

掘金Nginx日志

张波

虎牙直播-基础保障部



前言

- **Nginx**（读作 **Engine-X**）是现在最流行的负载均衡和反向代理服务器之一。仅 **Nginx** 每天就会产生上百 **M** 甚至数以十 **G** 的日志文件。当是有多少人关注过它背后的价值。



log_format main

```
'$host#_#$remote_addr#_#$upstream_addr#_#[$time_local]#_#$request#_#$status#_#$body_bytes_sent#_#$http_referer#_#$http_user_agent#_#$request_time#_#$upstream_response_time';
```

host	域名，如：www.yy.com
remote_addr	用户真实IP地址
upstream_addr	服务器节点ip
request	记录请求的方法、URI和HTTP协议信息
body_bytes_sent	status记录请求状态/返回码：如：404、200、206、502
http_referer	发送给客户端的字节数，不包括响应头的大小
http_user_agent	记录从哪个页面链接访问过来的
request_time	记录 客户端 浏览器相关信息
upstream_response_time	请求处理时间
	后端处理时间



log_format main

```
'$host#_#$remote_addr#_#$upstream_addr#_#[$time_local]#_#$request#_#$status#_#$body_bytes_sent#_#$http_referer#_#$http_user_agent#_#$request_time#_#$upstream_response_time#_#$local_addr#_#$hit_info#_#$cdn_user_agent#_#$bytes_sent';
```

host	CDN加速域名, 如: www.yy.com
remote_addr	用户真实IP地址
upstream_addr	回源ip
request	记录请求的方法、URI和HTTP协议信息
body_bytes_sent	status记录请求状态/返回码: 如: 404、200、206、502
http_referer	发送给客户端的字节数, 不包括响应头的大小
http_user_agent	记录从哪个页面链接访问过来的
request_time	记录 客户端 浏览器相关信息
upstream_response_time	请求处理时间
local_addr	后端处理时间
hit_info	CDN厂商的边缘节点IP
	1、HIT: 命中成功, 直接响应内容给客户端;
	2、MISS: 命中失败, 可能存在两种情况:
	(1) 边缘节点回CDN上层节点;
	(2) 直接回客户源站;
cdn_user_agent	记录 CDN厂商 浏览器相关信息
bytes_sent	发送给客户端的总字节数;



remote_addr

1 UV计算

2 ISP分布

3 地域分布

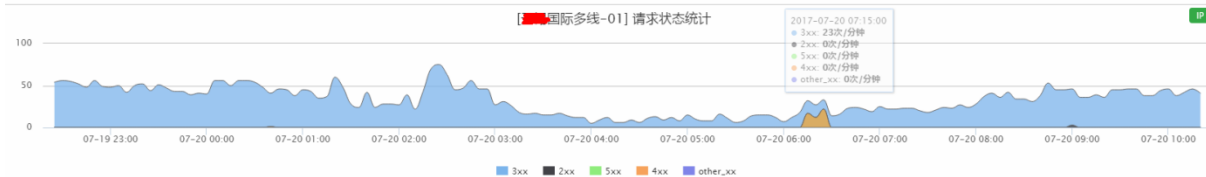
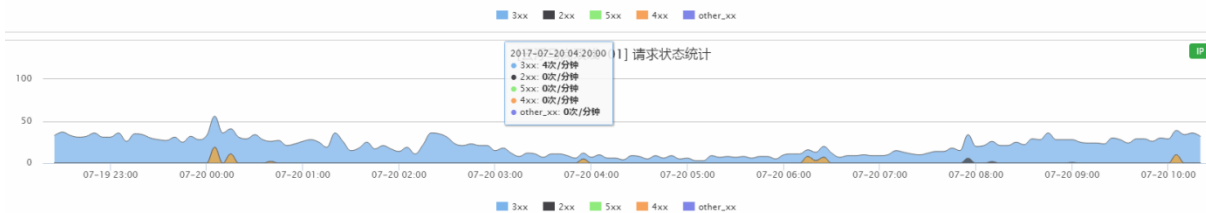


