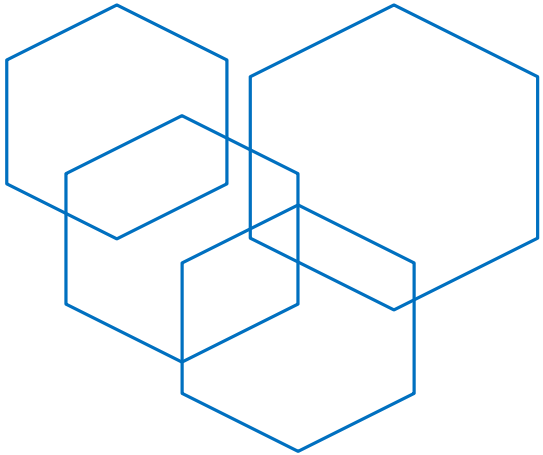


分布式云资源监控

郑伟

zhengw@ihep.ac.cn

中科院高能物理研究所

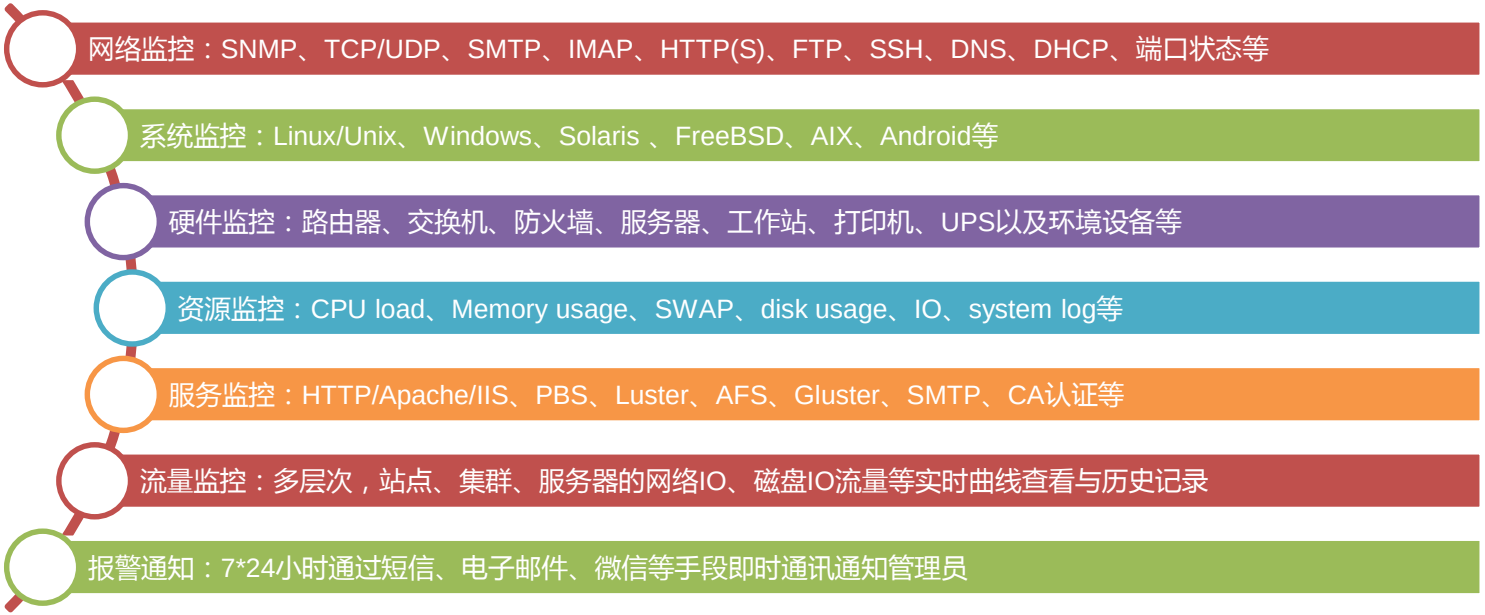
01 系统需求

随着云计算、大数据、高性能平台等技术应用，站点规模不断扩大、设备数量不断增加。传统的IT基础设施管理水平和运维技术逐渐难以应对不断扩展的网络环境和不断增长的应用需求。传统的数据中心扩展为多个分布式站点，物理资源和云资源混合运行模式。对传统的监控和运维系统提出新的需求，需要一个既能够完成对本地大规模网络设备、服务器、系统服务的集中统一监控系统，又能对分布各地的云资源进行信息收集，对监控的实时性、可靠性以及分布式扩展功能提出更高的要求。

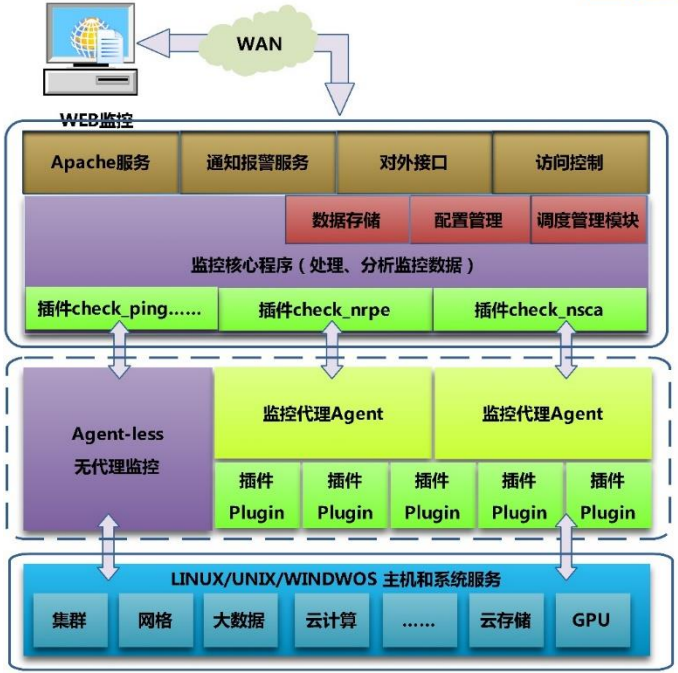
目前分布式云资源监控系统实现以IHEP为中心站点，多个分布式站点联合统一监控的监控模式，在运维上保证所有站点的有效可用，保障物理作业跨站点正常提交，实现降低运维成本和提升整体IT管理能力。

02 系统介绍

云资源监控系统，可以实现跨设备、跨平台、跨系统的数据采集，能够实时监控各个设备和服务的运行状态和性能，及时显示故障信息，方便快速掌握各个站点整体运行情况。



监控系统架构



监控逻辑层

包括具体实现系统功能，如监控对象管理、调度实现、WEB展示、报警通知等.

