

The BigPanDA self-monitoring alarm system for ATLAS

A. Alekseev, T. Korchuganova, S. Padolski

8th International Conference "Distributed Computing and Grid-technologies in Science and Education" (GRID 2018),
10-14 September 2018, Dubna, Russia

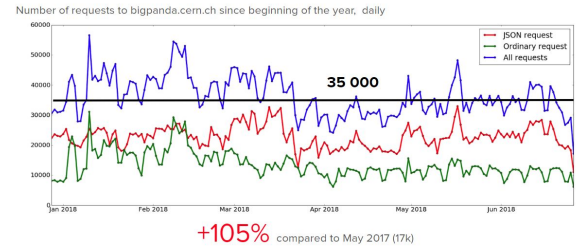
Outline

- Motivation
- Classification of BigPanDA errors
- State-of-the-art approaches for monitoring of distributed systems
- Architecture of BigPanDA self-monitoring alarm system
- Processing notification messages stream
- Structure of notification message
- Summary

Motivation

- **BigPanDA** is a monitoring system which provides a comprehensive and coherent view of the tasks and jobs is executed by the PanDA system, from high level summaries to detailed drill-down job diagnostics (*The BigPanDA monitoring system architecture, Korchuganova T. et al., Grid 2018*)
- Highly loaded service for analysis of Big Data in real-time
- ~35 000 requests to the system per day, including ~25 000 JSON requests
- Multicomponent/multi-module distributed service
 - 7 nodes + 1 node are working in load-balancer mode
 - Web-server (Apache), Load-balancer (Nginx), Database (Oracle), Distributed cache (Redis)
 - External authentication providers (CERN,Google,GitHub)

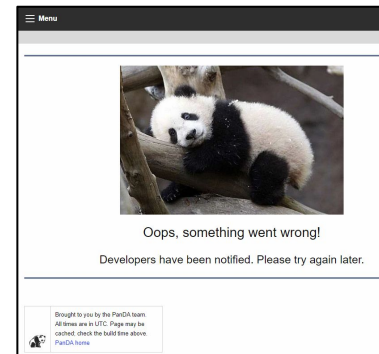
Usage statistics



BigPanDA is used as a processing engine!

Classification of BigPanDA system errors

- Internal BigPanDA system errors:
 - **View errors (e.g. wrong variable type, none values, etc)***
 - User errors (e.g. wrong url requests to bigpanda system)
- External systems errors:
 - **Database errors (e.g. exceeding the number of simultaneous sessions or Oracle database not available)***
 - Extended libraries errors (e.g. social-auth lib)
 - **Cache errors (e.g redis not available)***
 - Extended components errors/critical problems (e.g. Apache, Nginx performance issues)
- Superfluous requests (DoS-attack, irresponsible user behavior)



* **Critical**

Internal Server Error

The server encountered an internal error or misconfiguration and was unable to complete your request.

Please contact the server administrator, webmaster@mydomainname.com and inform them of the time the error occurred, and anything you might have done that may have caused the error.

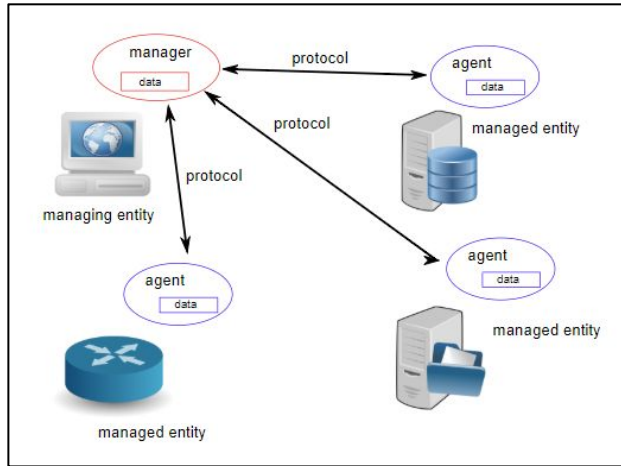
More information about this error may be available in the server error log.

Additionally, a 500 Internal Server Error error was encountered while trying to use an ErrorDocument to handle the request.

