# 赛题解析

[实时复杂自定义计算](http://contest.aicubes.cn/#/detail?topicId=96)

* 公司提供一个现成的静态脱敏后的实时+历史行情流数据服务或文件

支持实时数据、历史数据等多种形式的数据源

* 选手设计一套算子管理系统，可支持动态配置大量的复杂算子，支持不停机热更新

算子动态管理、热更新

* 以极速性能，计算出精确结果，并推送给终端用户或公司其他业务系统

极致性能、推送形式多样

* 原始数据组织形式按下述三个维度：1）X个字段; 2）Y只股票；3）Z种周期

抽象算子计算模型

# 调研

|  |  |
| --- | --- |
|  | 参考 |
| 实时计算 定义及现状 | 大数据实时计算：  https://xie.infoq.cn/article/b12ec8271c2f7d5c8c862beed |
| 实时计算 发展阶段及问题 | 实时计算在有赞的实践：  https://www.infoq.cn/article/ZwJ\_cyplQgeueAzTSyDR |
| Flink 介绍及使用 | Flink：  https://flink.apache.org/ |

主流框架存在 依赖多、接入复杂、极致性能等问题

# 设计优势

* 跨平台

支持Windows、Linux 等主流操作系统、可适配麒麟、统信等主流国产系统、支持容器化部署;

