物联网（英语：Internet of Things，缩写IoT）是互联网、传统电信网等信息承载体，让所有能行使独立功能的普通物体实现互联互通的网络。物联网一般为无线网，而由于每个人周围的设备可以达到一千至五千个，所以物联网可能要包含500兆至一千兆个物体。在物联网上，每个人都可以应用电子标签将真实的物体上网联结，在物联网上都可以查出它们的具体位置。通过物联网可以用中心计算机对机器、设备、人员进行集中管理、控制，也可以对家庭设备、汽车进行遥控，以及搜索位置、防止物品被盗等，类似自动化操控系统，同时透过收集这些小事的数据，最后可以聚集成大数据，包含重新设计道路以减少车祸、都市更新、灾害预测与犯罪防治、流行病控制等等社会的重大改变，实现物和物相联。

分布式系统（distributed system）是建立在网络之上的软件系统。正是因为软件的特性，所以分布式系统具有高度的内聚性和透明性。Spark 是在 Scala 语言中实现的，它将 Scala 用作其应用程序框架。与 Hadoop 不同，Spark 和 Scala 能够紧密集成，其中的 Scala 可以像操作本地集合对象一样轻松地操作分布式数据集。尽管创建 Spark 是为了支持分布式数据集上的迭代作业，但是实际上它是对 Hadoop 的补充，可以在 Hadoop 文件系统中并行运行。通过名为 Mesos 的第三方集群框架可以支持此行为。Spark 由加州大学伯克利分校 AMP 实验室 (Algorithms, Machines, and People Lab) 开发，可用来构建大型的、低延迟的数据分析应用程序。

我们运用物联网收集信息，然后用分布式系统存储和处理信息，最终反馈给我们的终端。在这之前，我们需要做一些可行性分析，来说明我们可以运用分布式系统来处理大量的终端收集的原始数据。注意这个数据可能是百兆千兆级别的，中间存在着大量的数据共享行为，这让我们的可行性分析变得有必要性。图一表示我们的分布式系统在处理单一数据源的时候展现的性能，图二表示我们在处理数据的时候所占内存空间的消耗，图三代表我们将分布式系统和传统系统的性能比较，图四代表两种大数据集下分布式系统的表现。

在可行性分析以后，我们就可以利用当前分布式系统中的数据资源进行一些用户行为分析，这些用户行为分析将实际提供优化。

最后，我们来回顾一下这个大数据战略是怎么样的。首先我们通过物联网来获取数据，然后在庞大的数据压力下，我们用分布式技术来应对这个问题。改变了传统模式下，系统无法应对海量数据的情况。然后我们就可以利用数据进行一些建模和用户行为分析，来进行有效的产出。比如成本和战略优化。

这个战略为物流信息化提供了一个可行解，参考的是顶会上相关的最新论文。