



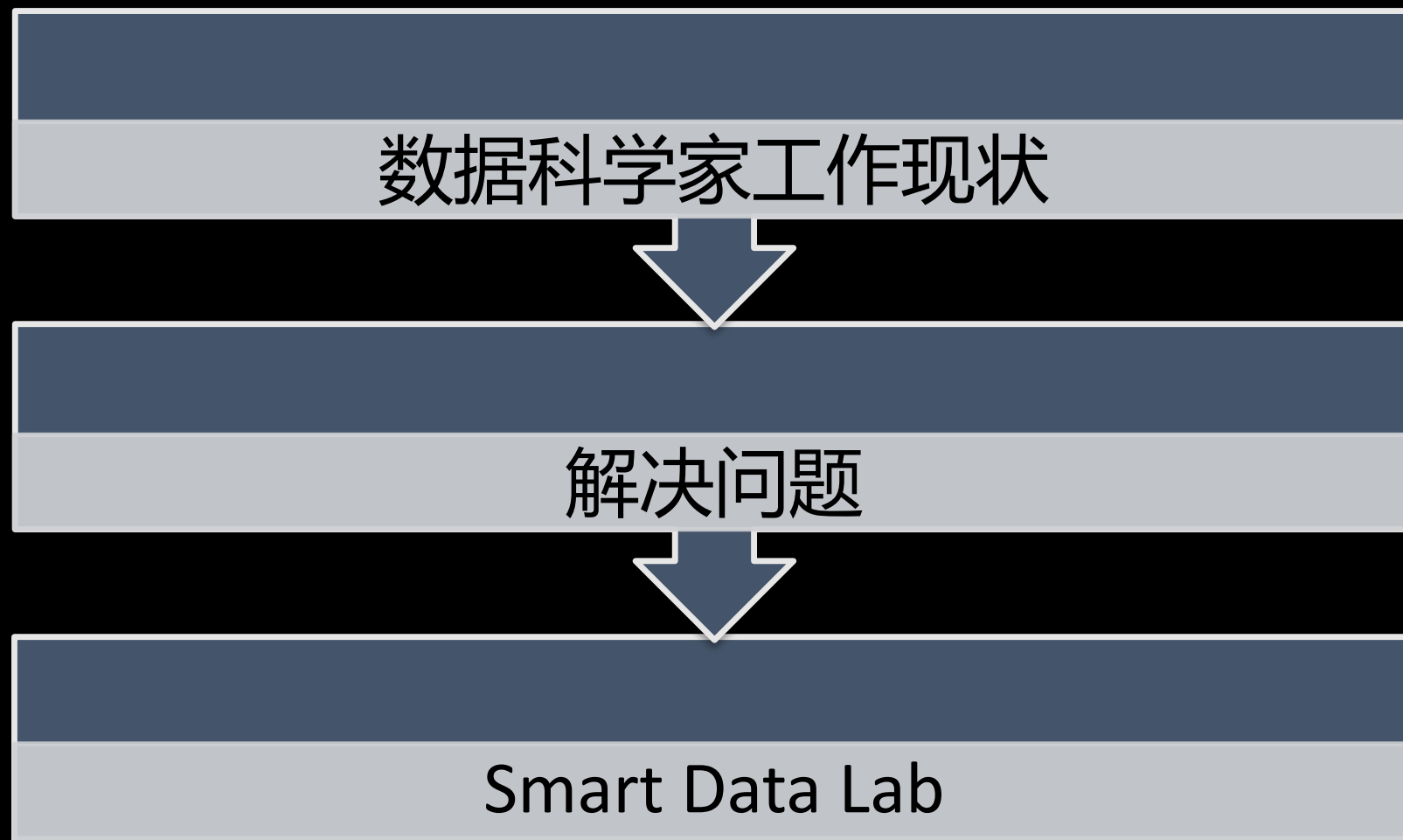
# Smart Data Lab—— 数据科学基础设施搭建的探索与实践

主讲人：TalkingData数据科学中心数据科学家 冯博

# Smart Data Lab

——数据科学基础设施搭建的探索与实践

主讲人：冯博·TalkingData