* 轻量化

计算引擎AOT编译，自包含环境，脚本一键运行

* 极致性能

基于内存的实时计算、只读形式共享数据内存块

* 动态算子管理

热更新算子、支持动态编译上传

* 接入简单

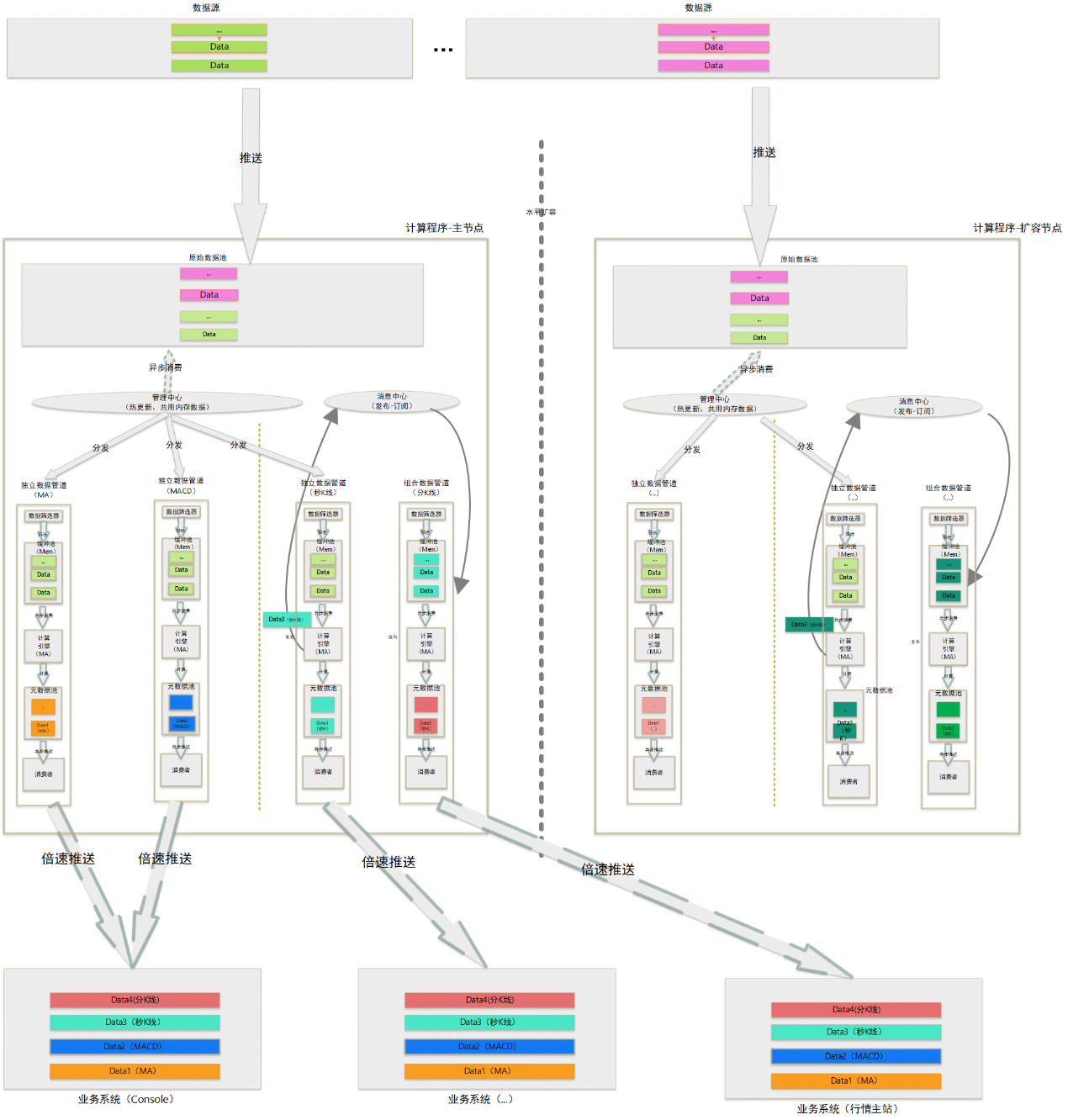
接口化抽象算子流程、支持自定义计算函数和推送规则

* 高可用

算子间域隔离、容器化部署可以动态扩容

