计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2020,29(11):87−91 [doi: 10.15888/j.cnki.csa.007502]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

运营商网络监控系统高可用性设计及应用①

吴　舸, 袁守正, 孙　鼎

(中国电信上海理想信息产业 (集团) 有限公司, 上海 201315)
通讯作者: 吴　舸, E-mail: wuge.sh@chinatelecom.cn

摘　要: 中国电信 NetCare 服务, 是中国电信统一建立面向企业客户的云网一体化监控平台 (监控对象包括客户端网络设备、云端应用及虚拟资源、专线及互联网线路), 为企业提供网络层、传输层以及应用层在内的监控管理和分析的业务. 该业务为中国电信政企客户提供服务, 对于系统的高可用性有着很高的要求. 影响一个系统的高可用性有硬件、网络、操作系统、数据库、中间件、应用本身等多个方面, 本文着重讨论 NetCare 系统应用层的高可用性实践, 通过实测及合理估算数据计算得出系统的整体可用性: 不包含异地灾备系统时, 系统可用性为 99.9978%; 包含异地灾备系统时, 系统可用性为 99.45%.

关键词: 监控系统高可用性; 应用层高可用性; 可用性影响分析; 可用性计算

引用格式: 吴舸,袁守正,孙鼎.运营商网络监控系统高可用性设计及应用.计算机系统应用,2020,29(11):87–91. http://www.c-s-a.org.cn/1003-3254/7502.html

High Availability Design and Application of Network Monitoring System of ISP

WU Ge, YUAN Shou-Zheng, SUN Ding

(China Telecom Shanghai Ideal Information Industry (Group) Co. Ltd., Shanghai 201315, China)

Abstract: China Telecom NetCare service is a unified cloud network integrated monitoring platform. The system monitors client's network device, cloud applications, virtual resources, VPN and Internet. It is developed by China Telecom to provide the monitoring management and analysis services of network layer, transmission layer and application layer for enterprises. It serves the government and enterprise customers of China Telecom, and has very strict requirements for availability of the system. The high availability of a system is influenced by hardware, network, operating system, database, middleware, application itself and other aspects. This study focuses on the application layer's high availability of the NetCare system, and calculates the availability of the system with actual measurement and reasonable estimation of data. Finally, we come to a conclusion according to the paper: when the remote disaster recovery system is excluded or included, the whole system's availability is 99.997 8% and 99.45% respectively.

Key words: monitoring system high availability; application layer high availability; availability impact analysis; availability calculation

中国电信 NetCare 服务, 是利用中国电信统一建立的基于通用网络监控技术和专用探针技术的监控平台, 对包括客户端网络设备、云端应用及虚拟资源、专线及互联网线路提供端到端监控和管理的业务, 其界面如图 1、图 2 所示. 该业务面向中国电信政企客户提供服务, 监控了大量的客户设备以及相关的线路资源, 对于系统的可用性要求为 99.99%. 基于安全方面的考虑, 企业的网络监控系统基本采用自建的方式, 其

① 收稿时间: 2019-12-16; 修改时间: 2020-01-07, 2020-01-14; 采用时间: 2020-01-19; csa 在线出版时间: 2020-10-29

架构设计是以有限的监控对象为基础设计的, 所以其系统的监控容量是有限的[1]; 中国电信 NetCare 服务基于电信级的安全服务构建, 以 SaaS 的方式提供服务, 在监控的设备数量、动态的监控数据体量、系统性能、系统可用性方面有着更高的要求, 整个系统采用分布式分层架构模式, 底层使用成熟的分布式数据库系统、消息队列系统, 相对于传统封闭式的监控系统, NetCare 系统能够通过底层分布式系统的资源的动态扩展更好地满足业务的发展需要[2]. 一个系统的可用性是由多方面的因素共同决定的, 通常会涉及硬件、网络、操作系统、数据库、中间件、应用本身等[3], NetCare 系统的高可用性方案中一方面在硬件、网络、操作系统、数据库、中间件方面引进了相应厂商的高可用性解决方案, 另一方面通过设计与实践提升了应用系统自身的高可用性.