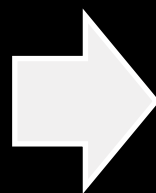


# 现状—数据预处理

查看数据

数据可视化

特征工程



# 现状—调参



运气



耐心



经验

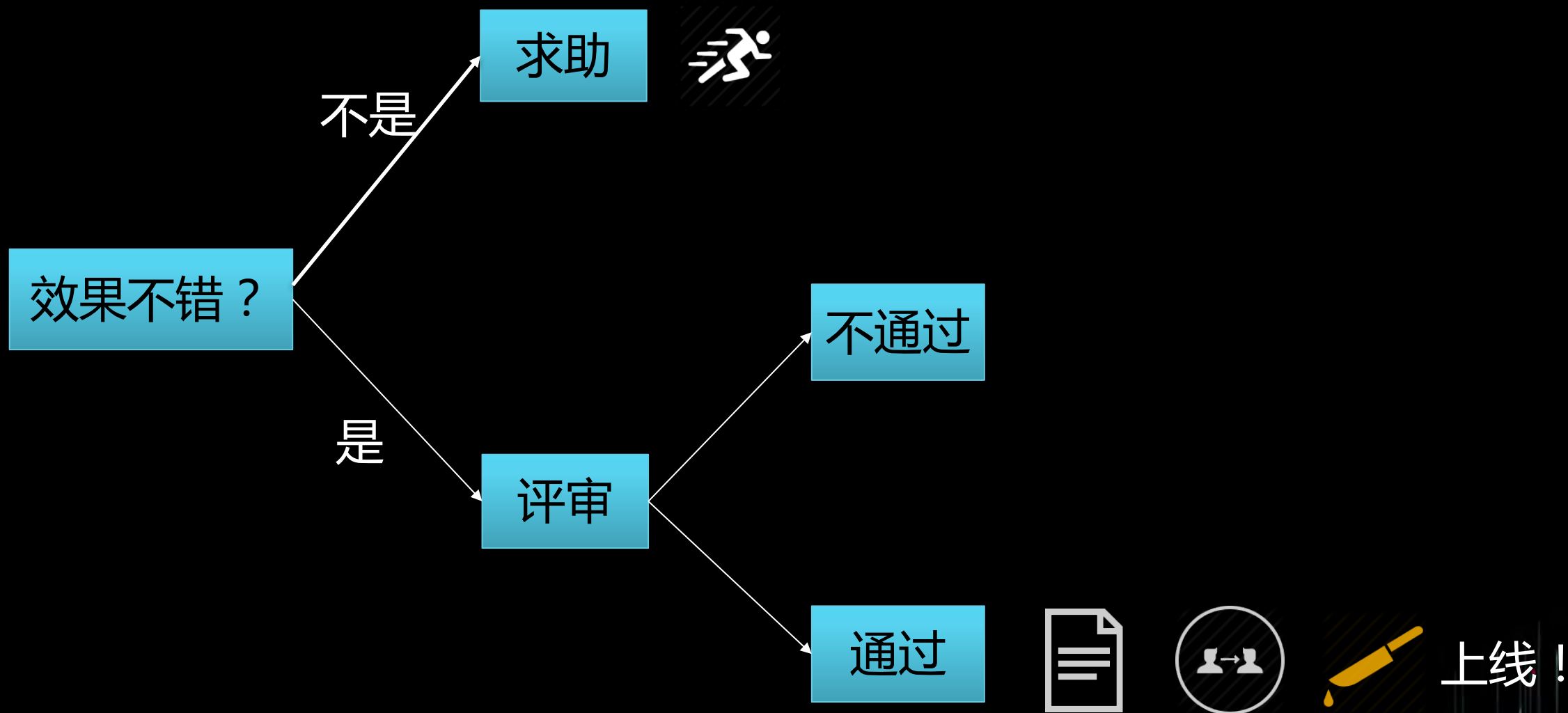


天赋



艺术

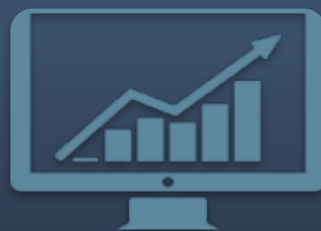