≡ 运营商维度		FAQ: 曲线指标的意义?	≡ 地域维度	
运营商名称	值		省份	值
电信	70,794		江苏	42,513
联通	5,052		广东	22,414
移动	1,452		北京	1,932
其他	648		山东	1,189
异常	489		浙江	1,161
阿里云	156		辽宁	1,029
鹏博士	122		河北	820
铁通	52		黑龙江	745
			河南	665
			四川	614

1 Idc分布



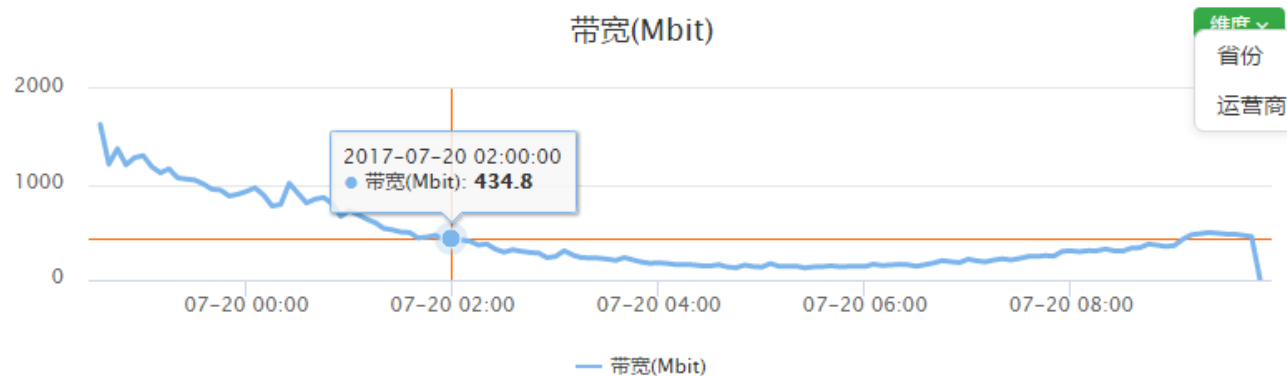
2 Ip分布



body_bytes_sent

1 带宽统计

2 下载速率



http_referer

1

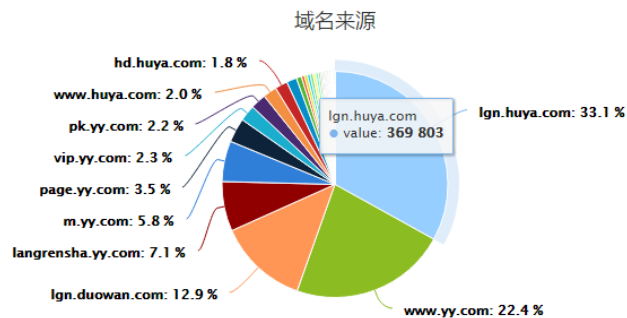
流量来源

2

转化率漏斗图

3

SEO



- lgn.huya.com
- www.yy.com
- lgn.duowan.com
- langrensha.yy.com
- m.yy.com
- page.yy.com
- vip.yy.com
- pk.yy.com
- www.huya.com
- hd.huya.com
- channelboard.yy.com
- udblogin.duowan.com
- wap.yy.com
- live.yyembed.yy.com
- show.vip.yy.com
- channel.yy.com
- find.yyembed.yy.com
- peipei.yy.com
- aq.yy.com
- bbs.duowan.com
- kfapp.yy.com
- channel.yyembed.yy.com
- yxgh.g.yy.com
- cqby.yy.com
- www.duowan.com
- hao.yy.com
- i.huya.com
- zc.yy.com
- g.duowan.com
- udbproxy.duowan.com
- pay.duowan.com
- im-kf.yy.com
- layer.huya.com
- illegal.yy.com
- kf.yy.com
- hgame.yy.com
- f2e.yy.com
- www.1931.com
- lgn.zhiniu8.com
- wan.yy.com
- wxgift.game.yy.com
- gpay.duowan.com
- ka.duowan.com
- jjyh.yy.com
- hd.vip.yy.com
- vippk.yy.com
- lgn.1931.com
- customlogin.yy.com
- fengyun.yy.com