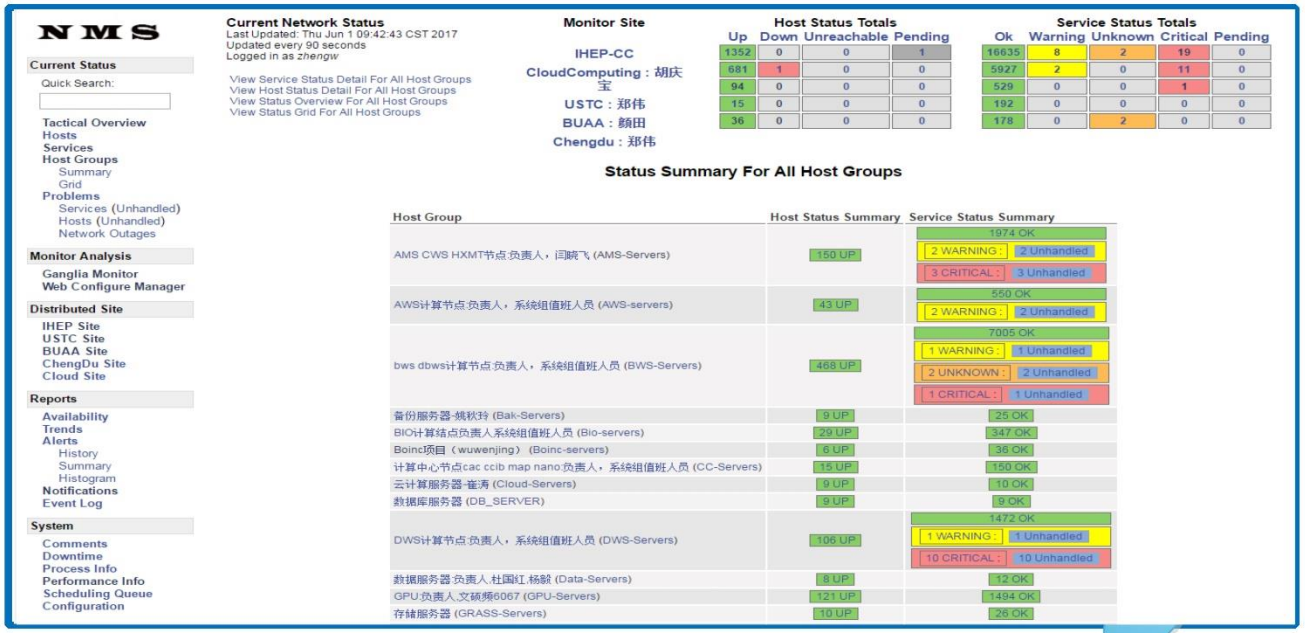
监控抽象层

是由各种检测插件组成的一个虚拟层次。插件作为连接监控逻辑服务与实际被监控主机和服务之间的桥梁功能，起到承上启下的关键作用。

监控实体层

监控实体可以是服务器、交换机、路由器、打印机等设备，也可以是应用服务如apache、共享存储、PBS服务等。应用的场景包括网格、集群、大数据、云计算、GPU应用等。

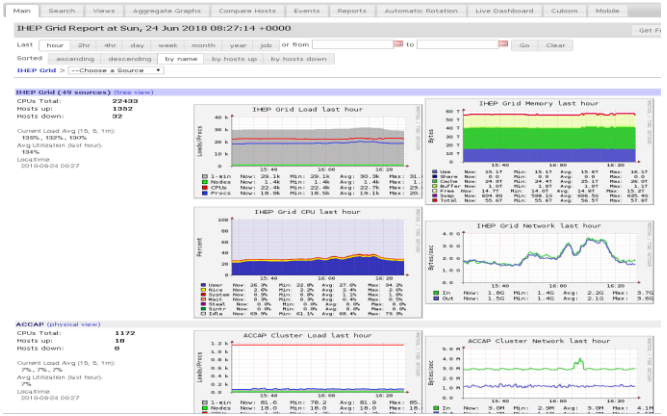
总体监控界面

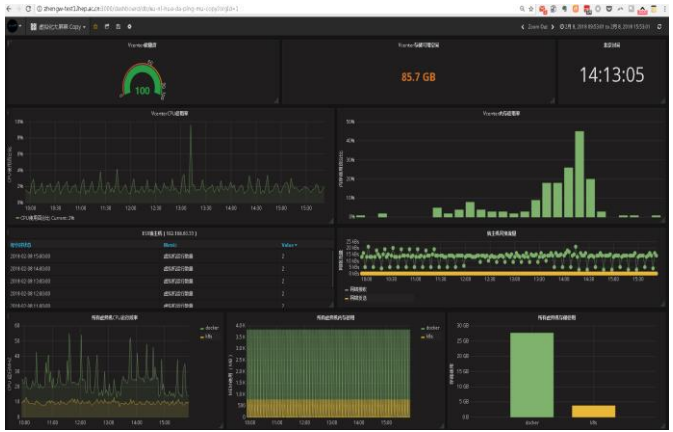


集中统一的监控界面，直观获取分布式云资源的整体状态，及时准确定位故障

05 监控Dashboard

服务器详细监控





监控作业调度详情

按下一次检查时间条目排序(升序)

