基于大数据的可视化日志分析

2020年5月13日上午11:00

Frank Zhang，甲骨文资深云解决方案专家

公益讲座11点准时开始，请大家先浏览云技术微信公众号技术文章
资料会在各群同步发布，已入群客户请勿重复入群！

扫码加入:
19c新特性讲座群

欢迎关注:
甲骨文云技术公众号

ORACLE

Log Analytics Cloud Service

基于大数据的可视化日志分析

Frank Zhang

Cloud Platform

2020

Safe harbor statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

在计算机世界中,“日志记录发生的所有的事”。
每个计算设备都在生成日志——应用程序、软件、硬件、平台、终端...

应用

2019-05-10T12:59:52.212: INFO: OrderApp-3212: Order type: Failed, cust: 933373, order: 3997396, region: APAC

中间件

<May 10, 2019 12:59:50 PM PST> <Error> <Server> <BEA-002608> <The Listen Thread closed because of an error>
java.sql.SQLException: Closed Connection
at oracle.jdbc.driver.SQLStateMapping.newSQLException(SQLStateMapping.java:70)
at processOrders.acme.com(SubmitOrder.java:112)

数据库

150510 12:59:45 [ERROR] /usr/sbin/mysqld: Incorrect key file for table '/tmp/#sql_21b2_0.MYI'; try to repair it
150510 12:59:45 [ERROR] Got an error from unknown thread, storage/myisam/mi_write.c:223
150510 12:59:45 [ERROR] /usr/sbin/mysqld: Sort aborted: Error writing file '/tmp/MYK74Kpi' (Errcode: 28)

网络

Jul 17 22:04:29 router last message repeated 2 times Jul 17 22:04:29 router dnsprobe[276]: Primary DNS server Is Down... Switching To Secondary DNS server Jul 17 22:05:08 router dnsprobe[276]: Switching Back To Primary DNS server

尽管日志数据很枯燥...但是日志信息中有巨大的机会...

- 包含所有信息的日志可能是垃圾，
- 也可能是金矿

进一步洞察，发现有价值的信息（金子）

应用

Order Number

2019-05-10T12:59:52.212: INFO: OrderApp-3212: Order type: Failed, cust: 933373, order: 3997396, region: APAC

中间件

Error Message

<May 10, 2019 12:59:50 [illegible] Server> <BEA-002608> <The Listen Thread closed because of an error>
java.sql.SQLException: Closed Connection
at oracle.jdbc.driver.SQLStateMapping.newSQLException(SQLStateMapping.java:70)
at processOrders.acme.com(SubmitOrder.java:112)

数据库

Time Stamp

Error Message

150510 12:59:45 [ERROR] /usr/sbin/mysqld: Incorrect key file for table '/tmp/#sql_21b2_0.MYI'; try to repair it
150510 12:59:46 [ERROR] Got an error from unknown thread, storage/myisam/mi_write.c:223
150510 12:59:46 [ERROR] /usr/sbin/mysqld: Sort aborted: Error writing file '/tmp/MYK74Kpi' (Errcode: 28)

网络

May 10, 2019 22:04:29 myhost.acme.com: router dnsprobe[276]: dns query failed
May 10, 2019 22:04:30 myhost.acme.com: router dnsprobe[276]: Primary DNS server Is Down... Switching To Secondary DNS server
May 10, 2019 22:04:30 myhost.acme.com: router dnsprobe[276]: Switching Back To Primary DNS server

Host Name

痛点很急迫，且日志量呈指数增长...

日志数据扩张

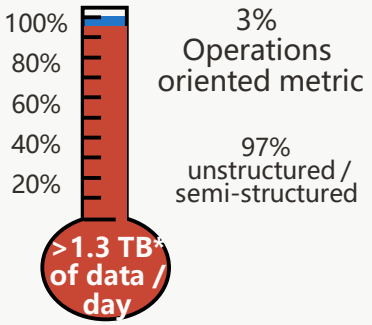
Server, storage, network, platform, and applications

- 日志数据分散在上千个IT基础设施、服务器、平台和应用程序中
- 日志内容和格式多样、专有
- 日志分析需要多领域的专业技能

弹性架构+敏捷开发

- 很难分析弹性架构创建的动态应用程序拓扑产生的的日志数据
- 使用基于规则的警报，难以使用敏捷开发者自创的新型日志

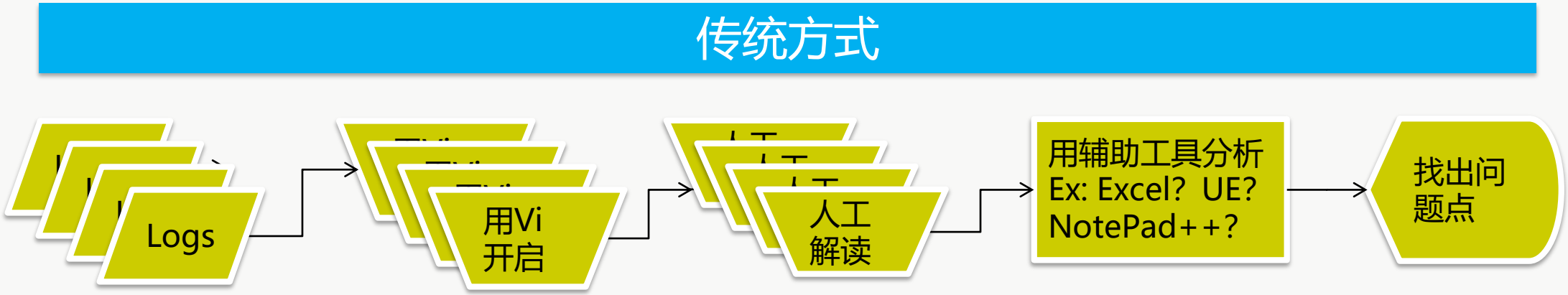
日志数据量



- IT基础设施和应用程序生成大量的日志数据,每年新增数据量大于40% **
- 日志关联更困难

* Data from IBM report: 5000 servers, 125 applications generate in excess of 1.3TB of data per day; 1 TB of log data equals 3 billion log entries
** Monash Research: Growth in machine-generated data

传统处理棘手故障的方式，故障排查（根因分析）非常耗时



I. 1人多系统，日志处理主要靠人眼+Ultra edit / Notepad等工具
II. 不能跨平台/系统进行关联性查询
III. 不能对时间段进行查询，不利于总结规律
IV. 不能利用日志进行统计/分析/报警等高附加值工作
V. 数据量稍微大一些就无能为力。

有没有更好的办法呢?

日志分析的场景

两大场景

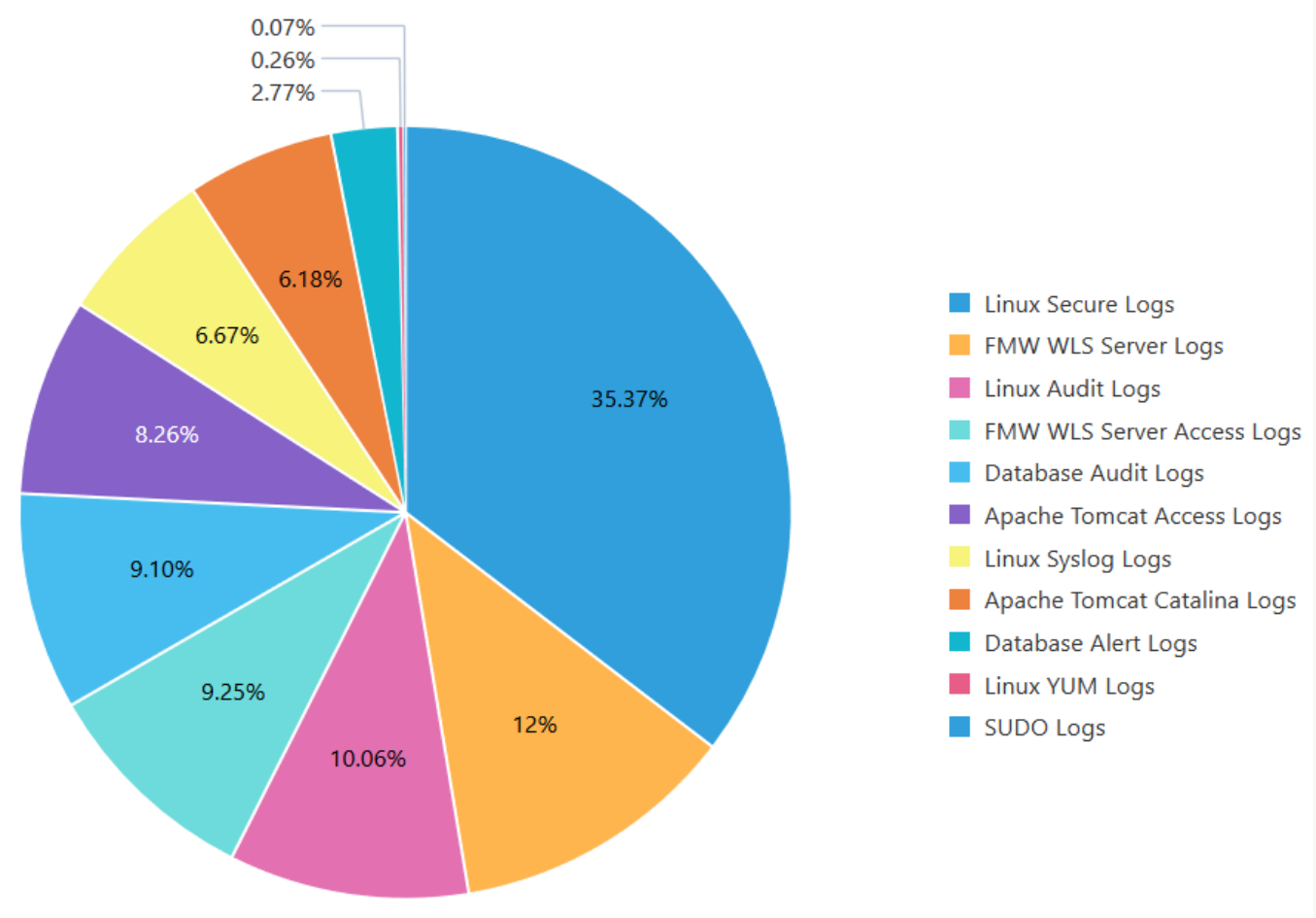
自动采集日志数据，
快速获取应有价值

从日志中获取“IT”
与“业务”两层面
的洞察力与可视化

日志几乎无所不能 —— 记录所有必要的信息

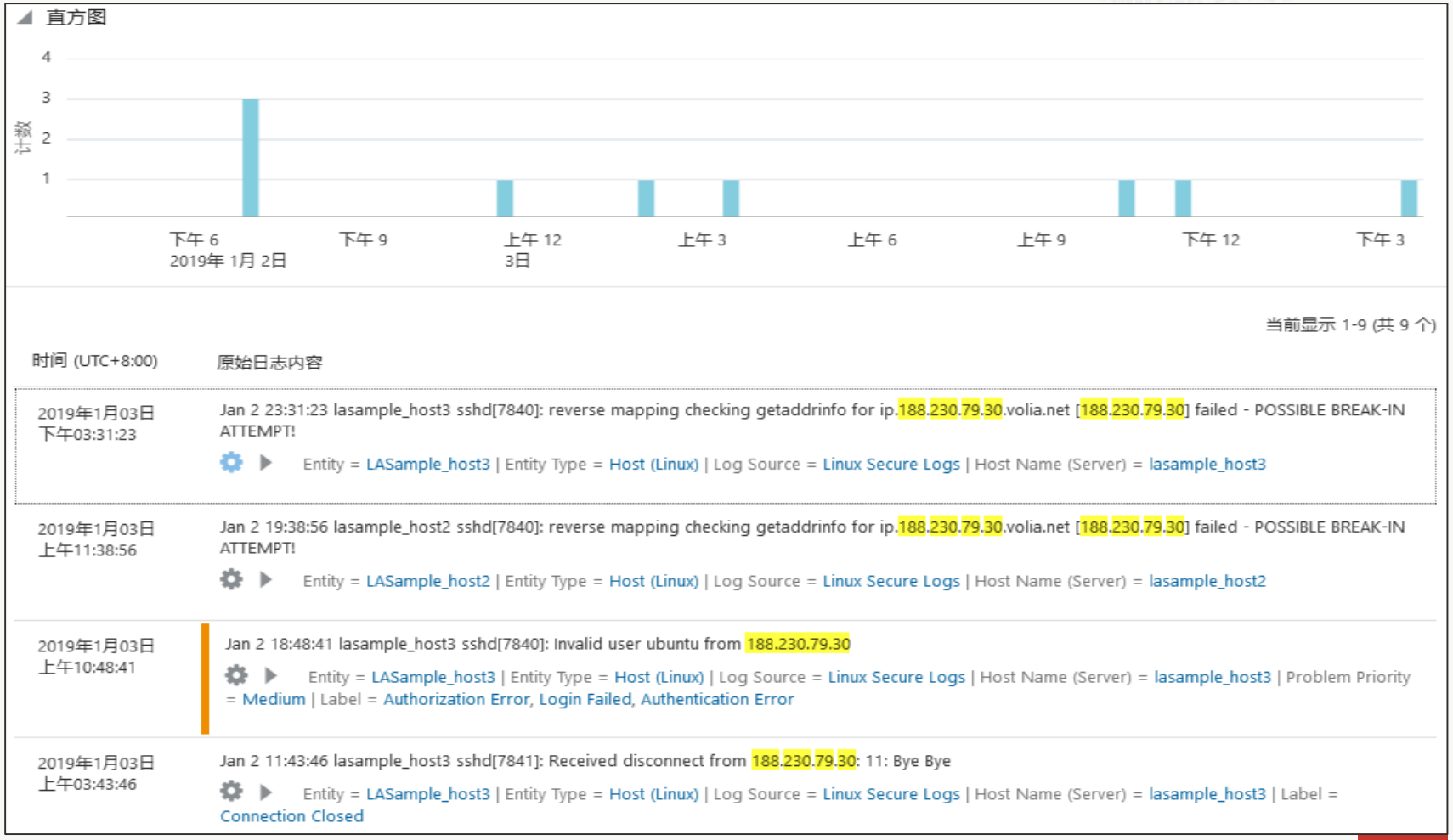
具体场景1：日志收集与统一存储

- 描述
  - 日志文件类型庞杂、增长量惊人、分布在各地，一般客户无法收集、存储和管理
- 举例
  - 部署在全国各省的应用的日志需要统一采集并分析
  - 某一个重要应用的所有日志需要及时备份、实时分析
  - 云上云下的统一应用运维（混合云、多云环境）



