Smart Monitoring System For Automatic Anomaly Detection and Problem Diagnosis

Xianping Qu
quxianping@baidu.com
March 2015
Who am I?

• Xianping Qu
  – Senior Engineer, SRE team, Baidu
  – quxianping@baidu.com

• Baidu SRE Team
  – Over 400 Engineers
  – Support over 300 products, developed by 20,000 engineers

• Areas of Interest
  – Monitoring system
  – Data Analysis
About Baidu (百度)

- **Search**
  - Webpage/Image/Video/News/Dictionary/Web Directory/...

- **Social**
  - Forum/Album/...

- **LBS**
  - Maps/Group Buy/...

- **Knowledge**
  - Wiki/Knows/Experience/...

- **Mobile**
  - Search/Mobile phone assistant/...

- **Cloud**
  - Personal Cloud storage/Baidu Cloud/...

- ...
Agenda

• Background

• Smart Monitoring System
  – Anomaly Detection
  – Alarm Filter
  – Problem Diagnosis

• Summarize
Exponentially growing servers

Trend of numbers of servers at Baidu

~20 times over 5 years
The scale of the data

<table>
<thead>
<tr>
<th>Total Data</th>
<th>&gt; 1.5 EB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Webpage count</td>
<td>&gt; 300 billion</td>
</tr>
<tr>
<td>Webpage updates</td>
<td>&gt; 1 billion / day</td>
</tr>
<tr>
<td>Monitoring data</td>
<td>&gt; 20PB</td>
</tr>
<tr>
<td>Monitoring data growth</td>
<td>&gt; 40TB / day</td>
</tr>
</tbody>
</table>

Monitoring and Diagnosis are vital!
Early methods for monitoring and diagnosis

- **Monitoring Items**
  - Machine level: 90 million
  - Service level: 50 million

- **Alarms**
  - Threshold
    - Accuracy
    - Drifting
  - Annoyance vs promptness

- **Diagnosis**
  - Curves, Logs, Trace
Some Examples In Detail
Diverse thresholds

Threshold against yesterday’s value

Threshold against accumulated value
Drifting thresholds

Workday

Weekend

Holiday
Annoyance vs Promptness of Alarms
Diagnosis

- Monitoring
- Changing Events
- Relationship
- Log
- ...

```
<table>
<thead>
<tr>
<th>task</th>
<th>thread</th>
<th>state</th>
<th>start time</th>
<th>duration</th>
<th>status</th>
</tr>
</thead>
<tbody>
<tr>
<td>tdi</td>
<td>thread_1</td>
<td>done</td>
<td>2015-03-10</td>
<td>14:30</td>
<td>0.5s</td>
</tr>
<tr>
<td>event</td>
<td>Downloader_http</td>
<td>done</td>
<td>2015-03-10</td>
<td>14:45</td>
<td>1.26 s</td>
</tr>
<tr>
<td>tdi</td>
<td>thread_2</td>
<td>running</td>
<td>2015-03-10</td>
<td>14:50</td>
<td></td>
</tr>
<tr>
<td>event</td>
<td>Downloader_http</td>
<td>done</td>
<td>2015-03-10</td>
<td>14:55</td>
<td>1.27 s</td>
</tr>
<tr>
<td>tdi</td>
<td>thread_3</td>
<td>running</td>
<td>2015-03-10</td>
<td>15:00</td>
<td></td>
</tr>
<tr>
<td>event</td>
<td>Downloader_http</td>
<td>done</td>
<td>2015-03-10</td>
<td>15:05</td>
<td>1.28 s</td>
</tr>
<tr>
<td>tdi</td>
<td>thread_4</td>
<td>running</td>
<td>2015-03-10</td>
<td>15:10</td>
<td></td>
</tr>
<tr>
<td>event</td>
<td>Downloader_http</td>
<td>done</td>
<td>2015-03-10</td>
<td>15:15</td>
<td>1.29 s</td>
</tr>
</tbody>
</table>
```

```
[12250 task_manager.cpp:504] insert task to be executed: 5.npsa-imbs-999.IM.all
[12254 detect_task.cpp:890] build threshold success: 10.npsa-imbs-999.IM.all: IMBS_FLOW
[1224 task_manager.cpp:338] abnormal detected: [10.npsa-imbs-999.IM.all: IMBS_FLOW]
[1224 task_manager.cpp:306] execute query: [5.npsa-imbs-999.IM.all: IMBS_FLOWcnt_avg]
[1224 task_manager.cpp:338] abnormal detected: [5.npsa-imbs-999.IM.all: IMBS_FLOWcnt_avg]
[12250 task_manager.cpp:504] insert task to be executed: 10.npsa-imbs-999.IM.all
[12254 detect_task.cpp:890] build threshold success: 10.npsa-imbs-999.IM.all: IMBS_FLOW
[1224 task_manager.cpp:338] abnormal detected: [10.npsa-imbs-999.IM.all: IMBS_FLOW]
```

```
12214 redis_manager.cpp:50c] Redis pipeline thread exited
```
Automatic and smart monitoring and diagnosis
Smart monitoring system: 酷贝