# 现状—交流和上线





# 问题：大数据环境下，数据预处理成本太高

解决：标准化



数据管理：

- 目录浏览 数据预览
- 文件名搜索



数据可视化

- 基本统计信息展示
- 自定义统计



协同

- 数据评论
- 责任人制度

# 问题：工具繁杂，资源分配困难等问题

解决：Smart Resource



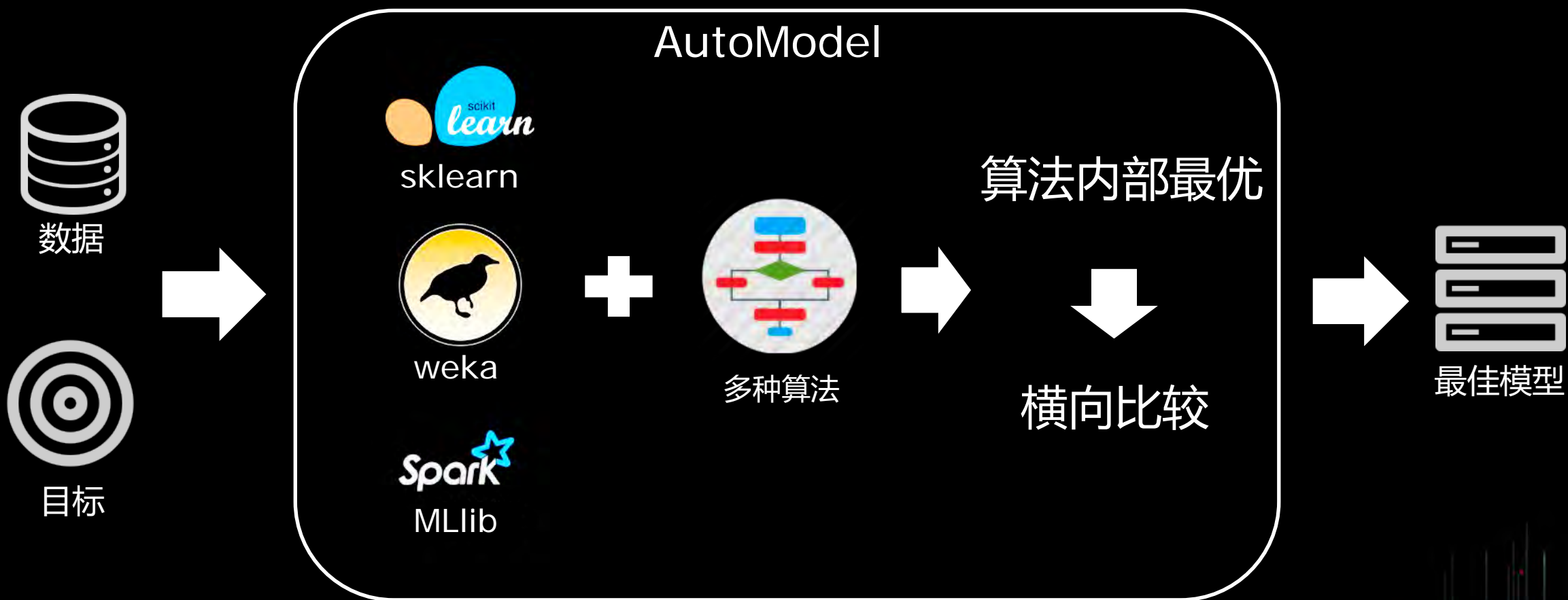
计算资源是黑盒

灵活



# 问题：调参占用太多时间

解决—AutoModel



# 问题：发布流程复杂

解决：一键发布



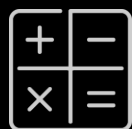
Restful API

# 问题：协作、沟通

解决：社交模块