Apache/2.2.21 (Unix) mod_ssl/2.2.21 OpenSSL/0.9.8e-fips-rhel5 mod_auth_passthrough/2.1 mod_bwlimited/1.4 FrontPage/5.0.2.2635 Server at mydomainname.com Port 80

State-of-the-art approaches for monitoring of distributed systems

- **Simple Network Management Protocol (SNMP)**



- **Logs**

```
1 #Software: Microsoft Internet Information Services 8.0
2 #Fields: date time s-sitename cs-method cs-uri-stem cs-uri-query s-port cs-username c-ip cs(User-Agent) cs(Cookie) cs
3 2017-02-15 13:54:26 AZUREOVERVIEW_522C GET / X-ARR-LOG-ID=8b90ffae-54b1-43c7-abaa-af0fcb759fac 80 - 141.101.76.74 Mo
4 2017-02-15 13:54:26 AZUREOVERVIEW_522C GET /css/site.min.css X-ARR-LOG-ID=3a3cd496-76fe-42a4-be99-7e34fb69aa63 80 -
```

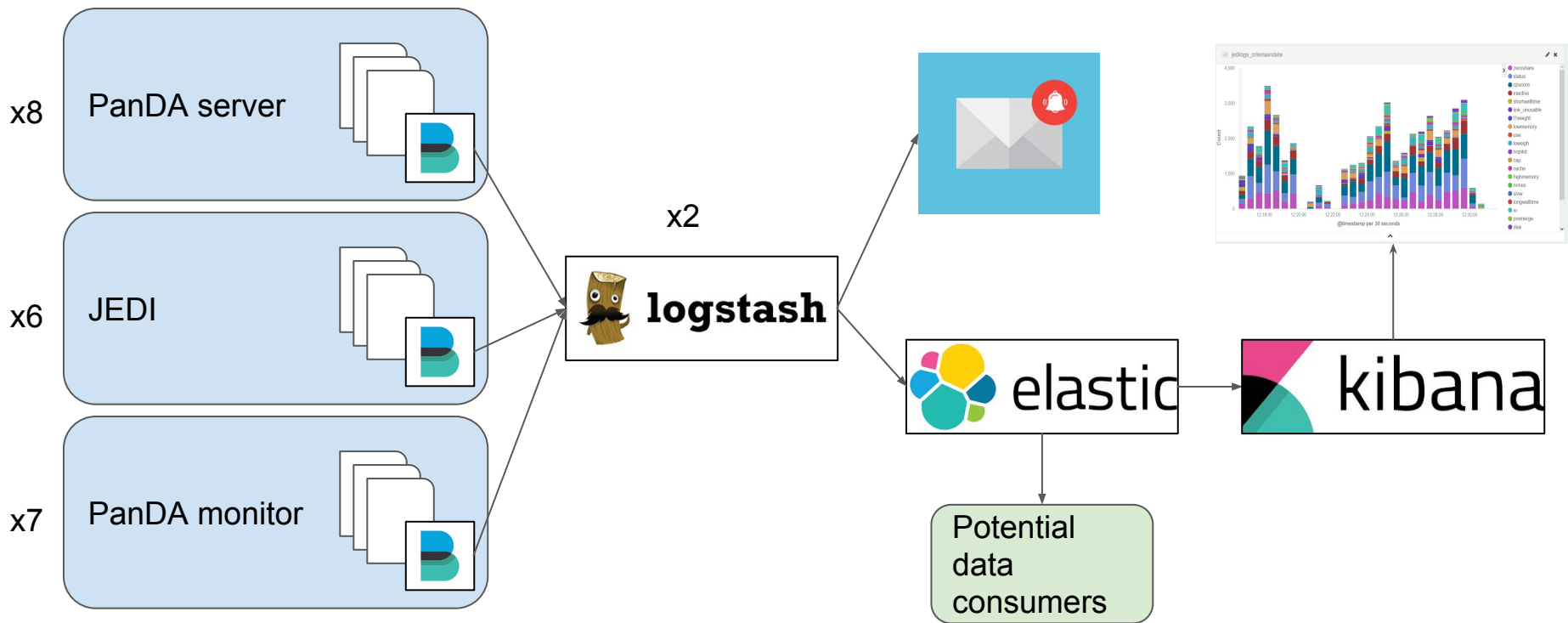
State-of-the-art approaches for monitoring of distributed systems