图 1 NetCare 系统首页界面

图 2 NetCare 系统监控板界面

1 影响系统高可用性的因素分析

1.1 NetCare 系统逻辑架构

如图 3, NetCare 系统架构主要由数据采集子系统、数据处理子系统、业务管理子系统、消息队列、数据缓存、数据存储 6 大部分组成. 为了分散系统的采集压力数据采集子系统部署多台采集机, 系统实测数据: 每台采集机在采集频率为 10 s 的情况下可以承担 2000 台设备的数据采集工作. 数据处理子系统采用多台数据处理机+数据缓存的方式提高数据归并、计算的性能, 系统实测数据: 每台数据处理机在采集频率为 10 s 的情况下可以同时支持 20000 台设备的采集数据的计算、归并, 并将数据写入分布式数据库中[4].

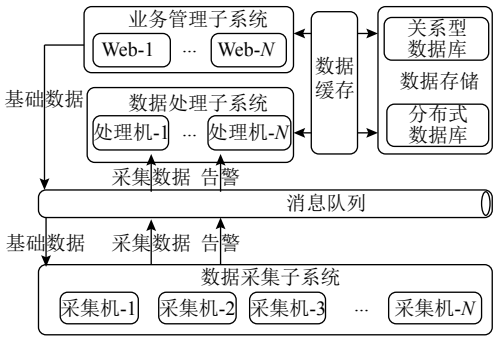

图 3 NetCare 系统逻辑架构

1.2 应用层高可用性影响分析

故障综合影响级别的定义见表 1.

表 1 故障综合影响级别

对系统可用性影响程度	故障发生可能性		
	高	中	低
高	严重	高	中
中	高	中	低
低	中	低	提示

严重: 严重故障对系统有着较高影响且有较高发生可能性, 严重故障会导致系统长时间 (一般以天为时间单位) 的故障停机, 对客户和运营单位造成巨大损失.

高危: 高优先级别故障对系统有着较高影响和中等发生可能性, 或中等影响和较高发生可能性. 高优先级别故障会导致系统较长时间 (一般以小时为时间单位) 的故障停机, 对客户和运营单位造成较大损失.

中危: 中优先级别故障对系统有着较高影响和较低发生可能性, 或中等影响和中等发生可能性, 或较低影响和较高可能性. 中优先级别故障会导致系统短时间 (一般以分钟为时间单位) 的故障停机, 对客户和运营单位造成中度损失.

低危: 低优先级别故障对系统有着中等影响和较低发生可能性, 或较低影响和中等发生可能性. 低优先级别故障会导致系统较短时间 (一般以秒为时间单位) 的故障停机, 对客户和运营单位造成轻微损失.

根据 NetCare 系统架构分析可以得到影响系统可用性的应用层分析表 2. 表 2 中罗列了影响 NetCare 系统可用性的应用层方面的主要因素[5], 细分的因素数量会更多, 为了后续高可用性的设计描述更加清晰, 本文对于各子系统架构的风险进行了概括的总结; 对于消息队列、数据存储、数据缓存由于采用了成熟的开源软件且其架构中都具有高可用性的设计及实现, 所以故障发生的可能性定为低.

