
Let's Visualize Log Files for Troubleshooting Java Applications #BOF7768

**Shin Tanimoto / Koji Ishida
Acroquest Technology Co., LTD.**

Who am I?



- 谷本 心 (Shin Tanimoto)
 - Acroquest Technology Co., LTD.
 - Java Troubleshooter
 - Board member of JJUG (Japan Java User Group)
 - Twitter : @cero_t
 - Facebook : shin.tanimoto



Who am I?



- 石田 浩司 (Koji Ishida)
 - Acroquest Technology Co., LTD.
 - IoT System Development
 - Myanmar Branch Technical Leader
 - Twitter : @kojiisd
 - Facebook : koji.ishida.399

Quiz

What is the origin of the word “log”?

Web

Images

Videos

Maps

Shopping

More ▾

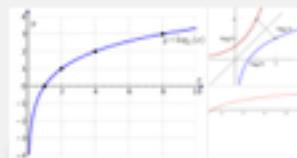
Search tools



SafeSearch ▾



Log



Logarithm



Hollow



Wood



Logo



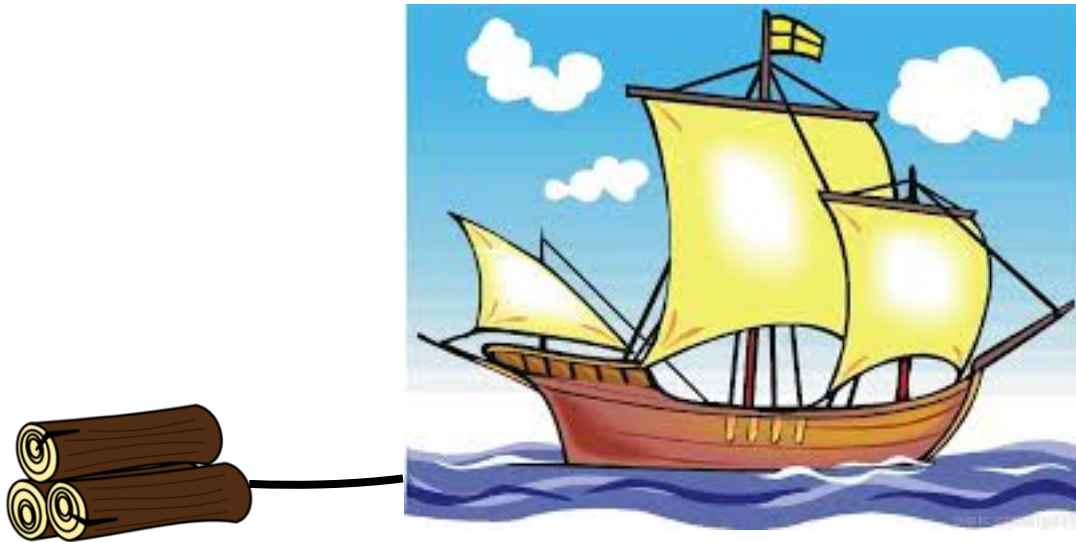
Wooden



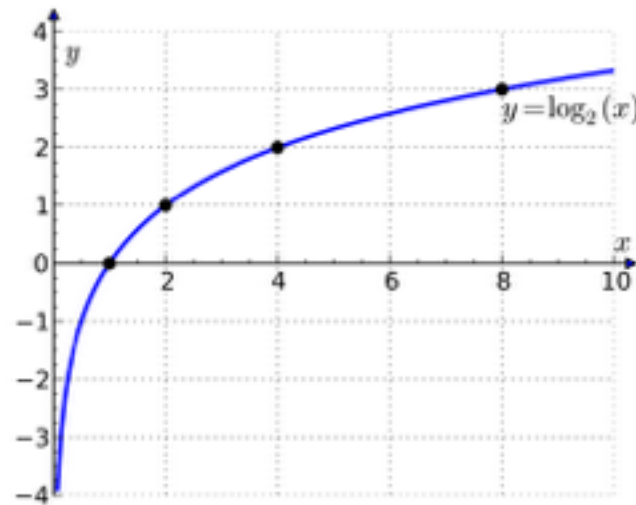
1. Ancient Greece people record the “date” using branches of the tree.



**2. In medieval Europe,
people measured “speed” of
ship with log (round wood).**



3. In the early 20th century United States, engineers used a logarithm table for “usage history” of computers.



- 1. Ancient Greece people's
“date” record.**
- 2. Medieval Europe sailors’
“speed” record.**
- 3. American engineers’
“usage” record.**

**1. ~~Ancient Greece people's~~
~~“date” record.~~**

**2. Medieval Europe sailors’
“speed” record.**

**3. ~~American engineers’~~
~~“usage” record.~~**

Log is “record”

Common sense: Log is important

**True common sense:
Watching log is painful!**

**Then log should be
watched and processed
by machine (ordinary)**

Let's Visualize Logs for Troubleshooting Java Applications #BOF7768

**Shin Tanimoto / Koji Ishida
Acroquest Technology Co., LTD.**

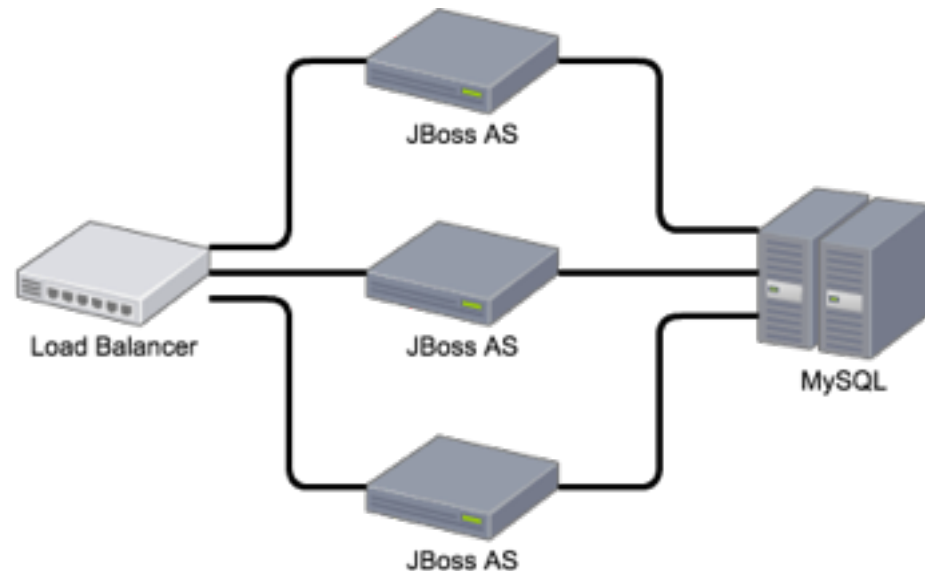
Table of contents

1. Troubleshooting case
2. Log processing
3. Log processing for business

#1 Troubleshooting case

1. Troubleshooting case

- Overview of the target system
 - Online booking web site of a famous hotel group
 - vacant room search / charge list / booking room
 - pv : > 10M / month
 - Components
 - Java SE 6
 - JBoss AS 6
 - Struts2
 - Hibernate3
 - MySQL 5.1



1. Troubleshooting case

- Overview of the target system
 - Issues...
 - Vacant room search needs $> 10s$
 - sometimes $> 30s$
 - Error occurs during booking process
 - TV programs pick up this hotel group, leading x20 traffic and system down

1. Troubleshooting case

- Strategy of troubleshooting
 1. Create integration test-cases using Selenium
 2. Source code reading and refactoring
 - Applying Findbugs
 3. Understanding + Test-cases = Maintainability!!

... resulted in

1. Troubleshooting case

waste of time 🙄

1. Troubleshooting case

- Why!?
 - Spaghetti code > 300KL
 - with lots of dead codes and copy codes
 - time wasting to read source code
 - While creating test-cases in integration test environment, system errors continue occurring in production environment
 - “The incidents are happening in the field!”

1. Troubleshooting case

**Care about the
issues in production**

1. Troubleshooting case

- It's time to Plan B
 - Refer to “logs” not source codes
 - Not to find theoretical issues, but real incidents
 - What should we focus on?
 - Errors
 - Performance
 - System resource

1. Troubleshooting case

- It's time to Plan B
 - What kind of logs should we collect?
 - To find Errors
 - Application logs
 - Access logs (Http status)

1. Troubleshooting case

- It's time to Plan B
 - What kind of logs should we collect?
 - To find Errors
 - Application logs
 - Access logs (Http status)
 - Every response status were 200 🙄

1. Troubleshooting case

- It's time to Plan B
 - What kind of logs should we collect?
 - To confirm performance
 - Access logs (Response time)
 - Slow query log of MySQL
 - To confirm system resource
 - sar
 - vmstat (or dstat)

1. Troubleshooting case

- It's time to Plan B
 - It is painful to read through all logs.
 - But we have to confirm whole logs to fix major issues first.

