

Hotdegree 指标说明

版本信息：

| 版本号 | 更新者 | 修订日期 | 说明 |
|-----|-----|------------|---------------------------|
| 1.0 | 缪璐璟 | 2014/09/27 | 创建 |
| 1.1 | 缪璐璟 | 2014/12/12 | 增加 19、20、21 指标 |
| 2.0 | 缪璐璟 | 2014/12/18 | 增加 LGW 业务指标 |
| 2.1 | 缪璐璟 | 2015/01/06 | 新增 cdn 业务中 url 指标 |
| 2.2 | 缪璐璟 | 2015/01/23 | 新增 rtmp 相关日志说明 |
| 2.3 | 缪璐璟 | 2015/01/26 | 新增日志来源为 uce 的日志说明 |
| 2.3 | 缪璐璟 | 2015/03/03 | 补充 rtmp 相关日志计算 |
| 2.4 | 缪璐璟 | 2015/03/16 | Rtmp 日志增加 instanceName 字段 |

目 录

| | | |
|----------|---|----|
| 1 | CDN 业务 | 4 |
| 1.1 | http 相关日志 | 4 |
| 1.1.1 | 日志采集 | 4 |
| 1.1.1.1 | 基于 HSM | 4 |
| 1.1.1.2 | 基于 UCE | 5 |
| 1.1.2 | 日志统计-基于 HSM | 5 |
| 1.1.2.1 | areaFlow_*.txt 日志统计 | 5 |
| 1.1.2.2 | areaHA_*.txt 日志统计 | 6 |
| 1.1.2.3 | allHAPercentage_*.txt 日志统计 | 6 |
| 1.1.2.4 | statusPercentage_*.txt 日志统计 | 7 |
| 1.1.2.5 | layerResponsetime_*.txt 日志统计 | 8 |
| 1.1.2.6 | areabackFlow_*.txt 日志统计 | 8 |
| 1.1.2.7 | hsmResponsetime_*.txt 日志统计 | 9 |
| 1.1.2.8 | hotContent_hour_*.txt 日志统计 | 9 |
| 1.1.2.9 | hotContent_day_*.txt 日志统计 | 10 |
| 1.1.2.10 | userResource_hour_*.txt 日志统计 | 10 |
| 1.1.2.11 | userResource_day_*.txt 日志统计 | 10 |
| 1.1.2.12 | userRefer_hour_*.txt 日志统计 | 11 |
| 1.1.2.13 | userRefer_day_*.txt 日志统计 | 11 |
| 1.1.2.14 | userDistrict_hour_*.txt 日志统计 | 12 |
| 1.1.2.15 | userDistrict_backflow_hour_*.txt 日志统计 | 12 |
| 1.1.2.16 | uv_hour_*.txt 日志统计 | 13 |
| 1.1.2.17 | uv_day_*.txt 日志统计 | 13 |
| 1.1.2.18 | gsd_*.txt 日志统计 | 13 |
| 1.1.2.19 | auth_*.txt 日志统计 | 14 |
| 1.1.2.20 | searchEngine_hour_*.txt 日志统计 | 14 |
| 1.1.2.21 | UserAgent_hour_*.txt 日志统计 | 14 |
| 1.1.2.22 | url_*.txt 日志统计 | 15 |
| 1.1.3 | 日志统计—基于 UCE | 15 |
| 1.1.3.1 | areaFlow_*.txt 日志统计 | 15 |
| 1.1.3.2 | areaHA_*.txt 日志统计 | 15 |
| 1.1.3.3 | allHAPercentage_*.txt 日志统计 | 16 |
| 1.1.3.4 | statusPercentage_*.txt 日志统计 | 17 |
| 1.1.3.5 | layerResponsetime_*.txt 日志统计 | 18 |
| 1.1.3.6 | areabackFlow_*.txt 日志统计 | 18 |
| 1.1.3.7 | hsmResponsetime_*.txt 日志统计 | 19 |
| 1.1.3.8 | hotContent_hour_*.txt 日志统计 | 19 |
| 1.1.3.9 | hotContent_day_*.txt 日志统计 | 20 |
| 1.1.3.10 | userResource_hour_*.txt 日志统计 | 20 |
| 1.1.3.11 | userResource_day_*.txt 日志统计 | 20 |

| | | |
|----------|---|----|
| 1.1.3.12 | userRefer_hour_*.txt 日志统计 | 21 |
| 1.1.3.13 | userRefer_day_*.txt 日志统计 | 21 |
| 1.1.3.14 | userDistrict_hour_*.txt 日志统计 | 22 |
| 1.1.3.15 | userDistrict_backflow_hour_*.txt 日志统计 | 22 |
| 1.1.3.16 | uv_hour_*.txt 日志统计 | 23 |
| 1.1.3.17 | uv_day_*.txt 日志统计 | 23 |
| 1.1.3.18 | gsd_*.txt 日志统计 | 23 |
| 1.1.3.19 | searchEngine_hour_*.txt 日志统计 | 24 |
| 1.1.3.20 | UserAgent_hour_*.txt 日志统计 | 24 |
| 1.1.3.21 | url_*.txt 日志统计 | 24 |
| 1.2 | Rtmp 相关日志 | 25 |
| 1.2.1 | 日志采集 | 25 |
| 1.2.2 | 日志统计 | 25 |
| 1.2.2.1 | rtmp_down_serviceFlow_*.txt 日志统计 | 25 |
| 1.2.2.2 | rtmp_up_serviceFlow_*.txt 日志统计 | 26 |
| 2 | LGW 业务 | 27 |
| 2.1 | 日志采集 | 27 |
| 2.1.1 | NAT 日志 | 27 |
| 2.1.2 | http 日志 | 28 |
| 2.1.3 | VPN 日志 | 28 |
| 2.1.4 | Hotdegree 建表 | 29 |
| 2.2 | 日志统计 | 30 |
| 2.2.1 | ExpressLane_service_flow_*.txt 日志统计 | 30 |
| 2.2.2 | ExpressLane_refer_hour_*.txt | 31 |
| 2.2.3 | ExpressLane_uv_hour_*.txt | 31 |
| 2.2.4 | ExpressLane_response_status_*.txt | 32 |
| 2.2.5 | ExpressLane_response_times_*.txt | 32 |
| 2.2.6 | ExpressLane_pvuv_*.txt | 32 |
| 2.2.7 | ExpressLane_user_agent_hour_*.txt | 33 |
| 2.2.8 | ExpressLane_url_*.txt | 33 |
| 2.2.9 | ExpressLane_uv_day_*.txt | 34 |