- Development of notification system using third-party libraries
- Using third-party monitoring software
 - Based on SNMP protocol (-) **to adapt BigPanDA project for SNMP**
 - Solarwinds Server and Application Monitor (SAM)
 - + great capabilities for visualization, easy to install and set up
 - commercial, windows platform
 - Zabbix
 - + open-source, integrating Django application metrics into Zabbix
 - need to install Agents, complicated customization for specific tasks
 - Nagios
 - + open-source, easy to set up
 - customization and updates are bit difficult
 - Based on analysis of logs (-) **online services**
 - Graylog
 - + open-source, rich opportunities for processing logs
 - **ELK-stack (Logstash, Elasticsearch and Kibana)**
 - + open-source, rich opportunities for processing logs, existing ATLAS infrastructure
 - filters should be developed, need to install Filebeat

ELK - stack

- **Logstash** is an special open source software for collecting, filtering and normalizing logs. It is used to collect log events from different log types using special filters
- **ElasticSearch** is a distributed open source software for storing and searching information. In our case, Logstash writes all log events into the ElasticSearch repository
- **Kibana** is an open source data visualization plugin for Elasticsearch. It is used for visualization of data from Elasticsearch cluster

Architecture of BigPanDA self-monitoring alarm system

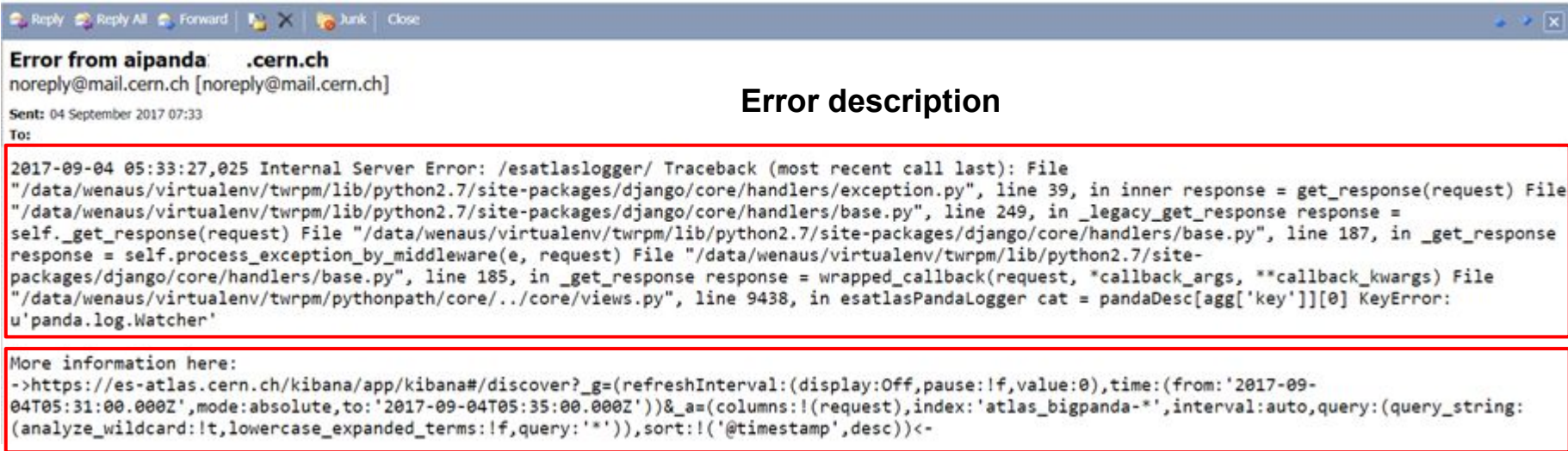


System design diagram

Processing notification messages stream

- **Application logs.** Filtered by “Internal Server Error” condition. Catch the following errors:
 - Oracle database related
 - Django (framework) related
 - Social-auth lib related
- **Web-server logs.** Log types “port80_access” or “bigpanda_access_ssl” and size of message “538” bytes
- **Filtration efficiency.** ~4000 notification candidates generated daily. Most of them are from broken connections. The system delivers only tiny part of errors (~10) which require a BigPanDA operator attention
- **An error description** collected from the different log types and aggregated into one message