http_user_agent

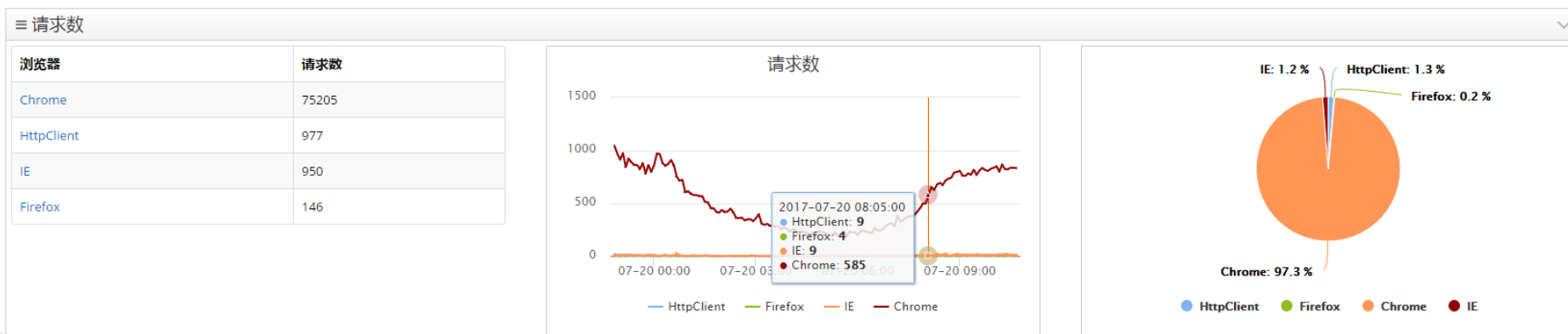
1 浏览器分布

2 操作系统分布

3 爬虫识别

≡ 操作系统维度

操作系统名称	值
Windows	12,847,897
Linux	807,369
Mobile	150,813



request_time(upstream_response_time)

1 延时分布

2 平均延时

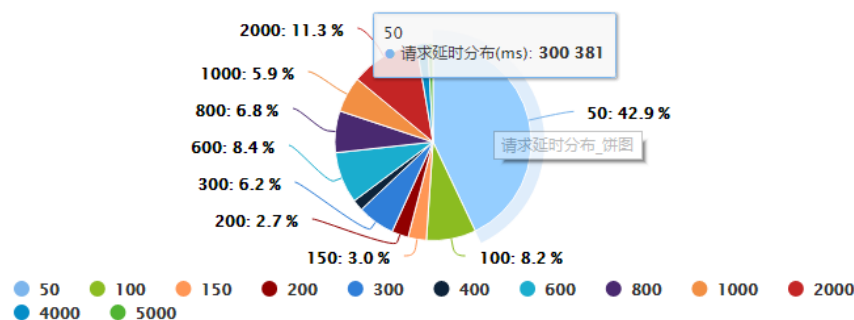
请求延时分布(ms)

线图 饼图



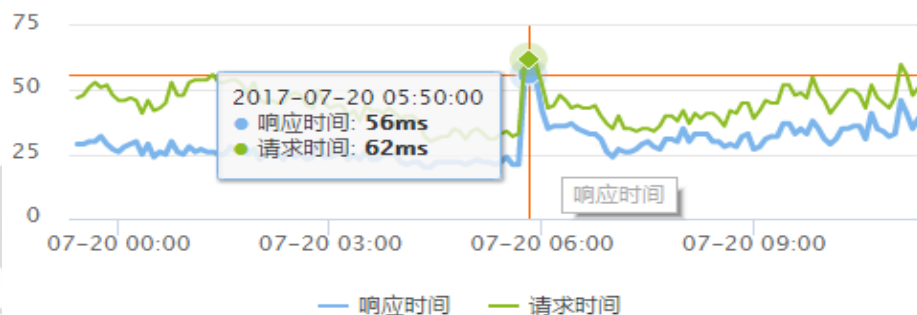
请求延时分布(ms)