1 CDN 业务

1.1 http 相关日志

1.1.1 日志采集

1.1.1.1基于 HSM

| | |
|-------------|---|
| 日志格式 | <p>Hsm日志字段：</p> <p>1 \$http_x_session_id 会话唯一标识 2 \$channel 频道标识 3 \$msec 请求时间unix时间戳，当前时间 4 \$request_method 请求方法（GET等） 5 \$request_uri 请求的uri路径 6 \$status 状态码 7 \$response_time 8 \$request_header_length 请求头长 9 \$segment_bytes_sent 发送字节数 10 \$upstream_cache_status 回源状态码 11 0 12 \$remote_addr 代理客户的IP 13 \$label 服务设备标识 14 1 15 \$http_referer refer信息 16 \$upstream_response_time 17 \$request_length 18 \$http_user_agent UA信息 19 \$host 访问主机名 20 \$ http_x_part_index 21 \$scheme 请求协议头 22 \$duration 访问时长 23 \$remote_user 24 \$server_addr 25 \$server_port 26 \$msec_first 开始时间 27 \$bytes_sent 本session累计流量 28 \$segment_type （0：标志日志开始；1：中间；2：访问结束） 29 \$upstream_addr 回源地址 30 \$status_auth 状态码</p> <p>当\$segment_type 为 2 时，msec 就是结束时间</p> |
| Hotdegree建表 | <p>建表对应字段：</p> <p>1 http_x_session_id STRING, 2 channel_name STRING, 3 msec DOUBLE, 4 request_method STRING, 5 request_uri STRING, 6 status_code SMALLINT, 7 response_time Double, 8 request_header_length INT, 9 segment_bytes_sent BIGINT, 10 is_hit STRING, 11 is_cache TINYINT, 12 remote_addr STRING, 13 cdn_id STRING, 14 device_type STRING, 15 http_referer STRING, 16 upstream_response_time STRING, 17 request_length STRING, 18 backlen STRING, 19 host STRING, 20 http_x_part_index STRING, 21 scheme STRING, 22 duration STRING, 23 remote_user STRING, 24 server_addr STRING, 25 server_port STRING, 26 msec_first DOUBLE, 27 bytes_sent BIGINT, 28 segment_type TINYINT, 29 upstream_addr STRING, 30 status_auth STRING</p> |

1.1.1.2基于 UCE

| | |
|-------------|---|
| 日志格式 | UCE日志字段： 1 \$http_x_session_id 会话唯一标识 2 \$channel 频道标识 3 \$msec 请求时间 unix 时间戳，当前时间 4 \$request_method 请求方法（GET 等） 5 \$request_uri 请求的 uri 路径 6 \$status 状态码 7 \$response_time 8 \$request_header_length 请求头长 9 \$body_bytes_sent 发送字节数 10 \$upstream_cache_status 回源状态码 11 1 12 \$http_x_forwarded_for 代理客户的 IP 13 \$label 服务设备标识 14 4 15 \$http_referer refer 信息 16 \$upstream_response_time 17 04 18 \$upstream_response_length 回源请求头长 19 \$host 访问主机名 20 \$http_x_part_index 21 \$scheme 请求协议头 22 \$duration 访问时长 23 \$remote_user 24 \$server_addr 25 \$server_port |
| Hotdegree建表 | 建表对应字段： 1 http_x_session_id STRING, 2 channel_name STRING, 3 msec DOUBLE, 4 request_method STRING, 5 request_uri STRING, 6 status_code SMALLINT, 7 response_time Double, 8 request_header_length INT, 9 body_bytes_sent BIGINT, 10 is_hist STRING, 11 is_cache TINYINT, 12 remote_addr STRING, 13 cdn_id STRING, 14 device_type STRING, 15 http_referer STRING, 16 upstream_response_time STRING, 17 request_length STRING, 18 backlen STRING, 19 host STRING, 20 http_x_part_index STRING |

1.1.2 日志统计-基于 HSM

1.1.2.1areaFlow_*.txt 日志统计

| | |
|------|--|
| 算法格式 | 1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次 2、五分钟计算一次，算法如下： 流量：sum (\$segment_bytes_sent +if (\$segment_type =2) \$request_length) 用户数：count (distinct \$http_x_session_id) 计算LHSM日志，去除探针 3、查看生成文件样式和数据准确性 时间 频道 域名 节点 省份 ISP 流量 用户数 2014-12-16 18:47 cy-cy cdn.cibn.com HTCDN-VOPE-L101-VID-02-hsm-1 411986555305 0 0 1142 2 |
|------|--|

1.1.2.2areaHA_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、五分钟计算一次，算法如下： 比率：计算304, 302, 404, 200状态码所占比率 总数：计算304, 302, 404, 200状态码总数 计算LHSM日志，去除探针</div> <div>3、查看生成文件样式和准确性 时间 频道 节点 比率 总数 2014-12-16 14:40 ott-video HTCDN-VOPE-L101-VID-01-hsm-1412046519 425 35.09 420</div> |
|------|---|

1.1.2.3 allHAPercentage_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、五分钟计算一次，算法如下： 时间范围：CASE <div>WHEN \$response_time < 0.02 THEN '-20'</div><div>WHEN 0.02<= \$response_time AND \$response_time < 0.04 THEN</div><div>' 20, 40'</div><div>WHEN 0.04<= \$response_time AND \$response_time < 0.06 THEN</div><div>' 40, 60'</div><div>WHEN 0.06<= \$response_time AND \$response_time < 0.08 THEN</div><div>' 60, 80'</div><div>WHEN 0.08<= \$response_time AND \$response_time < 0.10 THEN</div><div>' 80, 100'</div><div>WHEN 0.10<= \$response_time AND \$response_time < 0.12 THEN</div><div>' 100, 120'</div><div>WHEN 0.12<= \$response_time AND \$response_time < 0.14 THEN</div><div>' 120, 140'</div><div>WHEN 0.14<= \$response_time AND \$response_time < 0.16 THEN</div><div>' 140, 160'</div><div>WHEN 0.16<= \$response_time</div></div> |
|------|--|

| | |
|--|--|
| | <p>AND \$response_time < 1 THEN</p> <p>'160+'</p> <p>END</p> <p>所有用户：某段时间范围内所有用户数求和</p> <p>比率：某段时间范围内所有用户数求和/所有请求数</p> <p>计算LHSM日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 节点 时间范围 所有用户 比率</p> <p>2014-12-16</p> <p>13:50 ott-video HTCDN-VOPE-L101-VID-01-hsm-1412046519</p> <p>425 40,60 2 0.18</p> |
|--|--|