Structure of notification message



Error description

Detailed information

Structure of notification message



Total entries within 4 minutes

September 4th 2017, 12:31:22.000 /worldjobs/

September 4th 2017, 12:33:27.025 -

September 4th 2017, 12:33:25.000 /esatlaslogger/

Django log with error description

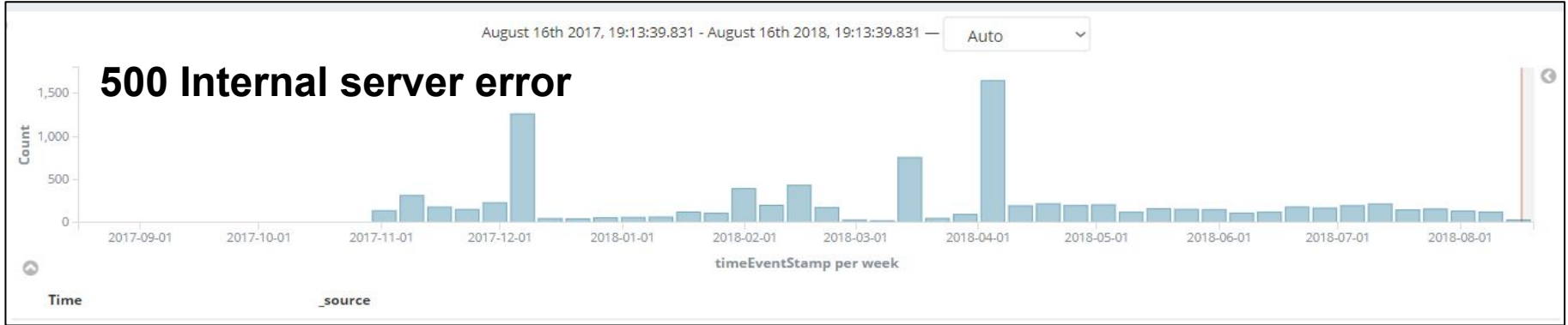
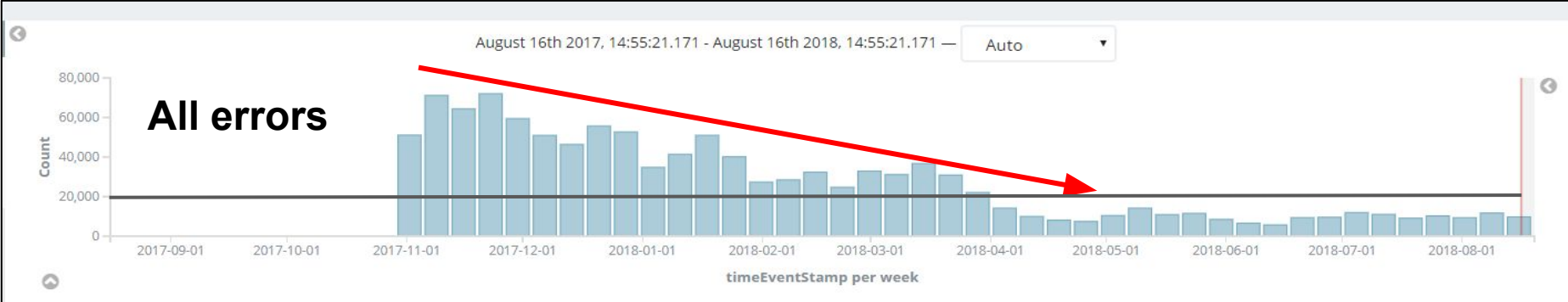
Table	JSON	View source
@timestamp	September 4th 2017, 12:33:27.323	
@version	1	
._id	AV5LYSGnBX3BKXyztfkI	
._index	atlas_bigpanda-2017.09.04	
._score	-	
._type	port80_access	
agent	"Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/60.0.3112.90 Safari/537.36 OPR/47.0.2631.71"	
auth	-	
beat.hostname		
beat.name		
beat.version	5.4.1	
bytes	8796	
clientip		
host		

