



图2-5 fan-in流

通常第二层比最外层扩展 Agent 要慢得多，因为写数据到第二层的服务器数量（最外层的 Flume Agent 的数量）比应用程序服务器数量的增长速度要缓慢得多。这是因为 Flume Agent 的第一层将承受增加应用服务器造成的大部分影响。

这种拓扑结构允许 Flume 根据需要回退，来控制写入到存储系统的速率，同时还允许应用程序无须担心地写入数据。当生产数据的应用程序部署在许多不同的数据中心时，这种拓扑结构也经常使用，并聚合数据到一个集群。通过在同一个数据中心写入到 Flume Agent，应用程序可以避免跨数据中心广域网链接写数据，从而确保数据最终会持久化到存储系统。Flume Agent 之间的通信可以配置为允许更高的跨数据中心的延迟，以确保 agent-to-agent 通信能够没有超时地成功完成。

通过不过度地使用任何一个 Agent 或层，并尽快移除每一层的数据，可以让更多的层承受更长和更大的峰值负荷。因此，层需要的数量主要取决于被推送到 Flume 部署的数据总量。因为最外层接收来自大量机器的数据，那么这层应该有最大数量的 Agent 来扩展网络架构。随着我们进一步讨论 Flume 拓扑结构，Agent 的数量可以明显减少。

20

如果生产数据的服务器数量持续增加，该层从应用程序服务器接收数据的 Flume Agent 数量也需要增加。这意味着在某种程度上，它可能需要增加后续层，尽管后续层上 Agent 的数量可以比外部层增加的速度更慢。这也确保在 Flume 设置内能增加缓冲容量，以适应数据产量的增加。图 2-6 展示了这样一个流的例子。