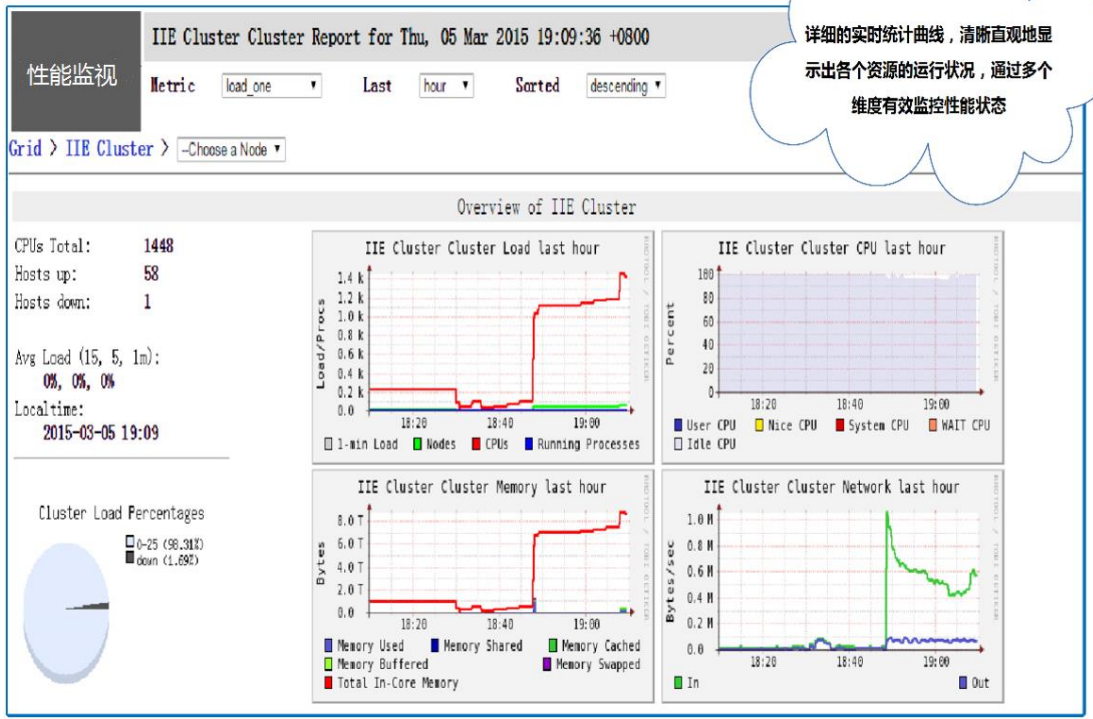
主机	服务	最近检查	下一次检查	类型	主动检查	动作
vm083187	SSH	2018年06月26日 06:40:40	2018年06月26日 06:45:40	正常	启用	
vm083153	SSH	2018年06月26日 06:40:40	2018年06月26日 06:45:40	正常	启用	
vm083103	SSH	2018年06月26日 06:40:40	2018年06月26日 06:45:40	正常	启用	
vm082183	SSH	2018年06月26日 06:40:40	2018年06月26日 06:45:40	正常	启用	
vm082167	SSH	2018年06月26日 06:40:40	2018年06月26日 06:45:40	正常	启用	
vm082144	SSH	2018年06月26日 06:40:40	2018年06月26日 06:45:40	正常	启用	
vm082137	SSH	2018年06月26日 06:40:40	2018年06月26日 06:45:40	正常	启用	
vm082132	SSH	2018年06月26日 06:40:40	2018年06月26日 06:45:40	正常	启用	
vm082127	SSH	2018年06月26日 06:40:40	2018年06月26日 06:45:40	正常	启用	
vm082119	SSH	2018年06月26日 06:40:40	2018年06月26日 06:45:40	正常	启用	
vm082114	SSH	2018年06月26日 06:40:40	2018年06月26日 06:45:40	正常	启用	
vm082093	SSH	2018年06月26日 06:40:40	2018年06月26日 06:45:40	正常	启用	
vm082150	SSH	2018年06月26日 06:40:41	2018年06月26日 06:45:41	正常	启用	
vm083184	SSH	2018年06月26日 06:40:42	2018年06月26日 06:45:42	正常	启用	
vm082174	SSH	2018年06月26日 06:40:42	2018年06月26日 06:45:42	正常	启用	
vm083228	SSH	2018年06月26日 06:40:43	2018年06月26日 06:45:43	正常	启用	

支持自适应，错误优先，加权算法等多种调度策略

06 性能监控系统

性能监控系统：

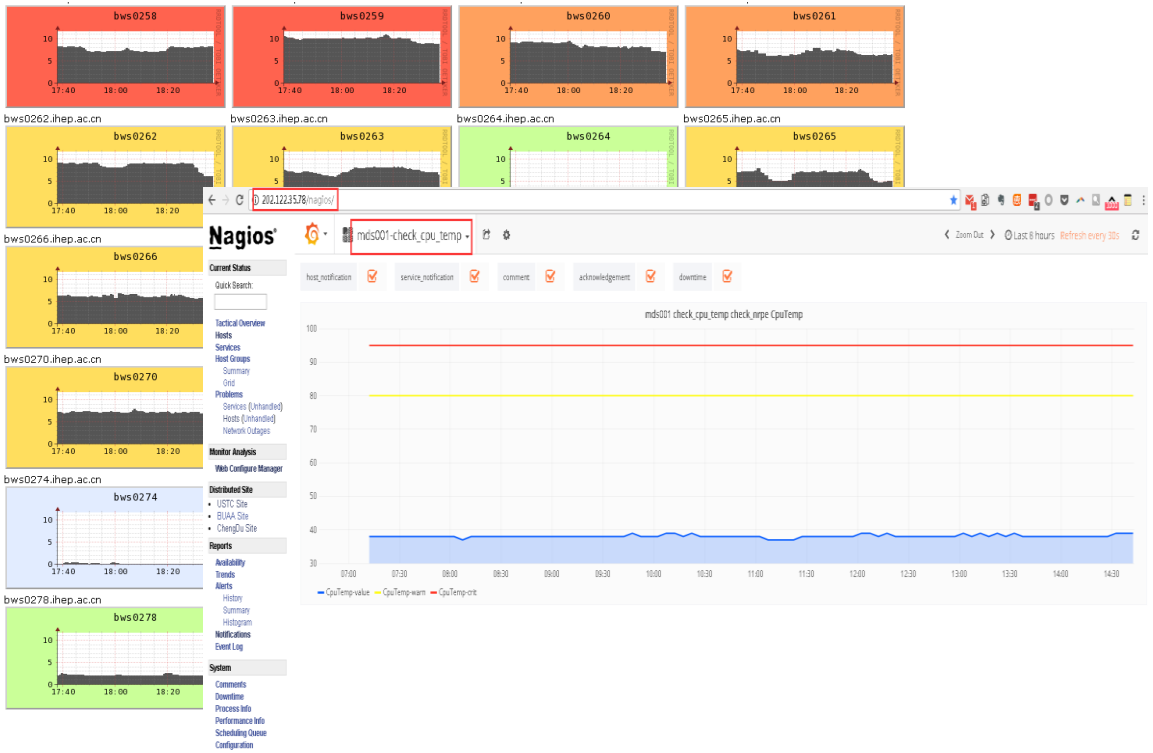
- 主要是用来监控系统性能，如：CPU 、 Memory、硬盘利用率、 I/O负载、网络流量情况等
- 每台主机的性能信息和状态信息绘制成曲线，通过这些曲线图形可以方便观察每个主机的工作状态
- 对整个大规模集群合理调整、分配系统资源，提高系统整体性能起到重要作用