Then, why not visualize logs?

ELK

Web

Images

Videos

News

Maps






More ▾

Search tools





SafeSearch ▾









Drawing





Fighting



Head



Cow



In Snow



1. Troubleshooting case

- ELK stack
 - Elasticsearch - full text search engine
 - Logstash - log collection agent
 - Kibana - Front-end or UI for elasticsearch

1. Troubleshooting case

- ELK stack



send
logs



1. Troubleshooting case

- ELK stack
 - demo

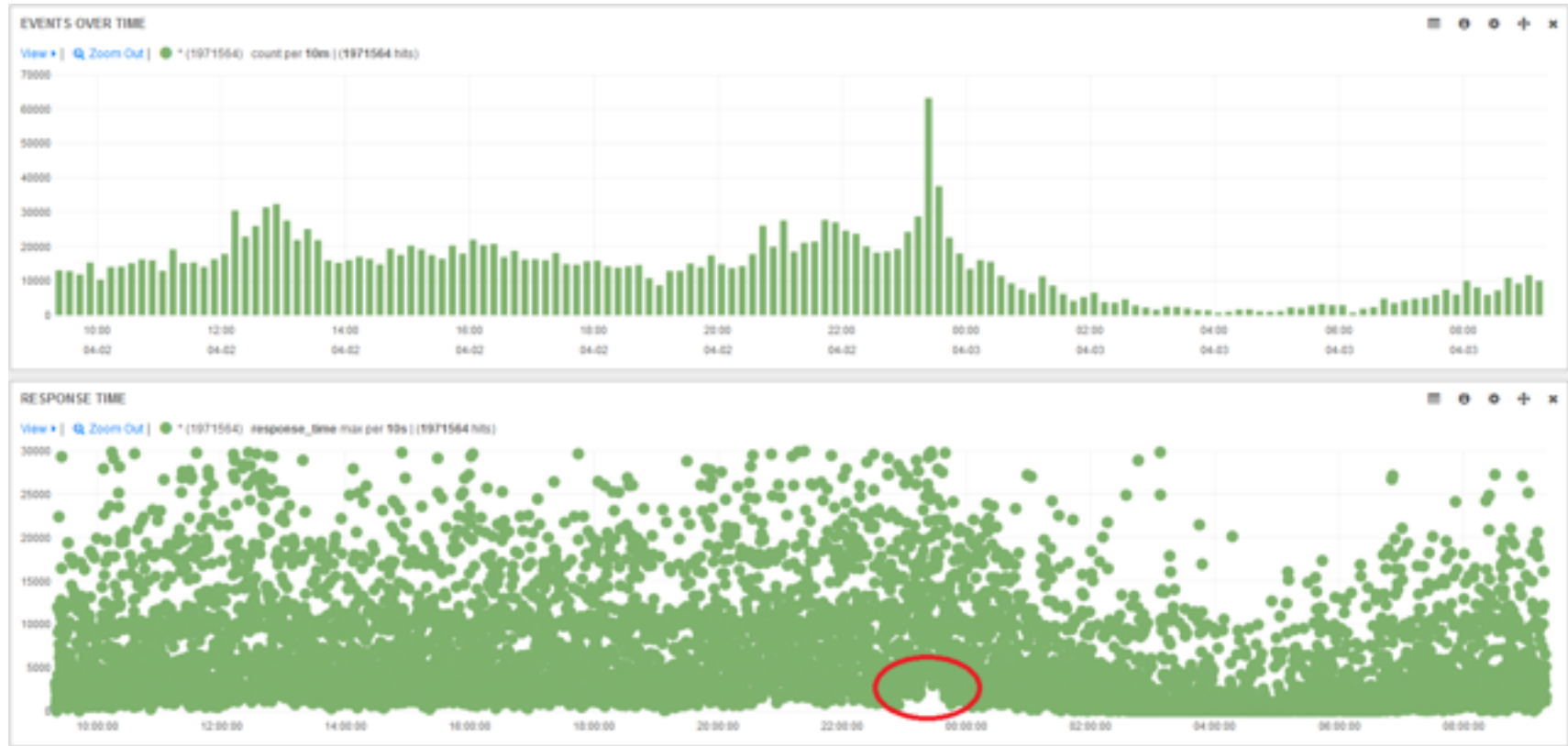
**Now we are ready for
troubleshooting.
Let's go on!**

Mission1: Performance issue of room search

Comparing access counts and response times

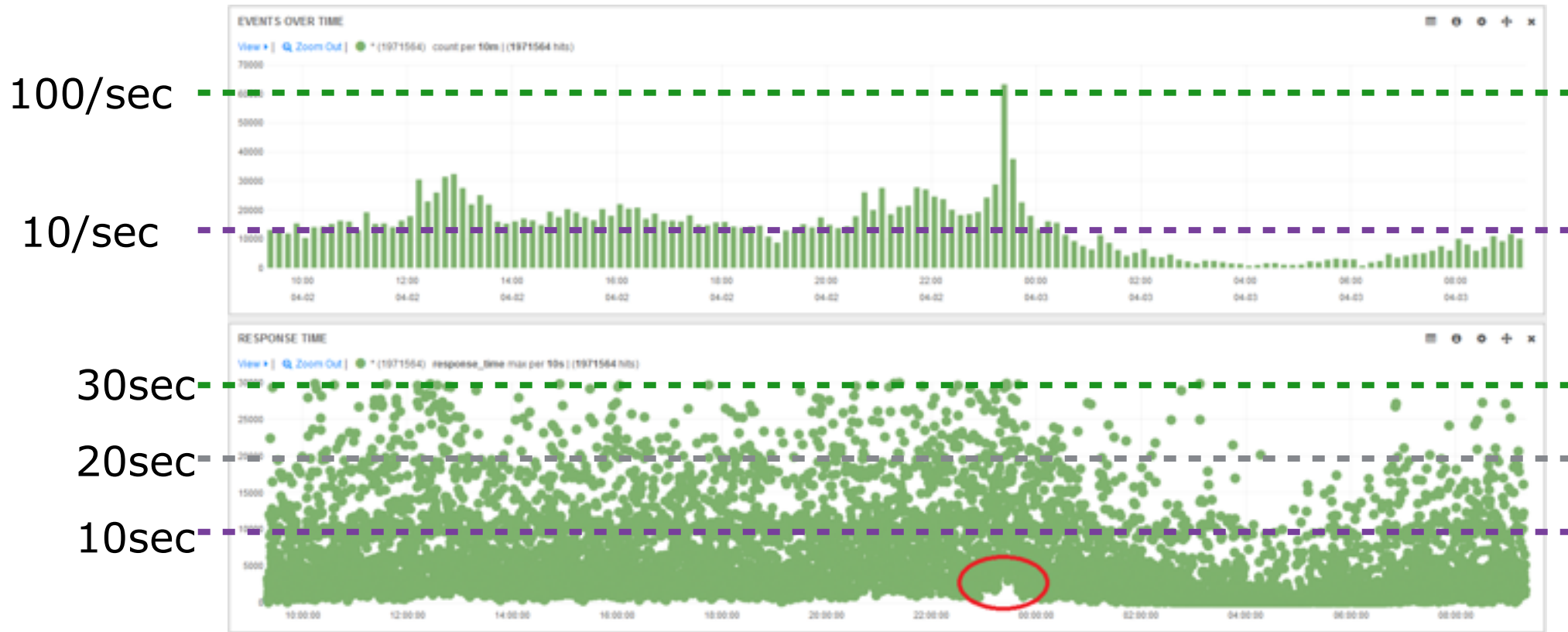
Mission 1 : Performance Issue of room search

- Access counts (upper) / response time (lower)