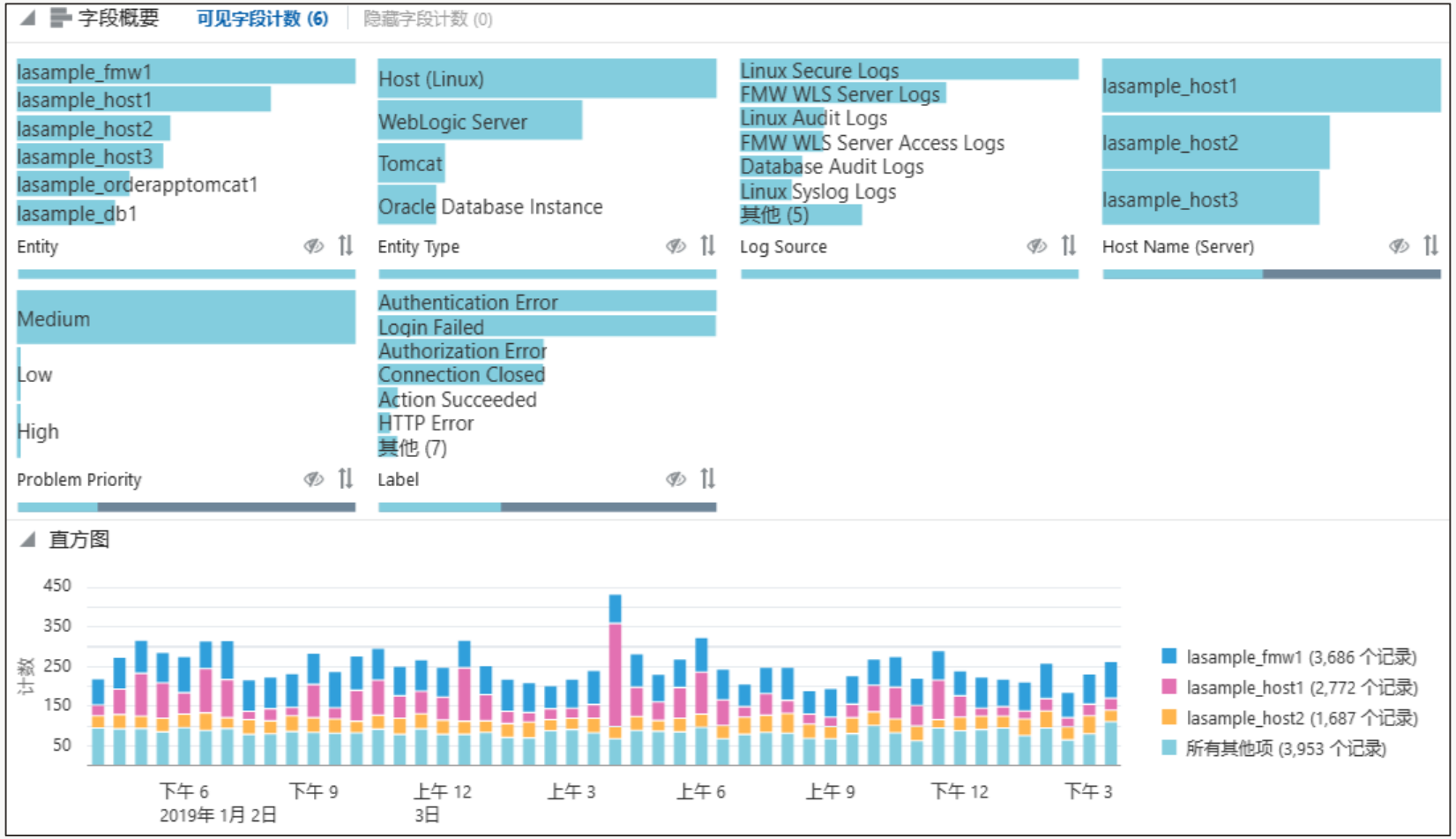
具体场景2：日志搜索

- 描述
  - 日志中的信息量非常巨大，包括调试、错误、提示、流程、业务、跟踪、进程等各种信息都可能记录在日志中，怎样快速找到我需要的信息？
- 举例
  - 昨天上午10：00左右所有有关IP 192.168.1.88的日志
  - 最近发生的Ora-00600之类的错误



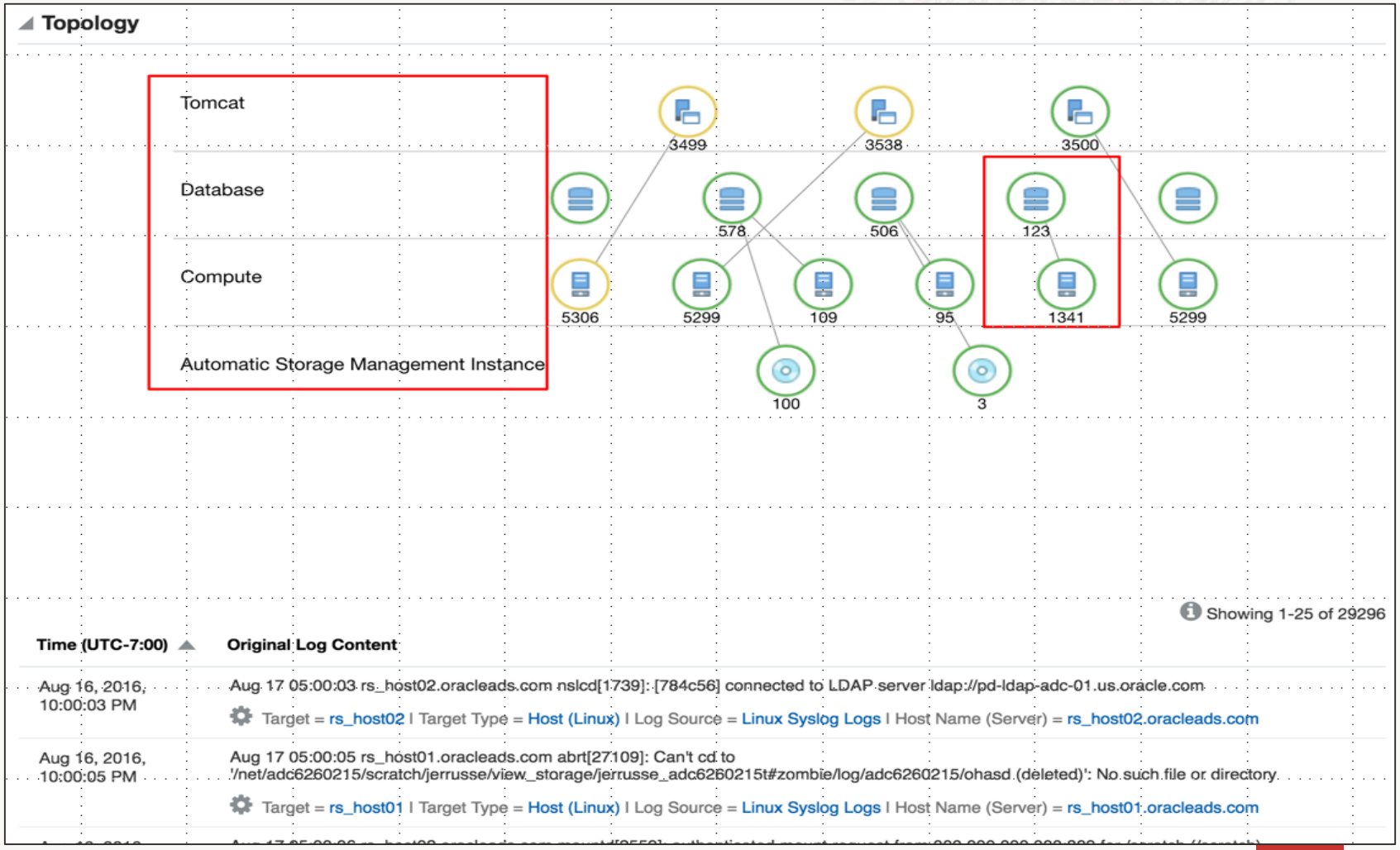
具体场景3：日志统计与分析

- 描述
  - 日志的长期积累（例如一年）是一大笔财富，从中可以发现很多高价值信息，但需要用到大数据分析技术，这不是普通客户能做到的。
- 举例
  - 企业有1000个数据库实例，最近一个季度最经常的报错是哪些，怎么分布（时间、地域）的？
  - 超大量的日志条目中，管理员需要关注哪些？



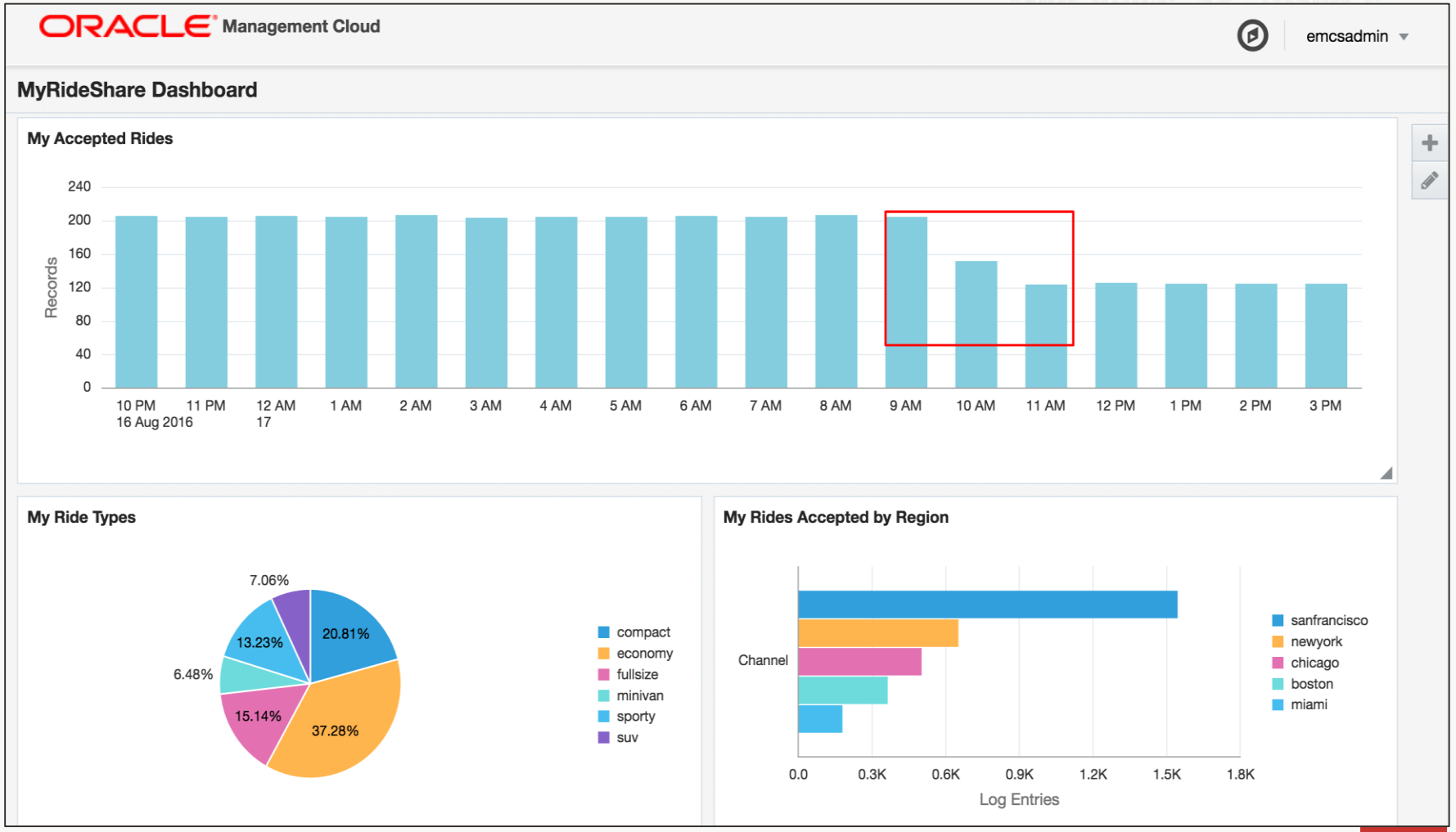
具体场景4：应用故障排查

- 描述
  - 一般地现有应用都涉及到主机、数据库、中间件（N台）、Web服务器（N台）、应用本身等多种IT设施（也包括网络），任何一个“点”发生故障都会造成业务故障，而探查这个“点”却不是那么容易
- 举例
  - 某电子商务的web页面突然无法访问，什么原因？
  - 感觉整体应用不稳定、有故障，但是说不出来究竟哪儿有问题。



具体场景5：一般性业务监控与简单的业务分析

- 描述
  - 不少客户会将业务状态等信息记录到日志中（中间件日志或个性化日志），当然日志中的敏感信息可以被脱敏，但不妨碍进行统计分析。
- 举例
  - 本月每天的订单成功量是多少，按时间、地域、业务属性（例如订单物品种类）的分布？



LACS：云、大数据、机器学习的最佳实践

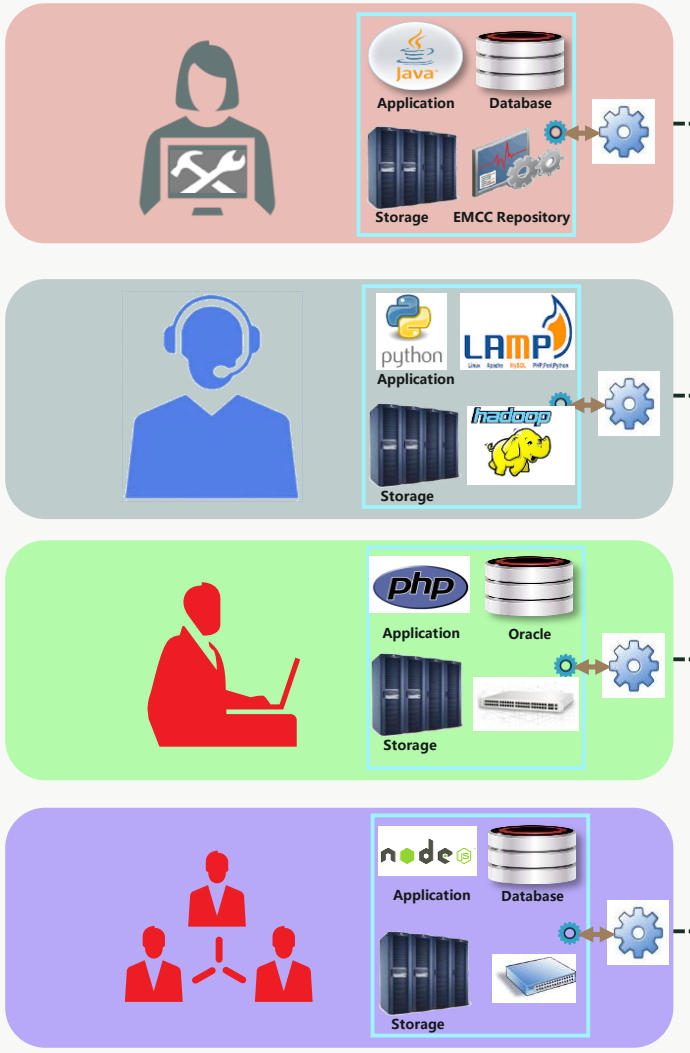
Log Analytics Cloud Service

客户数据中心
(含全球部署)