07 性能监控

性能监控主机快照

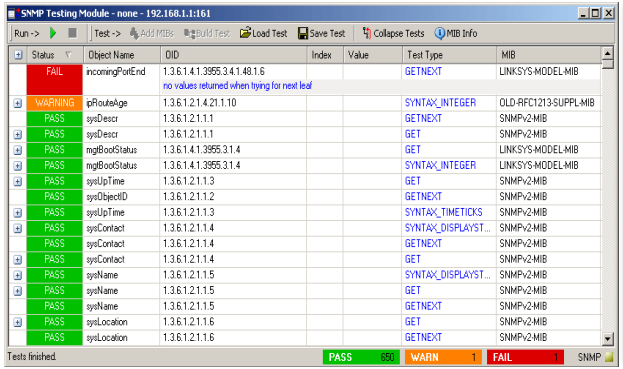
- 允许检查所有节点性能度量的历史
- 某个节点颜色变红时，意味着它的CPU平均负载比较大
- 除了CPU监控，还可以展示网络流量、内存、进程等度量的历史曲线。



08 底层硬件多协议支持

支持由单一的NRPE 代理监控模式，扩展为SNMP、IPMI、Sflow等多种监控协议支持

- SNMP:简单网络管理协议
  - 实现网络设备、Linux、Windows 的Cpu、内存、硬盘、网络流量等系统指标
- IDRAC\ILO\私有MIB
  - 通过设备管理口获取设备数据，磁盘阵列MIB信息分析,解决非通用设备SNMP协议不能解析数据问题
- IPMI :智能平台管理接口
  - 可以利用 IPMI 监视服务器的物理特征，如温度、电压、电扇工作状态，开关机
- Snmp-Trap:磁盘阵列、网络存储
- Sflow:(Sampled Flow,采样流)
  - 基于报文采样的网络流量监控



09 统一配置管理系统

所有站点配置统一管理

- 针对监控系统管理设计的WEB配置工具
- 可以方便的为系统创建，修改和删除配置文件
- 多站点集中配置，配置信息自动分发到分布式站点
- 解决分布式站点配置文件一致性，有效降低运维复杂度
- 系统管理员通过这个平台对被监控的主机和服务进行管理

10 分布式监控

- 分布式需求：之前本地和异地远程的云计算站点需要部署独立的监控系统进行运维管理。每一个站点都需要单独的系统管理员进行专门的日常监控和维护，大大增加了系统管理的难度和强度。
- 分布式监控：分布式扩展模式，多站点的集中统一监控, 对每个站点中各种设备和服务进行有效的监控和报警，提高运维人力的利用率，降低多站点运维强度和复杂性。

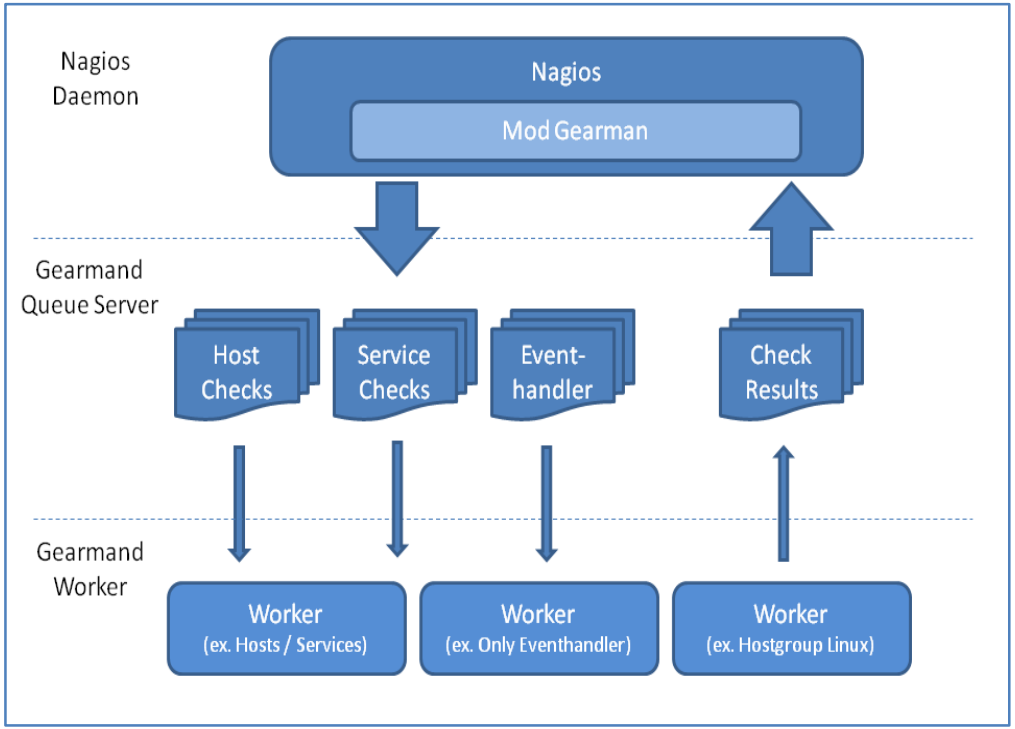
Monitor Site	Host Status Totals				Service Status Totals				
	Up	Down	Unreachable	Pending	Ok	Warning	Unknown	Critical	Pending
IHEP-CC	1387	0	0	1	15381	268	8	6	0
USTC：郑伟	78	0	0	0	156	0	0	0	0
BUAA：颜田	15	0	0	0	98	0	0	0	0
Chengdu：郑伟	35	1	0	0	173	0	2	5	0

