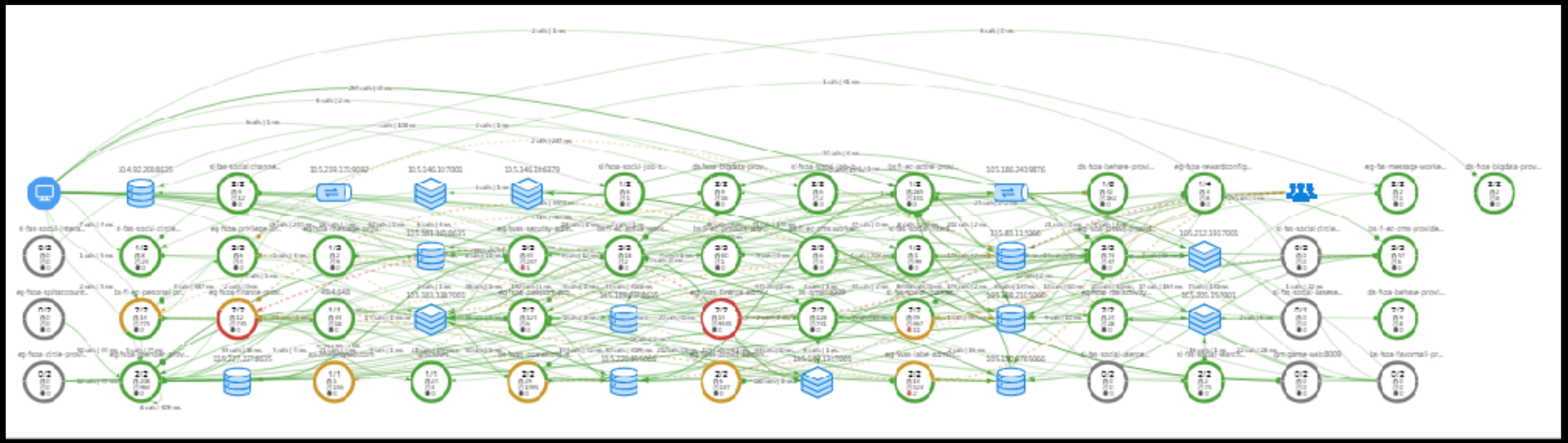
大型分布式系统跟踪系统
—调用链

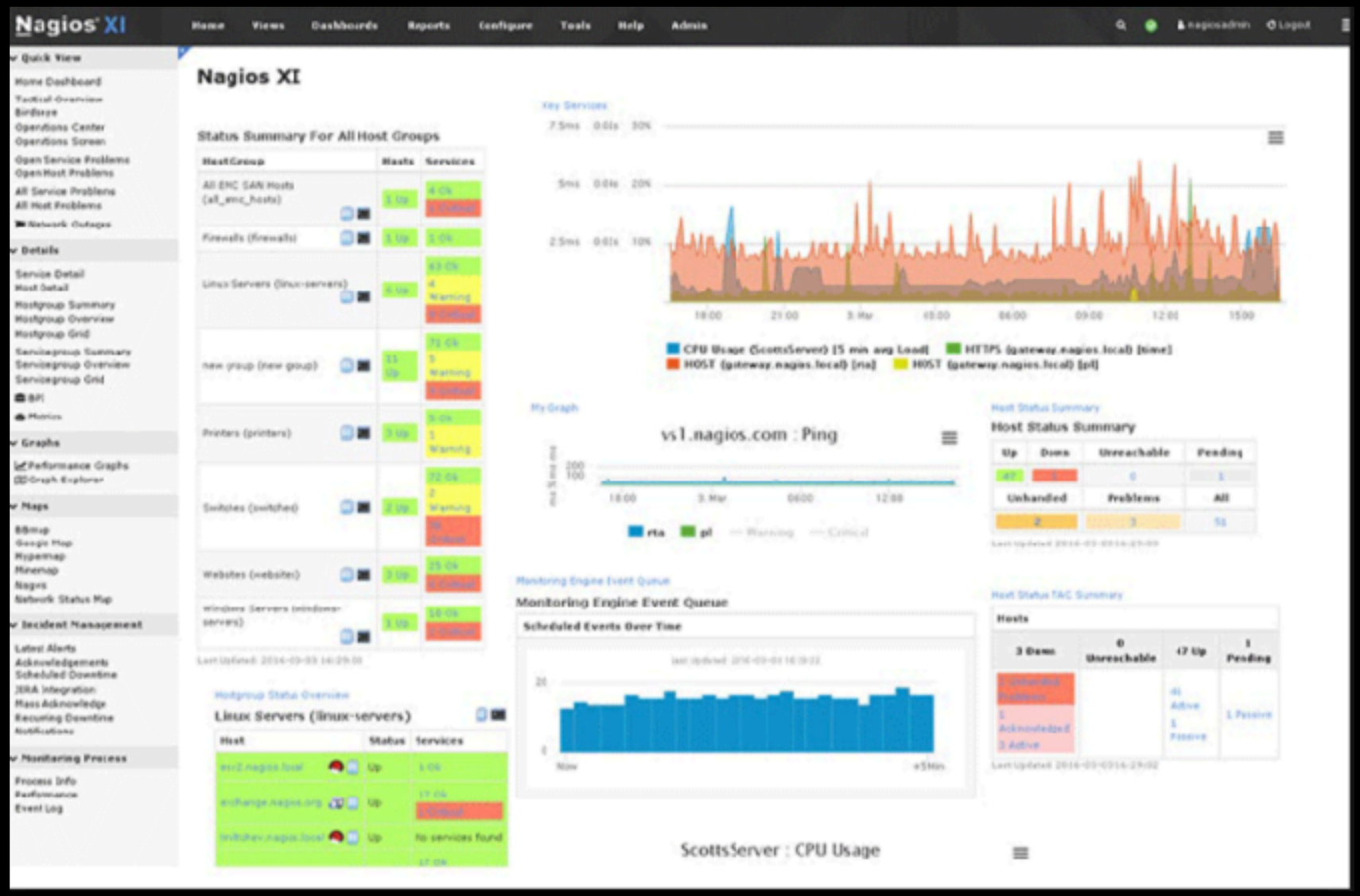
邵增卫

大型分布式系统

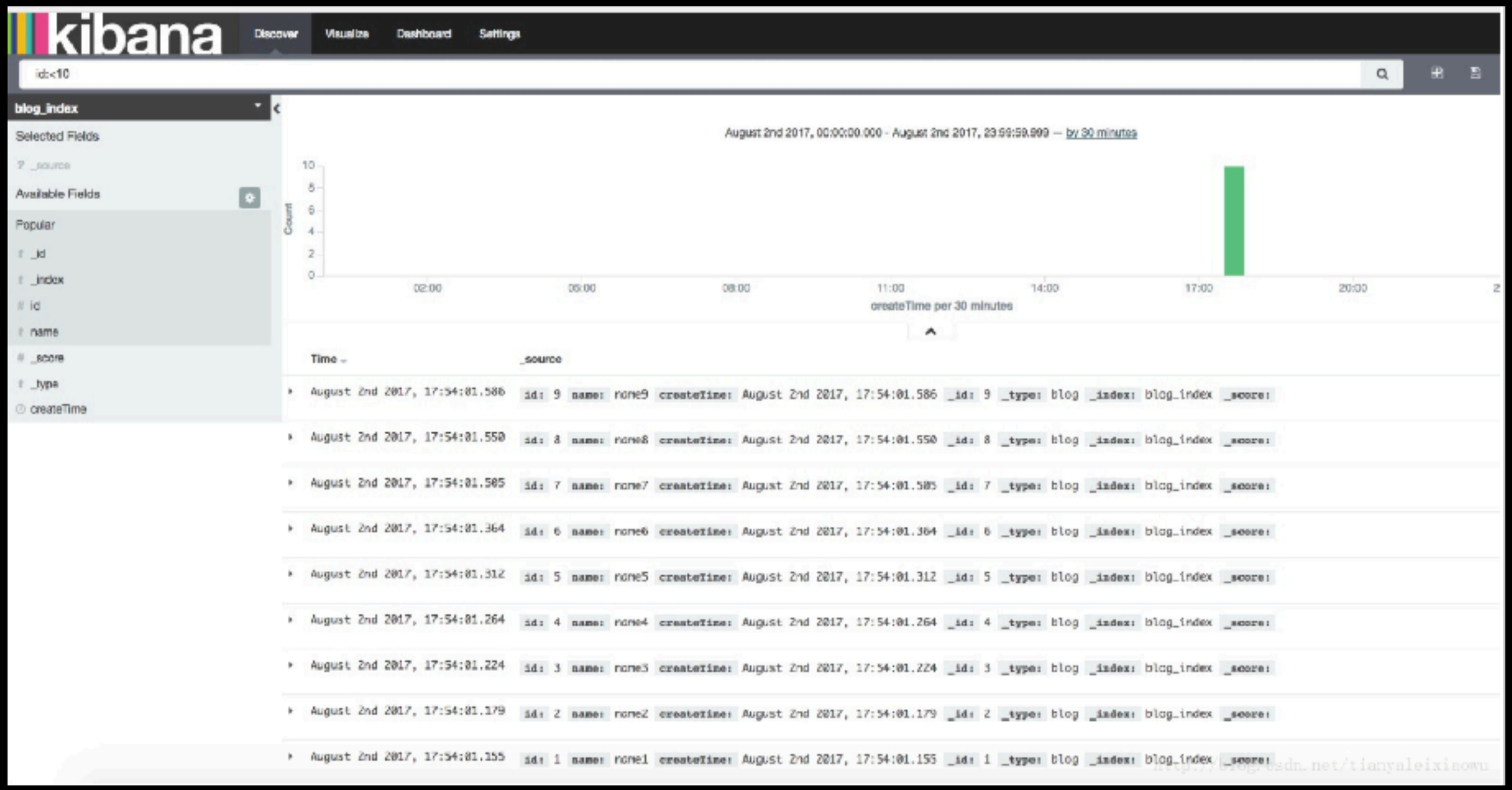


传统运维：监控系统，告警系统

- 网站主页反应慢，失败率高
- 购买商品失败频率高
- 商品搜索页面反应慢



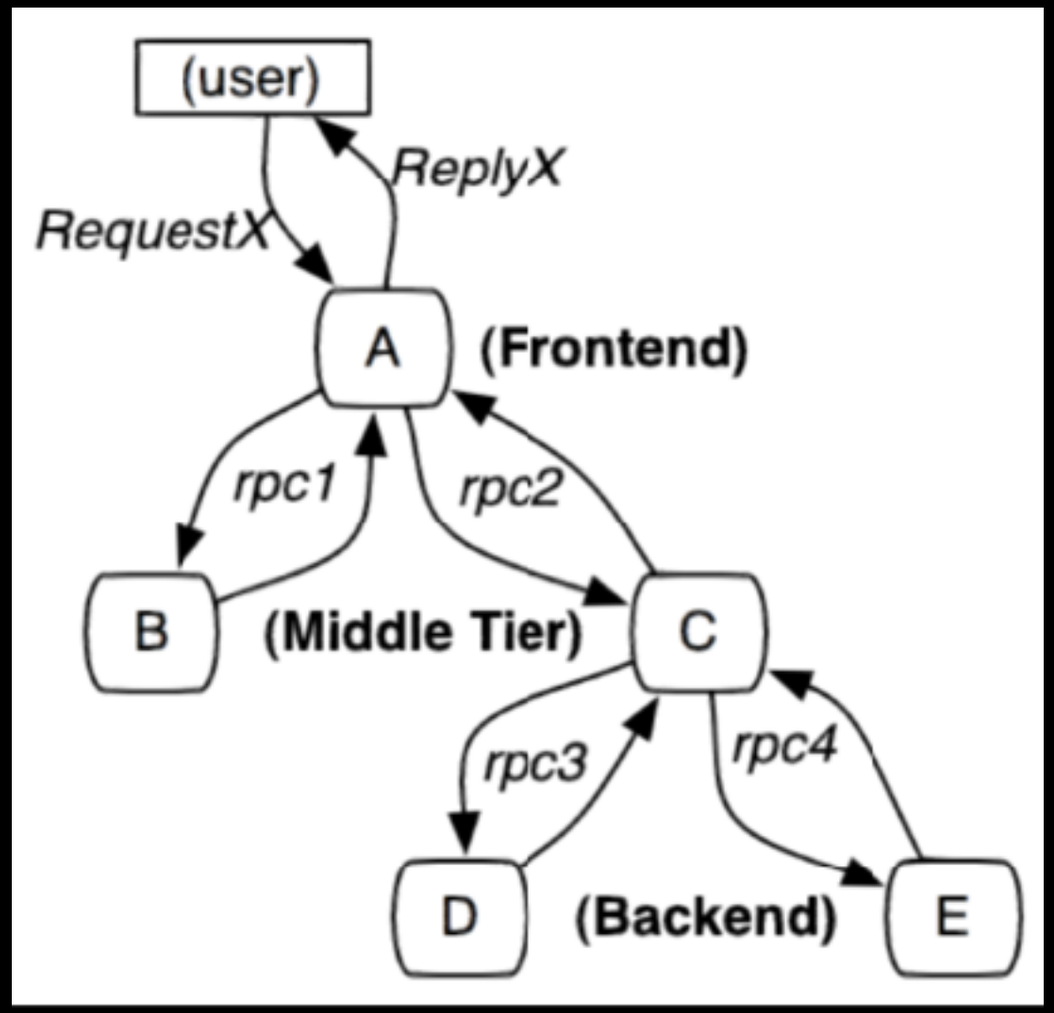
传统运维： 日志系统


kibana
Discover
Visualize
Dashboard
Settings
id:<10
blog_index
Selected Fields
_source
Available Fields
Popular
_id
_index
id
name
_score
_type
createTime