客户私有云

客户Oracle公有云

客户其他云



Logs + Operational Data

Logs + Operational Data

Logs + Operational Data

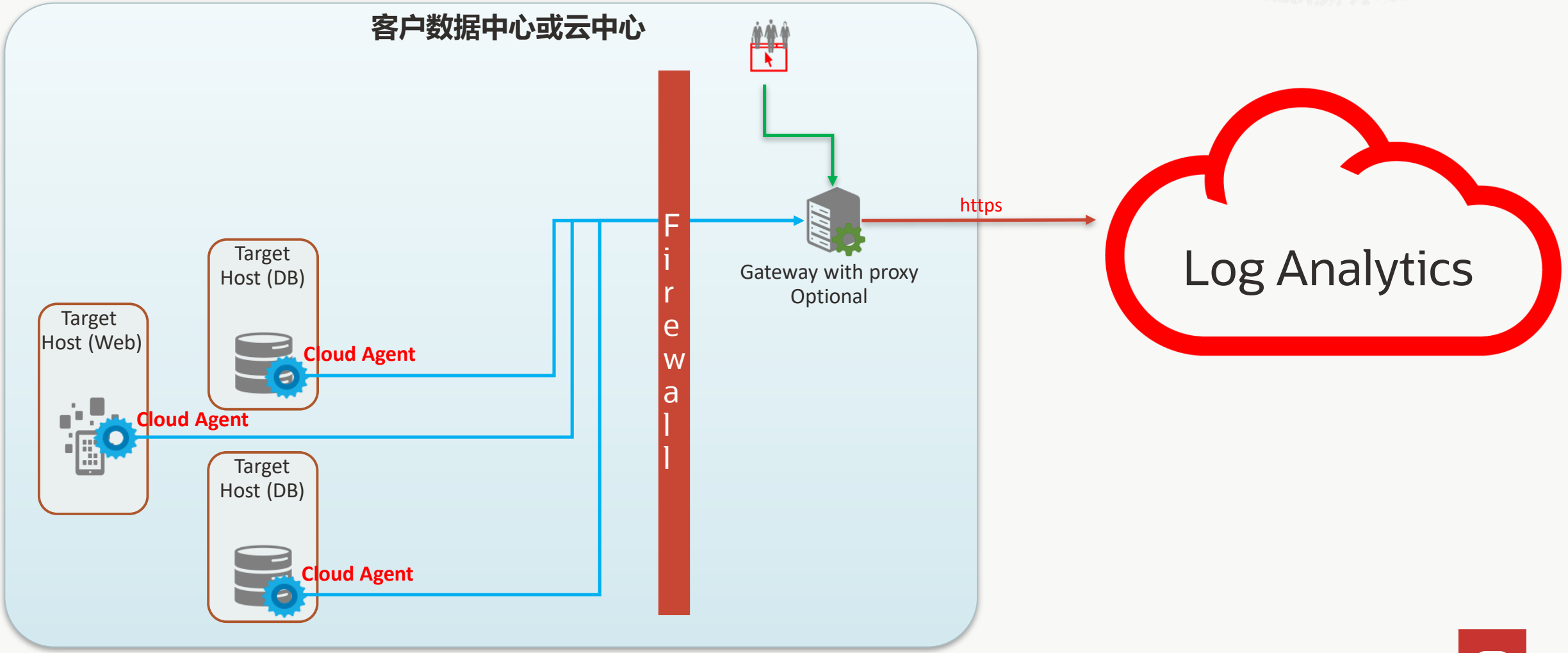
Logs + Operational Data

从应用程序和基础设施(本地和云)中，实时监控、集成、分析、搜索、探索、关联所有日志数据

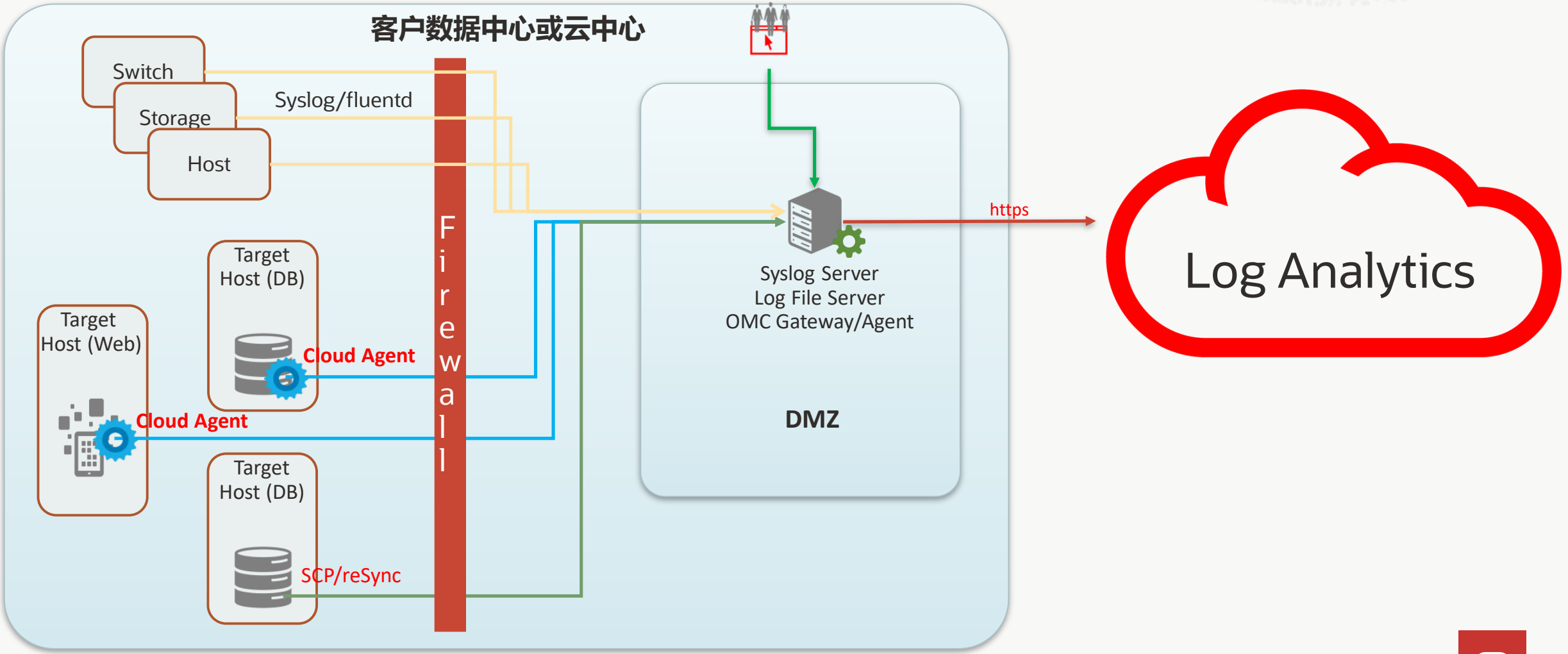
Log Analytics

一个利用现代的、安全的、预制的大数据平台、跨竖井集成数据的云服务

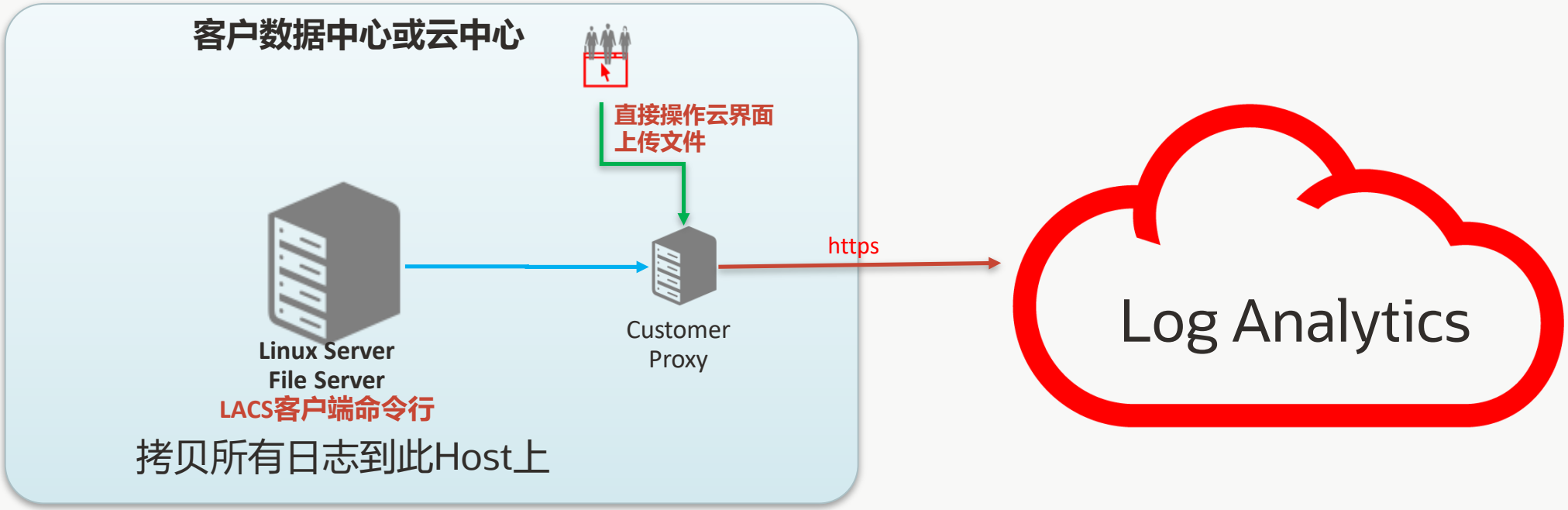
LACS：通过代理程序自动上传日志



LACS：通过代理程序与Syslog Server协作上传日志



LACS：手动上传


客户数据中心或云中心
直接操作云界面
上传文件
https
Log Analytics
Linux Server
File Server
LACS客户端命令行
Customer
Proxy
拷贝所有日志到此Host上

技术点一览

- 基本
  - 解析器，与正则表达式
  - 日志源，且与解析器关联
  - 目标，且与日志源关联
  - 日志人工上传
  - 字段扩充
  - 字段扩展（二次解析）
  - 日志脱敏
  - 日志标签
  - 流计算
- 高级
  - 数据清除与归档
  - 应用拓扑图
  - 可视化算法
  - 可视化查询
  - 仪表板定制

解析器：从非结构化日志中获取有用信息并结构化

编辑语法分析器

保存 取消

引导 手动

定义如何从给定类型的日志文件的日志记录中提取基本字段。

语法分析器 Apache Hive Log Format

类型 Regex

语法分析器说明

日志内容示例

```
2017-03-06T11:57:26,879  INFO [bfa7b56a-f553-417b-a7d9-849723e5447e main] CliDriver: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a
different execution engine (i.e. spark, tez) or using Hive 1.X releases.
2017-03-06T15:01:55,285  INFO [main] conf.HiveConf: Found configuration file null
2017-03-06T15:01:56,823  WARN [main] common.LogUtils: hive-site.xml not found on CLASSPATH
2017-03-06T15:01:57,380  INFO [main] SessionState:
Logging initialized using configuration in jar:file:/home/hadoop/hive/lib/hive-common-2.1.1.jar!/hive-log4j2.properties Async: true
```

语法分析表达式 {TIMEDATE}\s+(\w+)\s+\[([^\]]*)\]\s+(\S+):\s+(.*)

将整个文件作为单个日志记录来处理

日志记录跨越 单行 多行

条目开始表达式

{TIMEDATE}

字段 函数 语法分析器测试

数字	名称	语法分析表达式	数据类型	说明
1	Severity	(\w+)	STRING	Log severity level of the message. Examples: "DEBUG", "ERROR", "NOTIFICATION", etc.
2	Thread	([^\]]*)	STRING	A textual identifier for the process thread this log entry is associated with.
3	Class	(\S+)	STRING	The name or identifier of a class. Typically used to indicate a Java (or other programming language) Class
4	Message	(.*)	STRING	Primary content message body of the log entry

开箱即用的解析器：Oracle应用部分

E-Business Suite
- Concurrent Manager Logs
- Concurrent Request Logs
- Conflict Resolution Manager Logs
- Internal Concurrent Manager Logs
- Output Post Processor Logs
- Transaction Manager Logs
- Transaction Logs
- Workflow Notification Mailer Logs

Siebel
- Component Logs
- Gateway Server Logs
- Name Server Audit Logs

Fusion Applications
- Diagnostics Log

PeopleSoft
- Analytics Engine Server Logs
- Application Analytics Engine Server Logs
- App Server (APPSRV) Process Logs
- Monitor Server (MONITORSRV) Process Logs
- Watch Server (WATCHSRV) Logs
- Tuxedo Access Logs
- Tuxedo User Logs
- Integration Gateway Error Logs
- Integration Gateway Message Logs
- Master Scheduler Server Logs
  Process Scheduler App Engine Server Logs
- Process Scheduler Distribution Agent Logs
- Process Scheduler Master Scheduler Logs
- WLS Server Access Logs
- WLS Server Logs
- WLS Server STDOUT Logs
- WLS Servlet Logs

Oracle HTTP Server (OHS)
- Access Logs
- Admin Logs
- Server Logs
- OPMN Logs
- Diagnostics Logs

Fusion Middleware (WLS)
- Access Logs
- Diagnostics Logs
- Node Manager Logs
- Server Logs
- Standard Out Logs

BI Publisher
- Server Logs
- JBIPS Logs

IDM
- OAM Server Logs
- OAM Access Logs
- OAM Audit Logs
- OID Audit Logs
- OID Directory Control Logs
  OID Dispatcher Logs
- OID Replication Logs
- OID Server Logs
- OID Audit Logs
- OID Monitor Logs
- OID OPMN Logs

Database
- Alert Logs
- Audit Logs
- Trace Files
- Incident Dump Files
- Audit Logs stored in Database
- Audit Logs stored in DB for Unified Audit Trail