柱状图 线图



请求/转发响应时间

IDC IP



request

1

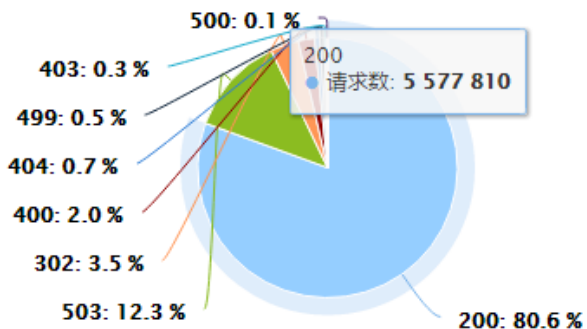
URI分布

2

返回码分布

3

请求类型分布



URI	请求总数	请求时间ms	响应时间ms	处理时间比‰	响应时间比‰	流量比‰	请求Apdex	响应Apdex
/lgn/jump/authentication.do	2355052	171	31	6652	4033	840	99	99
/lgn/oauth/authorize.do	1378344	60	35	1376	2617	7496	97	99
/lgn/oauth/x2/s/login_asyn.do	1168057	70	26	1365	1673	907	96	99
/lgn/oauth/wck_n.do	656432	24	24	264	850	369	99	99
/lgn/oauth/initiate.do	423754	5	5	41	131	106	99	99
/open/qrcode/loginqrcodeV2.do	229492	17	17	66	215	65	99	99
/lgn/oauth/wreqdomainck_n.do	182108	25	24	76	244	10	99	99
/	143438	33	-1	79	-1	46	99	100
/verify/x2/getsvcode.do	80251	12	11	16	50	37	99	99
/lgn/oauth/token.do	58394	19	19	18	60	23	100	99
/lgn/oauth/tokenValid.do	48198	18	18	14	47	1	99	99
/favicon.ico	45867	1	-1	0	-1	32	99	100
/lgn/oauth/delreqdomainck.do	30703	15	14	7	24	0	99	99
/lgn/oauth/p/noticeaccess.do	10661	21	21	3	12	3	99	99
/p/sec.do	9836	1	1	0	0	3	99	100

URI明细

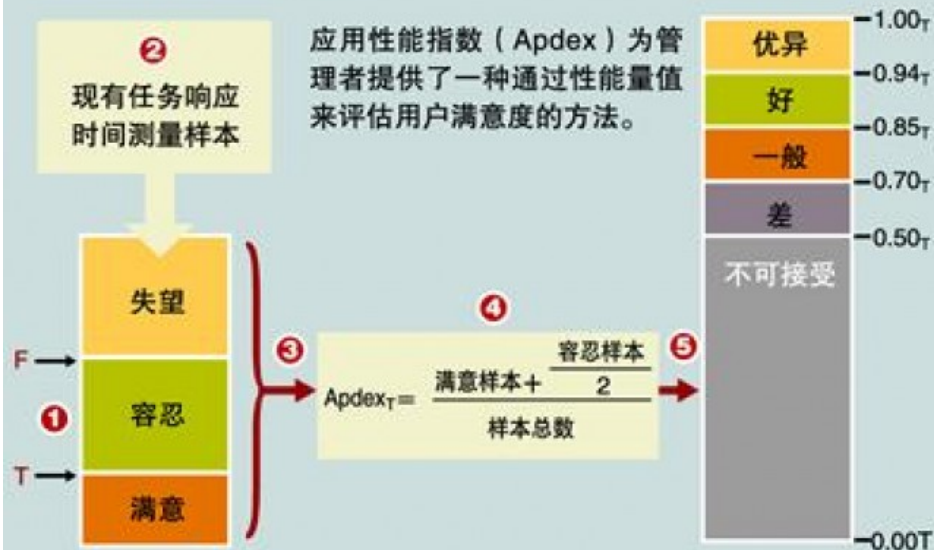
每页显示 10 条记录

URI	请求数	操作
/lgn/jump/authentication.do	2110992	删除
/lgn/oauth/authorize.do	1097644	删除
/lgn/oauth/wck_n.do	656139	删除
/lgn/oauth/x2/s/login_asyn.do	581409	删除
/lgn/oauth/initiate.do	423742	删除
/open/qrcode/loginqrcodeV2.do	229439	删除
/lgn/oauth/wreqdomainck_n.do	181708	删除
/verify/x2/getsvcode.do	72419	删除
/lgn/oauth/tokenValid.do	48168	删除
/lgn/oauth/token.do	39306	删除