August 2nd 2017, 00:00:00.000 - August 2nd 2017, 23:59:59.999 — by 30 minutes
Count
02:00
05:00
08:00
11:00
14:00
17:00
20:00
createTime per 30 minutes
Time
_source
August 2nd 2017, 17:54:01.586 id: 9 name: name9 createTime: August 2nd 2017, 17:54:01.586 _id: 9 _type: blog _index: blog_index _score:
August 2nd 2017, 17:54:01.550 id: 8 name: name8 createTime: August 2nd 2017, 17:54:01.550 _id: 8 _type: blog _index: blog_index _score:
August 2nd 2017, 17:54:01.505 id: 7 name: name7 createTime: August 2nd 2017, 17:54:01.505 _id: 7 _type: blog _index: blog_index _score:
August 2nd 2017, 17:54:01.364 id: 6 name: name6 createTime: August 2nd 2017, 17:54:01.364 _id: 6 _type: blog _index: blog_index _score:
August 2nd 2017, 17:54:01.312 id: 5 name: name5 createTime: August 2nd 2017, 17:54:01.312 _id: 5 _type: blog _index: blog_index _score:
August 2nd 2017, 17:54:01.264 id: 4 name: name4 createTime: August 2nd 2017, 17:54:01.264 _id: 4 _type: blog _index: blog_index _score:
August 2nd 2017, 17:54:01.224 id: 3 name: name3 createTime: August 2nd 2017, 17:54:01.224 _id: 3 _type: blog _index: blog_index _score:
August 2nd 2017, 17:54:01.179 id: 2 name: name2 createTime: August 2nd 2017, 17:54:01.179 _id: 2 _type: blog _index: blog_index _score:
August 2nd 2017, 17:54:01.155 id: 1 name: name1 createTime: August 2nd 2017, 17:54:01.155 _id: 1 _type: blog _index: blog_index _score:

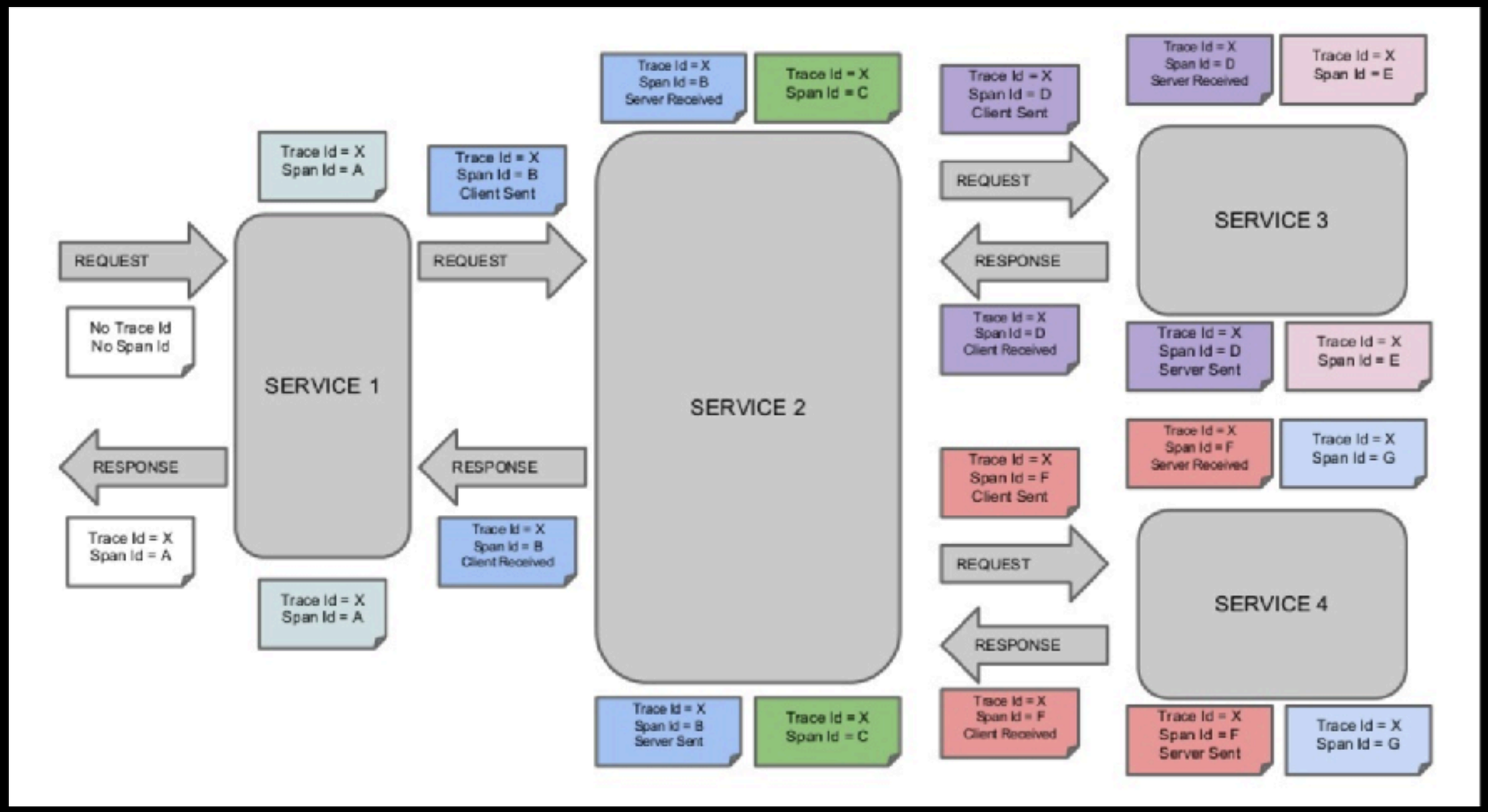
Time To Debug

Google Dapper

- Google Dapper 是google的大规模分布式系统的跟踪系统，经过两年在google系统中的实践使用，于2010年4月发表了论文Dapper。



调用链


REQUEST
No Trace Id
No Span Id
SERVICE 1
Trace Id = X
Span Id = A
RESPONSE
Trace Id = X
Span Id = A
Trace Id = X
Span Id = A
REQUEST
Trace Id = X
Span Id = B
Client Sent
Trace Id = X
Span Id = B
Server Received
Trace Id = X
Span Id = C
SERVICE 2
RESPONSE
Trace Id = X
Span Id = B
Client Received
Trace Id = X
Span Id = B
Server Sent
Trace Id = X
Span Id = C
Trace Id = X
Span Id = D
Client Sent
REQUEST
RESPONSE
Trace Id = X
Span Id = D
Client Received
Trace Id = X
Span Id = D
Server Received
Trace Id = X
Span Id = E
SERVICE 3
Trace Id = X
Span Id = D
Server Sent
Trace Id = X
Span Id = E
Trace Id = X
Span Id = F
Client Sent
REQUEST
RESPONSE
Trace Id = X
Span Id = F
Client Received
Trace Id = X
Span Id = F
Server Received
Trace Id = X
Span Id = G
SERVICE 4
Trace Id = X
Span Id = F
Server Sent
Trace Id = X
Span Id = G