1.1.2.4statusPercentage_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、五分钟计算一次，算法如下：</p> <p>状态码：CASE</p> <pre> WHEN \$status = 200 THEN '200' WHEN \$status = 206 THEN '206' WHEN \$status = 301 THEN '301' WHEN \$status = 302 THEN '302' WHEN \$status = 304 THEN '304' WHEN \$status = 403 THEN '403' WHEN \$status = 404 THEN '404' WHEN \$status = 502 THEN '502' WHEN \$status = 504 THEN '504' WHEN \$status >= 400 AND \$status <500 AND \$status <> 403 AND \$status <> 404 THEN '4xx' WHEN \$status >= 500 AND \$status <600 AND \$status <> 502 AND \$status <> 504 THEN '5xx' ELSE 'other' </pre> <p>END</p> <p>用户数：该状态码所有用户数之和</p> <p>比率：该状态码所有用户数之和/该时间所有请求数</p> <p>计算LHSM日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 节点 状态码 用户数 比率</p> <p>2014-12-09</p> |
|------|---|

| | |
|--|---|
| | 02:31 wasu-tv 8081-HTCDN-HBWH-L101-VID-01-uce-1399199 619818 200 10 6.99 |
|--|---|

1.1.2.5layerResponsetime_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、五分钟计算一次，算法如下：</p> <p>分段响应时间：Sum(\$response_time), sum(IF(Instr(\$upstream_response_time, ',') > 0, Cast(Substr(\$upstream_response_time, 1, Instr(\$upstream_response_time, ',')-1) AS DOUBLE), Cast (\$upstream_response_time AS DOUBLE)))</p> <p>图片计算002、003、004、005日志，视频计算102、103、104、105日志</p> <p>去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 分段响应时间 2014-12-09 22:24 wasu-tv 0.058,24</p> |
|------|--|

1.1.2.6areabackFlow_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、五分钟计算一次，算法如下：</p> <p>回源流量：sum(IF(Instr(\$upstream_response_length, ',') > 0, Cast(Substr(\$upstream_response_length, 1, Instr(\$upstream_response_length, ',') - 1) AS INT), Cast (\$upstream_response_length AS INT)) - \$request_header_length)</p> <p>回源用户：所有用户数之和</p> <p>图片计算004、005日志，视频计算104、105日志</p> <p>只取\$upstream_response_length>0, \$17=04,\$ \$upstream_cache_status in ('MISS','EXPIRED')</p> <p>去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 域名 节点 回源流量 回源用户 2014-12-09 05:47 cztvs-cztvs video.cztv.com 8081-HTCDN-ZJHZ-L201-VID-08-uce-1394175820674 17893069 1</p> |
|------|---|

1.1.2.7hsmResponsetime_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、五分钟计算一次，算法如下：</p> <p>回源流量: Avg(\$response_time) GROUP BY \$channel, \$label, cast(cast(\$msec AS int) div 60 AS string)</p> <p>回源用户: 所有用户数之和</p> <p>计算LIHSM日志，去除探针</p> <p>条件: \$response_time去掉为空的，且<=1</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 节点 平均响应时间 样本总数</p> <p>2014-12-09</p> <p>18:04 wasu-tv HTCDN-ZJCC-L101-VID-01-hsm-1417670466484 0.055 10</p> |
|------|---|

1.1.2.8hotContent_hour_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一小时计算一次，算法如下：</p> <p>用户请求数: 一个小时内该域名的请求总数</p> <p>流量: sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length)</p> <p>计算LIHSM日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 url 用户请求数 流量 域名</p> <p>2014-12-09</p> <p>17 cibn-flv /9d62b8d8e01a555dfca2d8dc74b06f54/5486c613/2014/12/03/13/pc/e8224d27c9741370689b4bd1bcbfadf4_pc_sd.flv?id=8419&ctype=sd 2 4453092 cibn.cdn.china.com</p> |
|------|---|

1.1.2.9 hotContent_day_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <ol style="list-style-type: none">1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次2、一天计算一次，算法如下： 用户请求数：一天内该域名的请求的总数 流量：sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length) 计算LIHSM日志，去除探针3、查看生成文件样式和准确性 时间 频道 url 用户请求数 流量 域名 2014-12-07 cibn-flv /795ff7374cd53a35a31a3df5e2aa4019/54847c45/2014/11/19/17/pc/47c0b225e131719d063d2427a6dfdbf9_pc_sd.flv?id=7888&ctype=sd 306 254536151 cibn.cdn.china.com |
|------|---|

1.1.2.10 userResource_hour_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <ol style="list-style-type: none">1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次2、一小时计算一次，算法如下： 用户请求数：一小时内内该域名的请求的总数 流量：sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length) 计算LIHSM日志，去除探针3、查看生成文件样式和准确性 时间 频道 请求ip 用户请求数 流量 2014-12-09 17 cibn-flv 175.0.131.166 3 27046623 |
|------|---|

1.1.2.11 userResource_day_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <ol style="list-style-type: none">1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次2、一天计算一次，算法如下： 用户请求数：一天内该域名的请求的总数 流量：sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length) 计算LIHSM日志，去除探针3、查看生成文件样式和准确性 时间 频道 请求ip 用户请求数 流量 2014-12-07 cibn-flv 218.92.115.5 230 142421566 |
|------|---|

1.1.2.12 userRefer_hour_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一小时计算一次，算法如下：</p> <p>url： Regexp_extract(\$http_referer, '(?<=//)((\\w)+\\.)+\\w+', 0)</p> <p>用户请求数：一小时内该域名的请求的总数</p> <p>流量：sum(\$segment_bytes_sent+if(\$segment_type=2)\$request_length)</p> <p>计算LIHSM日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 url 用户请求数 流量</p> <p>2014-12-09</p> <p>18 wasu-tv wasuott1.watone.net 449 14698615302</p> |
|------|--|

1.1.2.13 userRefer_day_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一天计算一次，算法如下：</p> <p>url： Regexp_extract(\$http_referer, '(?<=//)((\\w)+\\.)+\\w+', 0)</p> <p>用户请求数：一天内该域名的请求的总数</p> <p>流量：sum(\$segment_bytes_sent+if(\$segment_type=2)\$request_length)</p> <p>计算LIHSM日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 url 用户请求数 流量</p> <p>2014-12-08 cztvs-cztvs www.zjwmw.com 37 6639428</p> |
|------|---|