Listener
- Alert Logs
- Trace Files

ASM
- ASM Alert Logs
- ASM Trace Files

Cluster Ready Services
- Alert Logs
- Daemon Logs
- Disk Monitor Logs

Exadata Cell
- Alert Logs
- Management Server Logs
- Management Server Trace Logs

MySQL
- Audit XML Logs
- Error Logs
- General Query Logs
- Slow Query Logs

EM Cloud Control Agent
- Agent Logs
- AJTS Logs
- Host Target Event Logs JVMGC Logs
- PFU Logs
- STDOUT Logs

EM Cloud Control OMS
- OMC Logs
- Access Logs
- Diagnostics Logs
- STDOUT Logs

EM Cloud Services Agent
- Agent Logs
- AJTS Logs
- EMCTL Logs
- Host Target Logs
- JVMGC Logs
- Log Collector Logs,
- PFU Logs
- STDOUT Logs

OMC
- Compliance Assessment Result Logs
- Orchestration Service Output Logs
- Security Monitoring Analytics Event Format (XML) Source

开箱即用的解析器：non-Oracle应用部分

SAP
- Application Startup Logs
- Application Transport Logs
- Developer Trace Log
- Java Server Application Logs
- Java Server Default Trace Logs
- VMC Available Logs

Apache
- Access Logs
- Error Logs
- SSL Access Logs
- SSL Request Logs

Apache Tomcat
- Access Logs
- Catalina Logs
- Error Logs
- Host Logs

Microsfot IIS Logs
- For FTP Format Logs
- For IIS Format Logs
- For NCSA Format Logs
- For W3C Format Logs

NGINX
- Access Logs
- Error Logs

IBM WebSphere
- WAS Logs
- WAS System Logs

JBOSS
- EAP Log Source

IBM DB2
- Audit Logs
- Diagnostic Logs

Microsoft SQL Server
- Agent Error Logs
- Error Logs

Microsoft Active Directory
- Distributed File System Replication Logs
- Installation Wizard Logs
- Netsetup Logs
- NtFrsApi Logs

- **Linux**
- Audit Logs
- Cron Logs
- Mail Delivery Logs
- Secure Logs (+Ubuntu)
- Syslog Logs (+Ubuntu)
- Yum Logs
- IPTables Logs
- SUDO Logs
- Ksplice Logs (Oracle Linux)

Solaris
- ILOM Configuration Log
- Audit Logs
- Install Logs
- SMF Daemon Logs
- SU Logs
- Syslogs Logs

AIX
- Audit Log
- Cron Logs
- Dynamic System Optimizer Logs
- HACMP Cluster Logs
- SU Logs
- Syslog Logs

Windows
- Application Event
- Security Events
- System Events
- Setup Events
- DNS Logs

F5
- Big IP Logs

Citrix
- Netscaler Logs

HP ArcSight
- Common Event Format Source

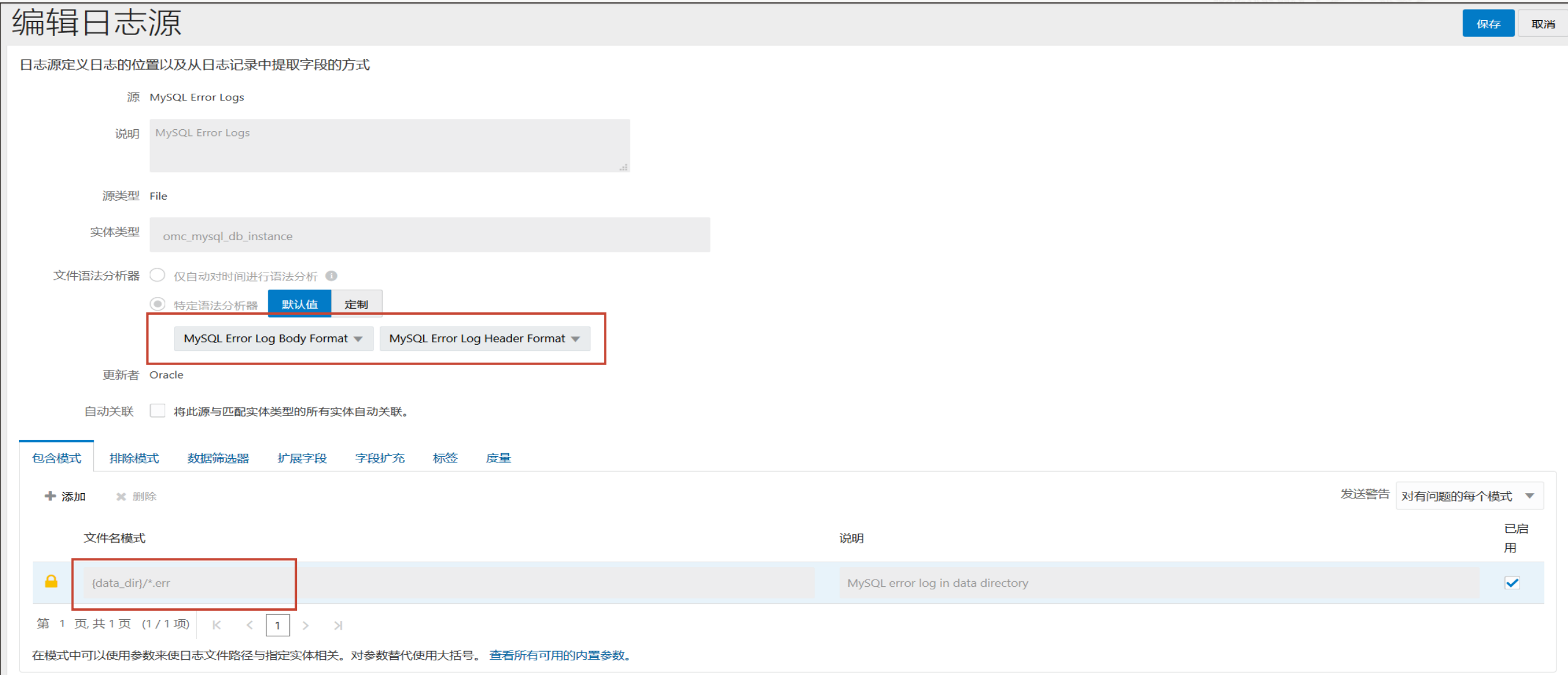
Bluecoat Proxy
- Squid Logs
- W3C Logs

Juniper
- SRX Syslog Logs

NetApp
- Syslog Logs

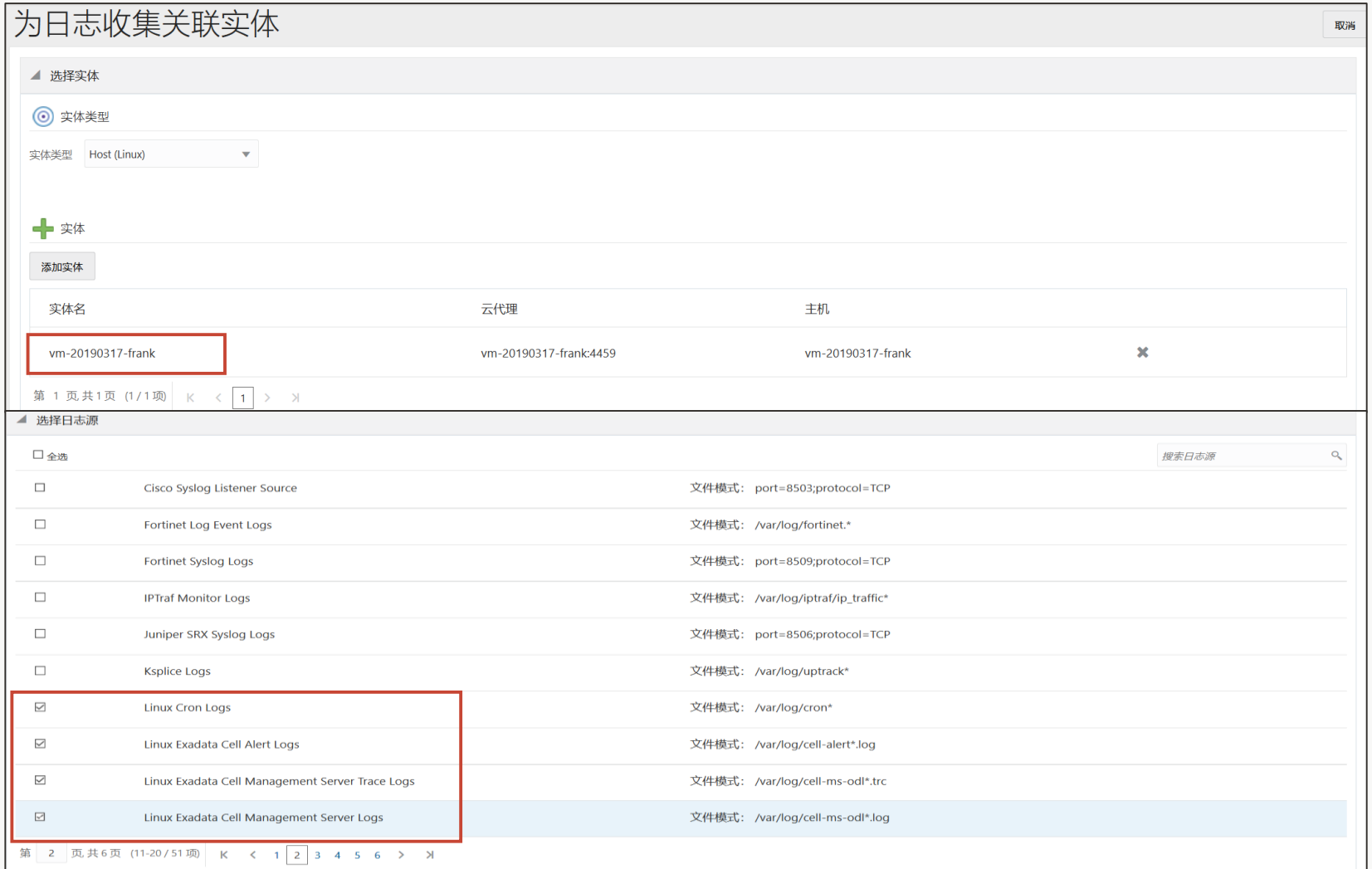
+ Logs in Text/XML/JSON Format
+ Logs in Table
+ Logs written to Syslog Listener

日志源：目标（或主机）上目录的位置



关联目标与日志源：从目标的指定位置自动收集日志

- 将日志与**实体（对象）**项关联
  - mySQL/Oracle/SQL Server
  - Tomcat/Jboss/Weblogic
  - Linux/AIX/Windows/Exadata
  - EBS/PSFT
  - ……
- 有利于从应用角度分析日志
  - 应用
  - 应用群
  - 数据中心
  - 架构



人工上传：即时分析日志，无需代理

- 命令行，交互式，UI
- odu-client upload --properties upload.prop
- .zip , .gz , .tgz , .tar

字段扩充：增加字段，例如注释等

- 有些日志条目中包括了错误号（error id），但是没有描述，以至于运维人员不知其然。可以增加一个映射，error id 到 文字描述，作为知识库用

Time (UTC+8:00)	Entity	Entity Type	Log Source	Host Name (Server)	Error ID	Action
Sep 19, 2016, 10:05:17 PM	frank171017db	Oracle Database Instance	frank Database Alert Logs new	'JUNHZHAN-CN'	ORA-00600	重大错误！请联系Oracle Support！
Sep 19, 2016, 10:05:15 PM	frank171017db	Oracle Database Instance	frank Database Alert Logs new	'JUNHZHAN-CN'	ORA-00600	重大错误！请联系Oracle Support！
Sep 19, 2016, 2:31:45 PM	frank171017db	Oracle Database Instance	frank Database Alert Logs new	'JUNHZHAN-CN'	ORA-00600	重大错误！请联系Oracle Support！
Sep 19, 2016, 2:31:42 PM	frank171017db	Oracle Database Instance	frank Database Alert Logs new	'JUNHZHAN-CN'	ORA-04031	无法分配内存！联系DBA！
Sep 19, 2016, 2:31:42 PM	frank171017db	Oracle Database Instance	frank Database Alert Logs new	'JUNHZHAN-CN'	ORA-04031	无法分配内存！联系DBA！
Sep 19, 2016, 2:31:42 PM	frank171017db	Oracle Database Instance	frank Database Alert Logs new	'JUNHZHAN-CN'	ORA-04031	无法分配内存！联系DBA！
Sep 19, 2016, 2:31:42 PM	frank171017db	Oracle Database Instance	frank Database Alert Logs new	'JUNHZHAN-CN'	ORA-04031	无法分配内存！联系DBA！
Sep 19, 2016, 10:05:55 PM	frank171017db	Oracle Database Instance	frank Database Alert Logs new	'JUNHZHAN-CN'	ORA-00313	打不开文件！联系System Admin！

