**关于信鸽大数据分析**

1. **调研背景**

**针对目前信鸽行业，以及本地数据量，对于目前的数据进行数据分析**

1. **研究目的**

**目前数据库已有数千万数据量，可以根据这些数据进行数据分析，对公棚或者鸽友进行用户画像，根据不同的用户画像进行不同的活动策略。**

1. **目前应用广泛的大数据框架**

# 1. Apache Hadoop

优点：它提供了经济高效的解决方案，高吞吐量，多语言支持以及与大数据服务中大多数新兴技术的兼容性。还通过出色的故障处理机制支持高可伸缩性，容错能力，更适合研发，高可用性

缺点：它包含安全漏洞的漏洞，不执行内存中的计算，因此会遭受处理开销，不适用于流处理和实时处理，这是处理大量小文件的问题。

# Apache Spark

优点：速度 使用方便 进阶分析 自然动态 多种语言 功能强大 增加对大数据的访问 对Spark开发人员的需求 开源社区

缺点：Spark带来了一些弊端，例如设置和实现的复杂性，语言支持限制，而不是真正的流引擎

# Apache Storm

优点：流计算框架，因此使用的是内存，延迟上有极大的优势

缺点：不会持久化数据

# Samza

优点：低延时、分区、避免回压

缺点：只支持JVM语言，灵活度上不足

# Flink

优点：混合式的计算框架，侧重点在于处理流式数据

缺点：目前不是很成熟

具体应用，应根据实际处理方式，流处理，批处理，图处理，查询处理来确定框架

1. **大数据所需技术**

**数据采集与预处理、数据存储、数据清洗、数据查询分析和数据可视化。**

**五．可应用场景**

**例如，可根据归巢数据进行统计，得到归巢率最高的公棚，然后在进行下一步有针对性的活动，可以统计成绩最好的比赛，根据当时的天气情况，温度等，可以统计哪些因素对赛鸽的影响较大，得到最适合放飞的天气。还可以根据不同地区，不同公棚的收鸽量进行特定计划。**