调用链数据结构

DEVOPS MEETUP
IT大咖说 知识共享平台

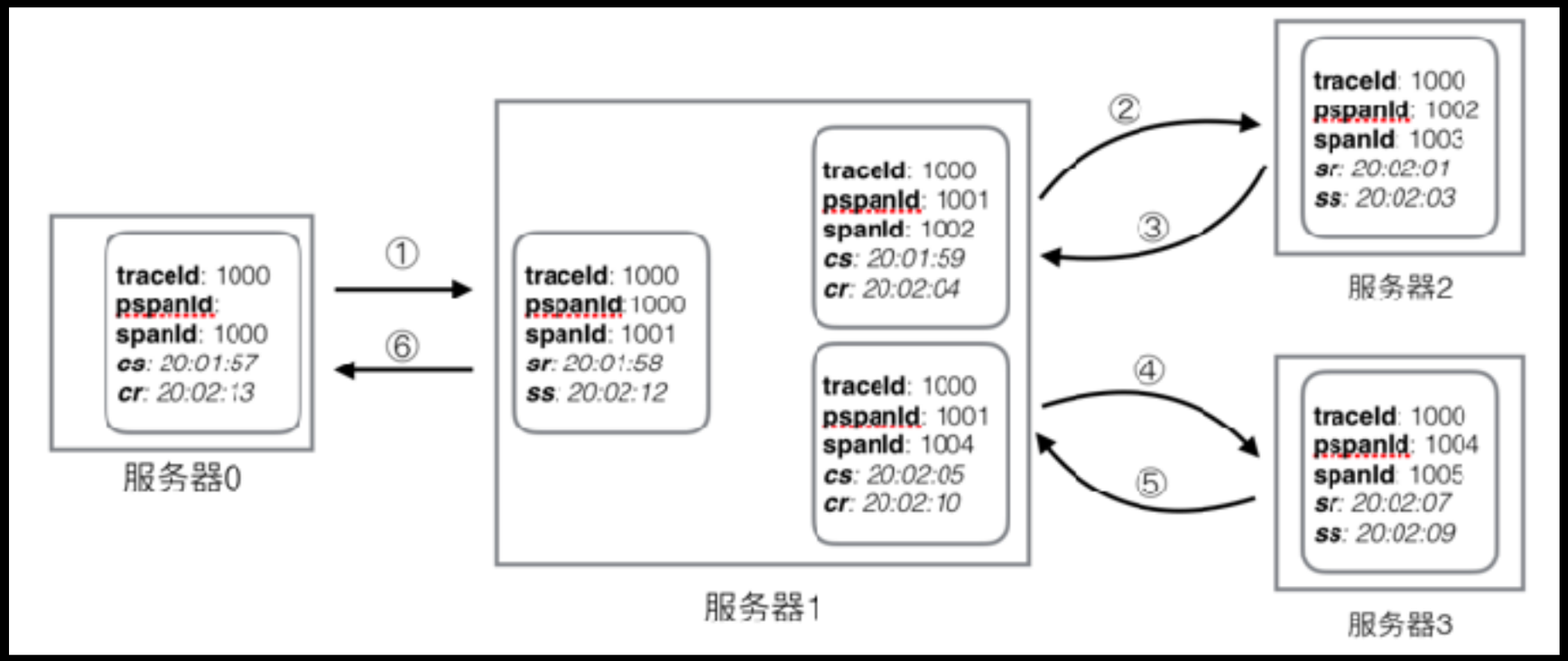

```
{
  "traceId": "bd7a977555f6b982",
  "name": "get",
  "id": "bd7a977555f6b982",
  "timestamp": 1458702548467000,
  "duration": 386000,
  "annotations": [
    {
      "endpoint": {
        "serviceName": "zipkin-query",
        "ipv4": "192.168.1.2",
        "port": 9411
      },
      "timestamp": 1458702548467000,
      "value": "sr"
    },
    {
      "endpoint": {
        "serviceName": "zipkin-query",
        "ipv4": "192.168.1.2",
        "port": 9411
      },
      "timestamp": 1458702548853000,
      "value": "ss"
    }
  ],
  "binaryAnnotations": []
},
```

```
{
  "traceId": "bd7a977555f6b982",
  "name": "get-traces",
  "id": "ebf33e1a81dc6f71",
  "parentId": "bd7a977555f6b982",
  "timestamp": 1458702548478000,
  "duration": 354374,
  "annotations": [],
  "binaryAnnotations": [
    {
      "key": "lc",
      "value": "JDBCSpanStore",
      "endpoint": {
        "serviceName": "zipkin-query",
        "ipv4": "192.168.1.2",
        "port": 9411
      }
    },
    {
      "key": "request",
      "value": "QueryRequest{serviceNa
      "endpoint": {
        "serviceName": "zipkin-query",
        "ipv4": "192.168.1.2",
        "port": 9411
      }
    }
  ]
},
{
```