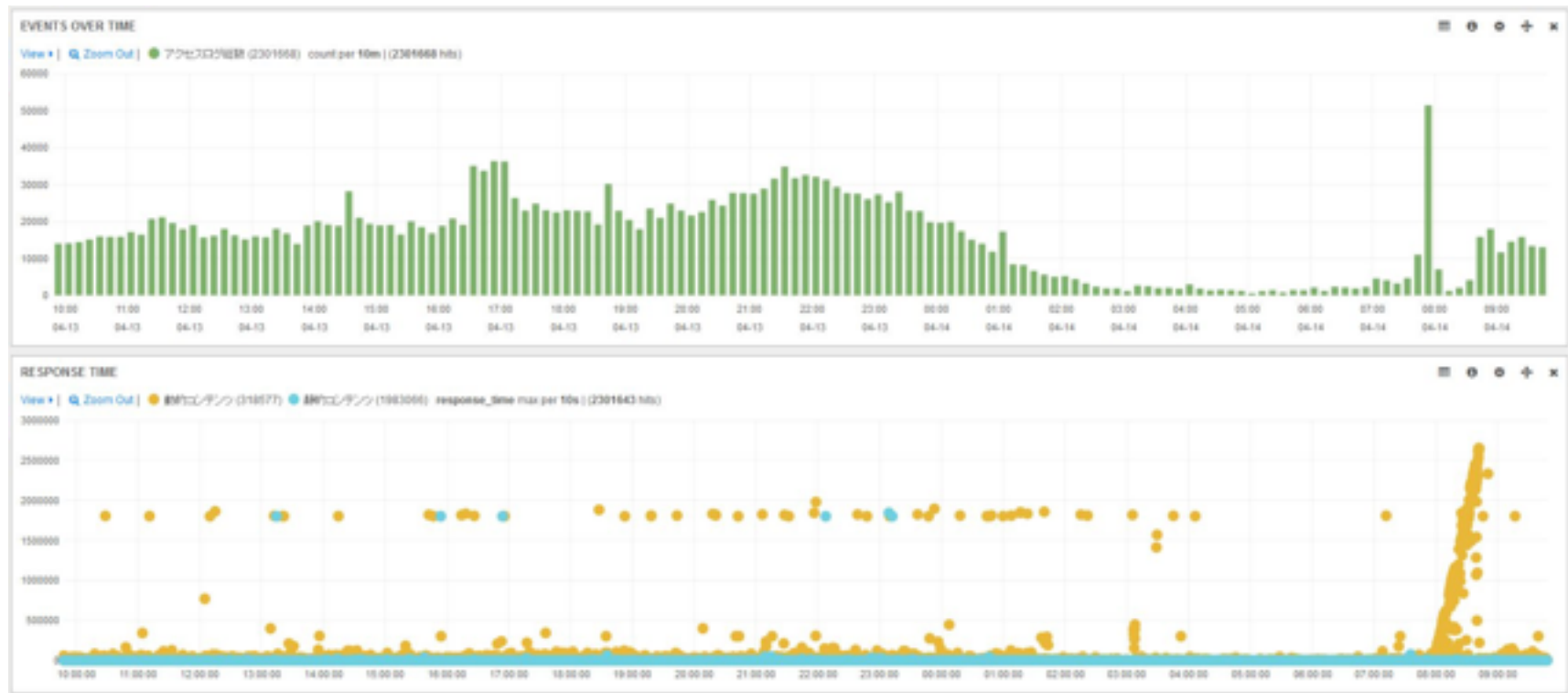
Mission 1 : Performance Issue of room search

- Access counts (upper) / response time (lower)



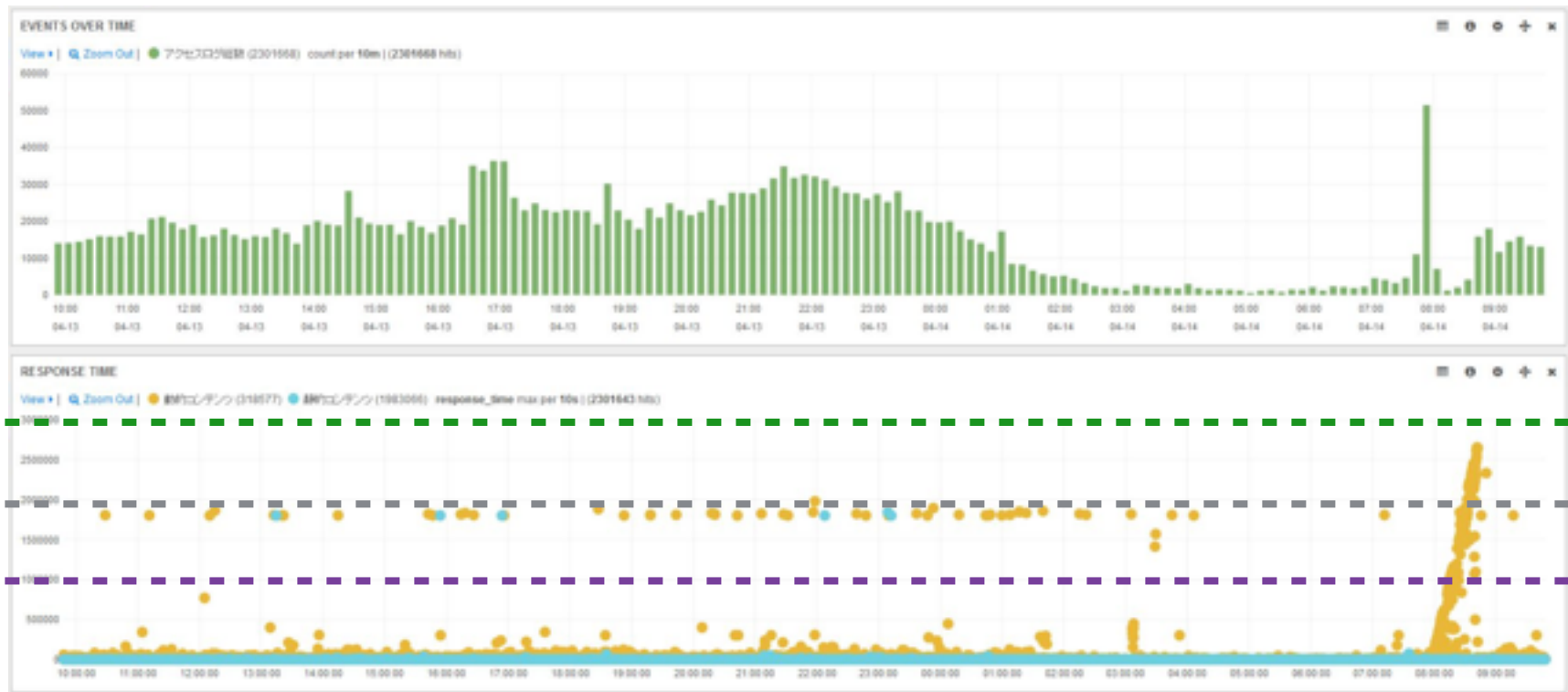
Mission 1 : Performance Issue of room search

- Huge performance issue



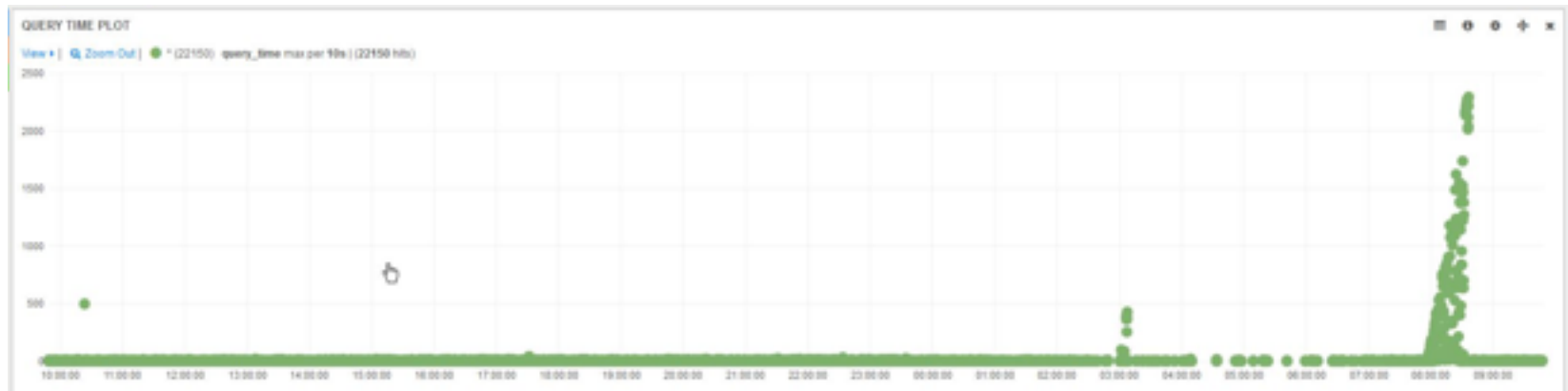
Mission 1 : Performance Issue of room search

- Huge performance issue



Mission 1 : Performance Issue of room search

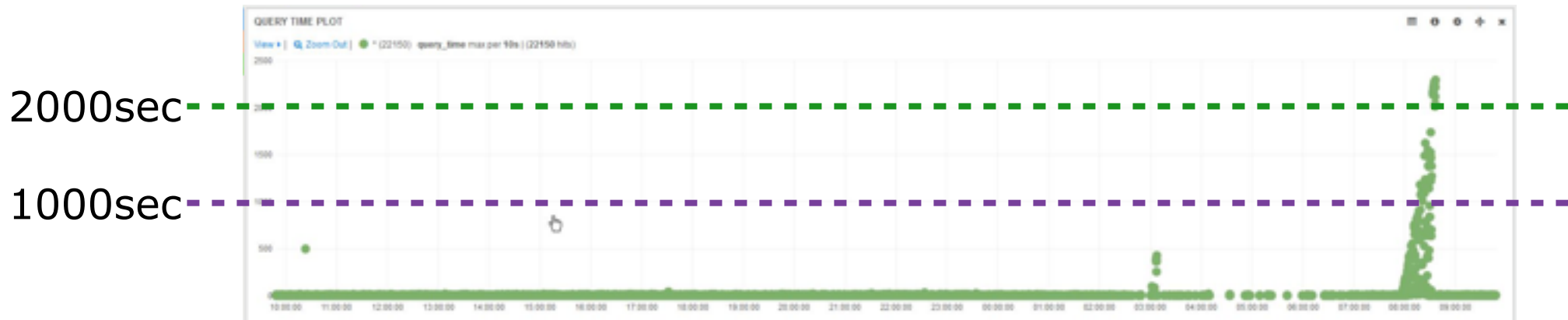
- Slow query log of MySQL



same shape!

