alphadoop

Doflamingo

An light-weight monitoring system for Apache Hadoop

TITLE **Kafka/ Zookeeper Monitoring Module built for Flamingo Ecosystem**

DURATION **March 13, 2016 ~ June 8, 2016**

CLIENT **EXEM** **PRESENTER** **ALPHADOOP**

CONTENTS

_ WHAT WE WILL DO

Collect Performance Metrics,
Visualize it, and
Integrate it with Flamingo.

_ WHAT WE WILL DO

Is all system working properly?

hadoop

kafka

Doflamingo

Of Course!

Check this out!

_ TECHNICAL DETAILS

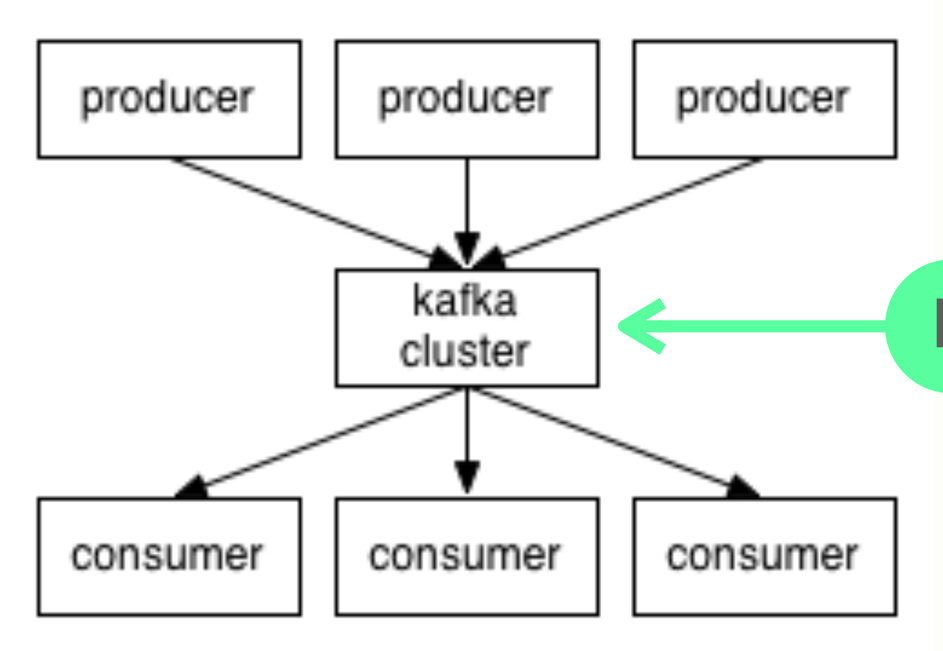
[A] WHAT IS KAFKA?

A distributed messaging system for log processing

- publish-subscribe model
- producer / broker / consumer
- built for scale-out & high availability

_ TECHNICAL DETAILS

[A] WHAT IS KAFKA?



producer
generate message that
fall into certain topics

broker

stacks up logs
based on topic

consumer
subscribe specific topics
& process the message

_ TECHNICAL DETAILS

[B] WHAT IS ZOOKEEPER?