表 2 影响系统可用性的应用层分析表格

序号	所属模块	主要故障描述	对系统可用性影响程度	故障发生可能性	故障综合影响
1	数据采集子系统	数据采集机故障: 导致分配在该采集机上进行数据采集的设备的性能数据无数据展示, 无法产生设备相关的告警	高	中	高
2	业务管理子系统	Web服务器故障: 导致面向客户的业务中断	高	中	高
3	数据处理子系统	数据处理机故障: 导致所有设备采集的性能数据、告警无法归并、计算, 无法存储到数据库中, 造成消息队列中的数据堆积, 无法产生设备相关的告警	高	中	高
4	消息队列	消息队列故障: 导致所有被采集设备的性能数据无数据展示, 无法产生设备相关的告警, 新增或变更的设备基础数据无法下发至采集机	高	低	中
5	数据存储	数据存储故障: 导致所有设备采集的性能数据、告警无法存储到数据库中, 造成消息队列的数据堆积; 业务管理子系统录入的数据无法持久化, 造成数据丢失	高	低	中
6	数据缓存	数据缓存故障: 导致数据处理子系统的数据处理机必须每次从数据存储中读取数据, 数据处理效率降低; 导致业务管理子系统的WEB服务器必须每次从数据库中读取数据, 能够支持的并发数量下降	中	低	低

2 系统高可用性设计与实践

2.1 网络层面的高可用实践

NetCare 系统使用 SD-WAN 服务为客户提供多通道高可用性技术, 可以在小于 1 s 的时间内切换通道, 保障网络的高可用性.

2.2 硬件层面的高可用实践

NetCare 系统部署在采用 VMWare 构建的私有云上, 通过 vSphere 建立包括 DRS、HA 功能的集群, HA 技术最高灵敏度可以在 30 s 内检测到虚拟机故障, 并重置虚拟机; DRS 可以将虚拟机从负载较重的主机迁移到负载较轻的主机上.

2.3 中间件、数据存储、数据缓存的高可用实践

NetCare 系统在实践中选取了开源的 ActiveMQ、MySQL、HBase、Redis 分别作为消息队列、数据存储、数据缓存的组件, 对应的高可用性设计也采用了开源软件自身的高可用性方案. ActiveMQ 采用了 Zookeeper+LevelDB 的部署方式. MySQL 采用了 Master-Slave 的部署方式. HBase 本身为高可用的分布式数据库. Rdeis 采用了 Redis-Cluster 的部署方式. 数据存储选用了 MySQL 和 HBase 两种类型的数据库, 主要是基于如下考虑: 关系型数据库用来存储设备、客户、服务包等基本信息, 这类信息数据量较小, 相对变动不大; 分布式数据库主要用来存储采集的动态数据, 这类信息的数据量巨大, 不适合采用关系型数据库存储[6].

2.4 数据采集子系统的高可用实践

NetCare 系统的数据采集子系统主要用于从设备或其他 API 接口采集动态的性能数据 (主要包括: 线路通断、端口流量、CPU、内存等), 采集机支持的最短采集频率为 10 s/次. 为了减少对采集对象的影响, 每个采集对象的数据仅由一台采集机进行采集.

为了确保采集的高可用性, 首先在线路上需要设置两条互相备份的采集链路 (Active-Standby 模式), 当一条链路不可用时, 采集机通过备份的链路进行数据采集.

由于单台采集机可以采集的对象的数量是有限的, 所以数据采集子系统中部署有多台采集机, 分散采集压力, 增强系统的可用性[7]: 当一台采集机出现故障时, 影响范围仅限于在该采集机上进行数据采集的设备. 表 3 是 NetCare 系统需要部署的采集机的数量实践数据, 其中数据是在单台采集机 CPU、内存均小于等于 60% 前提下的测试结果.

整个系统需要的采集机数量计算如下:

$$Ceil\left(\frac{M}{M1}\right)+\mathrm{MAX}\left(Ceil\left(M1\times\frac{F}{F1}\right), Ceil\left(M1\times\frac{S}{S1}\right)\right) \tag{1}$$

其中, $Ceil$ 为进位取整函数, MAX 为取最大数函数; M 是系统总的监控设备数量; $M1$ 为单台采集机测试的监控设备数量上限; F 为单台被监控设备实际每秒上报的 Netflow 的 flow 数量; $F1$ 为单台采集机测试的每秒上报的 Netflow 的 flow 数量; S 为单台被监控设备实际每秒上报的 SNMP Trap 包数量; $S1$ 为单台采集机实际每秒上报的 SNMP Trap 包数量.