# 开发方案

## 抽象数据流



如上图所示，计算程序表示一个独立的算子管理程序，是执行的最小单位。计算程序环境自包含，可以打包成镜像并容器化部署，支持横向、水平扩容。计算程序由以下部分组成：

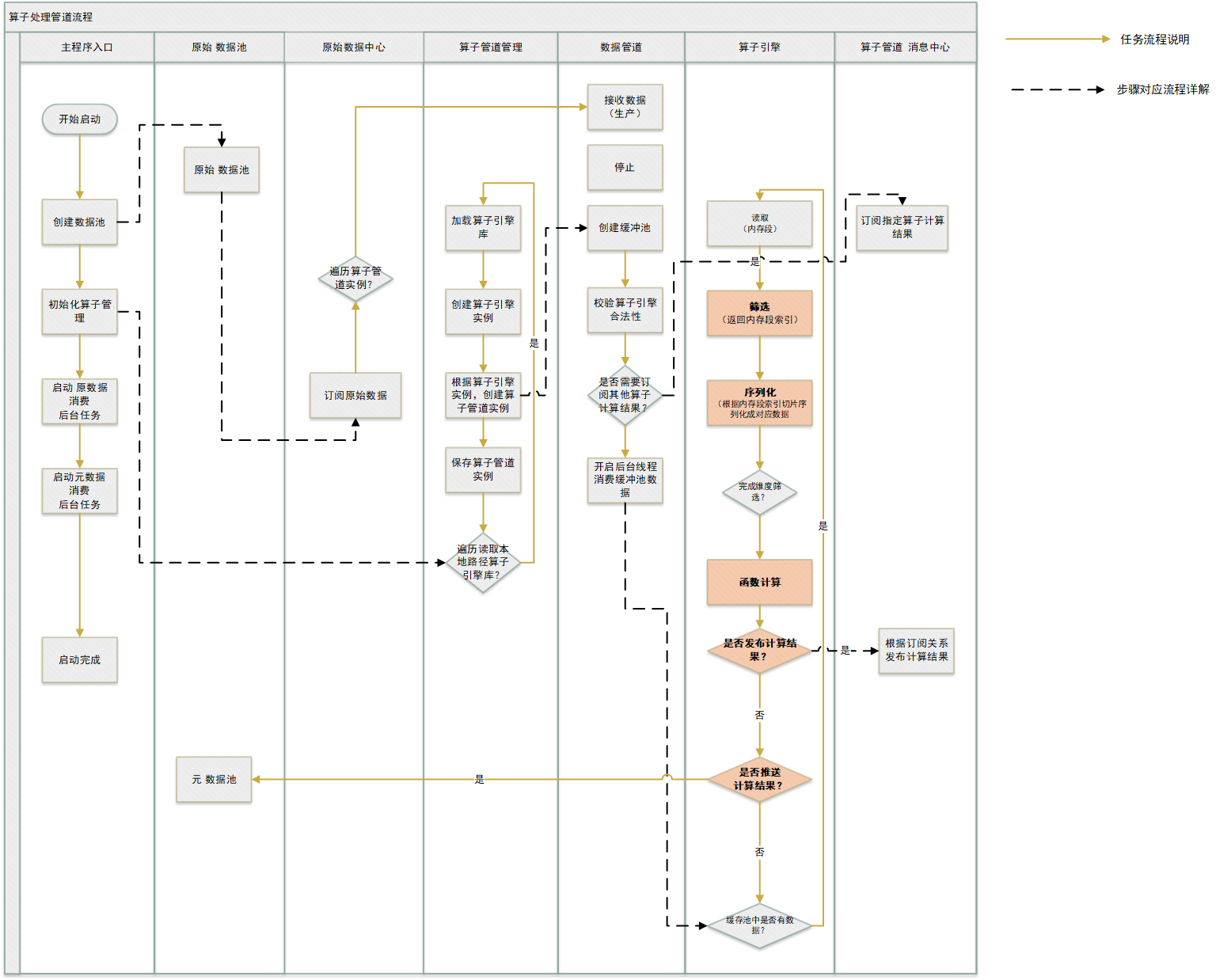
**原始数据池**：基于内存的队列，用于保存数据源的原始数据流，支持历史数据文件、实时数据等多种数据源。原始数据池用于解耦：数据接收---数据计算 两个步骤，保证接收速率；