从 1 到 10 / 共 38 条数据

Apdex量化应用性能

Apdex工作原理



- 1 管理员定义T，即目标应用响应时间。F (即 $4 \times T$) 代表容忍用户与失望用户之间的边界值。
- 2 管理员定义报告组，并从应用性能量值中提取数据。
- 3 Apdex报告工具统计3个性能区间的样本数量。
- 4 报告工具计算Apdex公式。
- 5 报告工具显示Apdex结果。T始终作为结果的一部分显示。

应用性能量化和染色

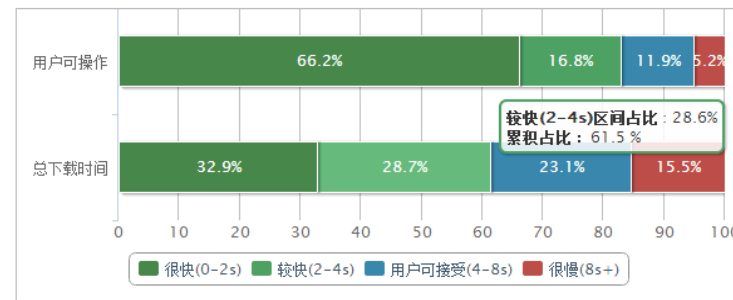
比如有100个目标时间T等于2秒的样本。60个样本低于4秒，30个样本在4到8秒之间，余下的10个样本高于8秒

Apdex的算式如下：

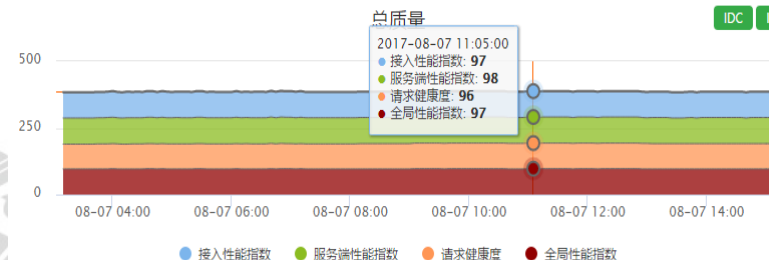
$$0.753 = (60 + 30/2) / 100$$

结果中的3代表目标响应时间。Apdex的结果始终显示与指数相关的目标时间。

Apdex报告提供了展示应用状况的独特视图。用户可以轻松掌握每个应用的真正相关性能。例如，两个指数为0.85T的应用提供了相同水平的用户体验。



URI	请求总数	请求时间 ms	响应时间 ms	处理时间 比%	响应时间 比%	流量 比%	请求 Apdex	响应 Apdex
/	49676	25	25	200	236	1	99	99
/serverinfos.do	33241	263	233	1360	1457	692	99	93
/updateManualLevel.do	22007	332	319	1134	1318	0	99	81
/test	17173	39	34	105	110	1	99	99
/getServiceInfo/getWsMd5Content.do	9213	779	384	1114	664	1319	57	63
/Report.do	7702	300	299	358	432	0	99	73
/server/list.do	6077	2113	1239	1993	1413	7735	69	27
/list.do	4552	525	516	371	441	13	94	34
/del.do	4498	192	179	134	151	0	99	98
/yip.do	4279	894	894	593	718	1	99	40
/tipinfo.do	4120	126	126	81	98	0	100	95
/add.do	3981	181	174	111	130	0	99	98
/ss/list.do	2788	966	882	418	461	143	80	0
/type/getAll.do	2637	51	51	21	25	0	100	99
/witch/list.do	2487	3276	3244	1264	1514	23	87	87