Cool ➔ Bay

Collection ➔ Detection ➔ Diagnosis
Collection

• **Service data**
  – PV, income, flow...

• **Machine & program data**
  – CPU, MEM, DISK, NETWORK...

• **Changing events**

• ...

Anomaly detection system

- offline part
  - Offline data
  - Data classification algorithm
  - 3-sigma
  - Holt winters
  - Local regression
  - Clustering, compared with regular training
  - Training Parameters
  - Loading user labels
  - Generate anomaly detection configuration

- online part
  - Real time data
  - Online anomaly detection
  - Detect algorithms
  - Viterbi decoder
  - Alarm
Detection: Strategy 1, 3-sigma rule

• Determine constant threshold algorithmically
  – Statistics on past data
  – Assume Gaussian distribution

\[
\bar{x} \pm 3\sigma
\]
Detection: Strategy 2, the segmented 3-sigma rule

- Multiple distributions: day vs night
  - Split data into 15-min segments
  - Day-on-day statistics
Detection: Strategy 3, KS-test

- Multiple distributions: workdays vs weekends / holidays
  - Differentiate using KS-test
  - Statistics within Workday/Weekend/Holiday
Other factors to consider

- Sometimes, the data is changing
  - e.g. compare the holiday PV with the workday PV

![Graph showing data comparison between workday, weekend, and holiday periods.](image-url)
Detection: Strategy 3, holt-winters

- Holt-winters algorithm
- Learn both seasonal and adjacent trend
Detection: Strategy 4, local regression

- The detection of local spurt or sudden drop
  - LOESS algorithm based on local regression
Deal with slow decline in data trend

- Smart cumulative method

- The signal cycle of raw data changed
  - 10s, 30s changed to 15mins, 1hour, 24hours

- Then use 3-sigma and holt-winters to detect
Challenge: generated threshold parameters vs. business requirement

• Incorporate engineers’ manual input:
  – Modify the parameters
  – Mark undetected anomaly
  – Mark false positive alarm

• Learn adjustment automatically
  – Alarm label $\rightarrow$ Adjust parameters $+$ / $-$
Alarm filter

- Viterbi decoder
- The formation of abnormal events, rather than a single abnormal points
Problem diagnosis

• Help Diagnose
  – Total-dimensions and sub-dimensions
  – Upgrade/operation event and time series data
  – Operation and maintenance module relationship
Total dimension and sub-dimension

- Based on the total dimension and sub-dimension
- Example: the total revenue and advertise revenue
- Sort by impact weight
Heat map

- Sort anomaly numbers by dimension degree
- Slice data by abnormal regional, browser, channel
- Comparison of multi-dimensional anomaly
Issues caused by upgrades

• Based on upgrade events and metrics

• Time approximation
Issues diagnosis from operation and maintenance perspective

- Module calling graph
- Event and module relations
Establish the overall service view
Now, “酷贝” is born!
Summary

- It is difficult to detect anomalies and find the root cause when an anomaly occurs.

- A smart monitoring system for automatic anomaly detection @Baidu is demonstrated.

- Including data model of incidents, proactive anomaly detection algorithms, correlation analysis, and visualization.
Future

• Characterization of workload spikes
• Dynamic resource allocation
• Capacity management
• Identification of performance problems
• ...

Thanks