Log Analysis using Open Source Scalable Systems

Gurvinder Singh
UNINETT

UNINETT

Motivation

- Distributed Systems
- In the moment of heat
- Centeralized interface to logs
- Easier access
- Detection of hidden pattern

http://www.themeparkreview.com/tatsumediaday/tatsumediaday57.jpg

Challenges

- Almost every component generates logs
- Different Formats and logging methods
- Different requirements for processing
- Dashboards
- Alerts

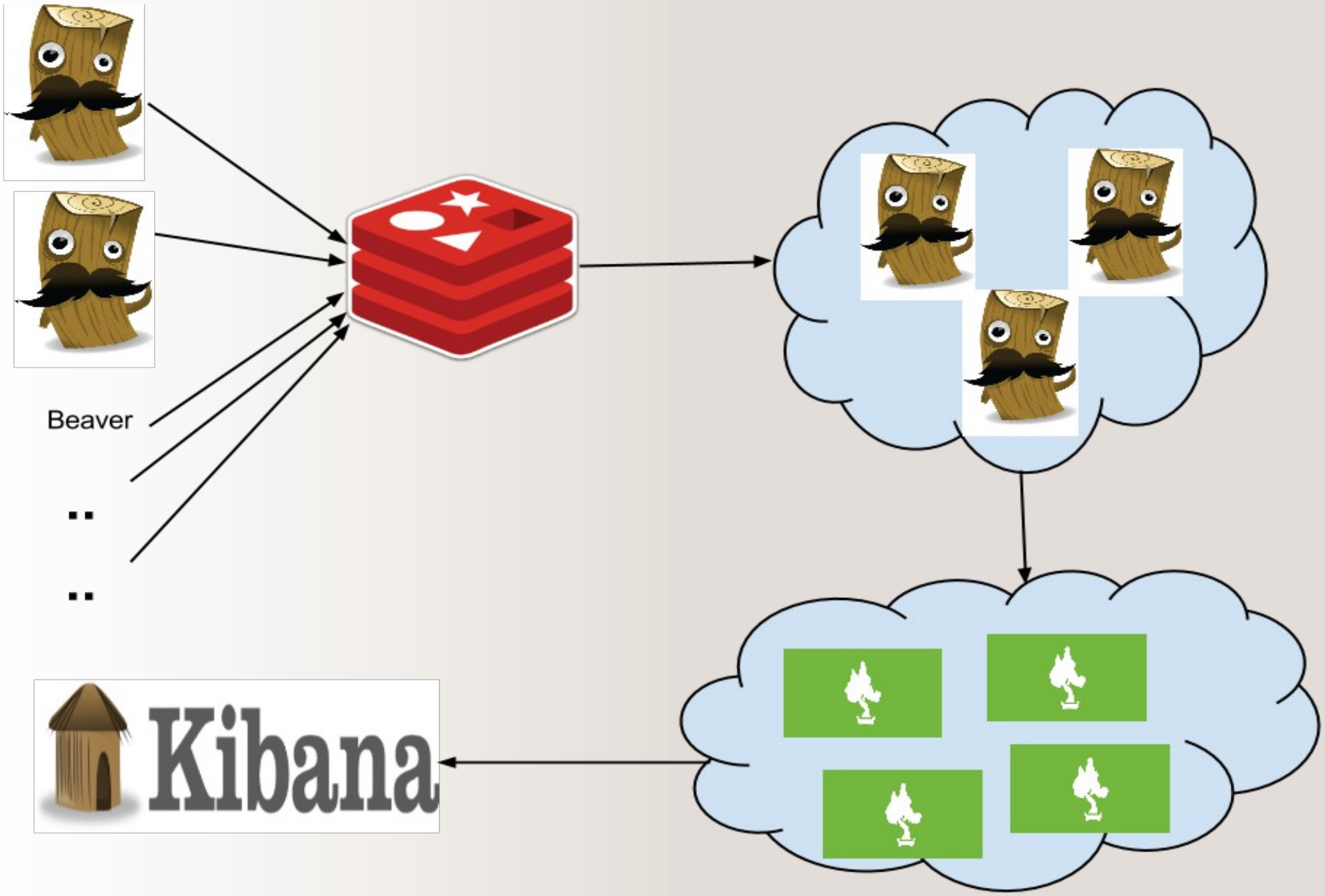
http://img72.imageshack.us/img72/3885/nephew2logs.jpg