1.1.2.14 userDistrict_hour_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一小时计算一次，算法如下：</p> <p>区域id和ispid: Ipsearch(\$http_x_forwarded_for)</p> <p>根据ip地址库查询对应的区域和isp的id</p> <p>用户请求数: 一小时内该域名的请求的总数</p> <p>流量: sum(\$segment_bytes_sent+if(\$segment_type=2)\$request_length)</p> <p>计算LHSM日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 省份id ispid 用户请求数 流量 设备标识</p> <p>2014-12-09</p> <p>18 cibn-flv 120000 20003 1 22908460 HTCDN-ZJNB-L102-V</p> <p>ID-01-hsm-1397552888672</p> |
|------|---|

1.1.2.15 userDistrict_backflow_hour_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一小时计算一次，算法如下：</p> <p>区域id和ispid: Ipsearch(\$http_x_forwarded_for)</p> <p>根据ip地址库查询对应的区域和isp的id</p> <p>用户请求数: 一小时内该域名的请求的总数</p> <p>流量: Sum(\$segment_bytes_sent)</p> <p>图片计算004、005日志，视频计算104、105日志</p> <p>去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 省份id ispid 流量 用户请求数 设备标识</p> <p>2014-12-08</p> <p>23 wasu-tv 530000 20003 133768 1 8081-HTCDN-GDGZ-L201</p> <p>-VID-11-uce-1399973298208</p> |
|------|--|

1.1.2.16 uv_hour_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <ul style="list-style-type: none">1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次2、一小时计算一次，算法如下： uv数: Count (DISTINCT \$remote_addr) 计算LHSM日志，去除探针3、查看生成文件样式和准确性 时间 频道 uv数 2014-12-09 00 wasu-tv 3627 |
|------|---|

1.1.2.17 uv_day_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <ul style="list-style-type: none">1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次2、一天计算一次，算法如下： uv数: Count (DISTINCT \$remote_addr) 计算LHSM日志，去除探针3、查看生成文件样式和准确性 时间 频道 uv数 2014-12-08 cztvs-cztvs 5491 |
|------|---|

1.1.2.18 gsd_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <ul style="list-style-type: none">1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次2、五分钟计算一次，算法如下： 解析次数: 该域名解析的总次数 成功次数: sum(issuccess) 图片计算201日志，视频计算201日志3、查看生成文件样式和准确性 时间 域名 解析次数 成功次数 201412121842 up.hy.baidu.com 2 0 |
|------|--|

1.1.2.19 auth_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、五分钟计算一次，算法如下： 流量：sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length) 请求次数：count (distinct \$http_x_session_id) 计算LIHSM日志，去除探针</div> <div>3、查看生成文件样式和准确性 时间 频道 设备 认证结果状态码 请求次数 流量 2014-12-18 11:56 hktv-hktv HTCDN-VOPE-L101-VID-02-hsm-1411986555 305 310 5 3087179</div> |
|------|---|

1.1.2.20 searchEngine_hour_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、五分钟计算一次，算法如下： 流量：sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length) 请求次数：count (distinct \$http_x_session_id) 计算LIHSM日志，去除探针</div> <div>3、查看生成文件样式和准确性 时间 频道 省份 isp 搜索引擎分布 访问次数 流量 2014-12-16 14 ott-video 330000 20004 soso 42059 16972664</div> |
|------|--|

1.1.2.21 UserAgent_hour_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、一小时计算一次，算法如下： 流量：sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length) 请求次数：count (distinct \$http_x_session_id) 计算LIHSM日志，去除探针</div> <div>3、查看生成文件样式和准确性 时间 频道 省份 isp 浏览器 访问次数 流量 2014-12-18 11 hktv-hktv 0 0 Chrome 973 493951013</div> |
|------|---|

1.1.2.22 url_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、一小时计算一次，算法如下： 文件大小：max (\$bytes_sent) where \$segment_type=2 请求次数：count (distinct \$http_x_session_id) 域名、url：配置文件中制定 计算LIHSM日志，去除探针</div> <div>3、查看生成文件样式和准确性 时间 域名 url 状态码 地区 isp 请求次数 文件大小 2015-01-06 21:03 vid.ts.com /1.ts 200 330000 20004 23 29350705</div> |
|------|--|

1.1.3 日志统计—基于 UCE

1.1.3.1areaFlow_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、五分钟计算一次，算法如下： 流量：sum (\$request_header_length+\$body_bytes_sent+if (\$http_x_part_index=1) \$request_length) 用户数：count (distinct \$http_x_session_id) 计算LIUCE日志，去除探针</div> <div>3、查看生成文件样式和数据准确性 时间 频道 域名 节点 省份 ISP 流量 用户数 2014-12-16 18:47 cy-cy cdn.cibn.com HTCDN-VOPE-L101-VID-02-hsm-1 411986555305 0 0 1142 2</div> |
|------|---|

1.1.3.2areaHA_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、五分钟计算一次，算法如下： 比率：计算304, 302, 404, 200状态码所占比率 总数：计算304, 302, 404, 200状态码总数 计算LIUCE日志，去除探针</div> <div>3、查看生成文件样式和准确性 时间 频道 节点 比率 总数 2014-12-16 14:40 ott-video HTCDN-VOPE-L101-VID-01-hsm-1412046519</div> |
|------|--|

| | |
|--|---------------|
| | 425 35.09 420 |
|--|---------------|

1.1.3.3 allHAPercentage_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、五分钟计算一次，算法如下：</p> <p>时间范围：CASE</p> <pre> WHEN \$response_time < 0.02 THEN '-20' WHEN 0.02<= \$response_time AND \$response_time < 0.04 THEN ' 20, 40' WHEN 0.04<= \$response_time AND \$response_time < 0.06 THEN ' 40, 60' WHEN 0.06<= \$response_time AND \$response_time < 0.08 THEN ' 60, 80' WHEN 0.08<= \$response_time AND \$response_time < 0.10 THEN ' 80, 100' WHEN 0.10<= \$response_time AND \$response_time < 0.12 THEN ' 100, 120' WHEN 0.12<= \$response_time AND \$response_time < 0.14 THEN ' 120, 140' WHEN 0.14<= \$response_time AND \$response_time < 0.16 THEN ' 140, 160' WHEN 0.16<= \$response_time AND \$response_time < 1 THEN ' 160+' END </pre> <p>所有用户：某段时间范围内所有用户数求和 比率：某段时间范围内所有用户数求和/所有请求数 计算L1UCE日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 节点 时间范围 所有用户 比率 2014-12-16 13:50 ott-video HTCDN-VOPE-L101-VID-01-hsm-1412046519 425 40, 60 2 0.18</p> |
|------|--|