**算子管理中心**：用于保存算子实例，提供算子增、删、改、查功能。管理中心 会将原始数据池中的数据，以只读内存片段指针的形式将数据分发给每个运行中的算子实例；

**数据管道**：表示一个独立、完整的数据处理流程。有两种数据管道类型：独立数据管道和组合数据管道两类。其中，组合数据管道订阅一个或多个独立数据管道的计算结果。独立数据管道 接收原始数据，经过计算引擎计算出元数据后，推送给业务方或者订阅了此元数据的组合数据管道；

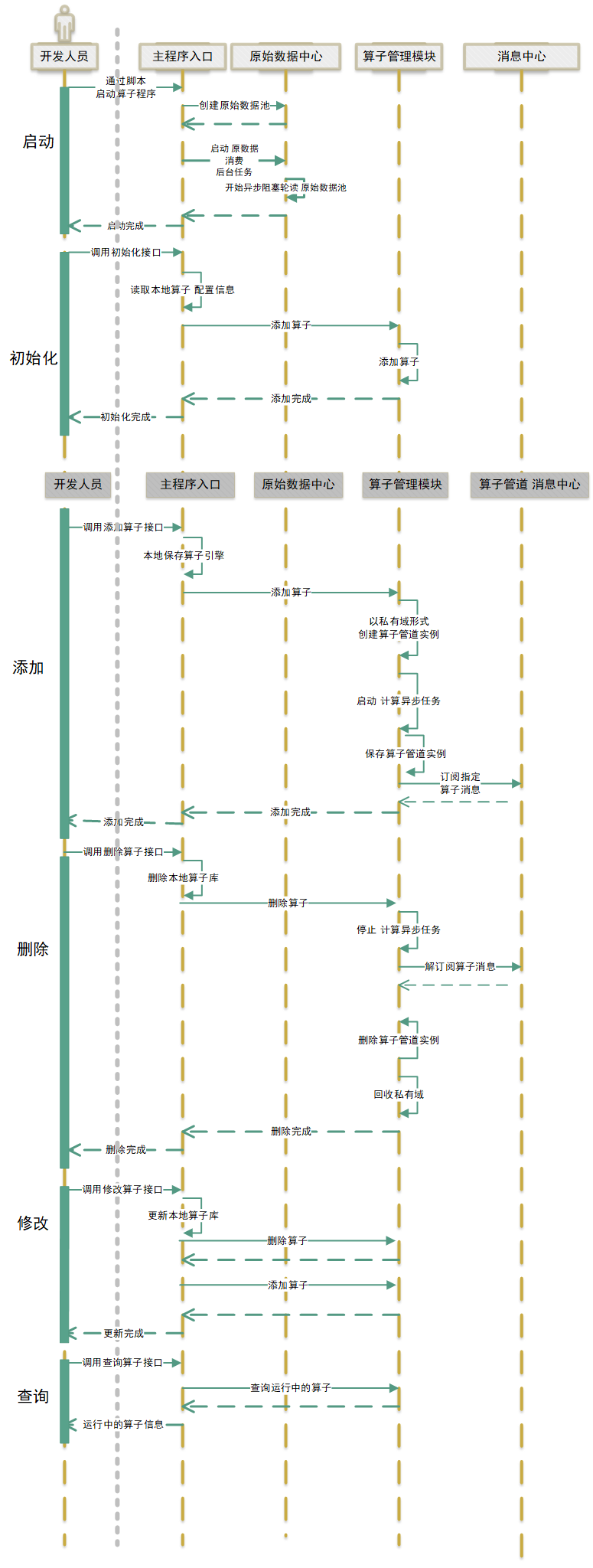
**业务系统**：表示元数据接收方，可以接收一个或者多个数据管道推送的元数据

## 抽象计算流程



如图所示，一个数据管道 包含 一个算子引擎。算子引擎 是数据计算的抽象流程的具体实现，可以根据业务场景 自由定制，并动态编译上传。

## 算子管理时序



如图所示，算子的增、删、改、查功能时序