A coordination service

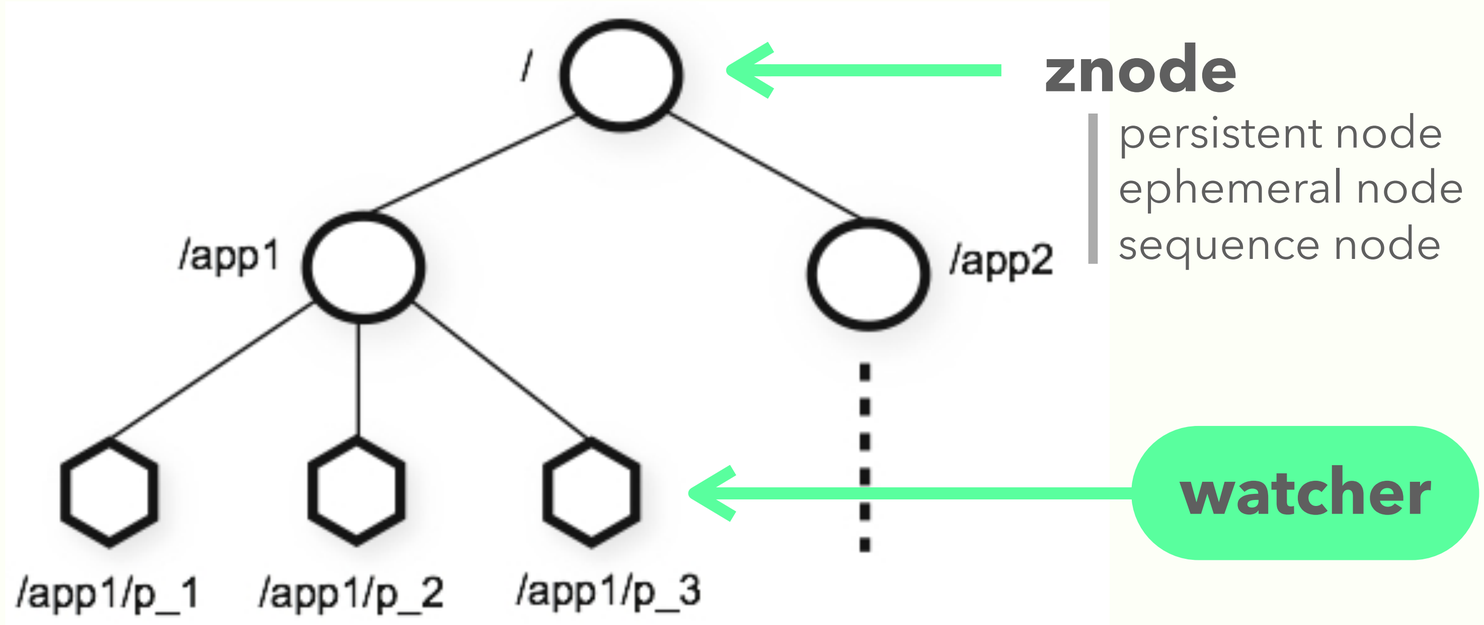
– Save important status/configuration info.

– Can be used for a global lock in the system

High Availability, Fast Date Access, Self Recovery

_ TECHNICAL DETAILS

[B] WHAT IS ZOOKEEPER?



http://bcho.tistory.com/1016

_ WHY WE NEED THIS PROJ

1. Hard to understand Hadoop

Distributed system – not intuitive

Unable to track fluctuant mass traffic

Eyes on only the upper level
– run and hope everything goes well

_ WHY WE NEED THIS PROJ

2. The Missing Link of Flamingo

Currently flamingo is able to monitor:

- Resources
- YARN application
- Map Reduce
- Nodes

_ REQUIREMENTS

1. **Built as a part of Flamingo system**
2. **Monitor and Report in Real-time**
3. **Utilize JVM ecosystem**
4. **Visualize the metrics, avoid numbers**
5. **Save metrics into Database**
6. **Special caution on log management**

_ HOW WE DO IT

Learn from other monitoring tools

Plenty of tools exists in the field – Learn from them and try to build up similar metrics

Build it into flamingo platform

There's flamingo's way of monitoring hadoop system. Add a new task into jobscheduler.

_ HOW WE DO IT

AGILE APPROACH

1 SPRINT = 2 WEEKS

TOTAL 5 SPRINTS along the semester

_ OBJECTIVES

O1: Set up an environment for Flamingo — SPRINT 1

O2: Define Kafka measurement metrics, visualization forms

O3: Implement API server which provides collected metrics — SPRINT 2

O4: Implement charts with Sencha

O5: Integrate with Flamingo Ecosystem — SPRINT 3

O6: Define Zookeeper measurement metric, visualization — SPRINT 4

O7: Implement a Zookeeper monitoring module on Flamingo — SPRINT 5

KAFKA MODULE → M1

ZOOKEEPER MODULE → M2

_ TECHNICAL CHALLENGES

Simulate distributed environment

Kafka and zookeeper can only be tested in multiple nodes. Need to mock clustering env.

