 Excel 模块编写器，然后将格式（Format）设置为“Microsoft Excel”并指定指出数据被写入的位置和文件的名称，在写入数据到 Excel 模板时，要选择模板文件作为目标文件。

写模板到 Writer 添加到工作区后，打开其对话框列在

“User Attributes”部分中创建，设置“Attribute Definition”为“Manual”然后输入最终的电子表格显示的名称，同时指定数据类型，当工作流程运行时，这些列将被具有同名的属性的值来进行填充。

需要向模板写入数据时需要有一个 Excel 文件，其数据需要不时刷新来反映变化，这对在将原始数据汇总成有意义的图表和报表时非常有帮助的。在 FME Writer 的 parameter 对话框中 template 中指定保留结果布局定义的表格作为模板。设置“Overwriter Existing File”为 Yes 就会被覆盖而不是被追加。除了这些亲身大致已经操作过的，FME 和 Excel 之间更复杂的关系还需要慢慢探索，FME 和 Excel 之间更多有趣的事，我要慢慢去学习，发现。

FME 与交通大数据案例：

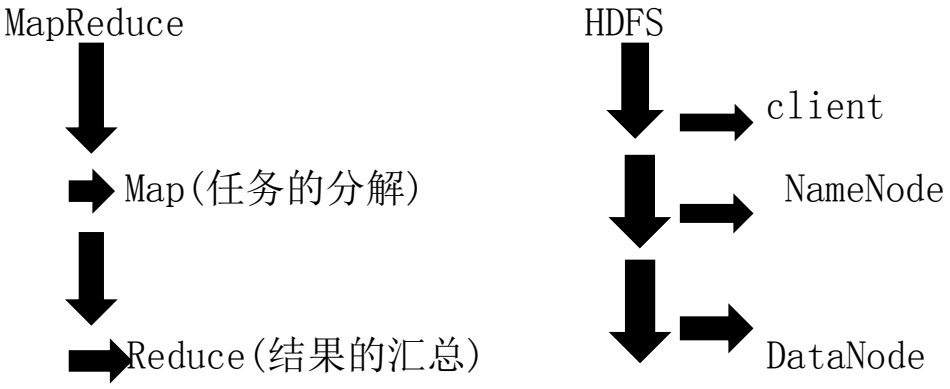
大数据面前无能为力时的状态。某市一个月的交通大数据，将近五六十 G 的出租车轨迹数据，但不包括地铁和公交刷卡数据。这两个数据加起来转成最简单的文件也有近上百 G 的数据量。如果你没有大平台真的没法处理。FME 可以做交通规划分析也可以做行为分析。可以知道某市的某个出租车司机住在什么地方，住在什么地方，有空间分布模式。还可以分析出哪一条路是最赚钱的，每一个出租车路径都知道，提取出来了，花了多长时间，走了多少路程。有了这些后就可以分析出哪些路程是最赚钱的，哪一个时间段是赚钱的。但首先要解决数据处理的问题。

这个数据量非常大，对于这个数据，我们处理之后在 Hadoop 大平台上（图三）处理。整个架构是这样的，相当于多个客户端同时扫，一共有 20 个节点，没有这个东西是处理不了这数据的。

Hadoop 的两大核心设计

Apache Hadoop 是一个用 java 语言实现的软件框架，由大量的计算机组成的集群中运行海量数据的分布式运算。可以让应用程序支持上上千个节点和 PB 的数据。

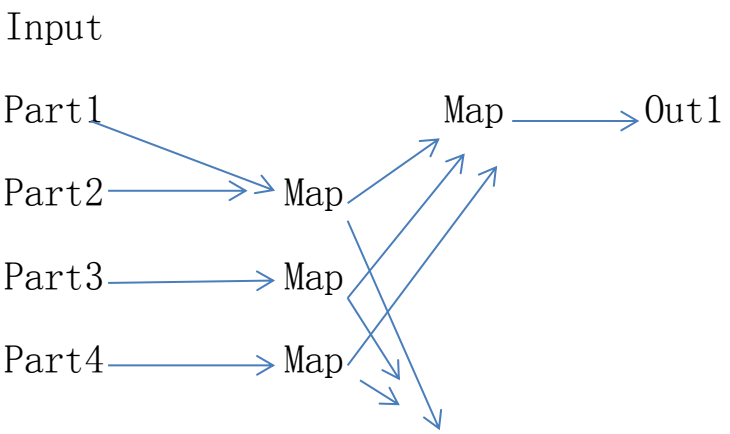
Hadoop 的两大核心设计



图三

MapReduce 是一种编程模型用于大规模数据集的运算，核心思想：“Map（映射）”和“Reduce(化简)”

MapReduced 的工作过程



Part5 ↗ Map → Out2

图四

Hadoop 分布式文件系统（HDFS）

	客户箱 1			DataNode	DataNode
数据 →	客户箱 2	→	NameNode →	DataNode	DataNode
	客户箱 3			DataNode	DataNode

图五

NameNode:分布式文件系统的管理者。

DataNode:文件存储的基本单元。

Client:需要获取分布式系统文件的应用程序。

首先要用 Hadoop 这种分布式计算做出来，从大数据转换为小数据后，关注的数据节点要从不同的尺度分析。不管是公交、出租车还是地铁，是要分不同尺度。

最后用 FME，从数据处理分析到最后的可视化，要建模，把思考用都用图的形式建模出来。

做了一个最简单的，做深入分析前要做一个可视化。比如假设从某市地铁刷卡出去，从早上六点到晚上十二点，六点到七点 1.8 万人，七点到八点增到 15 万，八点三十分这个节点增到 85-110 万，后面就更多了。这只是一天的地铁数据，后面还有

公交刷卡数据，还有出租车量加起来更大。

大致了解一下趋势是怎样的，做了简单的可视化，了解规律和过程，然后再做深入分析。在这个过程中，FME+ArcGIS 对它建成一个模型，建完模型后把数据输入进去，自动会算出来。这样就可以把一个星期做出来，同一天不同的时间，做横向和纵向对比，然后发现某种趋势再做深入分析。

出租车一天 24 个小时，分 24 个节点，168 个图层，一个一个做肯定要累死，就算很厉害的程序员，能不写程序尽量不写。这就需要建模了，FME 可视化建模就可以，包括爬取数据，不需要写程序，配置一下就好了。不管是数据量大还是数据量小。海量数据先用 Hadoop 处理，处理完就变成小数据了。对于小数据，就要建模，做分析，做可视化。

结束语：

自己的亲身经历和感受：应该把 FME 这东西学好，学好之后处理数据更会得心应手，学习 FME 是很有趣的事。以上论述是自己问老师、同学和自己去图书馆，网上学习后写的感受，如有缺点，请多多包涵多多指教。