## 应用架构



# 测试报告

详见: 性能测试 –> 实时计算-性能测试报告

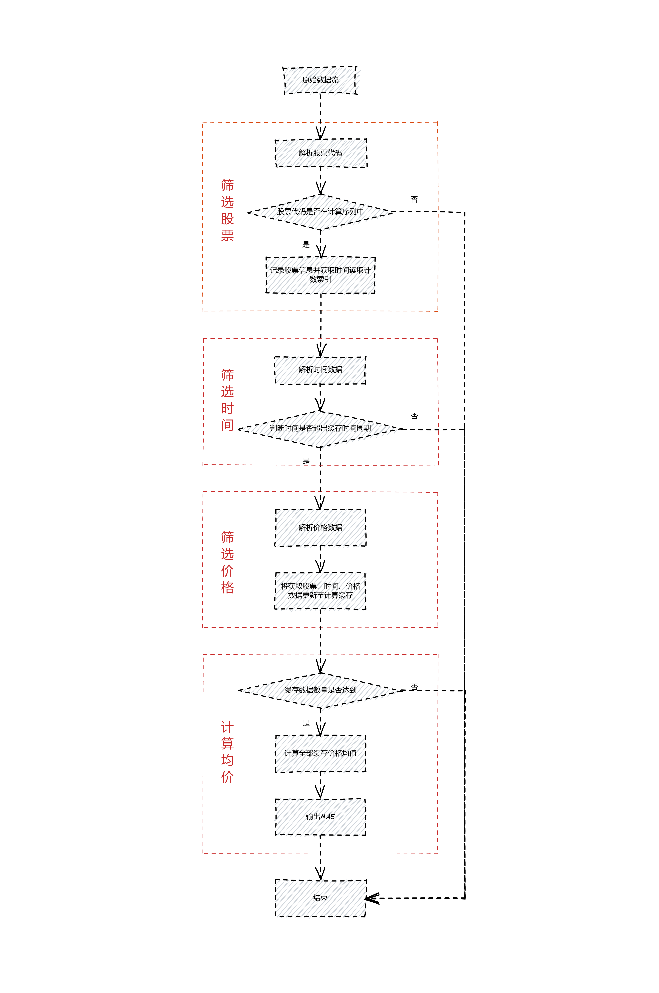
# 算子示例（MA）

## 接入步骤

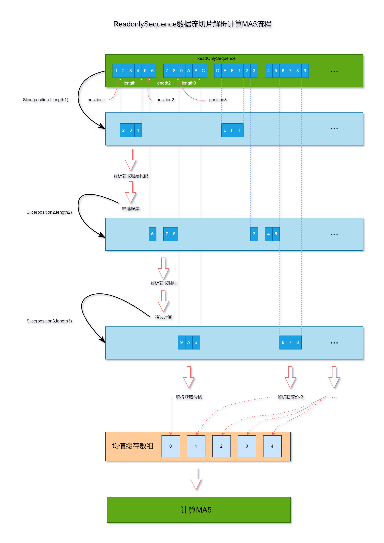
* 1）接入算子引擎SDK：Glink.Component.Abstractions.1.0.0.nupkg
* 2）实现接口：ICalculationEngine.cs
* 3）打包成动态库，并上传：Glink.Components.MA.dll
* 4) 通过HttpApi接口，运行算子引擎

## 算子开发方案

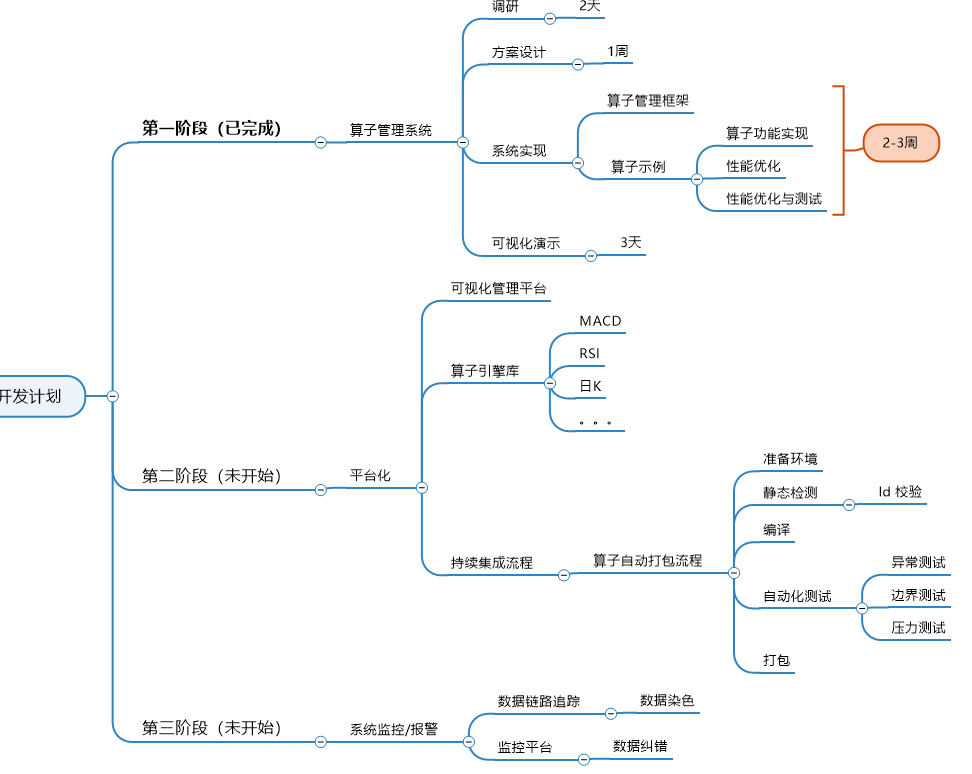
### 计算流程



### 内存序列处理示意图



# 开发计划/后续安排



# 演示

详见: 演示程序 –> 实时计算-演示流程.doc