账单计算手段

项	AA云	TT云	XX云
计算方式	基于访问日志统计带宽*系数	基于访问日志统计带宽*系数	基于网卡出口带宽统计
数据确定方式	仅参考厂商后台	仅参考厂商后台	仅参考厂商后台
对账手段	带宽与业务指标（UV）对比	带宽与业务指标（UV）对比	带宽与业务指标（UV）对比
对账精度	3-8%之间		
数据差异较大的处理方法	无	无	无

新账单计算手段

项	AA云	TT云	XX云
计算方式	基于访问日志统计带宽*系数		
数据确定方式	虎牙CDN独立复算数据		
对账手段	拨测带宽与CDN日志带宽对比 CDNAPI带宽与CDN日志带宽对比		
对账精度	拨测误差：<1% 总带宽误差： <0.5%		
数据差异较大的处理方法	拨测误差<1% 总带宽误差<1%		



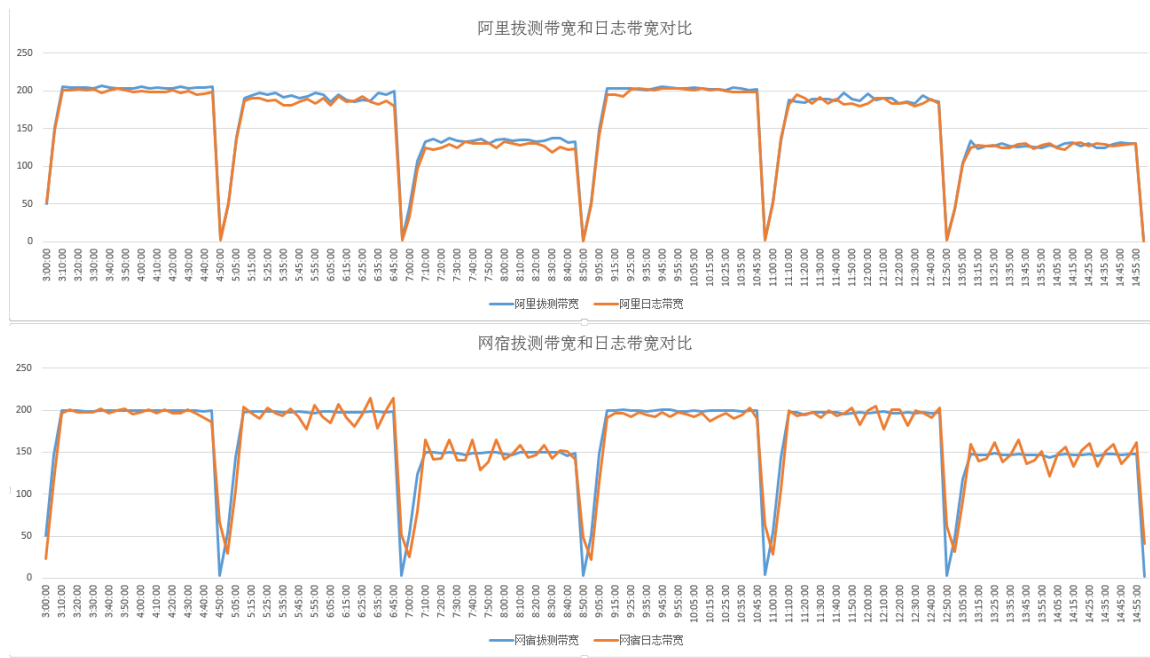
- 拨测实际下载带宽 VS CDN日志记录带宽
 - 覆盖90%+流量
 - 拨测时间段凌晨3点到下午3点
 - 拨测带宽保证一定的量
- CDN日志复算总带宽VS CDN API计费带宽
 - 次日6点以后下载全部日志后计算
 - API获取全部域名带宽数据统计
- 网卡流量 VS CDN日志计算带宽
 - 复算带宽系数