1.1.3.4statusPercentage_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、五分钟计算一次，算法如下：</div> <div>状态码：CASE</div> <div>WHEN \$status = 200 THEN '200'</div> <div>WHEN \$status = 206 THEN '206'</div> <div>WHEN \$status = 301 THEN '301'</div> <div>WHEN \$status = 302 THEN '302'</div> <div>WHEN \$status = 304 THEN '304'</div> <div>WHEN \$status = 403 THEN '403'</div> <div>WHEN \$status = 404 THEN '404'</div> <div>WHEN \$status = 502 THEN '502'</div> <div>WHEN \$status = 504 THEN '504'</div> <div>WHEN \$status >= 400</div> <div>AND \$status <500</div> <div>AND \$status <> 403</div> <div>AND \$status <> 404 THEN '4xx'</div> <div>WHEN \$status >= 500</div> <div>AND \$status <600</div> <div>AND \$status <> 502</div> <div>AND \$status <> 504 THEN '5xx'</div> <div>ELSE 'other'</div> <div>END</div> <div>用户数：该状态码所有用户数之和</div> <div>比率：该状态码所有用户数之和/该时间所有请求数</div> <div>计算LIUCE日志，去除探针</div> <div>3、查看生成文件样式和准确性</div> <div>时间 频道 节点 状态码 用户数 比率</div> <div>2014-12-09</div> <div>02:31 wasu-tv 8081-HTCDN-HBWH-L101-VID-01-uce-1399199</div> <div>619818 200 10 6.99</div> |
|------|--|

1.1.3.5layerResponsetime_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、五分钟计算一次，算法如下： 分段响应时间：Sum(\$response_time), sum(IF(Instr(\$upstream_response_time, ,') > 0, Cast(Substr(\$upstream_response_time, 1, Instr(\$upstream_response_time, ',')-1) AS DOUBLE), Cast (\$upstream_response_time AS DOUBLE))) 图片计算002、003、004、005日志，视频计算102、103、104、 105日志 去除探针</div> <div>2、查看生成文件样式和准确性 时间 频道 分段响应时间 2014-12-09 22:24 wasu-tv 0.058, 24</div> |
|------|---|

1.1.3.6areabackFlow_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、五分钟计算一次，算法如下： 回源流量：sum(IF(Instr(\$upstream_response_length, ,') > 0, Cast(Substr(\$upstream_response_length, 1, Instr(\$upstream_response_length, ',') - 1) AS INT), Cast (\$upstream_response_length AS INT)) - \$request_header_length) 回源用户：所有用户数之和 图片计算004、005日志，视频计算104、105日志 只取\$upstream_response_length>0, \$status=200,\$17=04,\$ \$upstream_cache_status in ('MISS', ' EXPIRED') 去除探针</div> <div>3、查看生成文件样式和准确性 时间 频道 域名 节点 回源流量 回源用户 2014-12-09 05:47 cztvs-cztvs video.cztv.com 8081-HTCDN-ZJHZ-L201 -VID-08-uce-1394175820674 17893069 1</div> |
|------|--|

1.1.3.7hsmResponsetime_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、五分钟计算一次，算法如下： 回源流量: Avg(\$response_time) GROUP BY \$channel, \$label, cast(cast(\$msec AS int) div 60 AS string) 回源用户: 所有用户数之和 计算LIUCE日志，去除探针 条件: \$response_time去掉为空的，且<=1</div> <div>3、查看生成文件样式和准确性 时间 频道 节点 平均响应时间 样本总数 2014-12-09 18:04 wasu-tv HTCDN-ZJCC-L101-VID-01-hsm-1417670466484 0.055 10</div> |
|------|--|

1.1.3.8hotContent_hour_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、一小时计算一次，算法如下： 用户请求数: 一个小时内该域名的请求总数 流量: sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length) 计算LIHSM日志，去除探针</div> <div>3、查看生成文件样式和准确性 时间 频道 url 用户请求数 流量 域名 2014-12-09 17 cibn-flv /9d62b8d8e01a555dfca2d8dc74b06f54/5486c613/2014/12/03/13/pc/e8224d27c9741370689b4bd1bcbfadf4_pc_sd.flv?id=8419&ctype=sd 2 4453092 cibn.cdn.china.com</div> |
|------|---|

1.1.3.9 hotContent_day_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一天计算一次，算法如下：</p> <p>用户请求数：一天内该域名的请求的总数</p> <p>流量：sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length)</p> <p>计算LIHSM日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 url 用户请求数 流量 域名</p> <p>2014-12-07 cibn-flv /795ff7374cd53a35a31a3df5e2aa4019/54847c45/2014/11/19/17/pc/47c0b225e131719d063d2427a6dfdbf9_pc_sd.flv?id=7888&ctype=sd 306 254536151 cibn.cdn.china.com</p> |
|------|---|

1.1.3.10 userResource_hour_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一小时计算一次，算法如下：</p> <p>用户请求数：一小时内内该域名的请求的总数</p> <p>流量：sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length)</p> <p>计算LIHSM日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 请求ip 用户请求数 流量</p> <p>2014-12-09 17 cibn-flv 175.0.131.166 3 27046623</p> |
|------|---|

1.1.3.11 userResource_day_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一天计算一次，算法如下：</p> <p>用户请求数：一天内该域名的请求的总数</p> <p>流量：sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length)</p> <p>计算LIHSM日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 请求ip 用户请求数 流量</p> <p>2014-12-07 cibn-flv 218.92.115.5 230 142421566</p> |
|------|---|

1.1.3.12 userRefer_hour_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一小时计算一次，算法如下：</p> <p>url： Regexp_extract(\$http_referer,'(?<=//)((\\w)+\\.)+\\w+', 0)</p> <p>用户请求数：一小时内该域名的请求的总数</p> <p>流量：sum(\$segment_bytes_sent+if(\$segment_type=2)\$request_length)</p> <p>计算LIHSM日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 url 用户请求数 流量</p> <p>2014-12-09</p> <p>18 wasu-tv wasuott1.watone.net 449 14698615302</p> |
|------|---|

1.1.3.13 userRefer_day_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一天计算一次，算法如下：</p> <p>url： Regexp_extract(\$http_referer,'(?<=//)((\\w)+\\.)+\\w+', 0)</p> <p>用户请求数：一天内该域名的请求的总数</p> <p>流量：sum(\$segment_bytes_sent+if(\$segment_type=2)\$request_length)</p> <p>计算LIHSM日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 url 用户请求数 流量</p> <p>2014-12-08 cztvs-cztvs www.zjwmw.com 37 6639428</p> |
|------|--|

1.1.3.14 userDistrict_hour_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、一小时计算一次，算法如下： 区域id和ispid: Ipsearch(\$http_x_forwarded_for) 根据ip地址库查询对应的区域和isp的id 用户请求数: 一小时内该域名的请求的总数 流量: sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length) 计算LHSM日志，去除探针</div> <div>3、查看生成文件样式和准确性 时间 频道 省份id ispid 用户请求数 流量 设备标识 2014-12-09 18 cibn-flv 120000 20003 1 22908460 HTCDN-ZJNB-L102-V ID-01-hsm-1397552888672</div> |
|------|---|