Apache log with user request

Summary

- ELK based self-monitoring system for BigPanDA was developed
- The system is in production since May 2017 and processes ~1 million log messages daily
- Logstash filters for BigPanDA logs processing were developed. Self-monitoring brings attention to the errors which require human intervention
- Error messages are sent to the BigPanDA developers immediately and available via Kibana dashboards
- Monitoring different components of BigPanDA monitoring system
 - Apache, Nginx
 - Oracle database
 - Social-auth library
- **Leads to better understanding of BigPanDA production behaviour and system insights.** A patch for Nginx load balancer, WSGI garbage collector, DDoS protection mechanism implemented using obtained insights
- Our development can be adapted to other Web-based distributed systems in HEP and beyond

Effect of implementation



Thank you for your attention!

Backup slide

- Notification messages with total sessions and total active sessions count
 - Based on PandaDB
 - Sessions count > 50

Error from aipanda .cern.ch

noreply@mail.cern.ch [noreply@mail.cern.ch]

Sent: 15 May 2018 06:26

To:

2018-05-15 04:26:29,203 Internal Server Error: Attention!!! Total session count: 57 Total active session count: 44

More information here:

```
->https://es-atlas.cern.ch/kibana/app/kibana#/discover?_g=(refreshInterval:(display:Off,pause:!f,value:0),time:(from:'2018-05-15T04:24:00.000Z',mode:absolute,to:'2018-05-15T04:28:00.000Z'))&_a=(columns:(request),index:'atlas_bigpanda-*',interval:auto,query:(query_string:(analyze_wildcard:!t,lowercase_expanded_terms:!f,query:'type:%22django%22%20R%20response:%22500%22')),sort:( '@timestamp',desc))<-
```

Backup slide

- Notification messages with internal server error description. Based on **port80_acces** or **bigpanda_access_ssl** logs

Internal server error from aipanda .cern.ch

noreply@mail.cern.ch [noreply@mail.cern.ch]

Sent: 10 September 2018 02:50

To:

Host: aipanda108.cern.ch

Log type: bigpanda_access_ssl

Time: 10/Sep/2018:02:40:38 +0200

Message: 188.185.71.30 - - [10/Sep/2018:02:40:38 +0200] "GET /report/?campaign=MC16c HTTP/1.0" 500 538 "-" "Python-urllib/2.6"

Backup slide

- Notification messages with full description of Social-auth library error. E.g Session value state error

Error from aipanda .cern.ch

noreply@mail.cern.ch [noreply@mail.cern.ch]

Sent: 08 September 2018 12:12

To:

2018-09-08 10:12:50,028 Internal Server Error: https://login.cern.ch/adfs/ls/?wtrealm=https%3a%2f%2foauth.web.cern.ch/08M1tQZiWET8R3U2QCcpo0cPr6VbwcOmlmwrvWdMpLmUZf4QRF1a90RjPxIP9cuzIP64GYkb_WC1JTbWlPTsAvDYq784baZg6t6kxbPew4tWBuUCpKIuwjRRQ3CsojI3lPA&wa=wsignin1.0 EXCEPTION:Session value state missing. SELF OBJECT: ACCESS_TOKEN_METHOD:GET ACCESS_TOKEN_EXPIRES:'] GET_ALL_EXTRA_DATA:False ID_KEY:id REDIRECT_STATE:False REFRESH_TOKEN_METHOD:POST REFRESH_TOKEN_URL:None SSL_PROTOCOL:None STATE_PARAMETER:True SESSION INFO: Code in data: 6f08b388b3d5421eb2ef43558ee5ab6f345ec1baf6184057uqwf01x1az29nlggdwmv7q4ywwmx5db _SessionBase__session_key: uqwf01x1az29nlggdwmv7q4ywwmx5db _session in the session