拨测数据与CDN厂商日志数据对比

阿里拨测日总带宽/阿里CDN日志日总带宽 = **1.0062**

网宿拨测日均带宽/网宿CDN日志日均带宽 = **0.9998**

腾讯拨测日总带宽/腾讯CDN日志日总带宽 = **1.0061**



三线路码率偏差对比

- AA:
2000Kbps: **0.992**; 4000Kbps: **0.927**;
 - BB:
2000Kbps: **0.996**; 4000Kbps: **0.926**;
 - CC:
2000Kbps: **0.992**; 4000Kbps: **0.925**;
- 不同码率偏差是不同，BB偏差最小



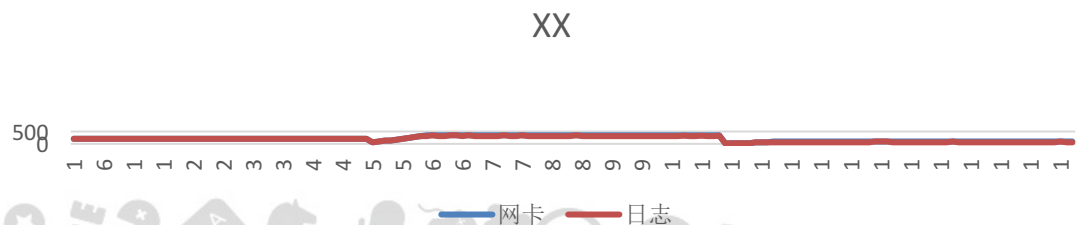
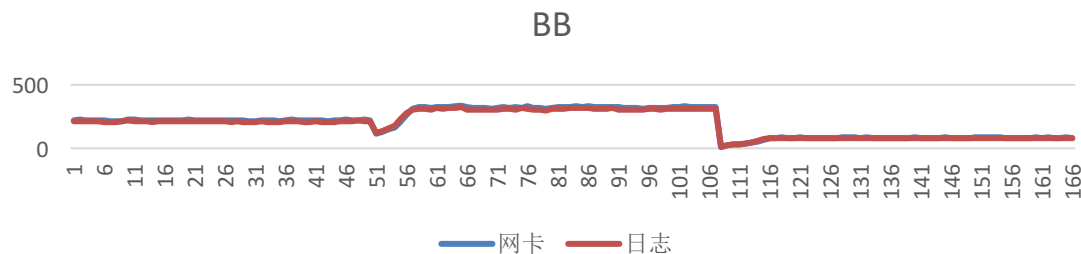
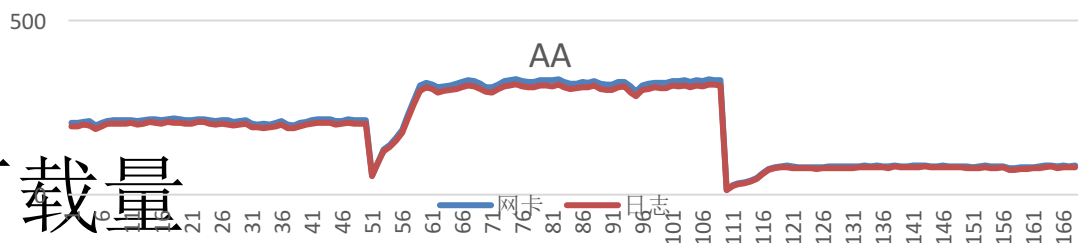
网卡系数