11 分布式监控-数据采集

采用**mod Gearman 数据传输中间件**

由三部分组成：

- 一个NEB模块，它同监控核心程序驻留在一起，将servicechecks，hostchecks和eventhandler加进Gearman 队列
- 一个或多个worker客户端，用于执行检查。worker可以被配置成只运行指定的主机或者服务组检查
- 至少需要运行一个Gearman Job Server



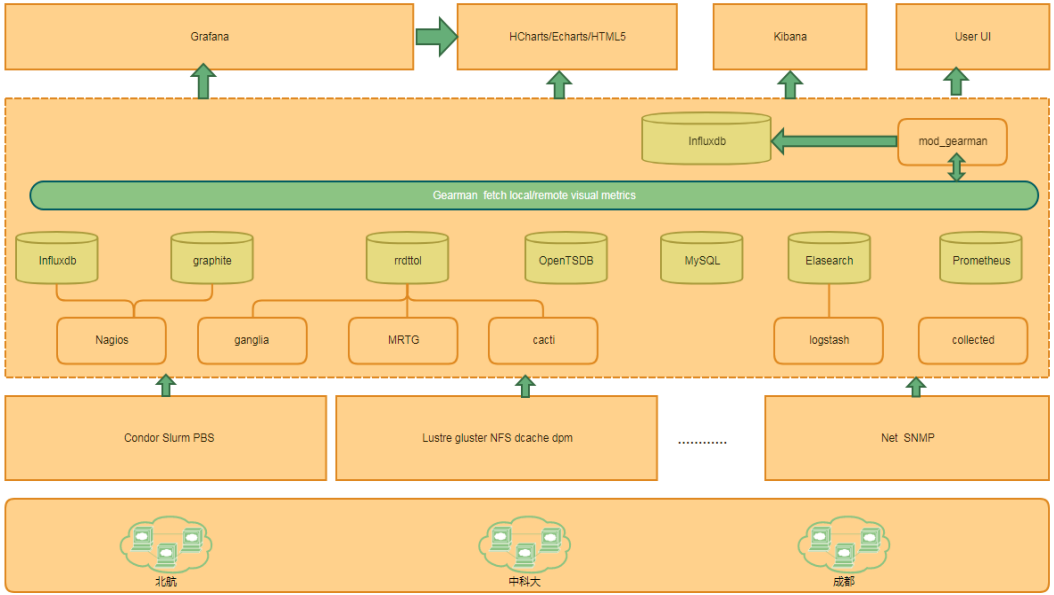
12 分布式监控数据存储和分析

数据集中存储

- 指标数据：时序数据库TSDB（Influxdb/Opentsdb）
- 统计信息：Mysql
- 日志:Elasearch

分析维度

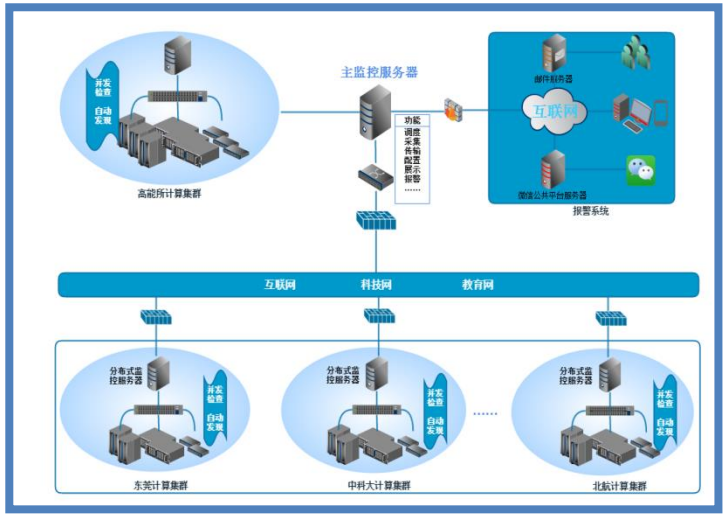
- 综合指标：健康度/可用性/可靠性
- 计算:作业排名、Cpu使用排名、计算时间……
- 存储:数据量、用户使用排名……
- 网络:交换机/服务器的出入流量……
- 基础设施:动力环境

