# 工作流综述

## 前言：

工作流也已经接触了很久一阵了。下面将对已有的工作流软件做一个综述。希望对后来者有一个借鉴的作用。

## 分类：

工作流大致分为两种，科学工作流和事物工作流。

举个例子：科学工作流的核心是数据的处理，一个最基础的模型就是从数据库或某个数据源中拿到数据然后经过各种处理，然后给人展示什么的。一个事物工作流的最基础的模型就是，一个新员工入职，需要多个部门联动，比如获得新人的信息，接受部门安排地方等等，这需要多个人联动协作，他们通过工作来知晓自己应该干什么。

然后再来具体讲一下都有哪些工作流。

事物工作流主要的软件有：K2,H3（这两个事商业版的，好像会收费的。）acitviti和JBPM（这两个好像有一定的渊源，不过可以理解为一个JAVA项目）

Activiti 是能运行BPMN2.0 描述的业务流程的开源工作流引擎。Activiti 是由原JBPM 两个核心开发人员Tom Baeyens 和Joram Barrez 离开Red Hat 加入Alfresco公司之后开发的。Activiti 第一个版本是Activiti5, 它继承了JBPM4 的所有优秀特性，其最大的优势是采用了PVM（流程虚拟机），而且与外部服务有很好的集成能力。

JBPM 是JBoss 公司发布的支持BPMN2.0 标准的开源工作流引擎。JBPM 的核心是一个轻量级的，可扩展的纯java 开发的工作流引擎，它可以执行以BPMN2.0 描述的业务流程，可以在任何java 环境中运行，也可以嵌入到应用中或者作为一个服务来使用。2017 年7 月JBPM 发布了最新版本JBPM7，JBPM7 在历史版本的基础上添加了一些新的特性，如具有案例管理能力、提供流程和任务管理API 等。

科学工作流有：Taverna ，Kepler。

Taverna是一款用于支持计算模型工作流平台，她可以支持复杂的，基于服务的数据密集型自动化计算过程。而且它已成功地应用于生物信息学、天文学、医学甚至音乐的研究。目前在生命科学中的应用最为广泛。Taverna的第一个版本在2004推出，并一直享有广泛的使用，这主要归功于Taverna提供了相当直观的模型的服务组合，还有越来越多的服务，这些服务主要是由Taverna社区提供的，并且免费使用。

Kepler是由美国国家科学基金资助，有伯克利大学和圣地亚哥超级计算中心联合开发的基于JAVA的科学工作流管理系统，其目标在于提供给科学家一个开源的科学工作流管理系统以帮助科学家进行流程设计，并在网络资源上进行调度和执行，达到提高工作效率的目标。Kepler在Ptolemy II的基础上开发而来，继承了Ptolemy II面向角色建模的特点，能在单个工作流中组合不同的计算模型。Kepler利用内嵌的秉性控制和工作流调度机制，讲科学工作流的设计，执行，运行时的交互，本地和远程的数据访问，本地和远程的服务调度相结合。

还有一类工具和科学工作流非常相似，叫数据分析工具。如果要仔细讨论一下科学工作流和数据分析工具间的区别的话，科学工作流的定位更像是为非计算机专业的科学家提供一定的帮助。而数据分析工具存在的目的是为了让数据分析人员更加专注于数据的分析，而不是编程上的琐碎。所以在数据分析工具中会提供很对分析工具，让数据分析人员更加专注于自己的分析思路。

所以在数据分析工具中，目前比较主流的工具有：KNIME, RapidMiner。他们更多的是注重了数据的处理。并且也都提供了十分丰富的工具。

KNIME最初的设计目标是建立一个智能，友好的界面，集数据处理，数据转换，数据分析和数据调查于一体的数据挖掘平台，目前也被用于一些场合的科学工作流建设。KNIME提供了一个GUI界面供用户创建数据流或数据通道，以及工作流，并运行KNIME中提供的分析步骤，并对分析结果进行可视化和交互式处理。

