需求分析

功能需求

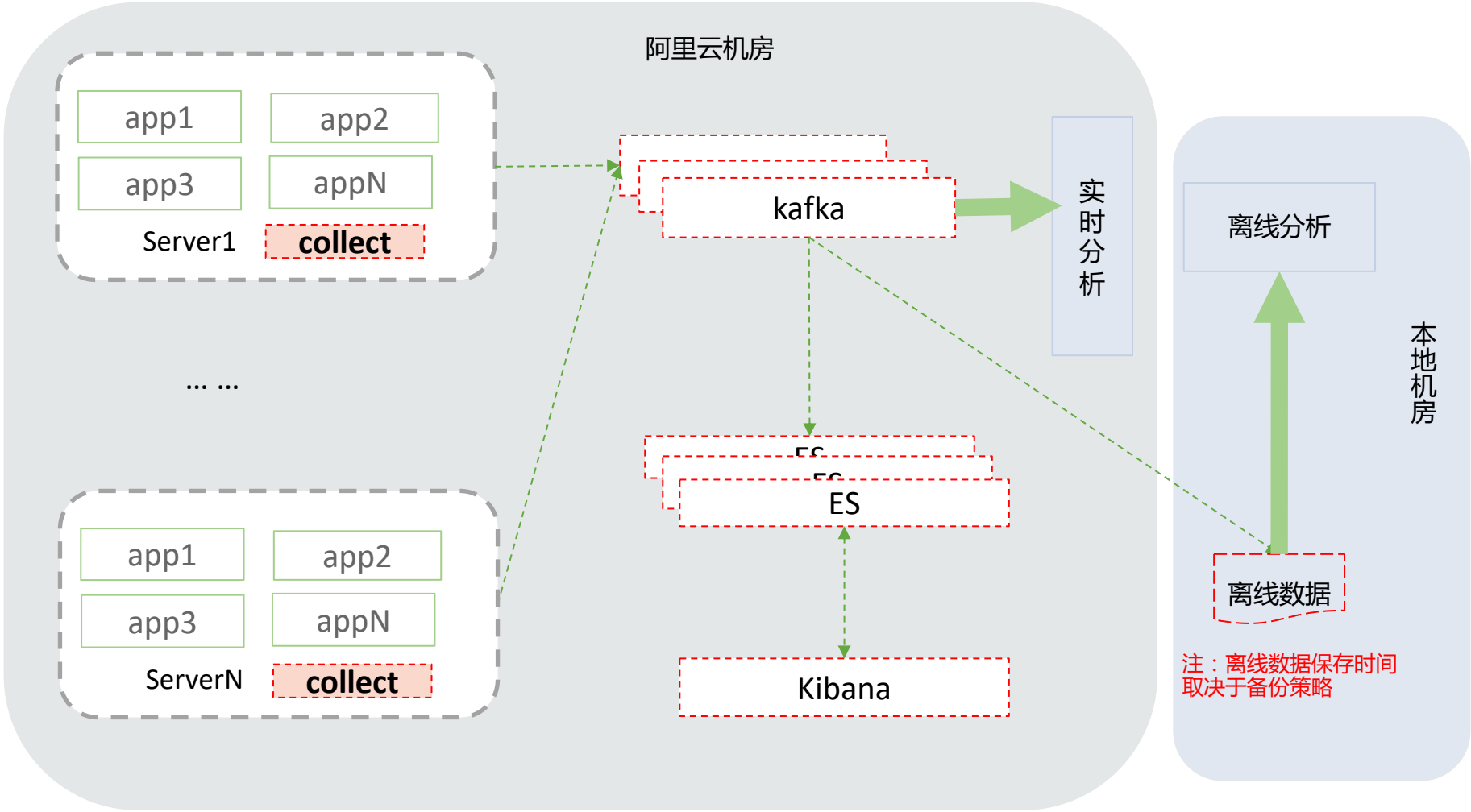
- **收集日志：**
  - 收集应用日志、需要补全信息的，都要补全
  - 支持多种类型日志处理
  - 处理日峰值500G-600G数据时，CPU、内存等系统资源占用不能高
- **消息中间件：**
  - 日志消息高可用
  - 支持消息生产和消费
  - 网络延迟尽可能小
- **查询日志管理平台：**
  - 查询搜索日志信息
- **大数据分析：**
  - 提供各种业务需要的数据报表
  - 支持多种数据源数据获取

非功能需求

- **可扩展性：**
  - 可根据数据和业务水平扩展
- **高可用：**
  - 主服务必须有备用服务
  - 关键日志文件必须有副本
- **低延时：**
  - 同步给大数据部门延时：2-7s
- **健壮性：**
  - 应用服务稳定可靠
  - 出现故障时要快速恢复

需求重点：考虑未来业务发展、使用尽少成本，产生最大价值

技术方案


阿里云机房
app1
app2
app3
appN
Server1
collect
... ...
app1
app2
app3
appN
ServerN
collect
kafka
实时分析
ES
Kibana
离线分析
本地机房
离线数据
注：离线数据保存时间
取决于备份策略

技术方案：概述

功能说明

- **日志收集程序：**
  - 负责收集和处理现有应用程序的日志
  - 将数据实时写入消息中间件（kafka）
- **消息中间件：**
  - 负责消息数据的高可用
  - 支持数据按照主题生产和消费
- **Elastic Search ：**
  - 负责消费日志消息并索引化
  - 日志实时分析搜索
  - 实时文件存储
- **Kibana ：**
  - 搜索，查看，并和存储在Elasticsearch索引中的数据进行交互
- 大数据部门：
  - 负责实时从消息中间件消费日志消息并进行数据分析
  - 支持从离线数据消费日志消息进行数据分析

日志流程说明

- **开发人员：**
  - 必须按照日志格式规范输出日志内容
  - 根据各需求相关方要求记录相应的日志文件中
- **运维人员：**
  - 负责日志数据接入
  - 负责相关服务的运维和管理
- **大数据分析人员：**
  - 关键业务数据以数据流的方式进行实时数据获取分析
  - 离线数据以文件方式获取做离线分析
  - 按照业务方要求生成报表
- **业务人员 ：**
  - 根据业务需求和开发人员确认日志关键字段等
  - 获取大数据部门出具的报表数据