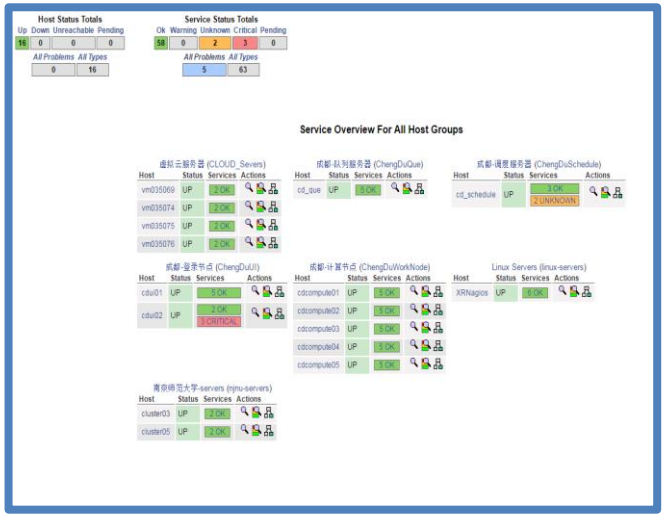


13 分布式监控架构

分布式监控架构和监控效果图

- 有效监控每个站点、每个主机、细化到每项服务
- 5min完成所有站点监控



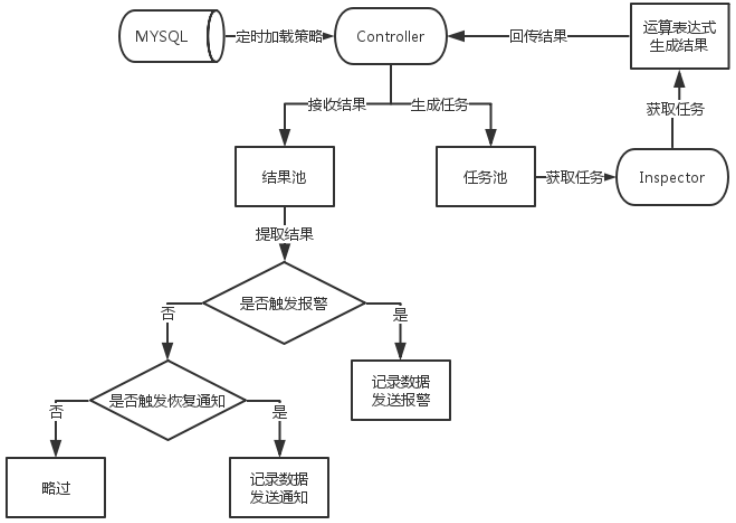


ChengDuWorkNode	Host Status	Services				
cdcompute05	UP	PING服务	SSH服务	check_disk_local	check_diskwrite	check_mem_hardware
cdcompute04	UP	PING服务	SSH服务	check_disk_local	check_diskwrite	check_mem_hardware
cdcompute03	UP	PING服务	SSH服务	check_disk_local	check_diskwrite	check_mem_hardware
cdcompute02	UP	PING服务	SSH服务	check_disk_local	check_diskwrite	check_mem_hardware
cdcompute01	UP	PING服务	SSH服务	check_disk_local	check_diskwrite	check_mem_hardware

14 报警系统-高效策略

高效的监控报警系统应当有一个灵活的、清晰的报警策略。报警策略的四个重要考量

1. 对报警进行分级、分类
2. 在添加报警时要能够批量添加、批量更改
3. 针对某一个或者某一组设备要有具备单独抽离控制的能力
4. 当发生大范围产生报警时，要具备有能力对报警进行合并，避免报警干扰



不同级别的策略产生的任务被存储在不同的channel。低级别报警需要达到一定数量才会合并成一条发出告警，如此以避免报警过多的干扰。

15 系统特色-与作业系统联动

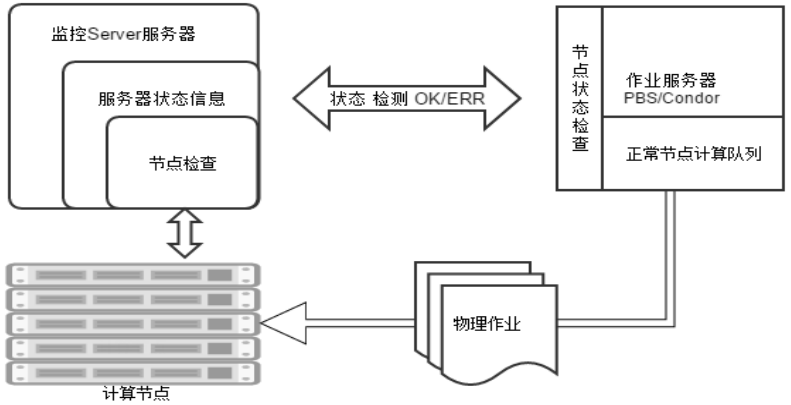
作业管理主要使用PBS和Condor作为实验作业和任务的管理系统。监控系统将节点的状态信息实时发送给作业管理系统，使作业能够**正确调度**到节点上，确保作业正常执行。

联动技术方案：主要通过提供监控正确节点确保作业管理系统中是健康节点，包括三项技术

- （NOTIFY）错误节点报送PBS/CONDOR从队列中移除
- （RECOVER）恢复节点通知PBS/CONDOR放回队列
- （SYNC）PBS/CONDOR同步两个系统中节点信息，确保队列节点被有效监控。