Mission 1 : Performance Issue of room search

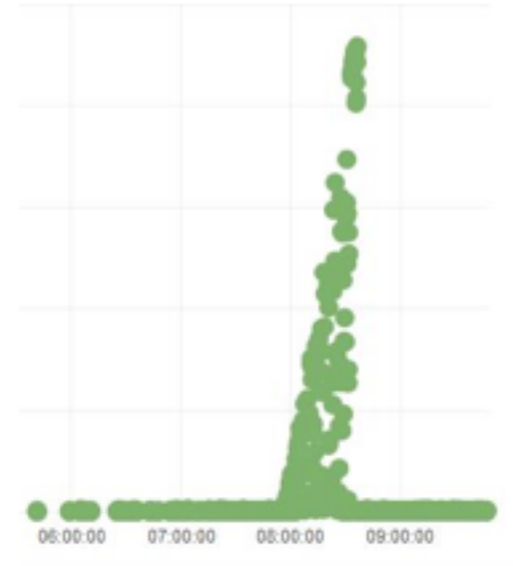
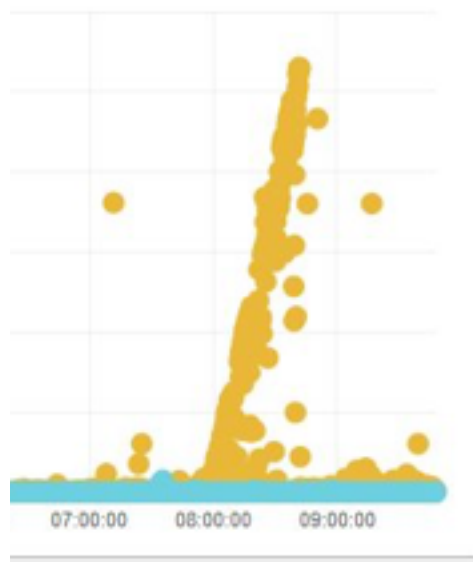
- Slow query log of MySQL



same shape! same scale!

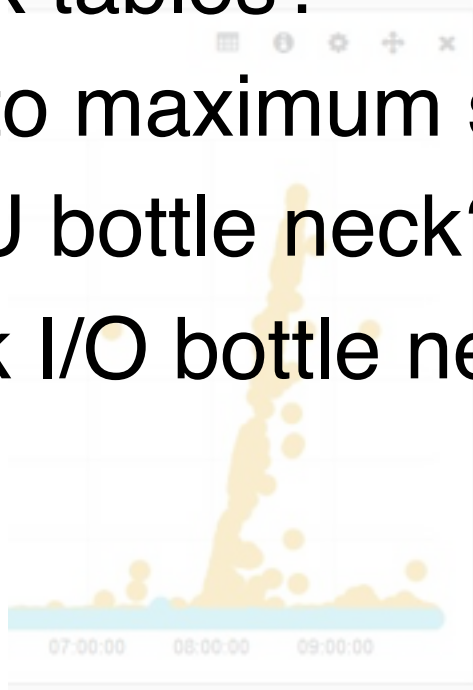
Mission 1 : Performance Issue of room search

- But where do these shapes come?



Mission 1 : Performance Issue of room search

- But where do these shapes come?
 1. Lock tables?
 2. Up to maximum size of connection pool?
 3. CPU bottle neck?
 4. Disk I/O bottle neck?

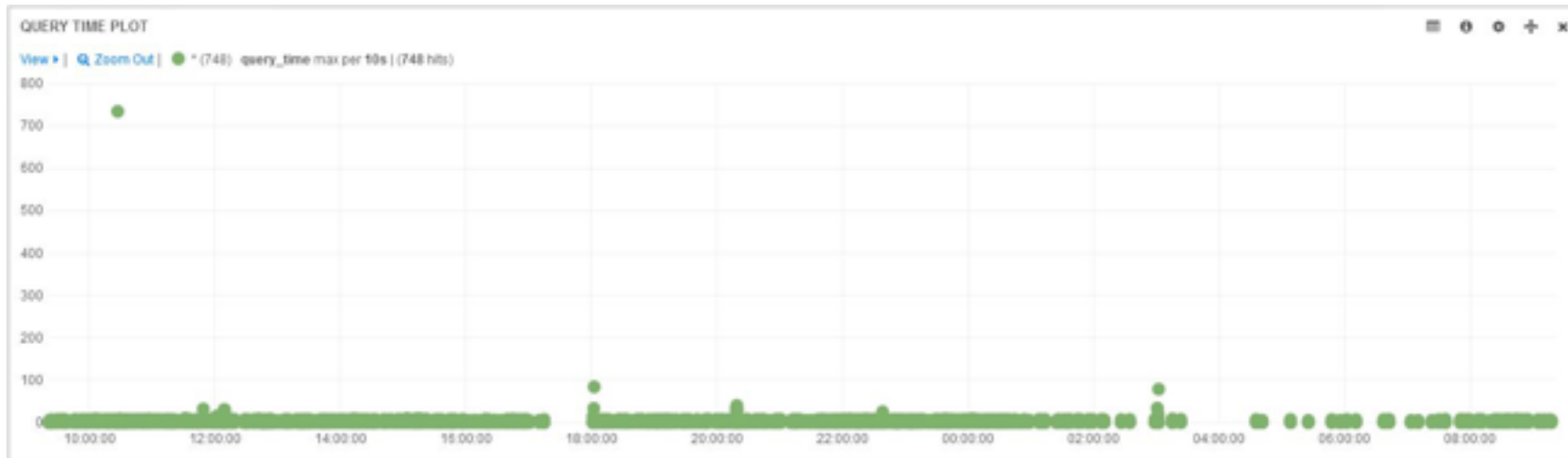


Mission 1 : Performance Issue of room search

- Confirm the stored procedure in detail
 1. Found 100,000 times of insert into “temporary table” query
 - (even in the search function ...)
 - causing high CPU and Disk I/O usage
 2. Optimized the stored procedure removing wasting process
 - Only a drop in the bucket 😞
 3. Modify the create temporary table state in the stored procedure to create that temporary table “on memory”
 - with memory tunings (tmp_table_size etc.)
 - resulted in ...

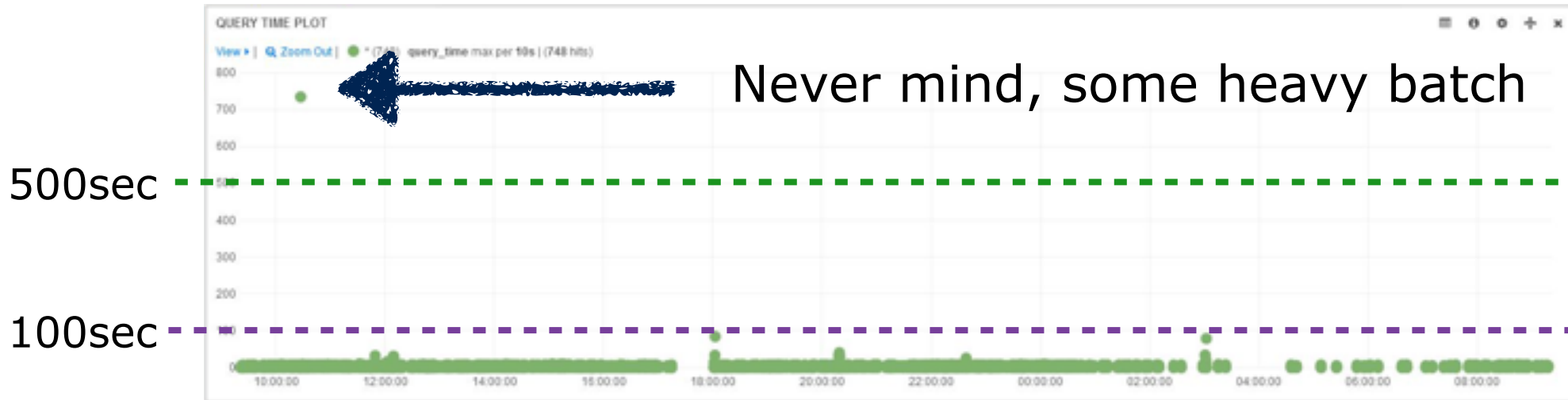
Mission 1 : Performance Issue of room search

- Performance issue was resolved!