字段扩充：IP地址 -> 地理信息

字段扩充

函数 Geolocation 查看详细信息

状态 已启用

输入字段

IP 字段 Destination IP

输出字段

洲 Client Host Continent

洲代码 Client Host Continent Code

国家/地区 Client Host Country

国家/地区代码 Client Host Country Code

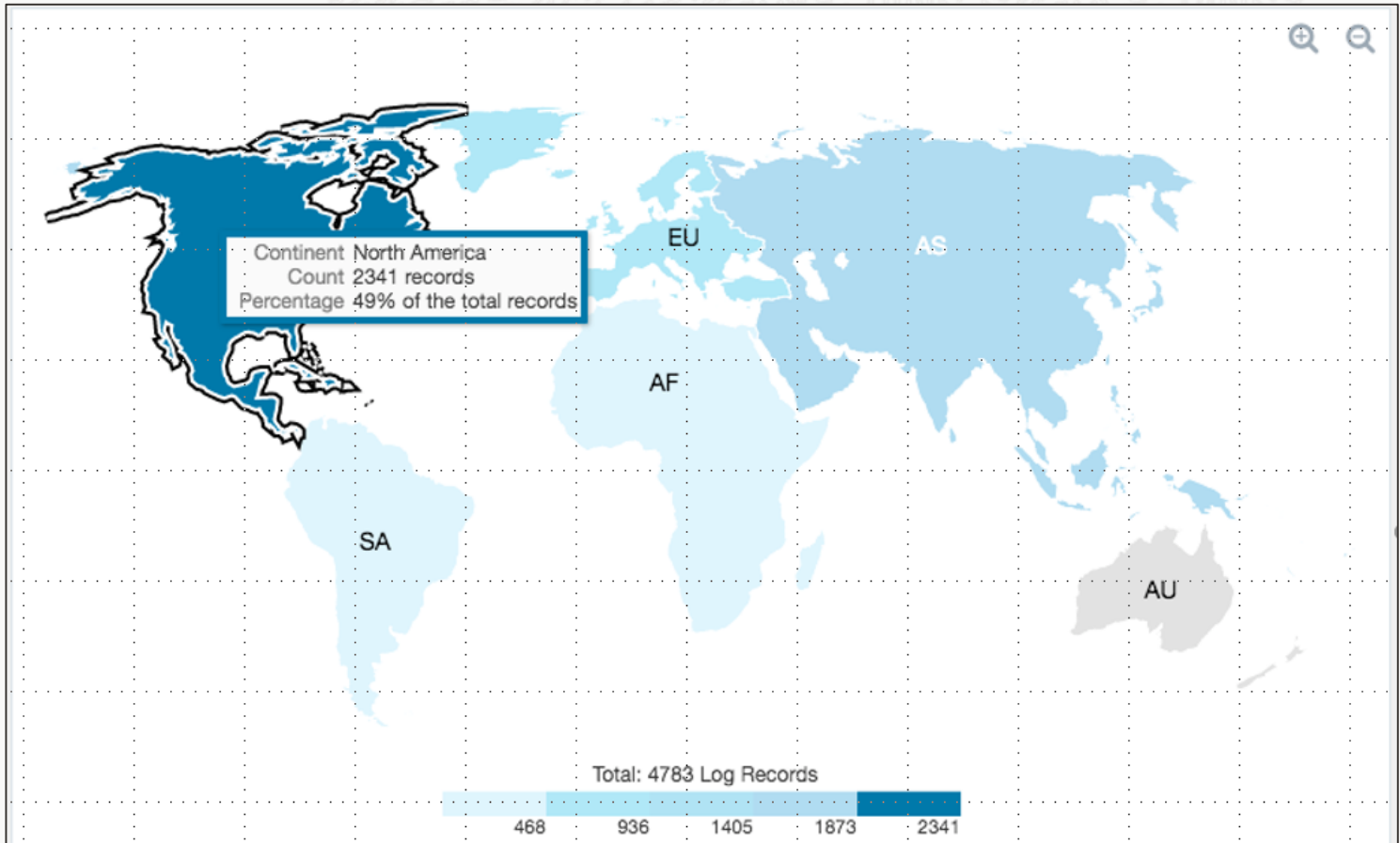
州/省/市/自治区 Client Host Region

省/市/自治区代码 Client Host Region Code

城市 Client Host City

坐标 Client Coordinates

添加 取消



2,717 log records have no value for Client Host Continent Code.

Showing 1-25 of 7500

Time (UTC-7:00)	Original Log Content
Aug 23, 2017, 3:23:54 PM	131.127.20.30 - Felix [23/Aug/2017:15:23:54 -0700] "GET /LocalStore/category3.com?PRODUCTID=CA-CH-33&JSESSIONID=CA555056 addtocart iPhone HTTP/1.1" 408 2056 19981 "http://www.LocalStore.com "Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24"

字段扩展：抓取非固定格式中的有用信息

- 可以称为日志的二次解析（一次性完成）
- 第一次解析：
  - 利用解析器自动捕获"key-value"
  - 但是，有些字段的内容是经常变化而没有固定格式的，但是其中又有感兴趣的信息
- 第二次解析：
  - 将第一次解析后的字段值中符合条件的信息通过正则表达式串查找出来

🔒	Message	login - login ok session 378 by root as root:other	Service 等于 auditd	login\s+-\slogin.+by\s+{User Name:[^\s]+}
🔒	Message	userName : TTY=pts/10 ; PWD=/home/userName/solr-4.6.1/example ; USER=root ; COMMAND=/usr/bin/vncserver start	Service 等于 sudo	^\s*\S+\s+:\s+TTY=.*?COMMAND={Security Resource Name:.*}$
🔒	Message	su irefresh failed for grdbms on /dev/pts/1	Service 等于 su	su\s+{User Name:\S+}\s+failed\s+for\s+{User Name (Originating):\S+}\s+on\s+{Terminal:\S+}

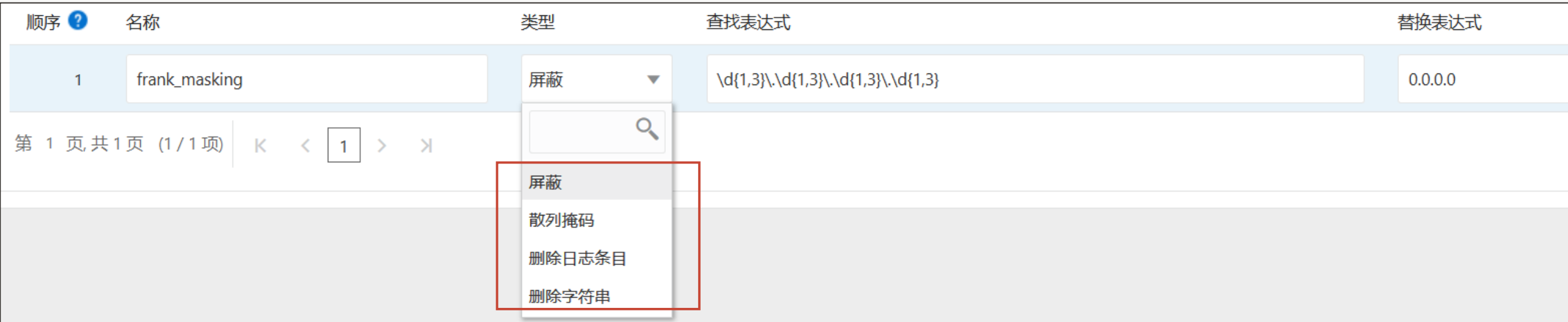
第一次解析后的字段
需要二次解析

不同条件下，其内容也不同

定义新的字段

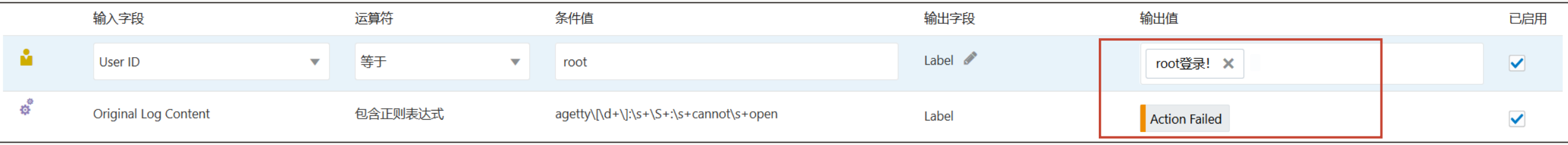
日志脱敏：敏感的业务或IT数据可以脱敏，以满足企业或法律法规要求

- 举例：
  - 业务日志中记录的金额
  - 主机日志中记录的IP地址
- 利用正则表达式查找、替换或删除日志条目中的敏感信息
  - ip: 192.168.88.99
  - 正则表达式查找：\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}
  - 替换为：0.0.0.0
  - 注意：替换的数据类型要和原始一样，否则解析时会出错

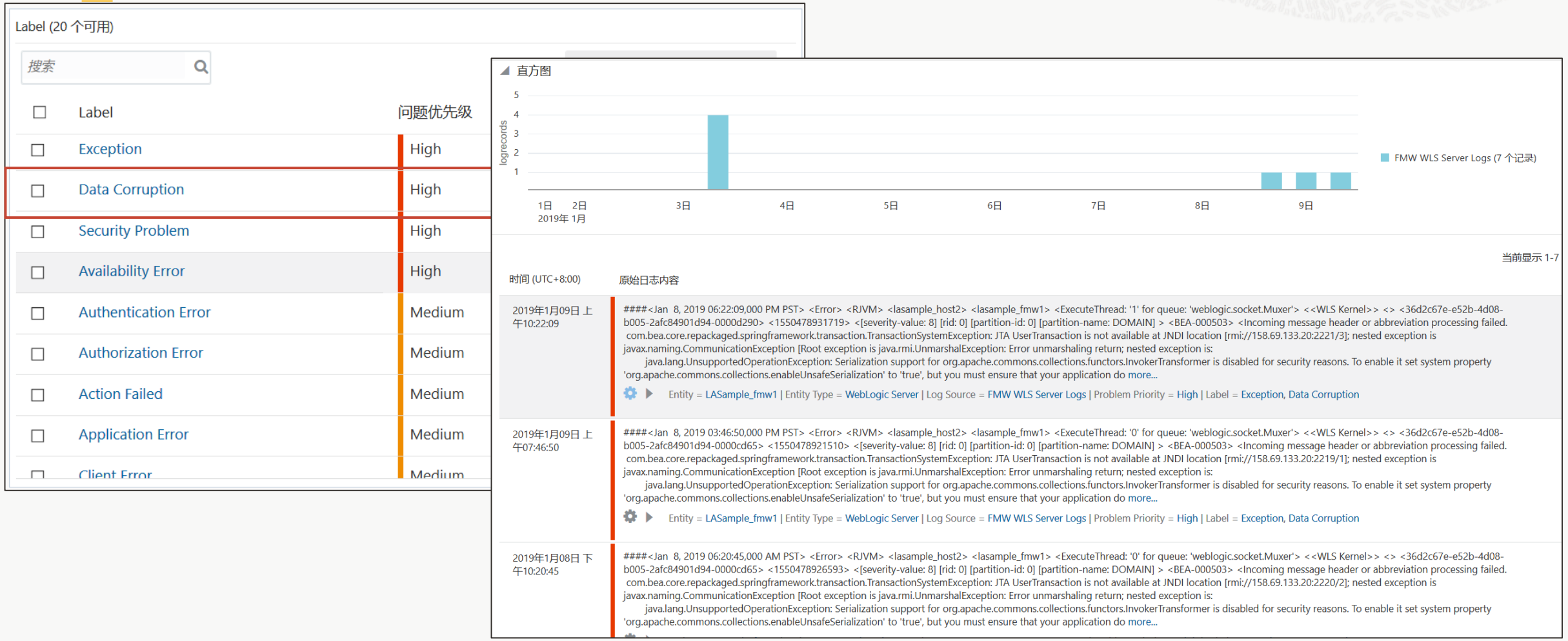


日志标签：快速分类符合条件的日志，为发现、诊断问题用；企业知识库

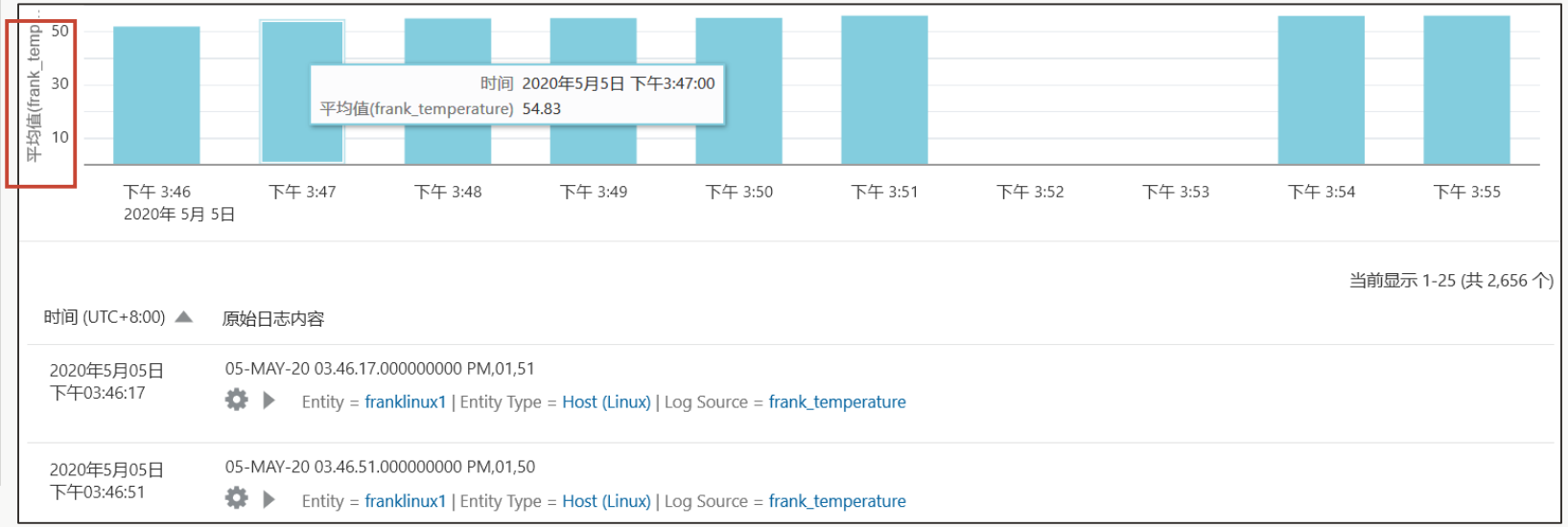
- 分析某个字段的值是否符合某些条件（比较操作符）
- 举例：
  - 如果转账金额大于10万，则及时打标签"重大金额"
  - 如果遇到严重错误，则及时打标签"重大错误"
  - 管理员定期查看这些标签下的日志条目