特点：

采用基于节点服务可用性的调度方式，更加准确、有效，保证作业调度尤其是分布式调度顺利完成



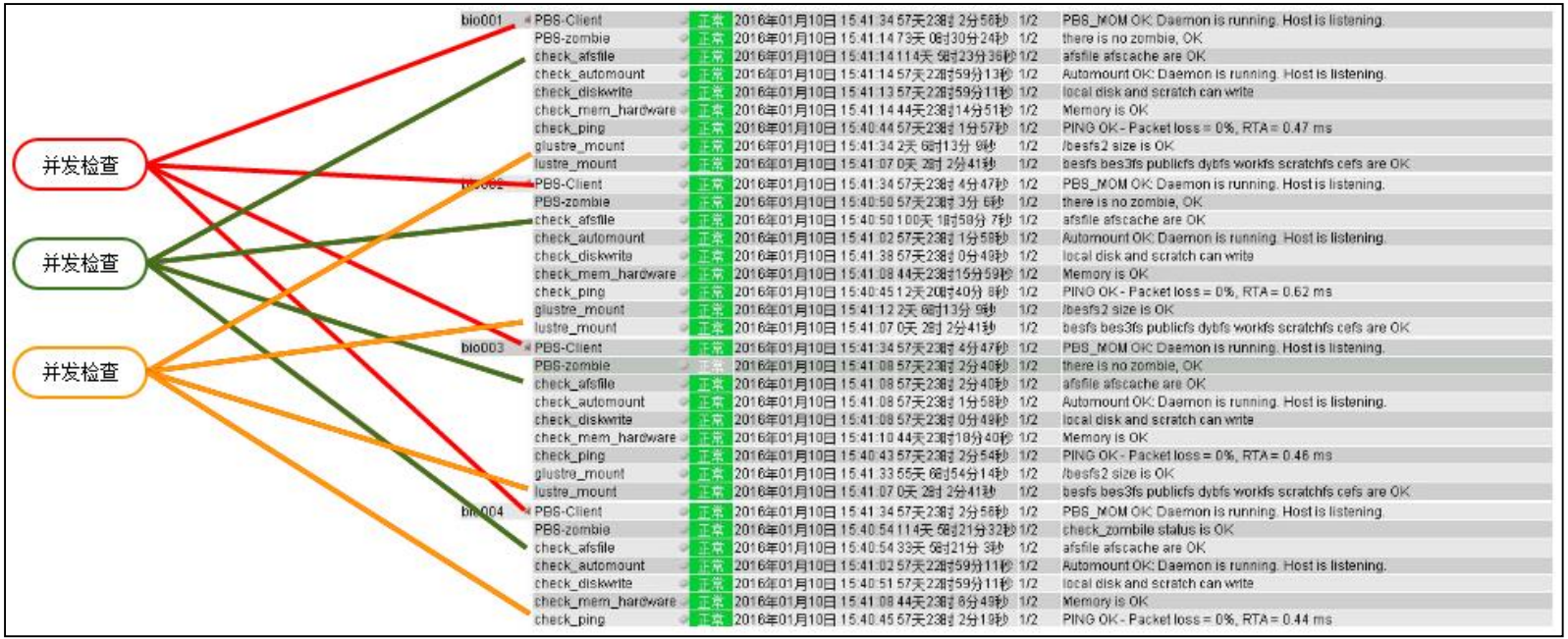
16 系统特色-高并发监控

- 性能技术瓶颈：并发检查数达到一定数量，其并发数无法再有效提高，造成较大的监控延迟，影响整体效率。
- 解决方案：并发检查机制，主要通过多线程技术、线程池、进程池技术实现**高并发**任务、以及采用多种调度策略。

顺序循环策略	随机轮询策略	故障优先策略	优先级策略
• 在初始化监控阶段将所有的监控对象按照顺序排序，进行顺序遍历所有监控任务，检查完最后的监控对象后再从头开始重新执行。优点是策略简单，运行稳定。	• 在初始化监控的阶段将所有的监控对象进行排序，按照给定的随机策略，随机抽取一定的数量并发执行，等检查完毕后再进行下一轮的随机检查。	• 对发现过故障的监控对象，在下一轮监控中优先重点检查。这样可以使对故障敏感的服务得到最快的状态反馈，同时刚发生过故障的对象，再次发生故障的概率也比较大，针对这样的问题可以提前有效监控。	• 给所有监控对象划分为多个队列或者组,给每个队列定义优先级，在高优先级队列中的对象或者服务器优先检查，并增加检查的频率。这样可以保证重要服务和监控对象及时有效的检查。

17 系统特色-秒级响应

- 性能两个条件约束Throughput（吞吐量）和Latecny（系统延迟）
- 通过实现最大并发数，优化检查插件的执行效率，优化操作系统的性能指标实现大并发、高吞吐量减少监控延迟，最终实现**秒级响应**。单服务器分钟级并发2万个