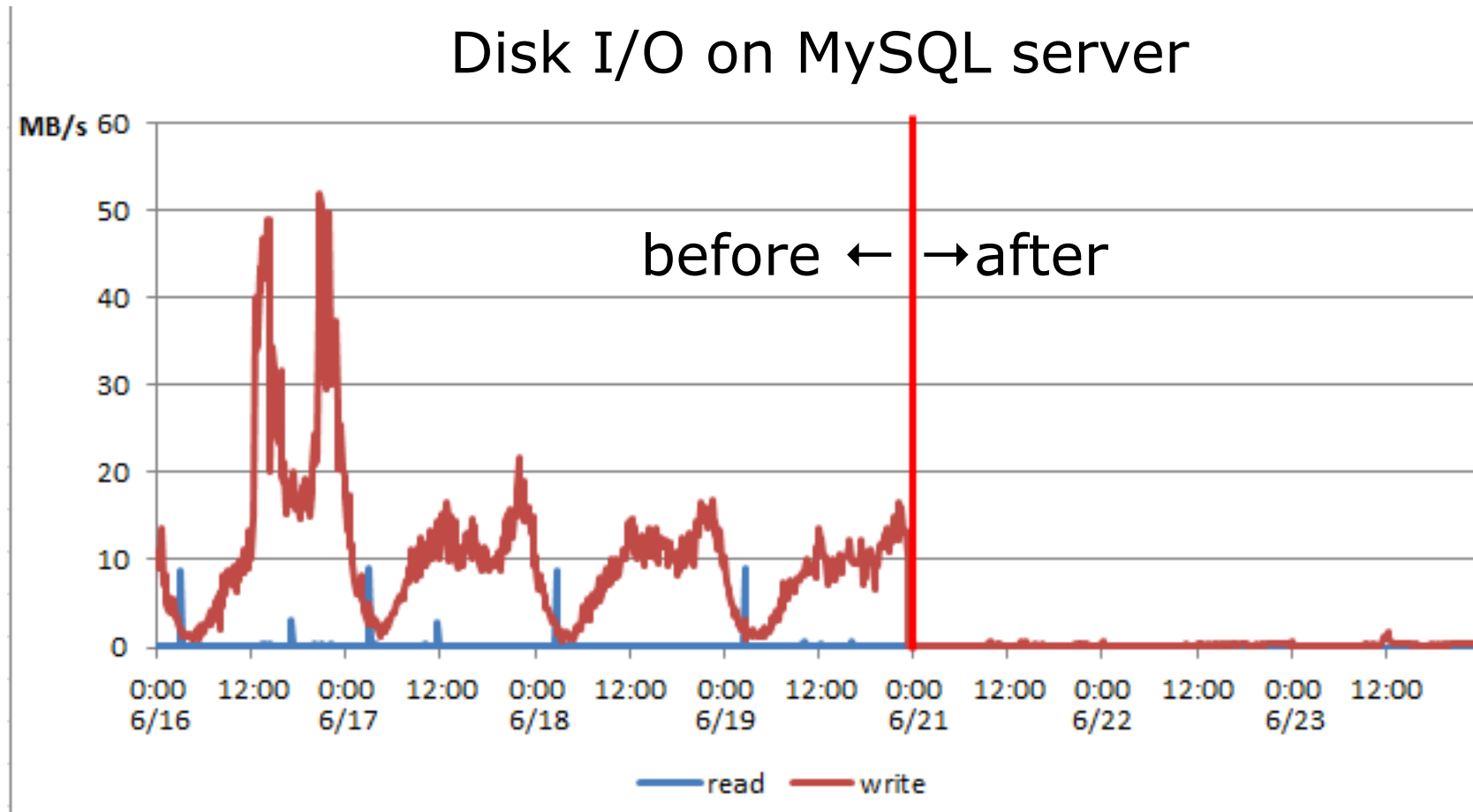
Mission 1 : Performance Issue of room search

- Performance issue was resolved!



Mission 1 : Performance Issue of room search

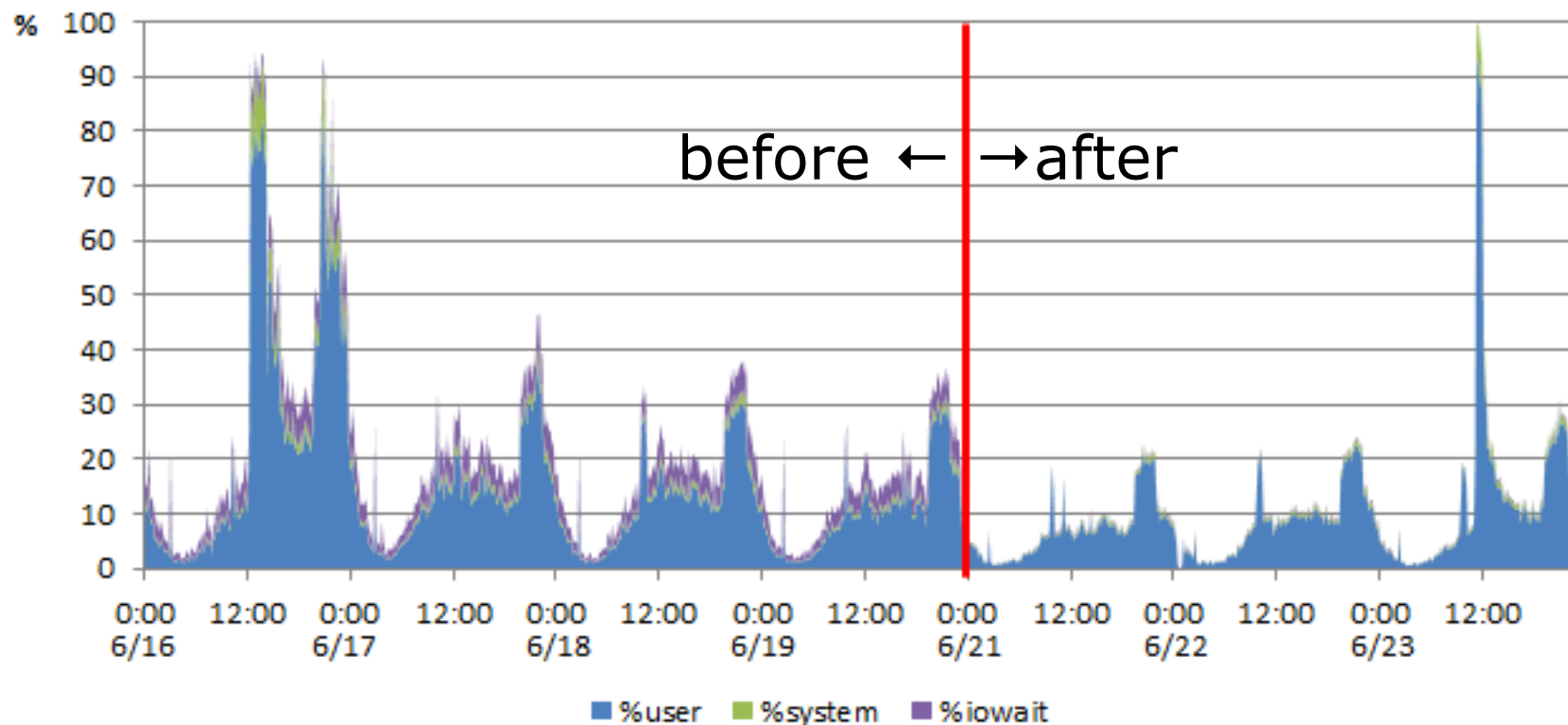
- Disk I/O improved!!!



Mission 1 : Performance Issue of room search

- I/O wait had gone!

