

在集群内，有多种方法可以组织 Flume Agent。第一个且最简单的一个是，部署 Flume Agent 的单个的层，以接收从应用程序服务器来的数据，且用相同的 Agent 直接写入数据到存储系统。这样的系统从存储系统中隔离应用程序的错误，并允许存储系统通过吸纳不断增长的输入速度，处理周期性爆发的数据。每次写入失败时，通过回退增长的时间量（直到一个最大周期），Flume 将调整存储系统的写入速率，如果能力低于处理当前写入速度，不至于使得存储系统崩溃。

单层中 Flume Agent 的数量通常只需要占应用程序服务器总数的一小部分，因为 Flume 的 Avro Source 和 Thrift Source 被设计用来接收大量的服务器的大量数据，尽管每个 Agent 的最大容量取决于具体的硬件部署、网络、延迟需求等。图 2-4 展示了设计为这样的拓扑。

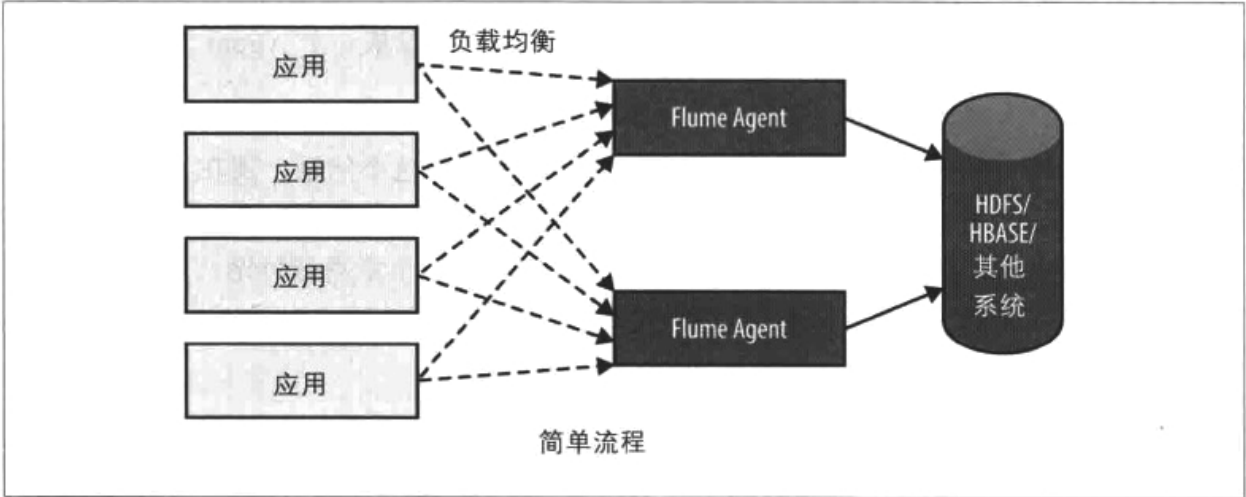


图2-4 使用Flume从大量应用程序服务器将数据聚合到HDFS

19 要设计一个 fan-in 拓扑，需要大量的 Flume Agent 接收产生数据的应用程序的数据，而少量 Agent 将数据写到存储系统。基于有多少服务器生产多少数据，Agent 可以组织到一个、两个或者更多的层，每一层的 Agent 使用 RPC sink-RPC source 的组合，从一层转发数据到另一层。

如图 2-5 所示，最外面的层有最大数量的 Agent 接收应用程序的数据，尽管 Flume Agent 的数量通常只占应用服务器数量的一小部分，但具体的数量取决于多种因素，包括网络、硬件和数据总量。当应用程序产生更多的数据或添加了更多服务器，很容易通过在最外层添加更多的 Agent，并配置它们向第二层的机器写数据来扩展。