1.1.3.15 userDistrict_backflow_hour_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <div>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</div> <div>2、一小时计算一次，算法如下： 区域id和ispid: Ipsearch(\$http_x_forwarded_for) 根据ip地址库查询对应的区域和isp的id 用户请求数: 一小时内该域名的请求的总数 流量: Sum(\$segment_bytes_sent) 图片计算004、005日志，视频计算104、105日志 去除探针</div> <div>2、查看生成文件样式和准确性 时间 频道 省份id ispid 流量 用户请求数 设备标识 2014-12-08 23 wasu-tv 530000 20003 133768 1 8081-HTCDN-GDGZ-L201 -VID-11-uce-1399973298208</div> |
|------|--|

1.1.3.16 uv_hour_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <ol style="list-style-type: none">1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次2、一小时计算一次，算法如下： uv数: Count (DISTINCT \$remote_addr) 计算L1HSM日志，去除探针3、查看生成文件样式和准确性 时间 频道 uv数 2014-12-09 00 wasu-tv 3627 |
|------|--|

1.1.3.17 uv_day_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <ol style="list-style-type: none">1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次2、一天计算一次，算法如下： uv数: Count (DISTINCT \$remote_addr) 计算L1HSM日志，去除探针3、查看生成文件样式和准确性 时间 频道 uv数 2014-12-08 cztvs-cztvs 5491 |
|------|--|

1.1.3.18 gsd_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <ol style="list-style-type: none">1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次2、五分钟计算一次，算法如下： 解析次数: 该域名解析的总次数 成功次数: sum(issuccess) 图片计算201日志，视频计算201日志3、查看生成文件样式和准确性 时间 域名 解析次数 成功次数 201412121842 up.hy.baidu.com 2 0 |
|------|--|

1.1.3.19 searchEngine_hour_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一小时计算一次，算法如下：</p> <p>流量：sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length)</p> <p>请求次数：count (distinct \$http_x_session_id)</p> <p>计算LIHSM日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 省份 isp 搜索引擎分布 访问次数 流量</p> <p>2014-12-16</p> <p>14 ott-video 330000 20004 soso 42059 16972664</p> |
|------|--|

1.1.3.20 UserAgent_hour_*.txt 日志统计

| | |
|------|--|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一小时计算一次，算法如下：</p> <p>流量：sum (\$segment_bytes_sent+if (\$segment_type=2) \$request_length)</p> <p>请求次数：count (distinct \$http_x_session_id)</p> <p>计算LIHSM日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 频道 省份 isp 浏览器 访问次数 流量</p> <p>2014-12-18 11 hktv-hktv 0 0 Chrome 973 493951013</p> |
|------|--|

1.1.3.21 url_*.txt 日志统计

| | |
|------|---|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、五分钟计算一次，算法如下：</p> <p>文件大小：um (\$request_header_length+\$body_bytes_sent +if (\$http_x_part_index=1) \$request_length)</p> <p>请求次数：count (distinct \$http_x_session_id)</p> <p>域名、url：配置文件中制定</p> <p>计算LIUCE日志，去除探针</p> <p>3、查看生成文件样式和准确性</p> <p>时间 域名 url 状态码 地区 isp 请求次数 文件大小</p> <p>2015-01-06</p> <p>21:03 vid.ts.com /1.ts 200 330000 20004 23 29350705</p> |
|------|---|

1.2 Rtmp 相关日志

1.2.1 日志采集

| | |
|------|--|
| 日志格式 | Hsm日志字段： 1 \$session_id 会话唯一标识 2 \$channel 频道 3 \$msec 请求时间戳 4 \$command 请求方法（GET 等） 5 \$tcurl 请求的 uri 路径 6 200 状态码 7 0.001 响应时间 8 0 请求头长 9 \$segment_bytes_sent 发送字节数 10 HIT 回源状态码 11 0 12 \$remote_addr 代理客户的 ip 13 - 14 1 15 - 16 0 17 \$bytes_received 18 - 19 - 20 - 21 rtmp 请求协议头 22 \$session_time 访问时长 23 - 24 - 25 1935 26 \$msec_first 27 \$bytes_sent 本 session 流量 28 \$segment_type 29 - 30 - 31 \$side |
|------|--|

| | |
|-------------|---|
| Hotdegree建表 | 建表对应字段： 1 session_id STRING,2 channel_name STRING,3 msec DOUBLE,4 command STRING,5 tcurl STRING,6 status_code SMALLINT,7 response_time Double,8 request_header_length INT,9 segment_bytes_sent BIGINT,10 is_hist STRING,11 is_cache TINYINT,12 remote_addr STRING,13 cdn_id STRING,14 device_type STRING,15 http_referer STRING,16 upstream_response_time STRING,17 request_length STRING, 18 http_user_agent STRING,19 host STRING,20 http_x_part_index STRING,21 scheme STRING,22 duration STRING,23 remote_user STRING,24 server_addr STRING,25 server_port STRING,26 msec_first DOUBLE,27 bytes_sent BIGINT,28 segment_type TINYINT,29 upstream_addr STRING,30 status_auth STRING,31 side STRING |
|-------------|---|

1.2.2 日志统计

1.2.2.1rtmp_down_serviceFlow_*.txt 日志统计

| | |
|------|--|
| 算法格式 | 1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次，计算服务的结果文件 2、五分钟计算一次，算法如下： 流量：\$segment_bytes_sent+\$bytes_received 总用户数：count(distinct session_id) 判断条件：\$side=sourth \$command=play 3、查看生成文件样式和数据准确性 |
|------|--|

| | |
|--|---|
| | 时间 频道 instanceName 流量 总请求数 总用户数 地区 isp 2015-03-04 14:51 live-live stione 18 12 1 330000 20001 |
|--|---|

1.2.2.2rtmp_up_serviceFlow_*.txt 日志统计

| | |
|------|--|
| 算法格式 | 1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次计算推送流量 2、五分钟计算一次，算法如下： 流量: \$segment_bytes_sent+\$bytes_received 总用户数: count(distinct session_id) 判断条件: \$side=sourth \$command=publish 3、查看生成文件样式和数据准确性 时间 频道 instanceName 流量 总请求数 总用户数 地区 isp 2015-03-04 14:53 live-live stone 404736033 3 1 330000 20001 |
|------|--|