日志标签：大量的开箱即用诊断结果！

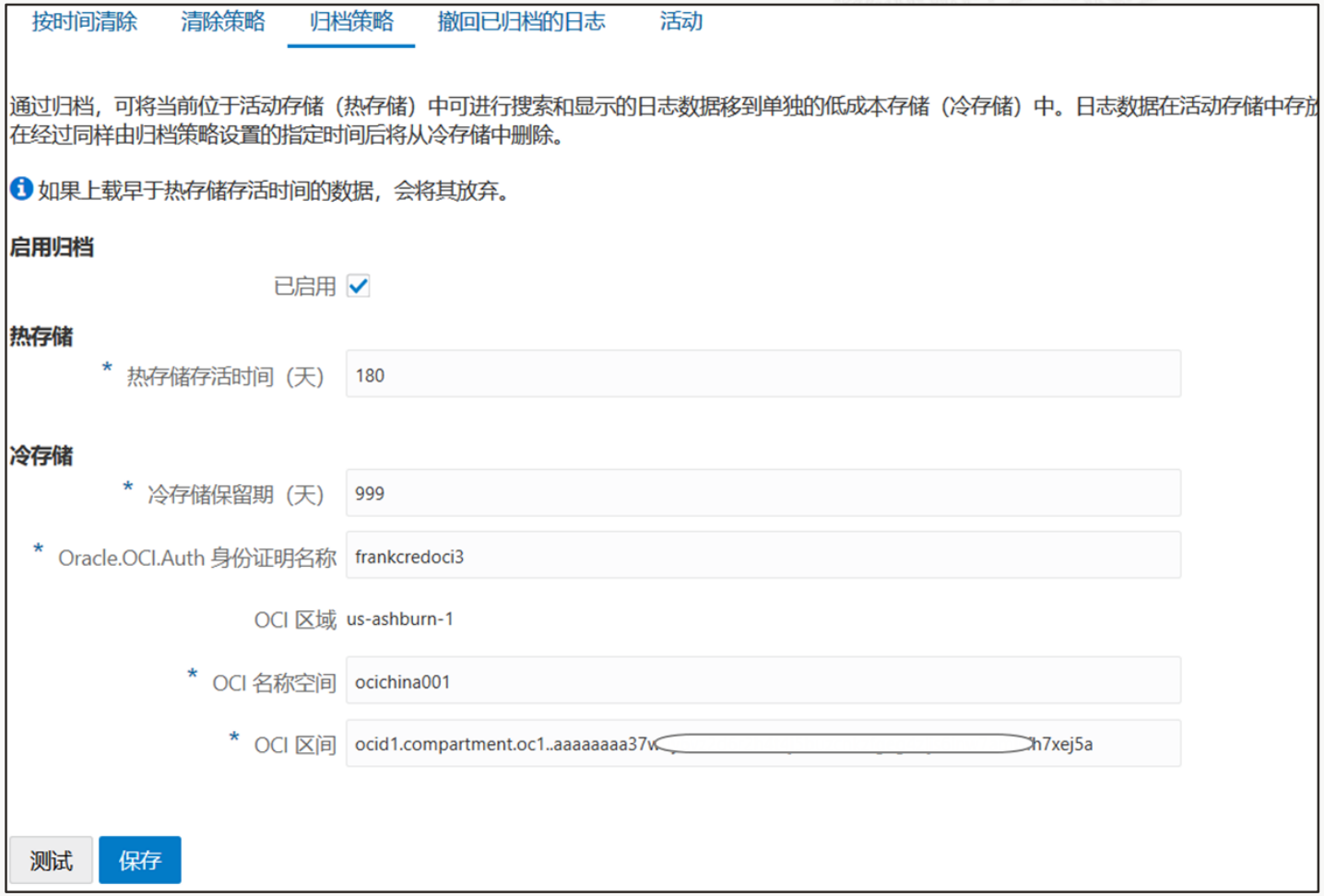


度量：流式计算，实时计算均值、总和、数量



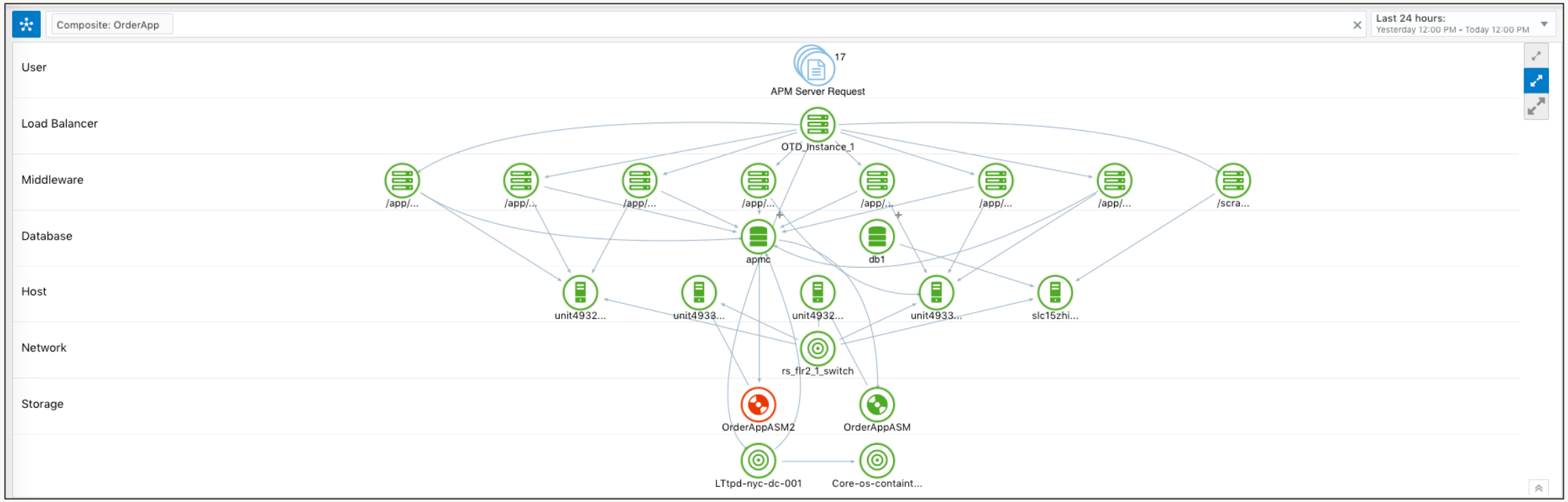
数据清除与归档：长期存储海量日志不再是问题

- 归档的成本远低于在线
- 按日志发生时间制定清除策略
- 按日志在线时长制定归档策略
- 已归档的日志可以随时重新在线，并在分析完毕后释放回归档



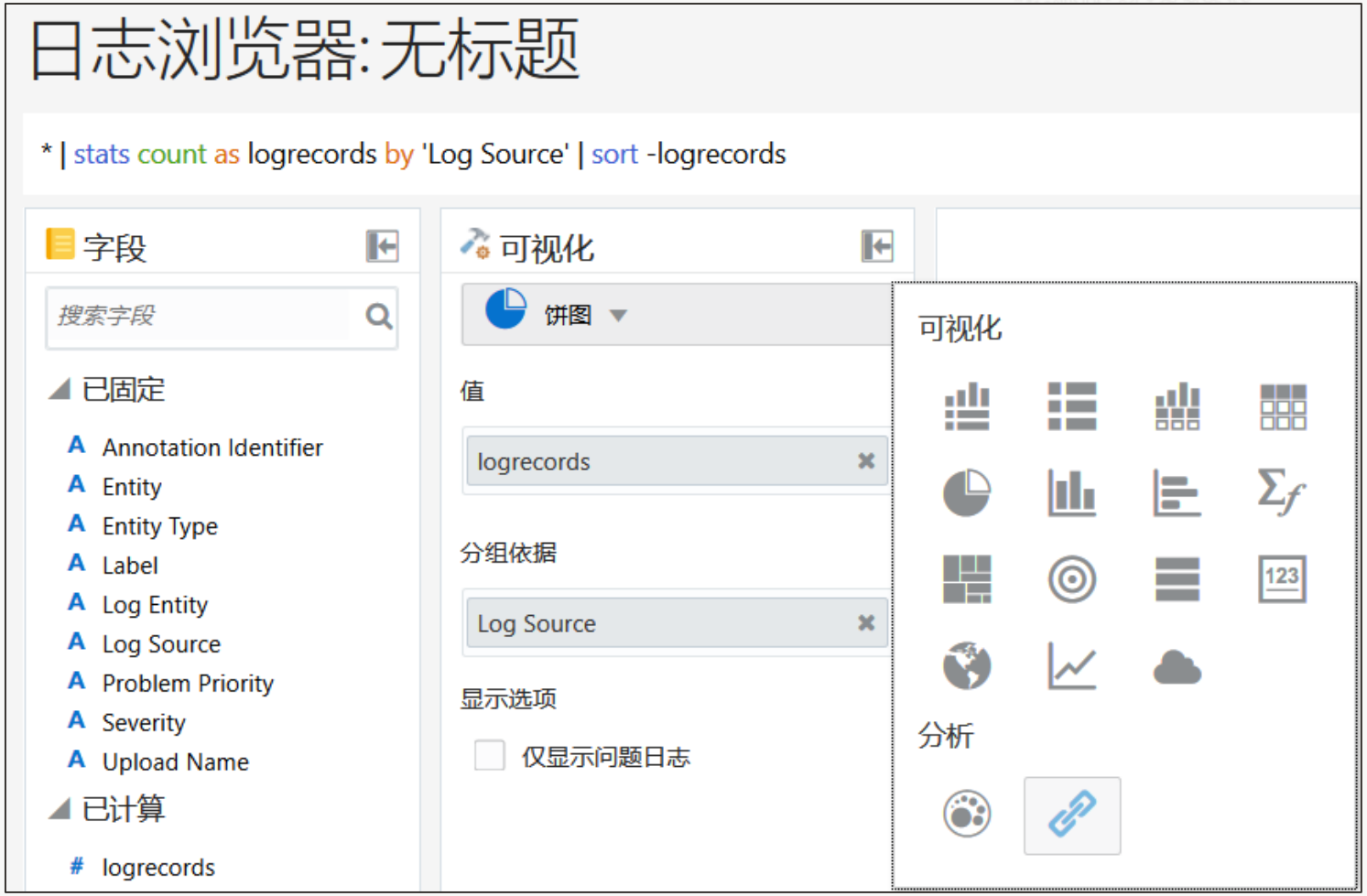
拓扑图：清晰定位应用系统故障

- 自动生成应用拓扑
- 自定义拓扑架构

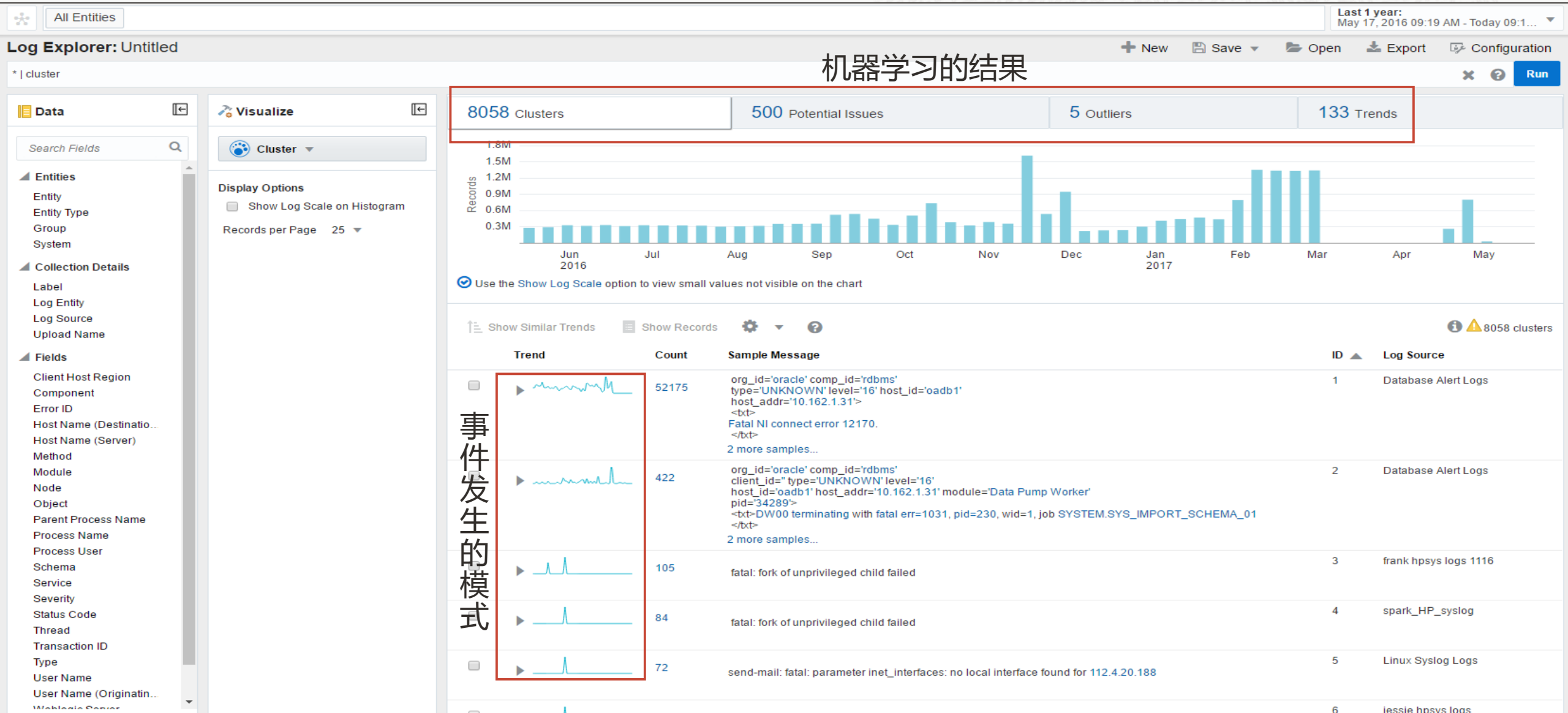


可视化算法：一键式

- 直方图
- 记录
- 表格
- 饼图
- 条形图
- 线形图
- 汇总计算
- 树状图
- 旭日图
- 唯一值（列表）
- 磁贴（数量）
- 地图
- 云图
- 聚类
- 链接



