



2016中国开源年会

China Open Source Conference 2016

时间：2016年10月15日-16日

地点：北京航空航天大学



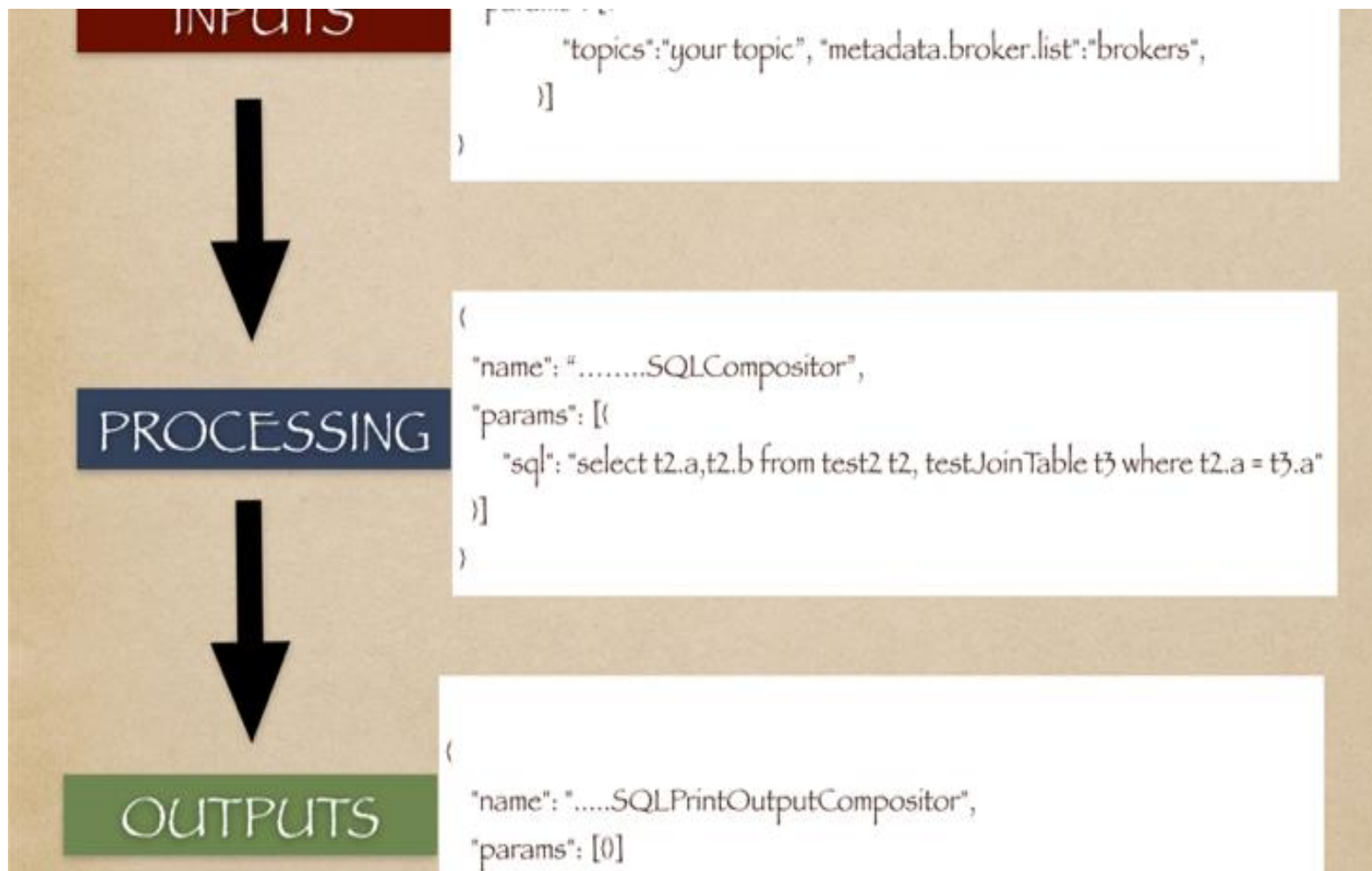
三个五分钟 构建流式，批处理，交互式应用

祝威廉

关于Spark的开发



- 用户友好 (User)
- Contributor 友好 (Developer)
- 做周边难 (third-party project based on Spark)