2 LGW 业务

2.1 日志采集

2.1.1 NAT 日志

| | |
|------|---|
| 日志格式 | <PRI> YYYY-MM-DD HH:MM:SS UTC +0800 [hostname] SN message_type protocol orig_sip orig_dip orig_sport orig_dport reply_sip reply_dip reply_sport reply_dport inbytes outbytes appname sgateway dgateway |
| 举例 | 1418769473570 2014-12-16 22:39:37 UTC -700 [HTCDN-ZJJX-L102-LGW-01] FAD2KD3114000022 nat-route 6 115.236.57.110 29850 101.69.204.168 22784 115.236.57.111 38943 115.236.57.110 29850 0 0 401 GW_SELF ZJHZ-10.1.24.203_10.1.17.203 |

| 字段 | 说明 |
|-------------------------------------|---------------------------------|
| YYYY-MM-DD HH:MM:SS UTC +0800 | 时间，24 小时 |
| message_type | 分为三种：http-proxy、vpn、nat-route |
| protocol | 四层协议（TCP、UDP） |
| orig_sip | 源IP(original direction)，客户端ip |
| orig_sport | 源端口(original direction)，客户端port |
| orig_dip | 目的IP(original direction) |
| orig_dport | 目的端口(original direction) |
| reply_sip | 源IP(reply direction)，源站ip |
| reply_sport | 源端口(reply direction)，源站port |
| reply_dip | 目的IP(reply direction)，客户端ip |
| reply_dport | 目的端口(reply direction)，客户端port |
| inbytes | 进栈字节数(original direction)，请求流量 |
| outbytes | 出栈字节数(reply direction)，服务流量 |
| appname | 识别应用业务标识 |
| sgateway | 源快线网关 |
| dgateway | 目标快线网关 |

2.1.2 http 日志

| | |
|------|---|
| 日志格式 | <PRI> YYYY-MM-DD HH:MM:SS UTC +0800 [hostname] SN message_type c-ip s-ip s-port cs-method cs-host cs-uri-stem sc-status sc-bytes cs-bytes c-useragent c-refer app-id nexthop-ip nexthop-delay trans-time |
| 举例 | 1418769473570 2014-12-16 22:39:37 UTC -700 [HTCDN-ZJJX-L102-LGW-01] FAD2KD3114000022 http-proxy 115.236.57.110 115.236.57.111 152 GET www.302.com "/route.txt" 200 1900 155 "Opera/9.27 (Windows NT 5.2; U; zh-cn)" "unknown" 302 10.1.17.201 0 5 |

| 字段 | 说明 |
|-------------------------------------|-------------------------------|
| YYYY-MM-DD HH:MM:SS UTC +0800 | 时间，24 小时 |
| message_type | 分为三种：http-proxy、vpn、nat-route |
| c-ip | client 的 ip |
| s-ip | proxy 绑定的 ip，源站 ip |
| s-port | proxy 绑定的 port，源站 port |
| cs-method | 请求方法 |
| cs-host | 请求 host 字段值 |
| cs-uri-stem | 请求 uri，不包含参数部分 |
| sc-status | 服务返回码（如 200、404 等） |
| sc-bytes | 返回内容字节数，服务流量 |
| cs-bytes | 请求字节数 |
| c-useragent | 请求的头部的 User-Agent 字段 |
| c-refer | 请求的头部的 refer 字段 |
| app-id | 识别应用业务标识 |
| nexthop-ip | 下一跳 IP（这里基本上为下一快线网关或者 SP 服务器） |
| nexthop-delay | 下一跳延迟，第一阶段暂不支持 |
| trans-time | 请求传输完成时间，第一阶段暂不支持 |

2.1.3 VPN 日志

| | |
|------|--|
| 日志格式 | <PRI> YYYY-MM-DD HH:MM:SS UTC +0800 [hostname] SN message_type event user-id vpn-id client-ip remote-ip life-time last-time inbytes outbytes |
|------|--|

| 字段 | 说明 |
|-------|----------------------------------|
| event | ESTABLISH, DELETE, EXPIRE, REKEY |

| | |
|-----------|--------------------------|
| type | VPN 类型 IPsec/pptp/l2tp 等 |
| user-id | 用户 ID |
| vpn-id | VPN ID |
| client-ip | Client IP |
| remote-ip | 分配给移动接入用户的内网地址 |
| life-time | SA 生命周期，可以为时间或传输字节数 |
| last-time | 连接时长 |
| inbytes | 进入 VPN 隧道字节数 |
| outbytes | 出 VPN 隧道字节数 |

2.1.4 Hotdegree 建表

| | |
|-------------|--|
| Hotdegree建表 | <p>建表对应字段：</p> <p>1 datetime STRING, 2 hostname STRING, 3 sn STRING, 4 message_type STRING, 5 c_ip STRING, 6 s_ip STRING, 7 s_port STRING, 8 cs_method STRING, 9 cs_host STRING, 10 cs_uri_stem STRING, 11 sc_status STRING, 12 sc_bytes STRING, 13 cs_bytes STRING, 14 c_useragent STRING, 15 c_refer STRING, 16 app_id STRING, 17 nexthop_ip STRING, 18 nexthop_delay STRING, 19 trans_time STRING</p> |
| NAT比对 | <p>1 2014-12-16 22:39:37 UTC -700</p> <p>2 [HTCDN-ZJJX-L102-LGW-01]</p> <p>3 FAD2KD3114000022</p> <p>4 nat-route</p> <p>5 6</p> <p>6 115.236.57.110 客户端 ip</p> <p>7 29850</p> <p>8 101.69.204.168 服务入口 ip</p> <p>9 22784</p> <p>10 115.236.57.111 源站 ip</p> <p>11 38943</p> <p>12 115.236.57.110 客户端 ip</p> <p>13 29850</p> <p>14 0 进栈字节数，请求字节数</p> <p>15 0 出栈字节数，返回字节数</p> <p>16 401 app_id</p> <p>17 GW_SELF 源网关</p> <p>18 ZJHZ-10.1.24.203_10.1.17.203 目标快线网关</p> |

| | |
|--------|--|
| http比对 | 1 2014-12-16 22:39:37 UTC -700 |
| | 2 [HTCDN-ZJJX-L102-LGW-01] |
| | 3 FAD2KD3114000022 |
| | 4 http-proxy |
| | 5 115.236.57.110 客户端 ip |
| | 6 115.236.57.111 源站 ip |
| | 7 152 源站 port |
| | 8 GET |
| | 9 www.302.com |
| | 10 "/route.txt" |
| | 11 200 状态码 |
| | 12 1900 返回字节数 |
| | 13 155 请求字节数 |
| | 14 "Opera/9.27 (Windows NT 5.2; U; zh-cn)" |
| | 15 "unknown" |
| | 16 302 app_id |
| | 17 10.1.17.201 下一跳 |
| | 18 0 |
| | 19 5 |