表 3 单台采集机上限测试数据

采集指标	频率	实测结果: 单台采集机上限
通断测试	1次/10 s/设备	监控设备: 2000
SNMP采集	20个OID/1分钟/设备	
Netflow上报	每秒发包数递增模式(30 flow/包)	26 400 flow/s
SNMP Trap上报	每秒发包递增模式	880 包/s
通断测试	1次/30 s/设备	监控设备: 5000
SNMP采集	20个OID/1分钟/设备	
Netflow上报	每秒发包数递增模式(30 flow/包)	16 800 flow/s
SNMP Trap上报	每秒发包数递增模式	560 包/s

系统对采集机的状态进行持续监控, 当采集机发生故障时, 可以从业务管理子系统将该采集机负责的设备转移到其他采集机上; 另外采集机上部署有关系型数据库, 当消息队列发生故障时, 可以临时存储采集上来的数据, 增强了数据采集子系统的可用性. 数据采集子系统的应用层故障时间主要取决于采集机监控频率且重新分配设备的所属采集机的时间.

2.5 数据处理子系统的高可用实践

NetCare 系统的数据处理子系统主要从消息队列中获取最新的采集数据, 从缓存中获取最近持久化的采集数据, 将两种数据进行归并、计算, 并持久化到数据存储中, 替换缓存中的数据. 多台数据处理机采用竞争消费的方式从消息队列中获取数据, 当一台数据处理机故障时, 其他数据处理机会分担数据处理任务. 单台数据处理机的实测的处理速率为近 600 条/s, 部署 3 台数据处理机, 因而数据处理子系统的应用层故障时间主要取决于消息队列竞争的多消费者之间切换的时间.

2.6 业务管理子系统的高可用实践

NetCare 系统的业务管理子系统主要面向客户、运营人员提供可视化的监控相关功能, 主要是 Web 服务, 采用两台 Web 服务器负载均衡的方式对外提供服务.

Web 应用程序框架为自研的边缘计算引擎[8], 采用前后端分离的方式, 该框架支持自动热迁移, 两台 Web 服务器的前后端可以分别互作备份, 当一台服务器的后端服务升级时, 两个前端服务可以共享依然活跃的后端服务, 可以有效减少系统维护的停机时间.

业务管理子系统应用层故障时间主要取决于 Web 应用容器的后端服务健康检测时间.

2.7 IDC 层面的高可用实践 (异地灾备)

以上均为同地的可用性设计及实践, 为了防患于未然, 需要设置异地的灾备系统, 由于资源限制, NetCare 系统主要采用同市区不同机房的异地灾备方式, 并未采用跨地域、跨电网、跨地震带的方式. 在发生机房故障时, 故障时间主要取决于域名的切换时间[9].

3 NetCare 系统的可用性计算

基于以上分析, NetCare 系统的各层故障停机时间分析如表 4.

表 4 NetCare 系统应用层停机时间分析表

序号	所属模块	估算依据	全年估算停机时间(s)
1	数据采集子系统	采集机的状态监控时间间隔为: 10 s 重新分配设备所属采集机并生效的时间间隔为: 5 min 单个采集程序全年故障预估次数: 1~2次	310~610
2	业务管理子系统	Web应用容器的后端服务健康检测时间间隔: 1 s 单个Web后端服务器全年故障预估次数: 1~2次	1~2
3	数据处理子系统	数据处理程序竞争消费数据的间隔: 1 s 单个数据处理程序全年故障预估次数: 1~2次	1~2
4	消息队列	Zookeeper的心跳时间间隔: 1 s 单个消息队列全年故障预估次数: 1~2次	1~2
5	数据存储	MySQL的主从心跳时间间隔: 1 s MySQL主备切换全年预估次数: 1~2次 Zookeeper的心跳时间间隔: 1 s 单个分布式数据库节点全年故障预估次数: 1~2次	2~4
6	数据缓存	Redis集群节点失效检测时间间隔: 1 s 单个Redis节点主备切换次数: 1~2次	1~2
7	网络层	SD-WAN多通道高可用切换时间: 1 s 网络通道全年故障预估次数: 1~2次	1~2
8	硬件层	vSphere的DRS故障检测时间: 30秒 虚拟机全年故障预估次数: 1~2次	30~60