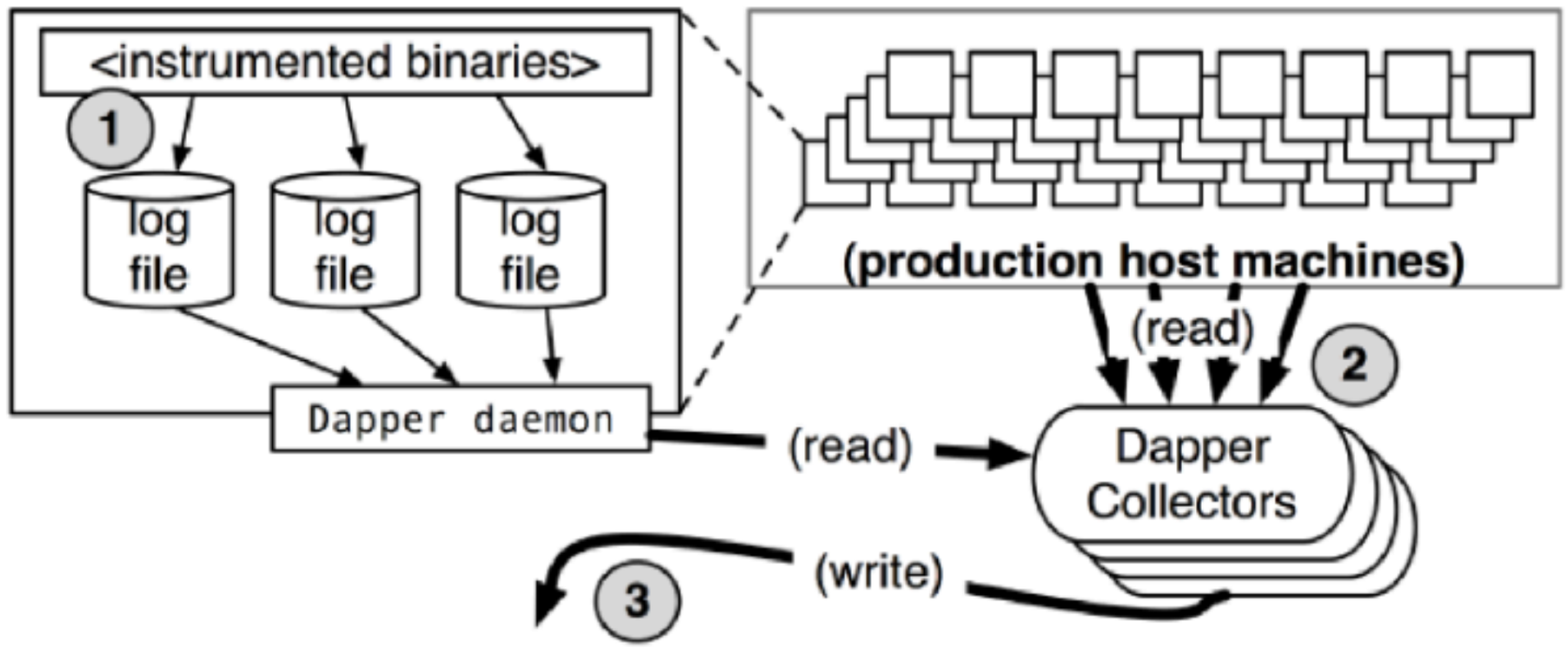
```
  "traceId": "bd7a977555f6b982",
  "name": "query",
  "id": "be2d01e33cc78d97",
  "parentId": "ebf33e1a81dc6f71",
  "timestamp": 1458702548786000,
  "duration": 13000,
  "annotations": [
    {
      "endpoint": {
        "serviceName": "zipkin-query",
        "ipv4": "192.168.1.2",
        "port": 9411
      },
      "timestamp": 1458702548786000,
      "value": "cs"
    },
    {
      "endpoint": {
        "serviceName": "zipkin-query",
        "ipv4": "192.168.1.2",
        "port": 9411
      },
      "timestamp": 1458702548799000,
      "value": "cr"
    }
  ],
```