2.2 日志统计

2.2.1 ExpressLane_service_flow_*.txt 日志统计

| | |
|------|---|
| 算法格式 | 1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次 |
| | 2、五分钟计算一次，算法如下： |
| | http-proxy |
| | 服务流量：sum sc_bytes |
| | 请求流量：sum cs_bytes |
| | 访问次数：所有访问数总和 |
| | NAT |
| | 服务流量：sum c_refer |
| | 请求流量：sum c_useragent |
| | 3、查看生成文件样式和数据准确性 |
| | http-proxy |
| | 时间 业务类型 设备ID 域名 APP-ID 访问次数 服务流量(B) 请求流量(B) |
| | 2014-12-19 04:30 http HTCDN-ZJJX-L102-LGW-01 www.302.com 302 49 93100 7595 |
| | NAT |
| | 时间 业务类型 设备ID - APP-ID - 服务流量(B) 请求流量 |

| | |
|--|---|
| | (B) 2014-12-19 04:29 nat HTCDN-ZJJX-L102-LGW-01 - 401 - 0 0 |
|--|---|

2.2.2 ExpressLane_refer_hour_*.txt

| | |
|------|--|
| 算法格式 | 1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次 2、一小时计算一次，算法如下： http-proxy 服务流量：sum sc_bytes 请求流量：sum cs_bytes 访问次数：所有访问数总和 3、查看生成文件样式和数据准确性 http-proxy 时间 业务类型 设备ID 域名 APP_ID refer 访问次数 服务流量(B) 请求流量(B) 2014-12-19 05 http HTCDN-ZJJX-L102-LGW-01 www.300.com 300 168 319200 28644 NAT 不统计 |
|------|--|

2.2.3 ExpressLane_uv_hour_*.txt

| | |
|------|--|
| 算法格式 | 2、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次 2、一小时计算一次，算法如下： http-proxy uv数：count(DISTINCT c_ip) 3、查看生成文件样式和数据准确性 http-proxy 时间 业务类型 设备ID 域名 APP_ID uv 2014-12-19 03 http HTCDN-ZJJX-L102-LGW-01 www.302.com 302 1 NAT 不统计 |
|------|--|

2.2.4 ExpressLane_response_status_*.txt

| | |
|------|---|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、五分钟计算一次，算法如下：</p> <p>http-proxy</p> <p>status: sc_status</p> <p>服务流量: sum sc_bytes</p> <p>请求流量: sum cs_bytes</p> <p>访问次数: 所有访问数总和</p> <p>3、查看生成文件样式和数据准确性</p> <p>http-proxy</p> <p>时间 业务类型 设备ID 域名 APP_ID status 访问次数 服务流量(B) 请求流量(B)</p> <p>2014-12-19</p> <p>06:06 http HTCDN-ZJJX-L102-LGW-01 www.302.com 302 200 31 58900 4805</p> <p>NAT</p> <p>不统计</p> |
|------|---|

2.2.5 ExpressLane_response_times_*.txt

| | |
|------|---|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、五分钟计算一次，算法如下：</p> <p>http-proxy</p> <p>response_time:sum(trans_time)</p> <p>3、查看生成文件样式和数据准确性</p> <p>http-proxy</p> <p>时间 业务类型 设备ID 域名 app_id response_time(时间内响应时间总和毫秒)</p> <p>2014-12-19</p> <p>04:30 http HTCDN-ZJJX-L102-LGW-01 www.300.com 300 567</p> <p>NAT</p> <p>不统计</p> |
|------|---|

2.2.6 ExpressLane_pvuv_*.txt

| | |
|------|--|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、五分钟计算一次，算法如下：</p> <p>http-proxy</p> <p>uv: count(DISTINCT c_ip)</p> |
|------|--|

| | |
|--|--|
| | <p>pv: 所有请求数之和</p> <p>3、查看生成文件样式和数据准确性</p> <p>http-proxy</p> <p>时间 业务类型 设备ID 域名 APP_ID uv pv</p> <p>2014-12-19</p> <p>06:04 http HTCDN-ZJJX-L102-LGW-01 www.300.com 300 1 96</p> <p>NAT</p> <p>不统计</p> |
|--|--|

2.2.7 ExpressLane_user_agent_hour_*.txt

| | |
|------|---|
| 算法格式 | <p>3、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一小时计算一次，算法如下：</p> <p>http-proxy</p> <p>服务流量: sum(sc_bytes)</p> <p>请求流量: sum(cs_bytes)</p> <p>3、查看生成文件样式和数据准确性</p> <p>http-proxy</p> <p>时间 业务类型 设备ID 域名 APP_ID userAgent 访问次数 服务流量(B) 请求流量(B)</p> <p>2014-12-19</p> <p>04 http HTCDN-ZJJX-L102-LGW-01 www.300.com 300 IE 5716 10860400 974578</p> <p>NAT</p> <p>不统计</p> |
|------|---|

2.2.8 ExpressLane_url_*.txt

| | |
|------|--|
| 算法格式 | <p>1、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、五分钟计算一次，算法如下：</p> <p>http-proxy</p> <p>url: 配置的指定url</p> <p>服务流量: sum(sc_bytes)</p> <p>请求流量: sum(cs_bytes)</p> <p>3、查看生成文件样式和数据准确性</p> <p>http-proxy</p> <p>时间 业务类型 设备ID 域名 APP_ID url 访问次数 服务流量(B) 请求流量(B)</p> <p>2014-12-19 06:01 http HTCDN-ZJJX-L102-LGW-01 www.163.com 191 /route.txt 16 30400 2496</p> |
|------|--|

| | |
|--|-------------------|
| | NAT 不统计 |
|--|-------------------|

2.2.9 ExpressLane_uv_day_*.txt

| | |
|------|---|
| 算法格式 | <p>2、日志采集生成对应的snappy文件到hdfs上，五分钟生成一次</p> <p>2、一天计算一次，每天2点计算前一天数据，算法如下：</p> <p>http-proxy</p> <p>uv: count(DISTINCT c_ip)</p> <p>pv: 所有请求数之和</p> <p>3、查看生成文件样式和数据准确性</p> <p>http-proxy</p> <p>时间 业务类型 设备ID 域名 APP_ID uv pv</p> <p>2015-01-20 http HTCDN-ZJXX-L102-LGW-01 upwt.qiniu.com</p> <p> 123 106 106</p> <p>NAT</p> <p>不统计</p> |
|------|---|