18 系统特色-云平台自动化服务监控

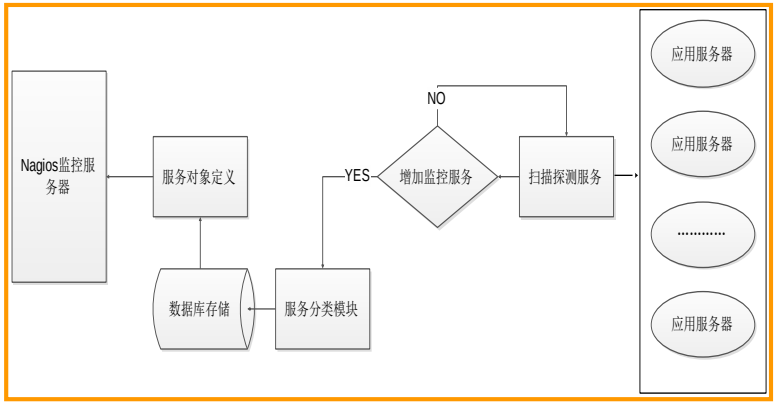
技术需求：云计算平台上的虚拟机开关频繁，无法像物理服务器正常有效监控

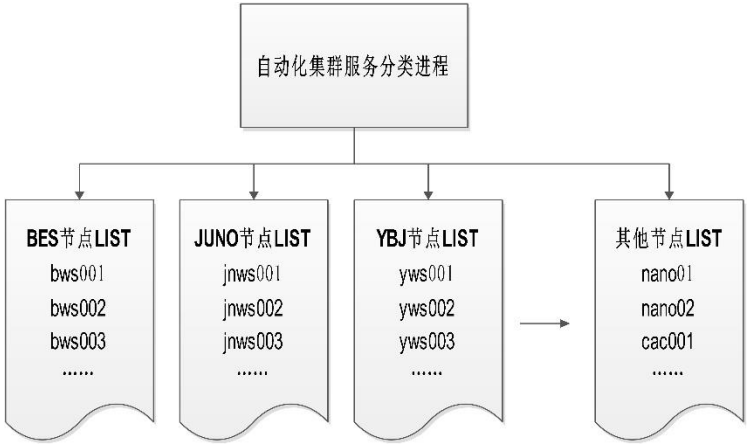
技术实现：

- 自动发现服务Daemon
- 数据库自动配置
- 从数据库生成文本配置程序
- 对外接口程序

特点：

- 服务自动发现技术首次应用到监控中
- 解决云主机监控难点，保证云平台正常运行





19 系统特色-分布式网关代理

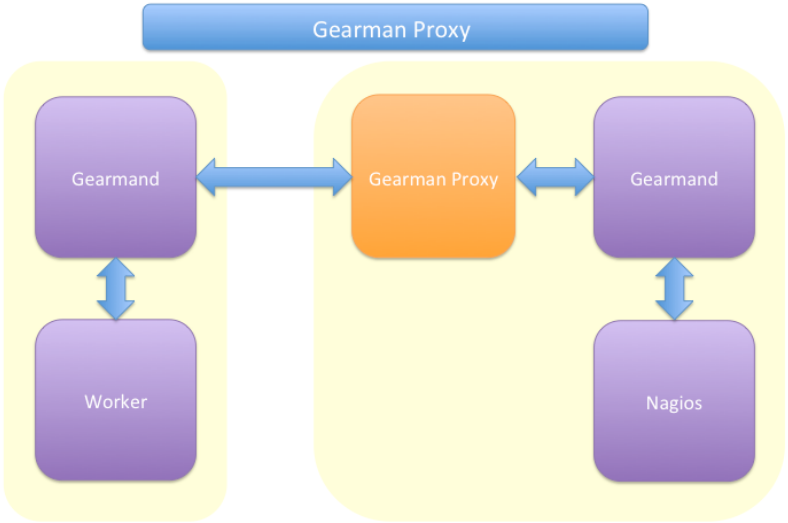
系统需求：分布式资源处于内网或特殊网络环境，无法与外界直接通讯，运维和监控不便

技术实现：

- 开发分布式网关代理
- 通过网关跳转访问
- 监控数据跨域传输
- 监控任务调度队列化管理

特点：

- 应用于复杂的网络环境，解决监控需求
- 解决站点监控难点，保证平台正常运行



20 系统特色-新媒体微信报警平台

- 监控系统通过开发的微信平台进行报警，增强报警效果，实现**移动运维**
- 微信应用数增加到8个,信息化、调度、CA等
- 微信使用人数达到23人
- 微信报警数量去年统计17500多条（含测试），节约大量的经费,更加便捷手机、台式机都能接收

21 系统特色-支持IPV6/IPV4双栈协议

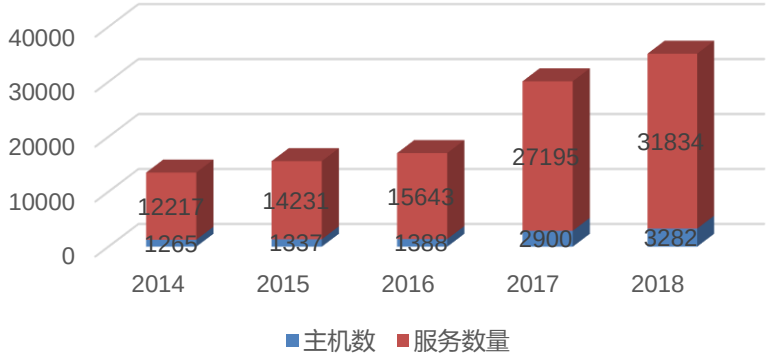
- 中共中央办公厅、国务院办公厅印发《推进互联网协议第六版(IPv6)规模部署行动计划》，到2025年末，我国IPv6网络规模、用户规模、流量规模位居世界第一位，网络、应用、终端全面支持IPv6
- IPV6监控
  - 服务器/客户端 支持IPV6地址协议
  - IPV6流量监控 IPV6 Ping
- IPV4+IPV6双栈支持
  - 所有监控服务同时支持V4和V6

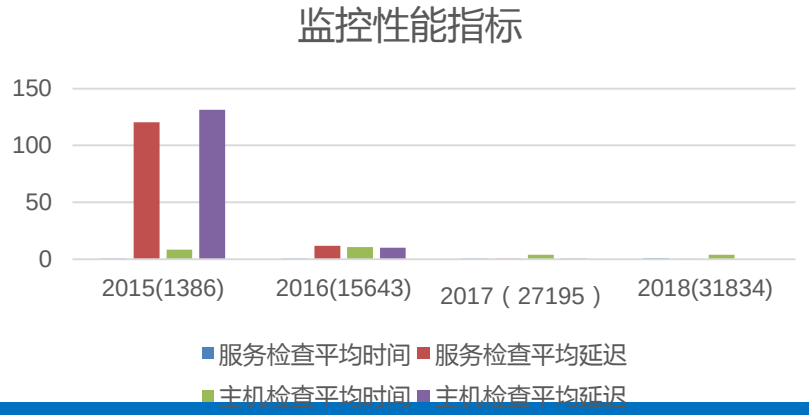
```
$ check_http -4 -H demo.funet.fi
HTTP OK: HTTP/1.1 301 Moved Permanently - 523 bytes in 0.006 second
    response time |time=0.005673s;;;0.000000 size=523B;;;0
$ check_http -6 -H demo.funet.fi
HTTP OK: HTTP/1.1 301 Moved Permanently - 523 bytes in 0.004 second
    response time |time=0.003978s;;;0.000000 size=523B;;;0
$ check_v46 check_http -H demo.funet.fi
OK: IPv6/demo.funet.fi OK, IPv4/demo.funet.fi OK |
    ipv6_time=0.004730s;;;0.000000 ipv6_size=523B;;;0
    ipv4_time=0.002237s;;;0.000000 ipv4_size=523B;;;0
```

22 当前规模与性能

分布式云资源监控，规模逐年扩大，系统更加完善和高效、性能进一步优化、运行稳定

- 规模：站点数4个，主机数3238 、 服务31834
- 性能：
  - 每个服务平均响应时间1s左右
  - 并发轮询时间2min-3min，完成所有服务检测
  - 性能超过大部分商业软件





23 下一步工作

- 北大、山大、华中师大、LHAASO稻城、南京师范大学等分布式站点将加入统一分布式云资源监控平台
- 阿里云平台环境正在测试

谢谢各位老师
Q&A