教程



成果



心得

社交



代码



竞赛

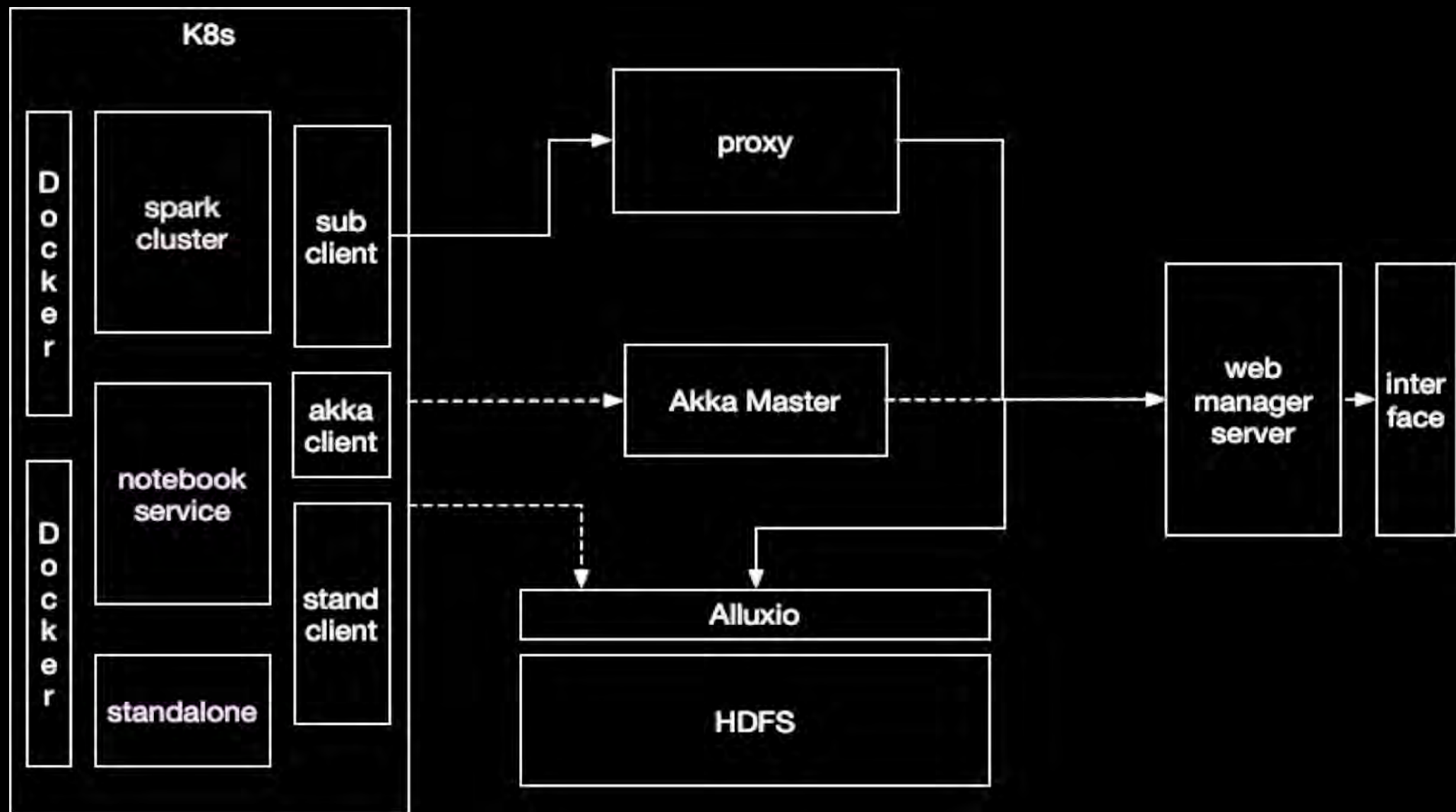


项目

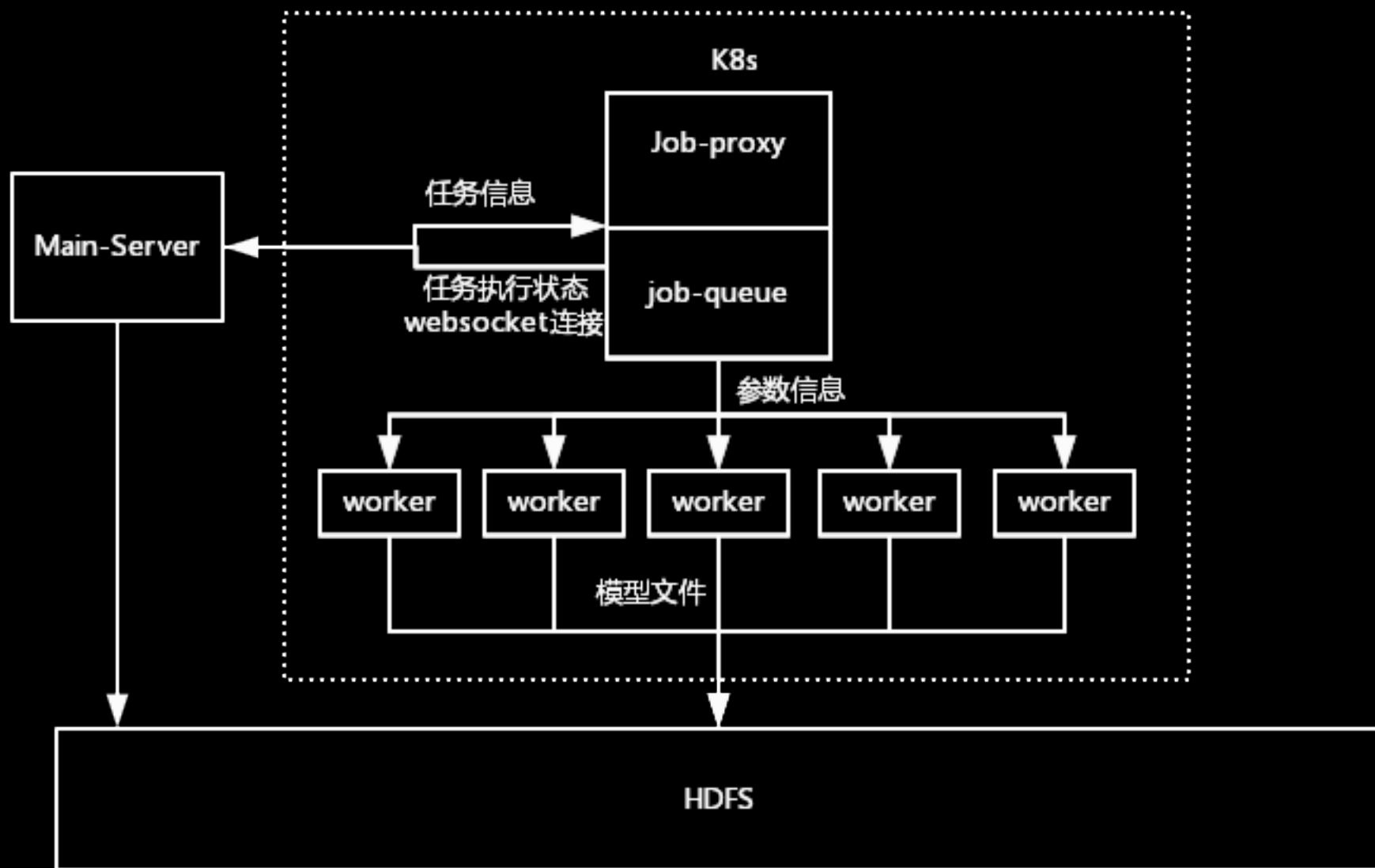
# SDL-功能架构



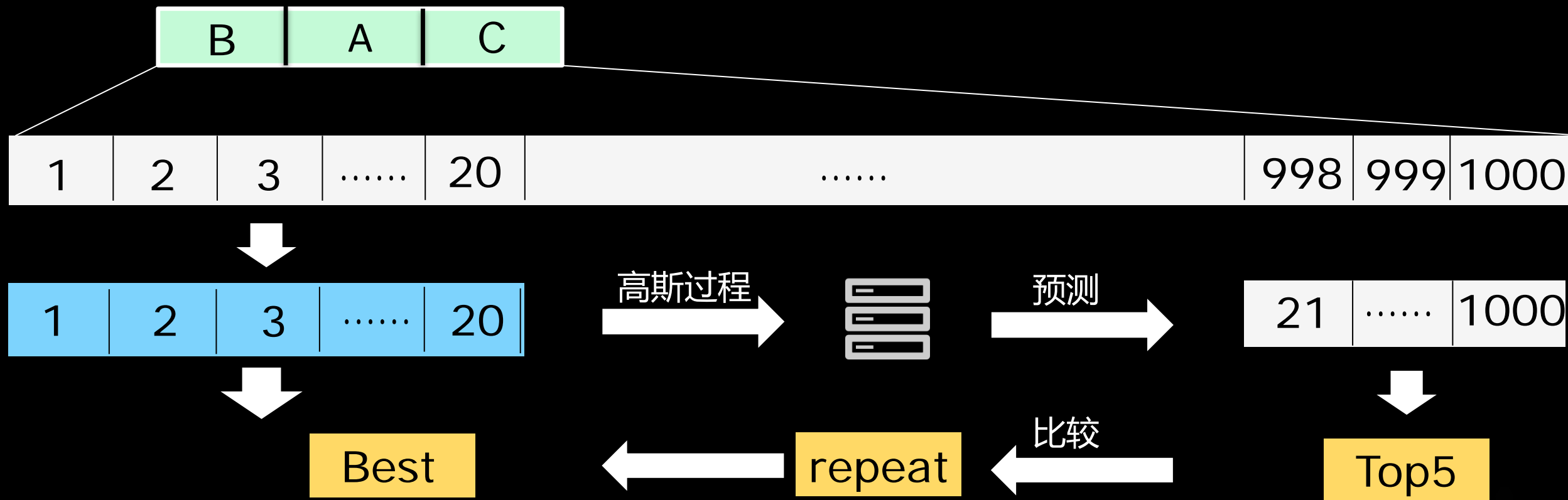