CPU usage on MySQL server



Mission 1

**“Performance Issue of
room search”
was completed!!**

Mission2: Errors on booking

Mission 2 : Errors on booking

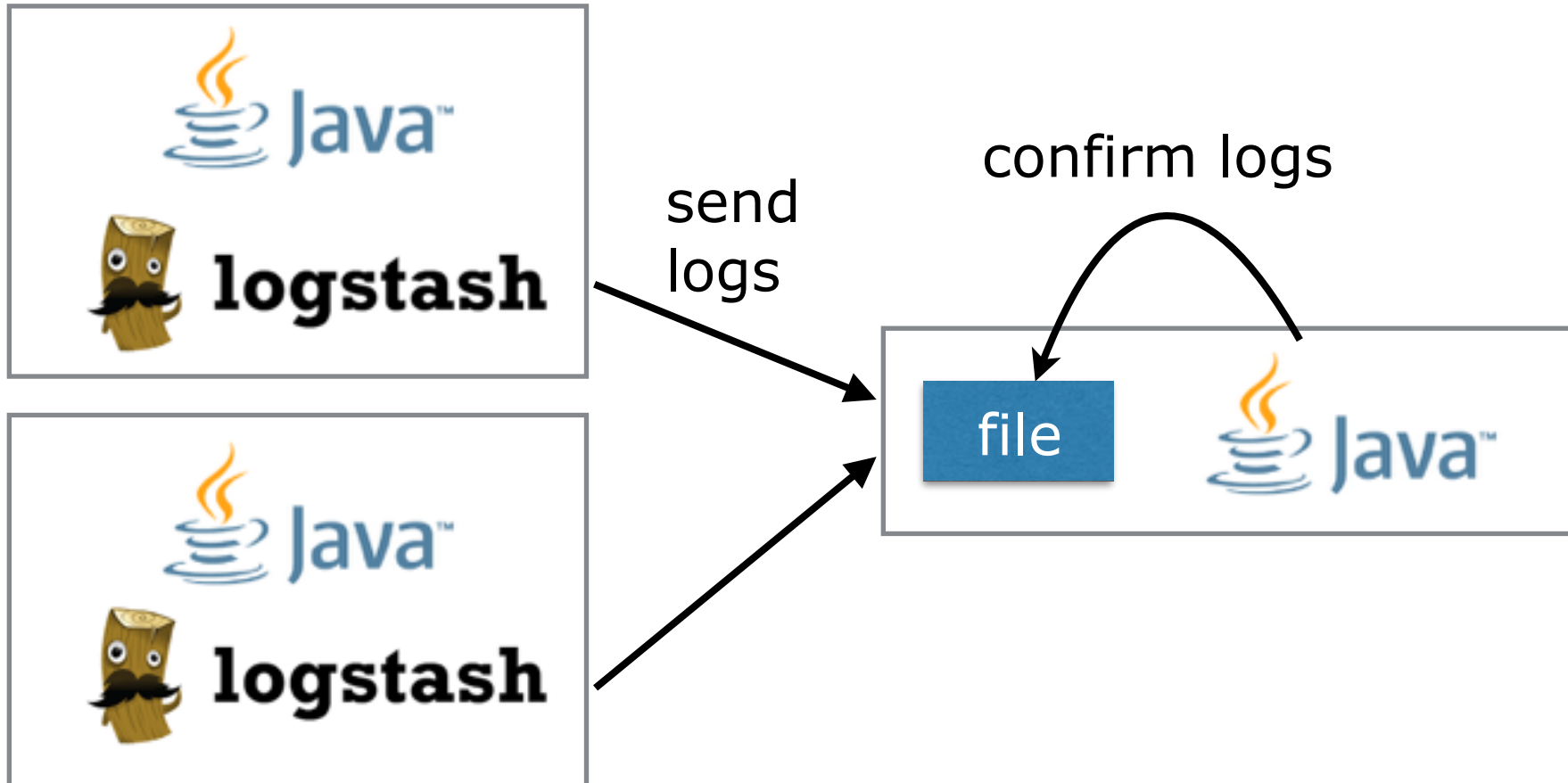
- Errors on booking
 - (1) Error occurs while processing booking requests
 - (2) Of course critical/severe error
 - Opportunity loss
 - Over booking
 - Card payment inconsistency
 - (3) Sometimes errors couldn't be even detected until the customers' claim.

Mission 2 : Errors on booking

- Strategy for troubleshooting
 - (1) Add the logging code at start / middle / end of the booking process.
 - (2) Create a batch to process logs to find the uncompleted booking processes.
 - Notify by e-mail when uncompleted processes were found
 - (3) Fix the issues in order of the frequency of the error occurrence.

Mission 2 : Errors on booking

- Strategy for troubleshooting



Mission 2 : Errors on booking

- This strategy worked very well
 - Apply patches every week
 - Errors were decreased to half every week.
 - Finally error occurs once in a week.
 - Yes, some errors still there... 🙄

Mission 2

**“Errors on booking”
completed!!**

- What we have learned from this troubleshooting
 - Analysing logs can help to resolve issues
 - Detecting errors
 - Finding the cause of the issues
 - We can use logs in various ways
 - Visualizing
 - Watching
 - Viewing

Table of contents

1. Troubleshooting case
- 2. Log processing**
3. Log processing for business

#2

Processing logs

**Watching logs to
detect errors is a
responsibility of
developers, isn't it?**

Watching logs is important but painful

Let's think about painless log processing system

#2 Processing Logs

- Logs can be used in various ways
 - Visualizing - as chart
 - Watching - and notifying by e-mail
 - Viewing - by human's eyes
 - Keeping - backup just in case

#2 Processing Logs

- Logs can be used in various purpose
 - Visualizing - To find “unknown” issues
 - Watching - To find “known” issues
 - Viewing - To find the cause of issues
 - Keeping - To use as necessary

#2 Processing Logs

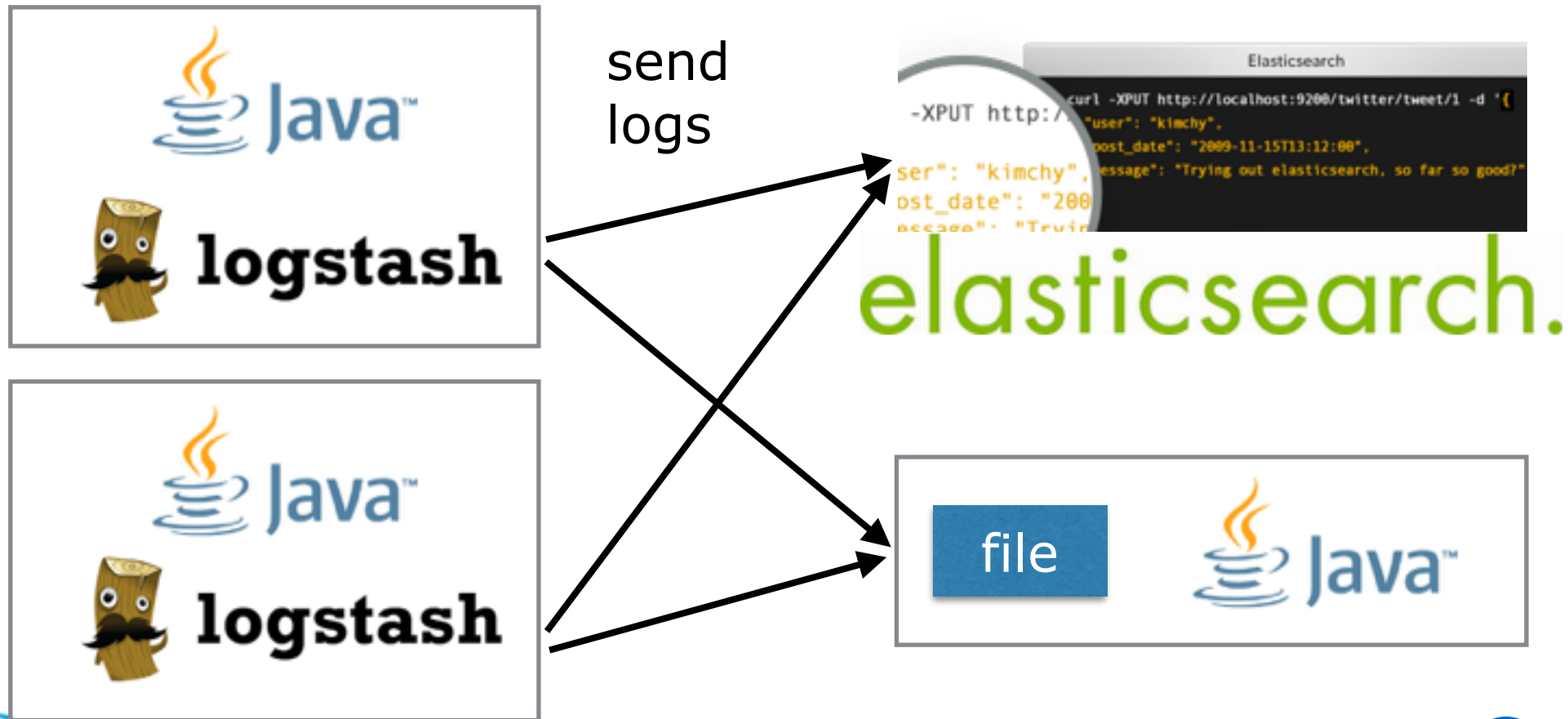
- Logs retention period are also various
 - Visualizing - last 2 or 4 weeks
 - Watching - last 24 hours
 - Viewing - last 2 or 4 weeks
 - Keeping - entire period

#2 Processing Logs

- Tools for processing logs are different
 - Visualizing - Elasticsearch
 - Watching - Zabbix or some custom batch
 - Viewing - Text editor
 - Keeping - File server

#2 Processing Logs

- Therefore, log processing system tends to be complexed



What is log?