可视化算法：聚类，无监督学习，自动分类，超过5000万条原始日志



可视化算法：链接，自动分析购物车异常

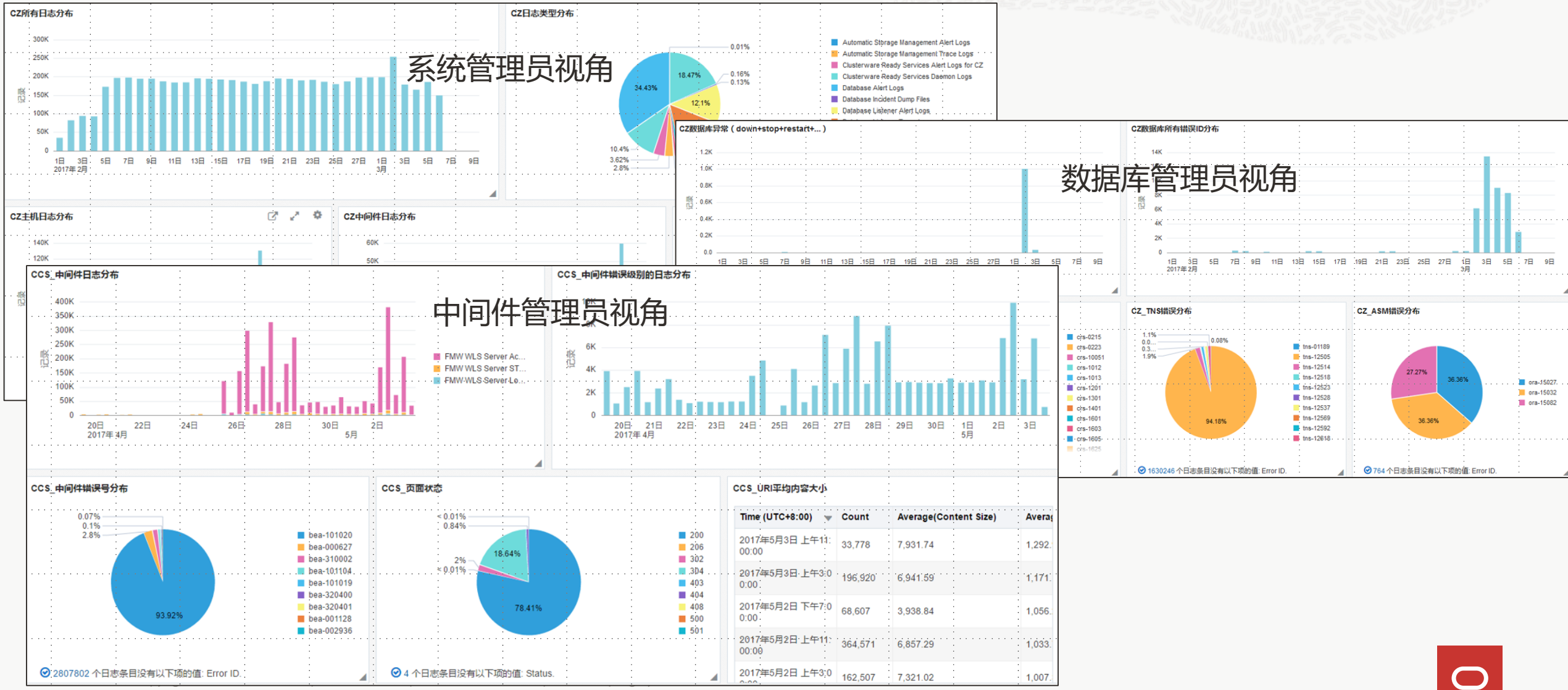
可视化查询语言：高级专家必备

命令	说明	语法
search	从资料档案库或查询字符串的以前阶段中搜索和筛选数据	SEARCH <logical_expression> | <logical_expression> [OR | AND] <logical_expression>
fields	从检索到的结果中添加或删除字段	FIELDS [+|-] <field_name> [, [+|-] <field_name>]
sort	按指定字段对结果排序	SORT [+|-] <field_name> [, [+|-] <field_name>]
distinct	返回所选字段的独特值	
stats	生成概要统计信息, 可以	
cluster	将类似日志记录分组到	
top	检索所选字段中具有最	
bottom	检索所选字段中具有最	
timestats	生成有关从时间序列图	
fieldsummary	返回指定字段的统计信	
highlightrows	突出显示与指定关键字或术语匹配的行。	HIGHLIGHTROWS keyword [, keyword, ...]
highlight	使用指定颜色突出显示关键字或术语	语法突出显示

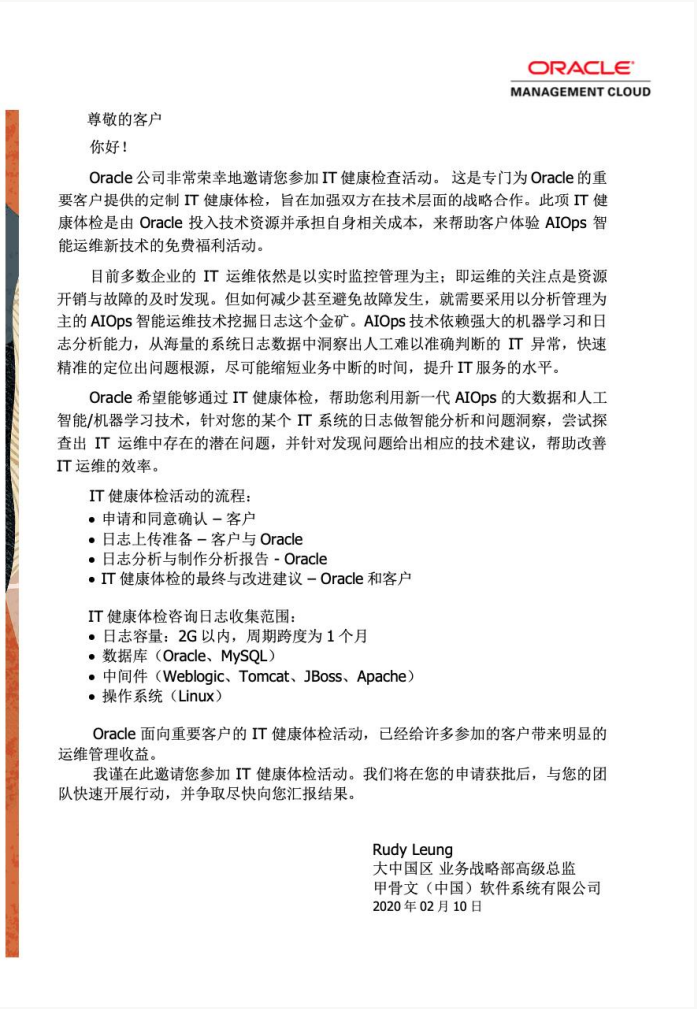
regex	检索与指定正则表达式匹配的结果	REGEX <field_name> [!]= '<value>'
rename	将字段重命名为其他名称	RENAME <field_name> as <new_field_name> [, <field_name> as <new_field_name>]*
eval	通过对表达式求值创建仅限运行时的新字段	EVAL <new_field_name> = <expression>
head	Head 命令返回前 "n" 个结果。limit 指定要返回的结果数。如果 <limit> 未指定, 则使用默认值 10。<limit> 为 -1 将返回所有行。	HEAD [limit = <limit>]*
tail	Tail 命令返回后 "n" 个结果。limit 指定要返回的结果数。如果 <limit> 未指定, 则使用默认值 10。<limit> 为 -1 将返回所有行。	TAIL [limit = <limit>]*
where	Where 命令用于评估布尔值表达式，以确定当计算结果为 true 时查询是否保留特定结果。	WHERE <boolean_expression>
link	Link 命令使用一个或多个字段将相关日志记录组合到组中。然后，可以按不同字段来分析组，以便确定模式或潜在的异常。	LINK [INCLUDENULLS = [true | false]] [INCLUDETRENDS = [true | false]] [SPAN =] <fieldName> [, <fieldName>), ...]
classify	Classify 分析链接结果以确定模式或异常。	LINK <fieldName> | CLASSIFY [TOPCOUNT = <count>] [BOTTOMCOUNT = <count>] <fieldName> [, <fieldName>] [AS <newFieldName>]
bucket	Bucket 命令将字段分组为存储桶。存储桶可以根据值自动创建，也可以指定。	BUCKET [<bucket_options>] <field_name> [<ranges>]

仪表板：拖拽式定制、权限分配

仪表板：拖拽式定制、权限分配



疫情期间，Oracle怎么能帮到您?



ORACLE MANAGEMENT CLOUD

尊敬的客户

你好！

Oracle 公司非常荣幸地邀请您参加 IT 健康检查活动。 这是专门为 Oracle 的重要客户提供的定制 IT 健康体检，旨在加强双方在技术层面的战略合作。此项 IT 健康体检是由 Oracle 投入技术资源并承担自身相关成本，来帮助客户体验 AIOps 智能运维新技术的免费福利活动。

目前多数企业的 IT 运维依然是以实时监控管理为主；即运维的关注点是资源开销与故障的及时发现。但如何减少甚至避免故障发生，就需要采用以分析管理为主的 AIOps 智能运维技术挖掘日志这个金矿。AIOps 技术依赖强大的机器学习和日志分析能力，从海量的系统日志数据中洞察出人工难以准确判断的 IT 异常，快速精准的定位出问题根源，尽可能缩短业务中断的时间，提升 IT 服务的水平。

Oracle 希望能够通过 IT 健康体检，帮助您利用新一代 AIOps 的大数据和人工智能/机器学习技术，针对您的某个 IT 系统的日志做智能分析和问题洞察，尝试探查出 IT 运维中存在的潜在问题，并针对发现问题给出相应的技术建议，帮助改善 IT 运维的效率。

IT 健康体检活动的流程：

- 申请和同意确认 – 客户
- 日志上传准备 – 客户与 Oracle
- 日志分析与制作分析报告 - Oracle
- IT 健康体检的最终与改进建议 – Oracle 和客户

IT 健康体检咨询日志收集范围：

- 日志容量：2G 以内，周期跨度为 1 个月
- 数据库（Oracle、MySQL）
- 中间件（Weblogic、Tomcat、JBoss、Apache）
- 操作系统（Linux）

Oracle 面向重要客户的 IT 健康体检活动，已经给许多参加的客户带来明显的运维管理收益。

我谨在此邀请您参加 IT 健康体检活动。我们将在您的申请获批后，与您的团队快速开展行动，并争取尽快向您汇报结果。

Rudy Leung
大中国区 业务战略部高级总监
甲骨文（中国）软件系统有限公司
2020 年 02 月 10 日

疫情期间如果您的业务或应用出现故障该怎么办?

是否能快速获取相关信息进行远程分析?

Oracle针对核心客户推出远程免费IT系统体检活动

IT体检邀请函 + **同行业体检报告**

IT体检报告（离线日志分析）

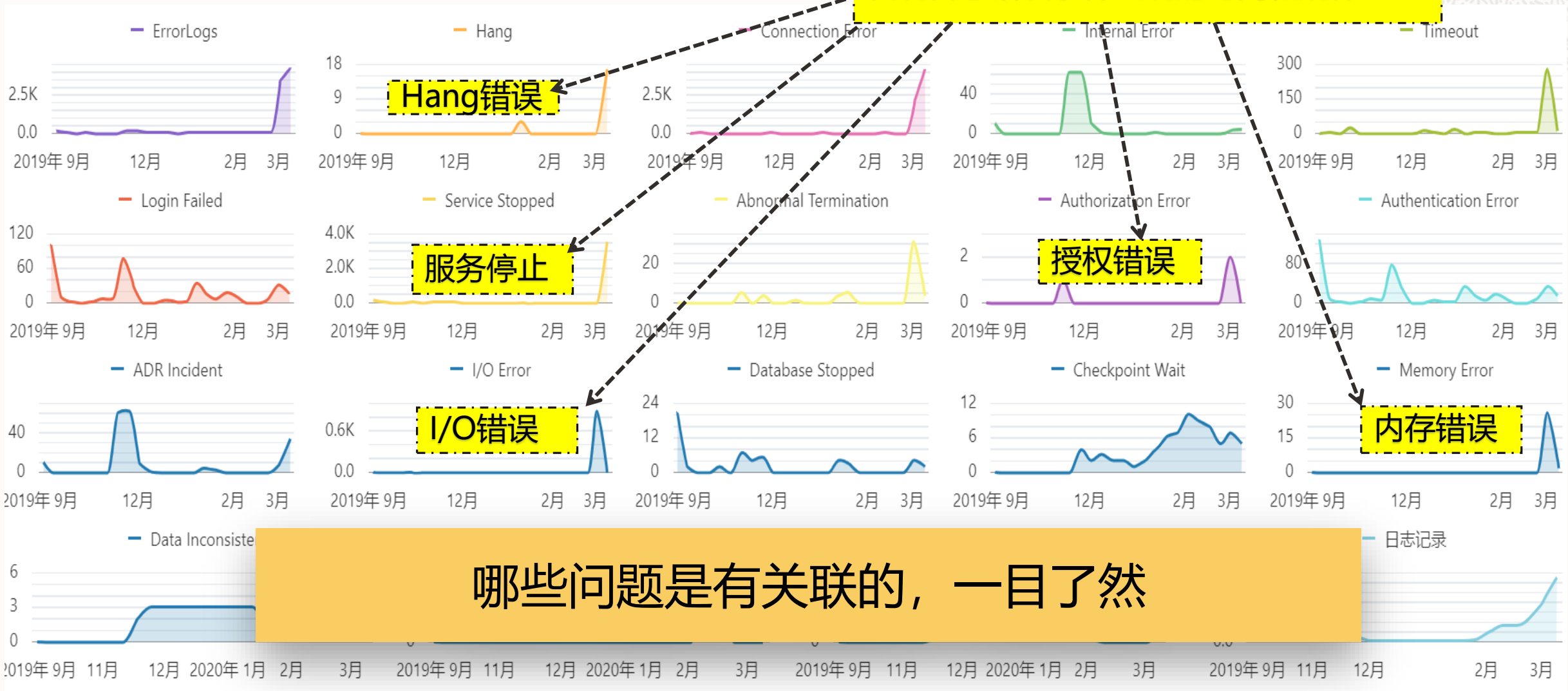
数据采集

主机	日志类型	采集周期	日志数量
	Database Listener Trace Logs	2019/09/24- 2019/09/26	292,421
操作系统日志	Linux Syslog	2020/02/01 – 2020/03/11	1,705,245
	Linux Secure Logs	2019/09/30 – 2020/03/11	650,568
	Linux Cron Logs	2019/09/30 – 2020/03/10	130,670
	Linux Mail Delivery Logs	2019/09/30 – 2020/03/11	12,364

客户提供半年的日志，包括数据库和操作系统，共计400多万条日志

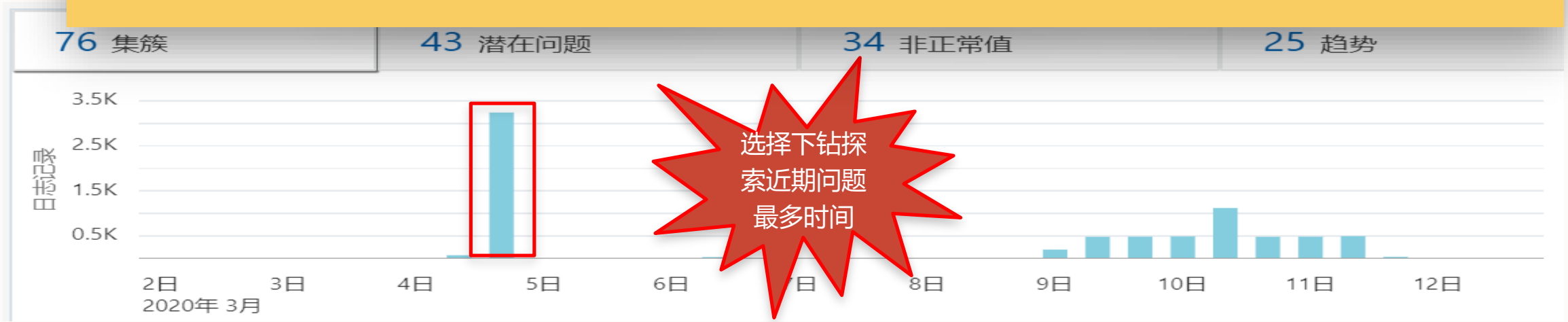
借助LACS，半个小时内就发现问题，2天完成IT体检报告整理

直观展示系统问题的关联性



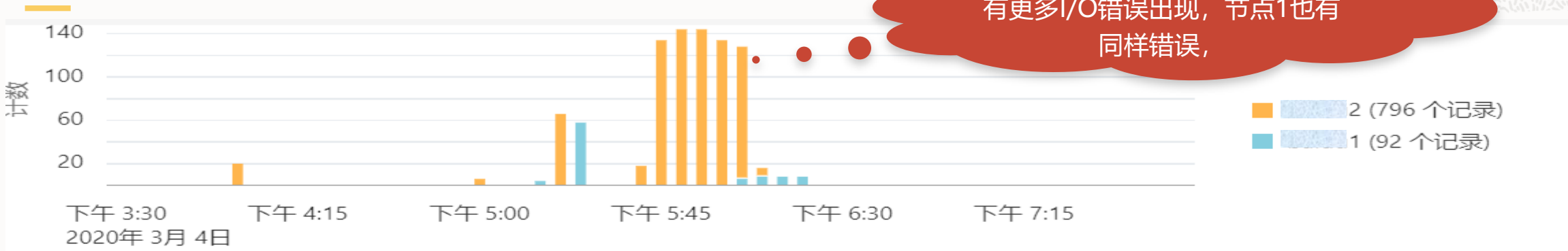
数据洞察

秒级即可洞察出63个潜在问题和36个异常值



分析（1）

首先由节点2报出I/O错误，随后有更多I/O错误出现，节点1也有同样错误，



存储阵列出现问题。根据错误提示，检查相应磁盘位置是否有软硬件错误。

当前显示 1-25 (共 888 个)