数据采集

- 植入点
- 采样率
- 跟踪收集的损耗&对应用性能的影响



数据传输



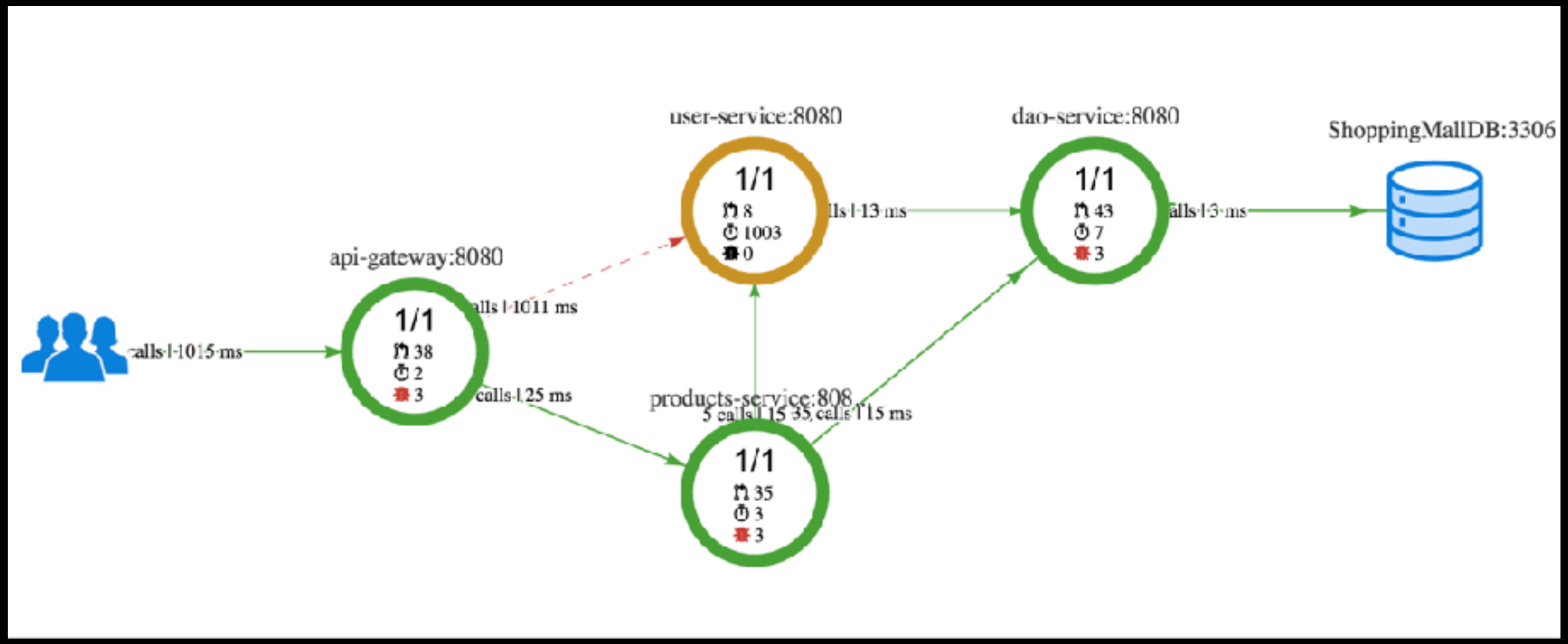
trace id	span 12	span 23	span 34	span 45	span 56	...
123456	*nil*	*nil*	**<data>**	**<data>**	*nil*	...
246802	**<data>**	*nil*	*nil*	*nil*	**<data>**	...
357913	*nil*	**<data>**	*nil*	*nil*	*nil*	...
...	...	...	...	...	...	...

(Central Bigtable repository for trace data)

后台数据分析

- 与异常监控集成
- 推断服务之间的依赖
- 不同服务的网络使用率，吞吐量
- 问题定界定位

华为APM-应用拓扑


user-service:8080
1/1
8
1003
0
dao-service:8080
1/1
43
7
3
ShoppingMallDB:3306
calls 13 ms
calls 3 ms
api-gateway:8080
1/1
38
2
3
calls 1011 ms
calls 1015 ms
calls 125 ms
products-service:808
5 calls 15 35 calls 115 ms
1/1
35
3
3

华为APM-调用链

应用	方法	参数	状态
api-gateway	org.apache.catalina.core.standardhostvalve.invoke	/product/product/buy/...	失败
api-gateway	org.springframework.web.servlet.frameworkservlet.dopost		成功
api-gateway	com.netflix.hystrix.hystrixcommand.queue		成功
api-gateway	com.netflix.hystrix.hystrixcommand$2.call		成功
api-gateway	org.apache.http.impl.client.closeablehttpclient.execute		失败
api-gateway	org.apache.http.protocol.httprequestexecutor.execute	/product/buy/3421122...	失败
products-service	org.apache.catalina.core.standardhostvalve.invoke	/product/buy/3421122...	失败
products-service	org.springframework.web.servlet.frameworkse...		失败
products-service	com.huawei.cloud.controller.productcontr...		失败
products-service	sun.net.www.protocol.http.httpurlco...		成功
user-service	org.apache.catalina.core.standa...	/user/validate	成功
user-service	org.springframework.web.s...		成功
user-service	com.huawei.cloud.api....		成功
user-service	com.huawei.clou...		成功

Thanks

Q&A