StreamingPro, 基于Spark 的周边产品，让Spark更易用



StreamingPro目标

- 开发能玩转大数据
- 产品能玩转大数据
- 分析师/算法工程师也能玩转大数据



StreamingPro特点

- 无编码
- 配置化/SQL Graph
- 全平台覆盖(流式/批处理/交互式查询)



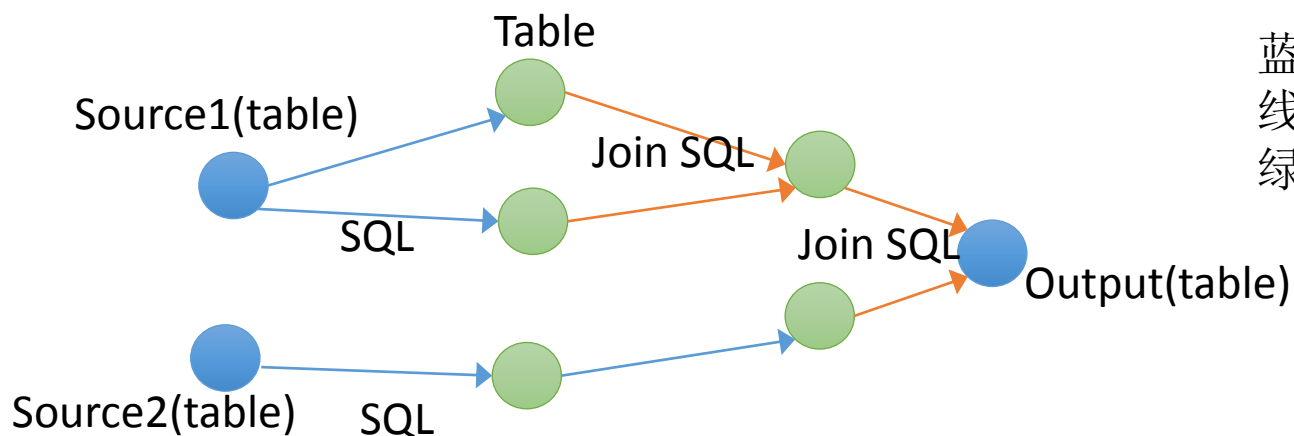
数据处理/查询之 一主二辅

极大降低大数据处理入门门槛，同时兼顾灵活



SQL Graph

—主



蓝圆： Table(输入和输出)
线： SQL 或者Script
绿圆： tempTable



Script/UDM 辅)

(二

- 支持脚本(Scala script),对于复杂文本处理有奇效,可替换SQL,输入输出都是表
- 用户自定义模块,如果现有模块无法满足你,开发自己的模块

实践出真知

好不好用拿出来溜溜



- 构建一个批处理程序
- 构建一个流式程序
- 构建一个即系查询程序
- 构建一个机器学习程序(Bonus)

构建一个流式处理程序



Demo Time <http://www.jianshu.com/p/d10edd6c7cf9>

只要三步



- 写配置
- 提交
- 完

构建一个批处理程序



Demo Time

<https://gist.github.com/allwefantasy/bf28ff336dcaa61370489a8187e397c2>

构建一个交互式查询程序



Demo Time

<https://gist.github.com/allwefantasy/bf28ff336dcaa61370489a8187e397c2>

构建一个批处理程序



Demo Time

<http://www.jianshu.com/p/ab2d34674ffb>

机器学习相关



StreamingPro 将模型训练抽象为：计算+存储
一个典型的ETL过程

StreamingPro 将预测抽象为：UDF函数
也是一个典型的ETL过程

<http://www.jianshu.com/p/ba78244d463d>

StreamingPro 带来的好处



继承了Spark的所有优点

SQL/Script Graph 降低使用门槛，丰富用户群，
降低成本

屏蔽底层Spark的版本

欢迎大家加入



<https://github.com/allwefantasy/streamingpro>

感谢大家



Github用户： [allwefantasy](#)

公众号：祝威廉