#2 Processing Logs

- Is log a “file” ?
 - Not necessary
 - Sometimes output into standard output
- Is log a “record” ?
 - Not necessary
 - Sometimes processed in real time

Is log an “event?”

#2 Processing Logs

- Is log an “event”?
 - Yes.
 - In computing, an event is an action or occurrence recognised by software that may be handled by the software. Computer events can be generated or triggered by the system, by the user or in other ways.
(by wikipedia)

Is log a “stream?”

#2 Processing Logs

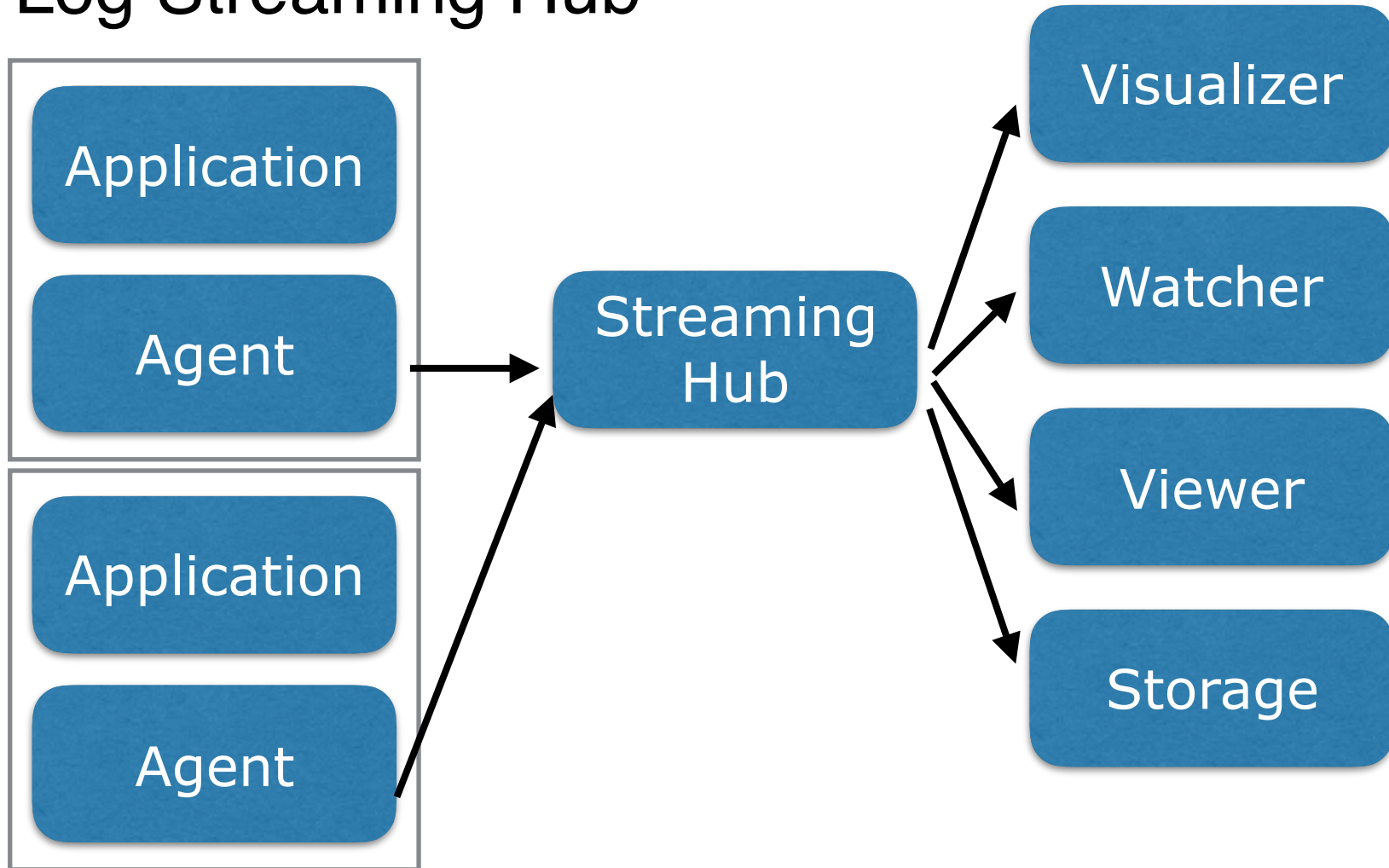
- Is log a “stream”?
 - Yes.
 - In computer science, a stream is a sequence of data elements made available over time. A stream can be thought of as items on a conveyor belt being processed one at a time rather than in large batches.
(by wikipedia)

