



部署到工厂的边缘设备需要自己部署：arm盒子或者在工厂部署linux小服务器，这个盒子要在网络上与工厂的传感器设备是互通的（如果是实验室环境，初期你可以使用安装arm版本的ubuntu模拟边缘设备）

盒子端手机传感器的代码可以使用docker swarm/docker compose部署本地api，手机api数据，盒子作为边缘端，根据每天/每小时收集的数据量大小（如果是手机频谱/传感器温度/转速的传感器）每天的数据量非常庞大（几百~上千条/每秒）

这部分边缘端部署的“机器学习预测分类算法KNN之类”过滤出来一部分异常数据存在本地，同时通过消息队列kafka或者RabbitMQ 同步到总部的 hadoop集群，

总部的hadoop集群可以根据工厂边缘端手机的设备数据在hbase中构建关系表，提取分析使用hbase生成的map-reduce，然后对hadoop中dfs的传感器收集的数据进行故障分析与预测（spark streaming）

就是本地一份异常传感器数据集，同时云端一份。边缘端数据分析主要是简单过滤和聚类，预测和整体数据的分析在云端。上传和下发主要通过docker的代码镜像，数据的上传下发主要通过消息队列

亮点可能就是：边缘端自己通过算法可以减少需要进行副本保存的数据量，同时保持了实时性和可用性。