- 网卡流量/下载量

- AA 1.031

- BB 1.049

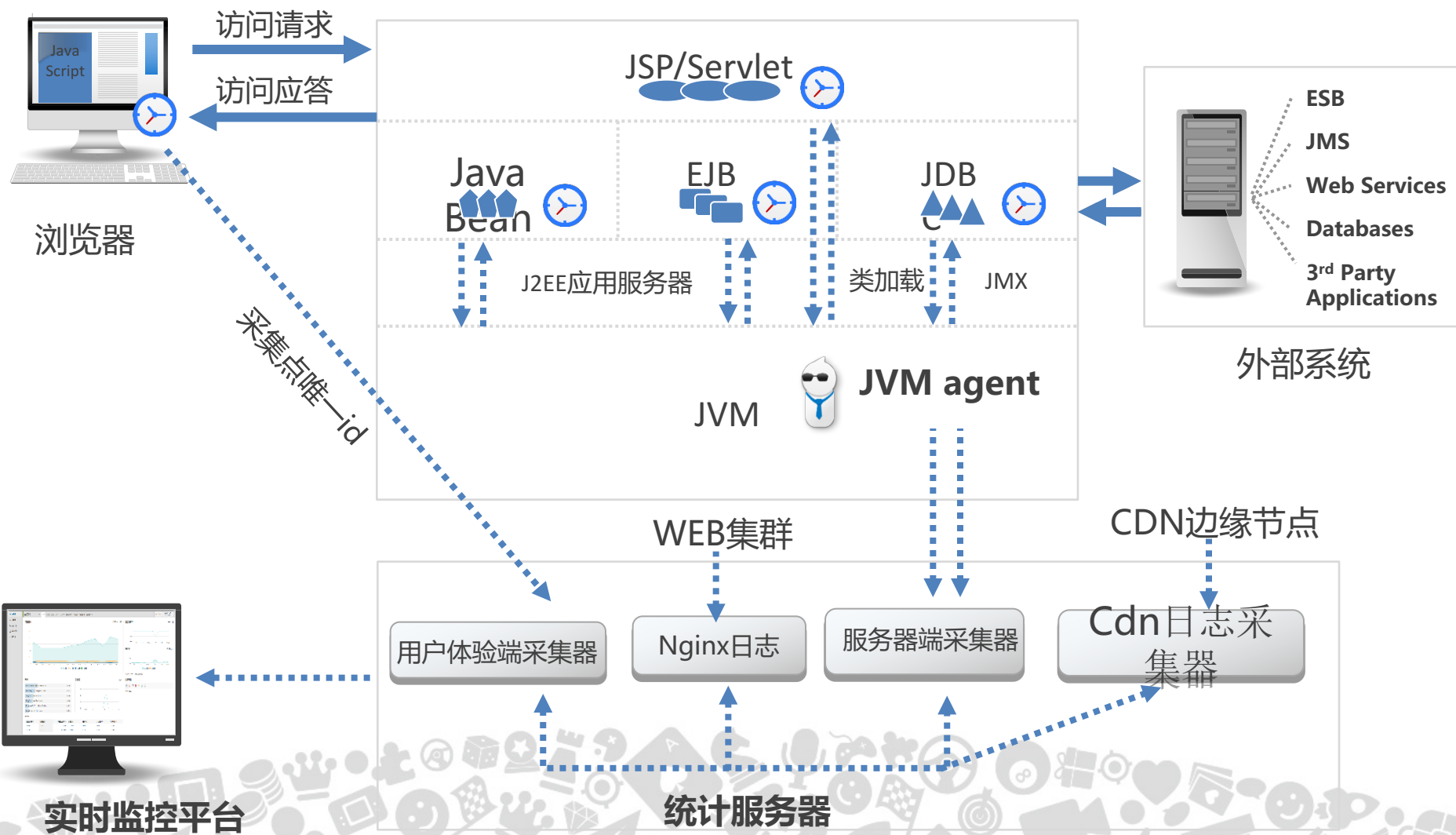
- XX 1.043



计费带宽数据对比

厂商	日期	厂商API数据		厂商日志复算数据				拨测日志误差率 拨测客户端带宽/ 拨测CDN日志复算 带宽
		峰值带宽 (M)	峰值时间点	峰值带宽 (M)	峰值时间点	站外带宽 (M)	计费误差率	
XX云	2018-xx-06	xxx	2018-xx-06 21:45	xxx	2018-10-06 21:40	xx	6xx.53	0.74%
BB云	2018-xx-06	xxx	2018-xx-06 18:35	xxx	2018-10-06 18:35	xx	1xx.77	0.10%
AA云	2018-xx-06	xxx	2018-xx-06 18:35	xxx	2018-10-06 18:35	xx	2xx.64	1.93%





- 1 ELK
- 2 Druid.io, Kylin
- 3 Storm, Spark

