服务要有容错设计，为失败而设计。

服务主要异常场景：
(1)服务内部出错、异常；
(2)服务处理延迟；
(3)服务处理过载；
(4)网络链路延迟或中断；
(5)服务依赖链中部分依赖SLA不达标，造成整体服务不可用；
(6)服务链条过长，造成SLA整体不可控；

解决的思路：隔离（物理或逻辑）、自我保护、失效转移或恢复、降级。
1、隔离手段：依据服务重要性分级或流量特点、用户画像等，从物理上隔离服 务。主要使用分流技术；将服务使用的资源（CPU、线程、IO等）隔离，主要使用舱壁模式；
2、自我保护手段：快速失败(failfast)、流控、超时、熔断；
3、失效转移或恢复手段：失效检测、重试、转移(failover)、回退恢复（failback)；
4、降级手段：依据依赖服务的重要性或依赖程度（强、弱）,同步变异步，降级开关、拒绝部分服务等。

降级方案、限流方案设计如下：
1. 每个系统需要分析调用量前10的服务（URL、ESB、RSF服务），并综合考虑其响应时间和耗时。原则上所有调用量大且降级后对销售或作业不造成较大影响的服务，都需要考虑。
2. 降级、限流的目的是保护系统，减少本系统的压力、或降低对后端系统的压力、或降低对网络的压力。
3. 限流方案不能造成正常销售或作业执行工作，降级后不能对销售造成较大的影响，对销售的较小影响是可以接受的。先限流，再降级。
4. 降级手段有：功能禁用、增加功能的缓存时间、使用本地缓存而不是调用外部服务、减少某些业务特性以降低业务复杂度、不调用后端依赖服务、异常时采用默认数据或兜底数据，同步变异步调用，减少JOB执行频率或变更业务峰值JOB触发调用时间等；限流手段有：随机拒绝请求、拒绝低优先级系统调用，拒绝低级别用户调用，根据白名单或黑名单规则拒绝特定用户请求调用，对失败率高或响应超时系统调用拒绝调用，利用线程池队列排队处理调用，拒绝超出处理能力调用等。

5. OLAP应用，如对OLTP系统的物理机器或网络资源造成了争用，同样需要设计降级方案。