RapidMiner是一个机器学习和过程挖掘的图形化分析软件。它是开源，免费的并基于java实现。它的设计理念和KNIME类似，对于任何一个复杂的数据分析问题，可以用若干的模块化的分析工具，通过组合嵌套等方式来实现。RapidMiner采用XML来描述算子树模型的知识发现的过程。RapidMiner具有数据的输入、输出灵活的特点，它可以兼容不同的文件格式。RapidMiner目前包含了超过100种的分类、回归、聚类模块。

## 我接触过的工作流：

工作流作为一个辅助性的工具，他们有个共同的特点就是都是基于JAVA开发的。下面我将仔细介绍一下我接触过的工作流，以及使用心得。

### Kepler工作流：

这个工作流目前好像并没有处于继续开发的状态中，我猜测已经停止继续开发了。官网上提供了一个十分详尽的说明文档。Kepler工作流的特点是他融入了一个导演和角色的概念，导演决定了这个工作流是怎么运行的，而演员是每一个部件，执行自己的任务。所以在开发的时候导演就变得非常难以设计的地方。另外，在在面对一些工作流的需求时变得非常难用。举个例子：操作等待超时后启动另一个活动。最关键的一点就是你只能依赖那个唯一的说明文档，在网上查不到一点资料。

### KNIME：

KNIME数据分析工具和kepler工作流一样，提供了一个可视化的编程环境，可以很方便的对数据进行操作，但是其实在流程的管理上是几乎没有的，而且因为他设计的初衷就是为了做数据分析，所以每一步都需要对你的数据进行处理，但其实在我的需求中，每一步都可能是执行一些命令，所以就需要在这个流程中一直考虑数据的处理，特别是在调用一些外部工具时，就会觉得这个更加难用。这个工具的有点事有中文的说明，因为它目前还处于开发过程中，它提供了插件的功能，你可以根据需求不断增加插件。不过比较坑的一点就是同样没有资料。但因为中文比较多，其实很多的时候理解起来还是挺快的。

### Activiti 商业工作流：

首先我不得不承认，我并没有像其他人那样去开发工作流，因为Activiti商业工作流的准确的名称是Activiti 工作流引擎，你可以理解为一个框架，所以Activiti只注重于流程的设计，至于每一步要完成什么样的工作，这部分完全是需要人手工编程实现的，也就是说这个需要一定的编程工作。Activiti工作流引擎本质上是在数据库中建立N多张表来控制各个工作流的进度的。所以他有很强的适用性。

首先要表扬一个优点，那就是网上资料非常多，官方也提供了一份非常详尽的英文文档，所以在使用上会很容易，只不过因为要自己开发所以麻烦些。一般而言，Activiti都用来作为OA系统的后端，帮助管理员更加系统的管理整个系统的运转。所以一般Activiti都要和spring相结合，因为我的项目不需要前端，所以我也没有做和spring相结合的东西。

## 总结：

工作流我都用了一圈了，为了让后人少走弯路。我总结一下。

首先，科学工作流应该不会再你的考虑范围内了。因为这玩意主要是给其他学科的科学家使用的。而且随着计算机的普及，科学家对这些的依赖好像也不是很强了。目前没有几个工作流还在更新了。

可以考虑的是数据可视化分析工具，和商业工作流。

数据可视化工具也像工作流一样是通过连线完成编程的，这个的优点主要体现在你对于数据的操纵会十分方便，你有更多的精力放在数据处理后的结果上。而免去了复杂的编程内容。

商业工作流主要用于处理事件的并发，在多个事件并发的时候，商业工作流可以清晰的告诉你都有哪些工作流启动了，他们都在进行到哪一步，以及目前这步如果有负责人的话这个负责人是谁。所以这两个类软件的用途的主要区别在于两点：一是否有大量数据的分析以及操作，二是是否在这个过程中需要有人参与。