**Log is an event,
Log is a stream**

Log streaming hub

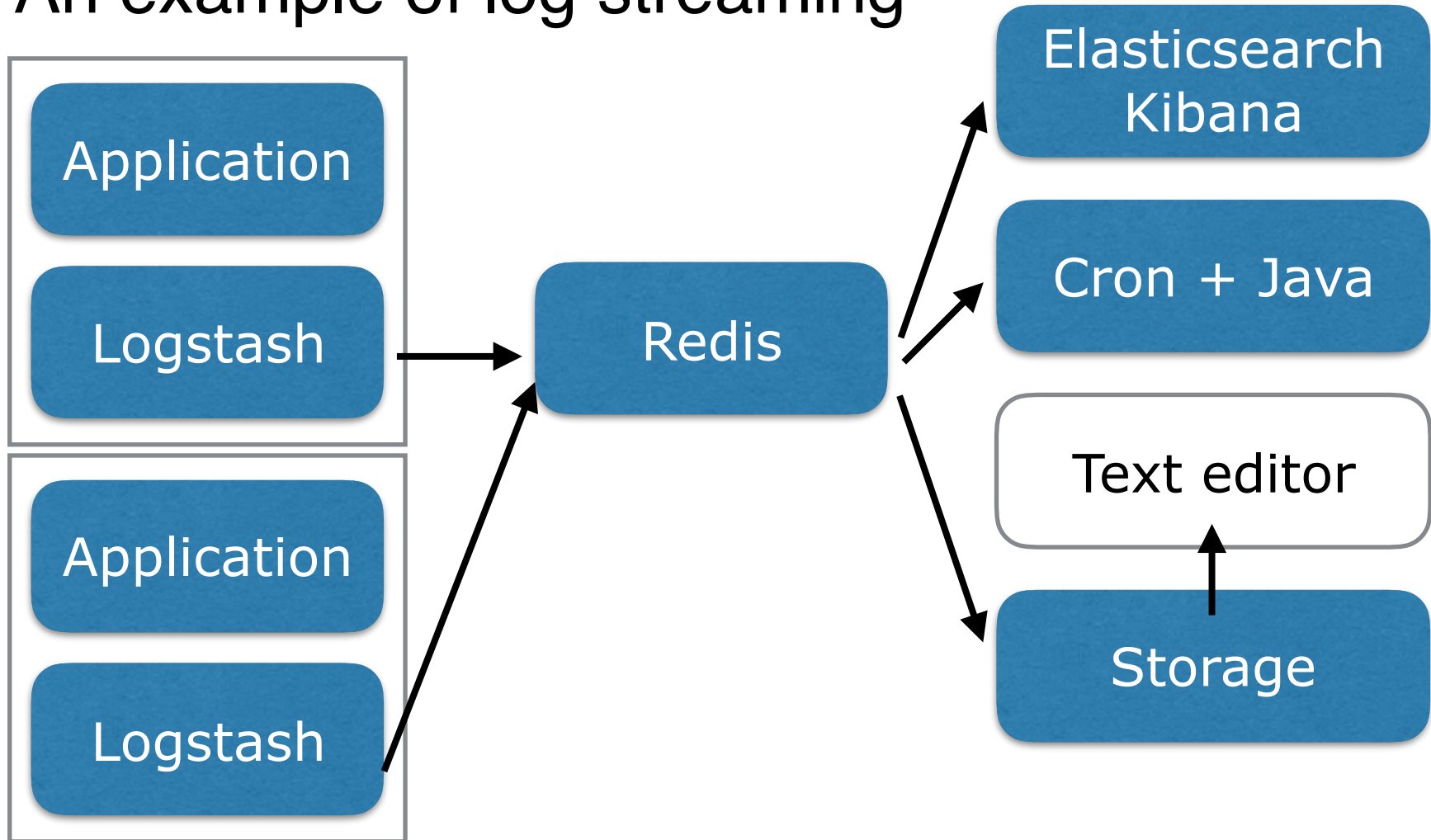
#2 Processing Logs

- Log Streaming Hub



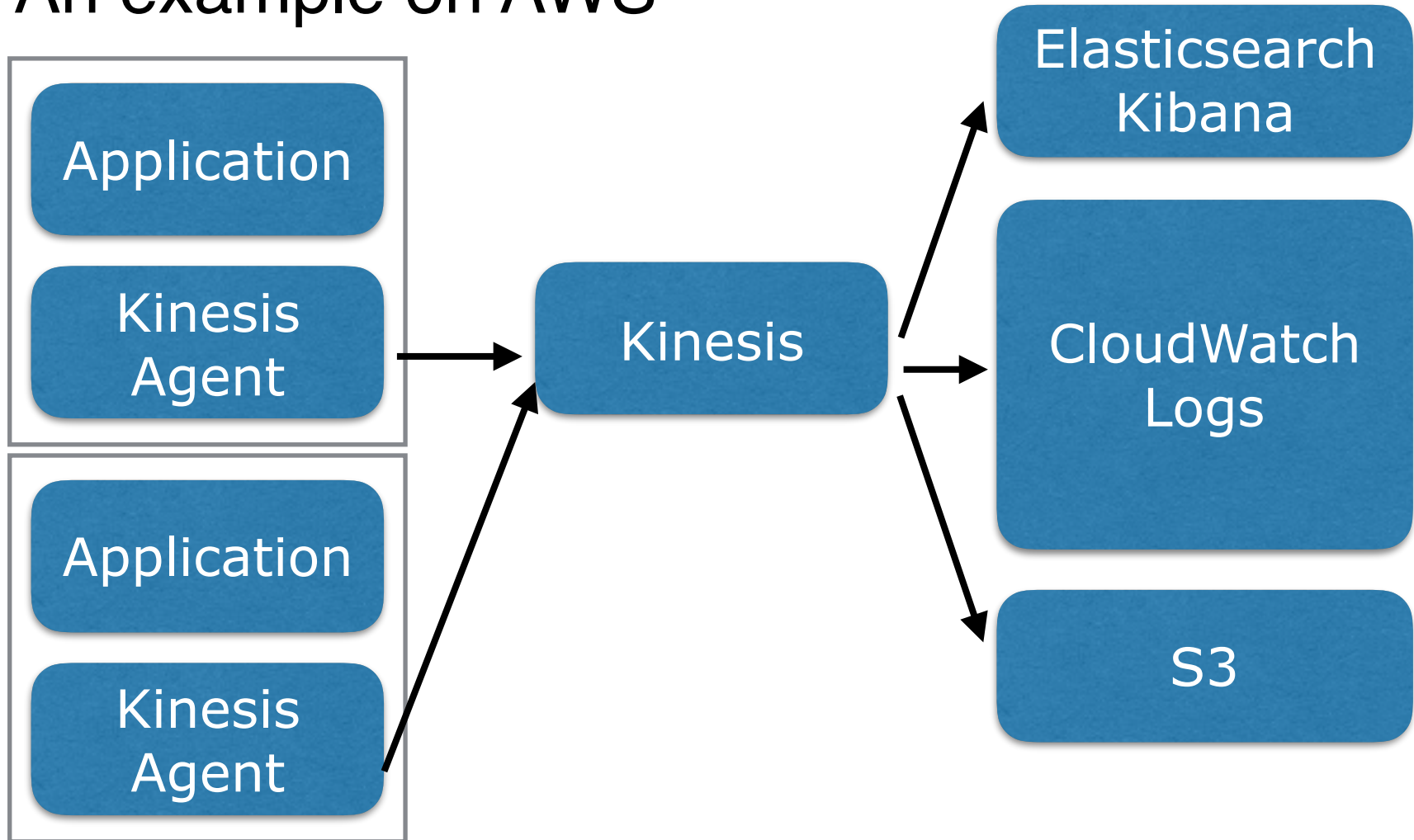
#2 Processing Logs

- An example of log streaming



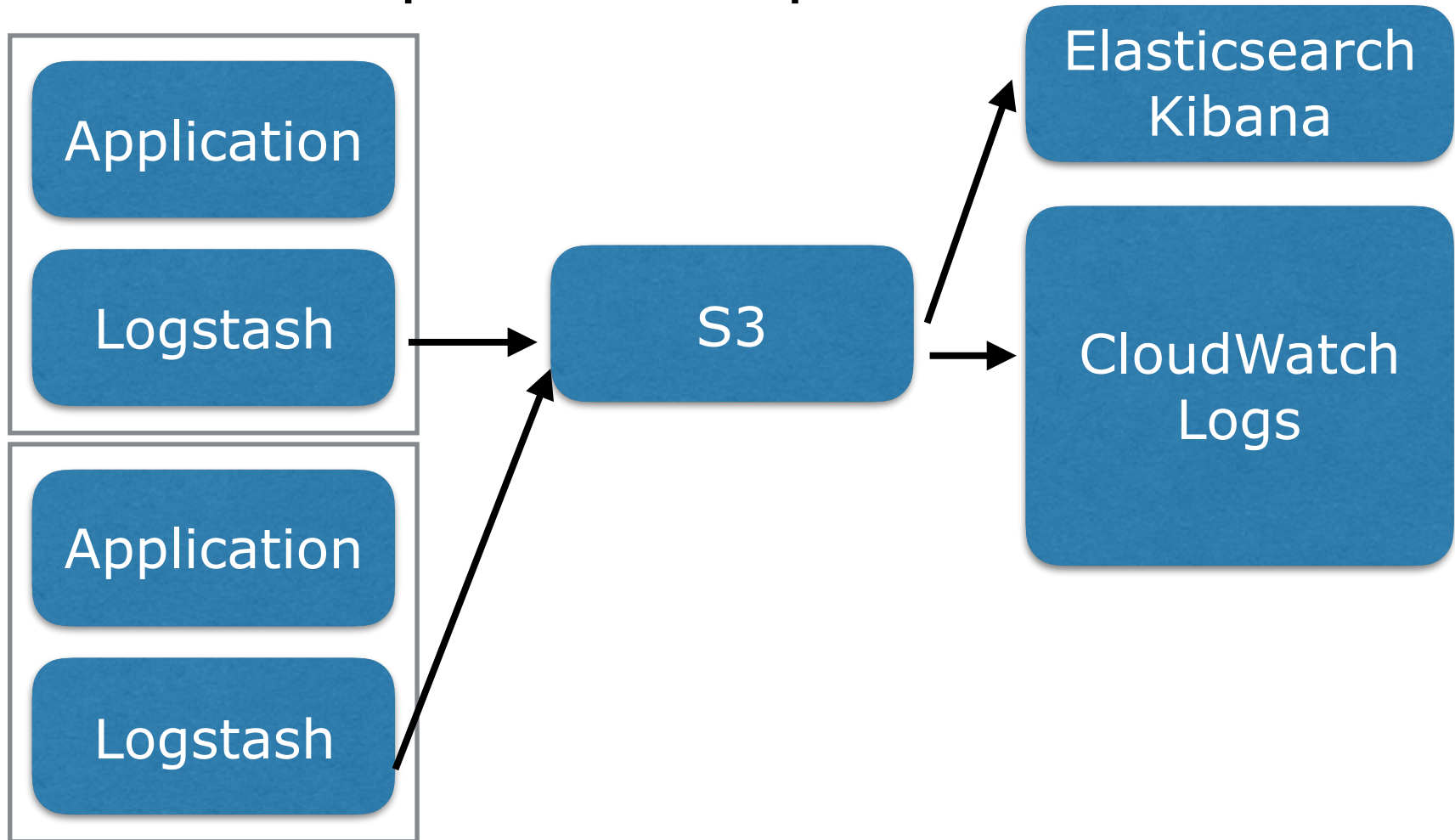
#2 Processing Logs

- An example on AWS



#2 Processing Logs

- Another simplified example on AWS



#2Processing Logs

- Visualizing logs = Visualizing systems
 - Get the system resources
 - Application errors
 - Access Results

**Will your manager
pay for developing
these systems?**

Which kind of visualization do they want?

Table of contents

1. Troubleshooting case
2. Log processing
- 3. Log processing for business**

#3

Processing logs for business

(briefly)

#3 Processing Logs for Business

- Visualizing businesses from logs
 - User behaviours
 - Conversion rates
 - booking count / unique user count
 - Search words
 - ...
- This kind of visualization touches management layer

#3 Processing Logs for Business

- Visualizing businesses from logs
 - demo

These are ordinal in Amazon / eBay

**Now we can apply
the idea for smaller
business web sites!**

Enjoy processing logs for business!



Acroquest Technology