# SDL-系统架构



# SDL—AutoModel架构



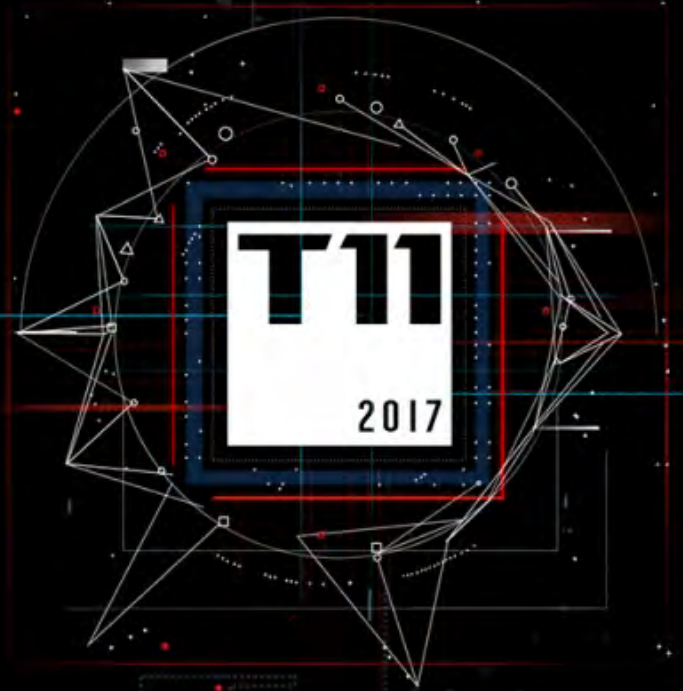
# SDL—AutoModel原理



# SDL——服务用户

- 6家国内大数据公司
- 5家国内知名高校
- TalkingData暑期夏令营
- 全部数据分析师和数据科学家





**THANKS**