时间 (UTC+8:00)　原始日志内容

2020年3月04日
下午04:00:37

Mar 4 16:00:37 bj db02 kernel: blk_update_request: I/O error, dev sdb, sector 131090

(Linux) | Log Source = Linux Syslog Logs | Host Name (Server) =
bel = I/O Error

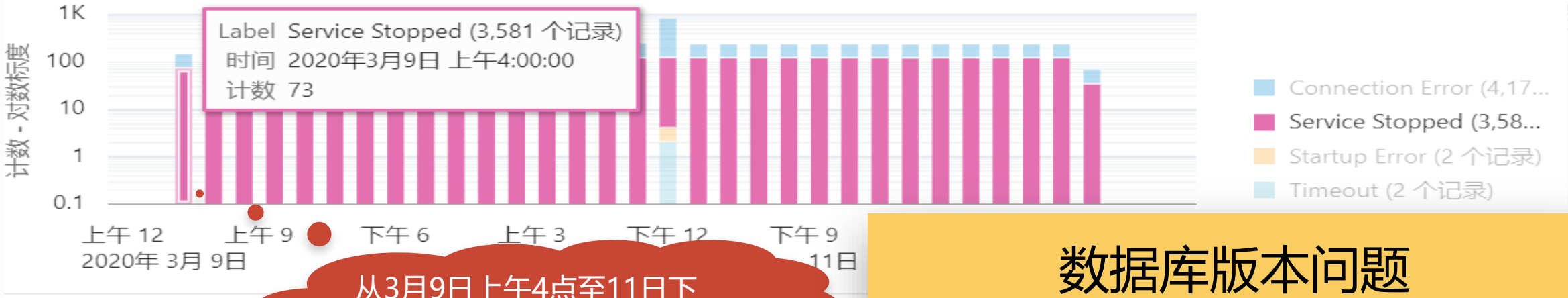
2020年3月04日
下午06:20:09

Mar 4 16:20:09 bj db01 kernel: blk_update_request: I/O error, dev dm-8, sector 0

Entity = u1 | Entity Type = Host (Linux) | Log Source = Linux Syslog Logs | Host Name (Server) = bjkjy-bi-x8db01 | Problem Priority = High | Label = I/O Error

存储硬件问题

分析（2）



从3月9日上午4点至11日下午4点大量出现服务停止错误

数据库版本问题

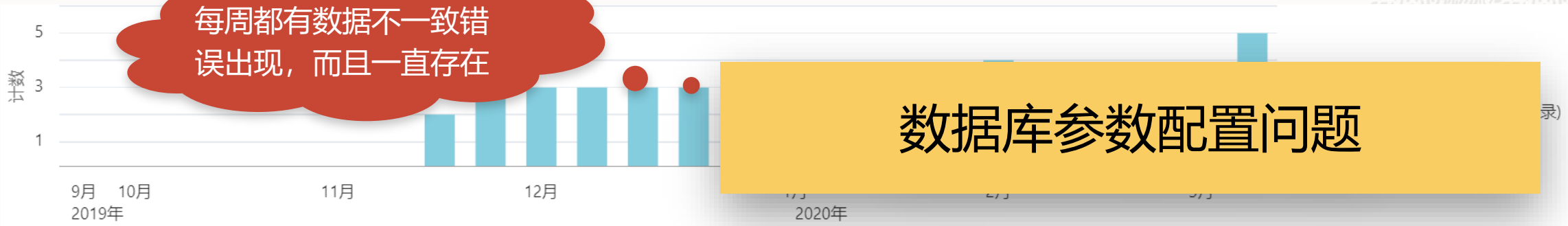
当前显示 1-25 (共 4,177 个)

时间 (UTC+8:00)	原始日志内容
2020年3月11日 下午04:33:36	11-MAR-2020 16:33:36 * service_died * +ASM1 * 12537 Entity = [redacted]2 | Entity Type = Oracle Database Listener | Log Source = Database Listener Trace Logs | Problem Priority = High | Label = Service Stopped, Connection Error
2020年3月11日 下午04:32:36	11-MAR-2020 16:32:36 * service_died * +ASM1 * 12537 Entity = [redacted]2 | Entity Type = Oracle Database Listener | Log Source = Database Listener Trace Logs | Problem Priority = High | Label = Service Stopped, Connection Error

参照文档 ID 2162994.1修改数据库参数或升级版本，消除错误隐患

分析（3）

每周都有数据不一致错误出现，而且一直存在



数据库参数配置问题

上午01:06:17

```
cursor valid_2? 0 makecr 1 line 16481
ORA-08103: object no longer exists
Dump of memory from 0x0000002C2F011FA0 to 0x0000002C2F01234D
2C2F011FA0 4752454D 4E492045 46204F54 4553555F  [MERGE INTO F_USE]
2C2F011FB0 45425F52 49564148 4120524F 49535520  [R_BEHAVIOR A USI]
2C2F011FC0 2820474E 454C4553 50205443 4F535245  [NG (SELECT PERSO]
2C2F011FD0 4B4E5F4E 414E202C 202C454D 49545241  [N_NK, NAME, ARTI]
2C2F011FE0 5F454C43 202C4449 54414C50 54202C45  [CLE_ID, PLATE, T]
2C2F011FF0 454C5449 4150202C 202C4854 45505954  [ITLE, PATH, TYPE]
2C2F012000 434F202C 5F525543 454D4954 4552202C  [, OCCUR_TIME, RE]
2C2F012010 5F594C50 53524550 4E5F4E4F 52202C4B  [PLY_PERSON_NK, R]
2C2F012020 594C5045 4F4C465F 202C524F 54
2C2F012030 2C544E45 4F4C4620 202C524F 48
2C2F012040 202C524F 202C4449 41445055 44
2C2F012050 20455441 4D4F5246 555F4620 5F
2C2F012060 41484542 524F4956 29205 more...
```

Entity = [redacted]1 | Entity Type = Oracle
Label = Data Inconsistency | Error ID = ORA-08

[redacted]m | Problem Priority = High |

根据Trace Log，检查数据操作问题

错误ID	ORA-08103
消息	object no longer exists
原因	The object has been deleted by another user since the operation began, or a prior incomplete recovery restored the database to a point in time during the deletion of the object.
操作	Delete the object if this is the result of an incomplete recovery.

2020年3月11日
上午01:06:17

```
<msg time='2020-03-11T01:06:17.216+08:00'
 type='UNKNOWN' level='16' host_id='bj
 host_addr='172.[redacted]5' pid='333404'>
 <txt>Errors in file /u01/app/oracle/diag/rdbms
ORA-08103: object no longer exists
 </txt>
</msg>
```

Entity = [redacted]1 | Entity Type = Orac

[redacted]m | Problem Priority = High |

分析（4）

长期存在资源受限问题

应用开发SQL语句问题

时间 (UTC+8:00)　　原始日志内容

2020年3月11日 上午07:48:50

```
<msg time='2020-03-11T07:48:50.524+08:00' org_id='oracle' comp_id='rdbms'
 type='UNKNOWN' level='16' host_id='bj                              m'
 host_addr='172.31.52.45' module='JDBC Thin Client' pid='309607'>
 <txt>ORA-01555 caused by SQL statement below (SQL ID: 4x8d612bc25dz, Query Duration=17322 sec, SCN: 0x0b74.a287a8bc):
 </txt>
</msg>
```

Entity = u1 | Entity Type = Oracle Database Instance | Log Source = Database Alert Logs | Host Name (Server) = bj m | Problem Priority = High | Label = Resource Constraint | Error ID = ORA-01555

ORA-01555参考日志信息，根据SQL ID检查SQL语句执行效率，或者调整数据库参数

2020年3月10日 上午11:51:53

```
<msg time='2020-03-10T11:51:53.859+08:00' org_id='oracle' comp_id='rdbms'
 type='UNKNOWN' level='16' host_id='bj                              m'
 host_addr='172.31.52.45' module='MMON_SLAVE' pid='32557'>
 <txt>Errors in file /u01/app/oracle/diag/rdbms/exa8/EXA81/trace/EXA81_m000_32557
ORA-12751: cpu time or run time policy violation
 </txt>
</msg>
```

Entity = u1 | Entity Type = Oracle Database Instance | Log Source = Database Alert Logs | Host Name (Server) = bj m | Problem Priority = High | Label = Resource Constraint | Error ID = ORA-12751

ORA-12751参考文档**ID 761298.1**内容，本案此时刻CPU同时存在耗尽问题

分析（5）

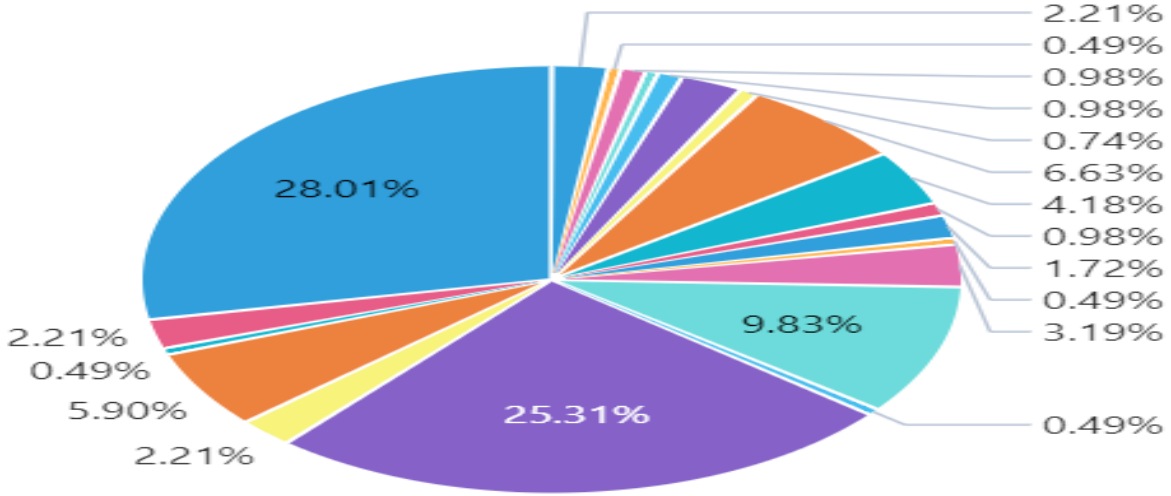
长期存在以root用户或公钥登录失败的记录

操作系统安全配置问题

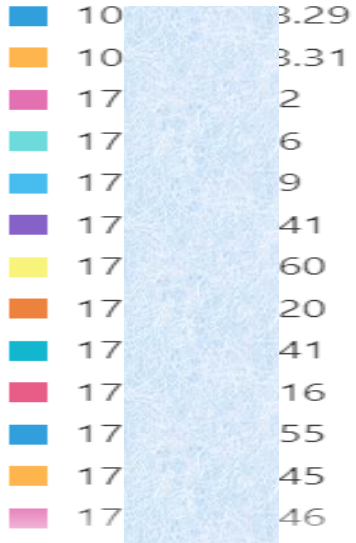
2020年3月11日 下午03:17:50

Mar 11 15:17:50 bjk[redacted]b02 sshd[30864]: Failed password for root from [redacted].45 port 14970 ssh2

Entity = [redacted]2 | Entity Type = Host (Linux) | Log Source = Linux Secure Logs | Host Name (Server) = bj[redacted]db02 | Problem Priority = Medium | Label = Login Failed, Authentication Error | Host IP Address (Client) = 17[redacted]5



检查这些访问地址是否是符合安全要求



问题汇总及建议方案

诊断一　日志量陡然增加问题

影响: **占用空间，影响数据库效率**

方案:
1) 根据文档建议，为BUG打相应补丁
2) 建立定期检查，清理日志空间

诊断二　宕机错误日志分析

影响: **造成数据库无法运行**

方案:
1) 确认宕机原因，本案为I/O(即存储)问题
2) 分析宕机前后的异常日志
3) 观察后续日志是否有相同错误出现

诊断三　安全问题

影响: 威胁数据安全

方案:
1) 分析源主机的安全性
2) 增加系统的安全机制

诊断四　长期存在的日志错误

影响: **影响数据库的可靠性**

方案:
1) 查找错误日志根源
2) 查找解决方案，并实施
3) 建立定期检查机制

日志分析成熟度模型——请对照检查

- “零”配置

Level 1
日志收集

- 持续不断地收集各种日志，无论本地还是公有云、无论IT还是业务

- “零”配置

Level 2
快速问题定位

- 搜索、过滤日志
- 从不同的IT或应用组件中进行日志关联分析
- 通过机器学习检测异常行为
- 跨平台日志上下文

- 少量配置

Level 3
IT运营洞察

- 通过解析器从繁杂的日志中抽取有效信息
- 通过小视图、仪表板展现、降低实施成本
- 大量分析函数、方法
- 客户化能力

- 个性化应用/业务日志的定制

Level 4
业务活动洞察

- 抽取、解析应用日志中和业务相关的信息
- 丰富的可视化能力展现业务运行状态
- 自动化关联业务应用与IT基础设施日志

Our mission is to help people
see data in new ways, discover insights,
unlock endless possibilities.

我们的愿景是以新的方式帮助人们认识
数据、洞察数据，并开启无限可能。

扫码加入:

19c新特性讲座微信群

欢迎关注:

甲骨文云技术公众号

纯技术分享 无广告