NetCare 系统总的全年预估最长停机时间:

$$\frac{610+2+2+2+4+2+2+60}{60}=11.4(\text{分钟}) \quad (2)$$

NetCare 系统的可用性为[10]:

$$\frac{(365\times24\times60-11.4)}{365\times24\times60}\times100\approx99.9978(\%) \quad (3)$$

以上为本地系统的可用性估算, 异地灾备系统的切换取决于域名的生效时间 (最长为 48 小时), 则考虑本地系统无法提供服务, 启动异地系统的情况下, NetCare 系统的可用性为:

$$\left(1-\frac{48}{365\times 24}\right)\times 100\approx 99.45(\%) \quad (4)$$

4 结论与展望

影响系统的可用性主要由平均无故障时间和平均维修时间决定, NetCare 系统采用了本文描述的相关高可用性设计后, 提升了系统的平均无故障工作时间, 整个系统的分布式架构减少了相关模块的耦合度, 单个模块故障对整个系统的影响范围得到了有效控制, 缩短了故障的处理时间, 整个系统可用性达到了 99.9978%. 影响系统可用性的因素比较多, 系统建设时需要尽可能多的识别, 提高系统的可用性同时也意味着成本的增加, 所以对于所有影响因素要进行综合考虑, 需要在系统可用性的实际需求、建设方的资金投入、系统建设周期之间找到一个切实的平衡点.

参考文献

1 陈卫丽. 证券公司集中监控系统设计. 计算机系统应用, 2019, 28(8): 115–119. [doi: 10.15888/j.cnki.csa.007007]

2 张瑞聪, 任鹏程, 房凯, 等. Hadoop 环境下分布式物联网设备状态分析处理系统. 计算机系统应用, 2019, 28(12): 79–85. [doi: 10.15888/j.cnki.csa.007181]

3 熊盛武, 王鲁, 杨婕. 构建高性能集群计算机系统的关键技术. 微计算机信息, 2006, 22(3): 86–88. [doi: 10.3969/j.issn.1008-0570.2006.03.031]

4 魏东平, 李奉娟, 苑志朋. 基于 OSGI 分层动态的软件设计与开发. 计算机系统应用, 2017, 26(9): 98–102. [doi: 10.15888/j.cnki.csa.005934]

5 王月瑶, 胡琴敏, 刘伟, 等. 智能分类算法在游戏故障告警中的应用. 计算机系统应用, 2018, 27(7): 133–138. [doi: 10.15888/j.cnki.csa.006417]

6 黄杰圣, 李传目. ORACLE 中对大数据量表的处理方法. 计算机系统应用, 2003, 12(12): 71–72. [doi: 10.3969/j.issn.1003-3254.2003.12.021]

7 孙超, 王永贵, 常夏勤, 等. 面向电力大数据的异构数据混合采集系统. 计算机系统应用, 2018, 27(12): 62–68. [doi: 10.15888/j.cnki.csa.006667]

8 袁守正, 姚磊, 周骏, 等. 中国电信工业互联网平台“边缘计算引擎”设计及实现. 电信技术, 2019, (4): 65–71. [doi: 10.3969/j.issn.1000-1247.2019.04.018]

9 宋丽君, 徐建民. 关于异地灾备中心数据解决方案研究. 数字技术与应用, 2016, (10): 209.

10 田立中, 周昭涛. 多层次集群系统的可用性指标计算. 中小企业管理与科技, 2007, (11): 66. [doi: 10.3969/j.issn.1673-1069.2007.11.040]