REQUEST → EXEM

Can we have sample environment or at least a tutorial that we can follow to setup distributed system?

_ TECHNICAL CHALLENGES

Selecting the important metrics

New to monitoring job and hadoop so we don’t know what are the important metrics

HOW WE WILL SOLVE THE CHALLENGE

Survey other services: what they are monitoring and ordering of metrics which implicitly denotes importance

Interview on developers – maybe EXEM engineers?

_ THE EFFECT OF OUR WORK

The ultimate control tower

Flamingo now monitors not only nodes,
but also modules that compose pipeline.

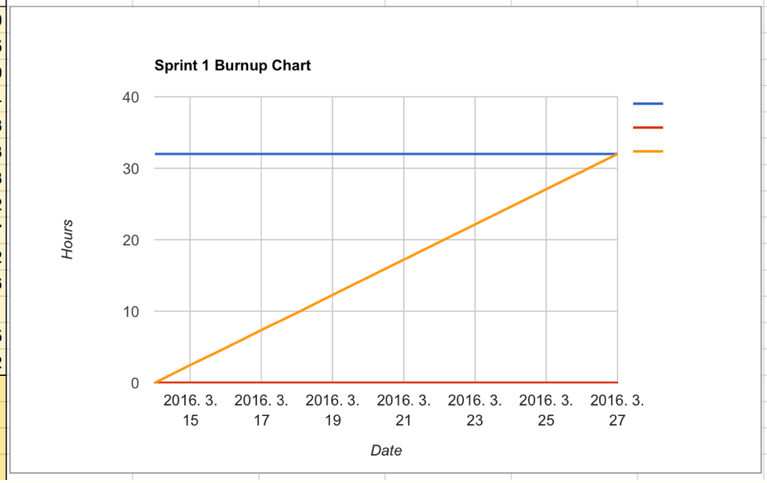
Opening up new possibility

The gathered metrics can be used for further
optimization or anomaly detection feature.

_ WHEN THINGS WILL BE DONE

We are on the cloud!

Ptoject	Class		Date	Goal	Completed	Expected
Sprint 1		Mon	2016. 3. 14	32	0	0
		Tue	2016. 3. 15	32	0	2.5
		Wed	2016. 3. 16	32	0	4.9
		Thu	2016. 3. 17	32	0	7.4
		Fri	2016. 3. 18	32	0	9.8
		Sat	2016. 3. 19	32	0	12.3
		Sun	2016. 3. 20	32	0	14.8
		Mon	2016. 3. 21	32	0	17.2
		Tue	2016. 3. 22	32	0	19.7
		Wed	2016. 3. 23	32	0	22.2
		Thu	2016. 3. 24	32	0	24.6
		Fri	2016. 3. 25	32	0	27.1
		Sat	2016. 3. 26	32	0	29.5
		Sun	2016. 3. 27	32	0	32
		Mon	2016. 3. 28			
		Tue	2016. 3. 29			
		Wed	2016. 3. 30			
		Thu	2016. 3. 31			



FOLLOW THE LINK →

_ WHO WILL DO WHAT

TEAM _ ALPHADOOP

SEUNGHYO KANG *the hadoop master*

← **Metric Analysis**

RESTful Server →

JARYONG LEE *the spring master*

YOUNGJAE CHANG *the sencha master*

← **Visualization**

_ WE ARE RESPONSIBLE FOR:

1. Built as a open source software

Fork and request merge into flamingo

License/ Copyrights are same with flamingo

2. Bye-bye after spring semester

A/S are not supported after June 21, 2016

_ WE ONLY HAVE THESE:

LIMITED TIME: 10 WEEKS

No delay accepted – when semester ends, project should be ended

LIMITED DEVELOPERS: 3 PEOPLE

No one will help us
– no money to hire someone!

END