Components

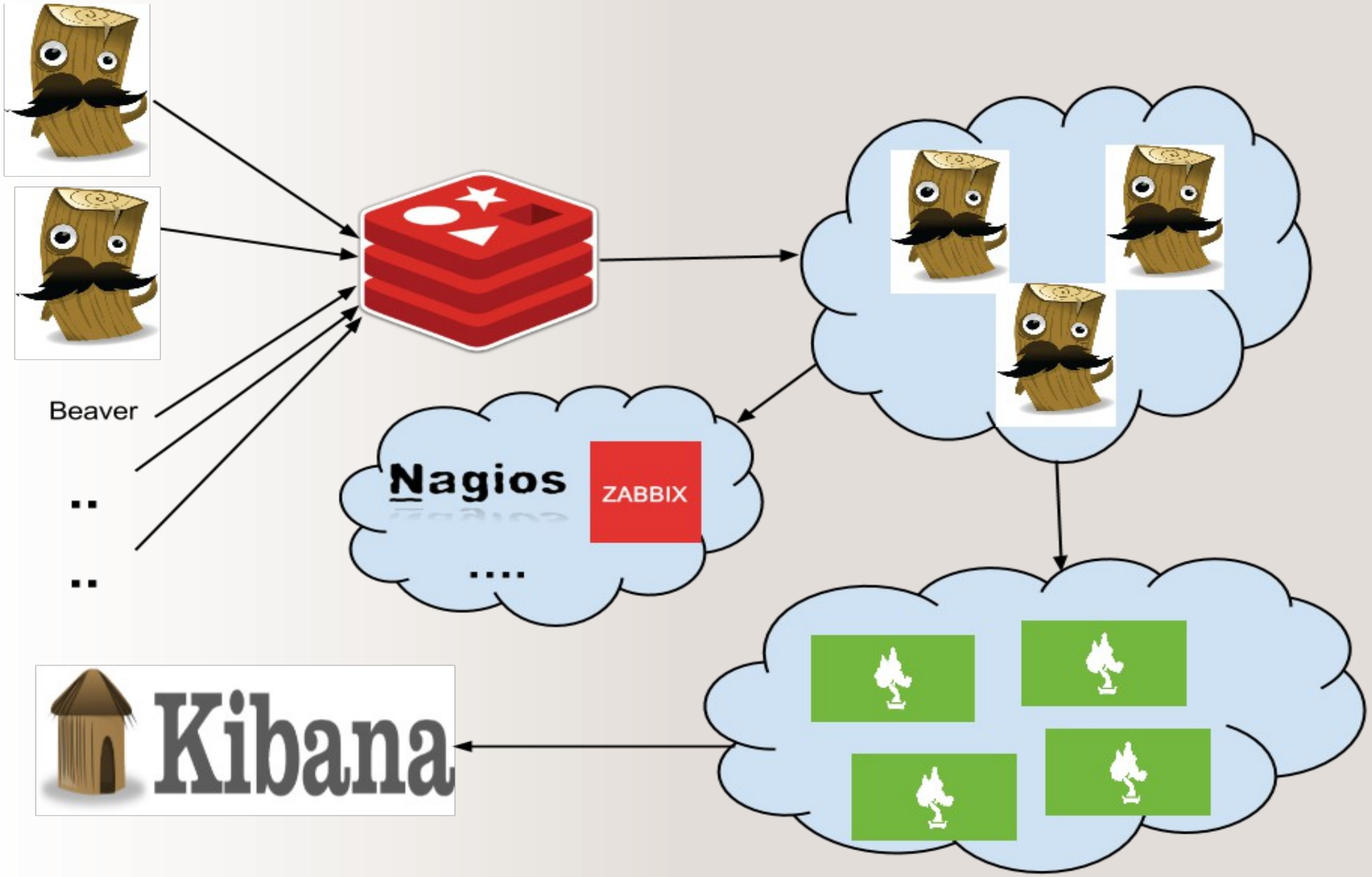
- Logstash
- Elasticsearch
- Kibana
- Redis
- Beaver
- Logstash-forwarder
- .....

Kibana

Most Common Architecture

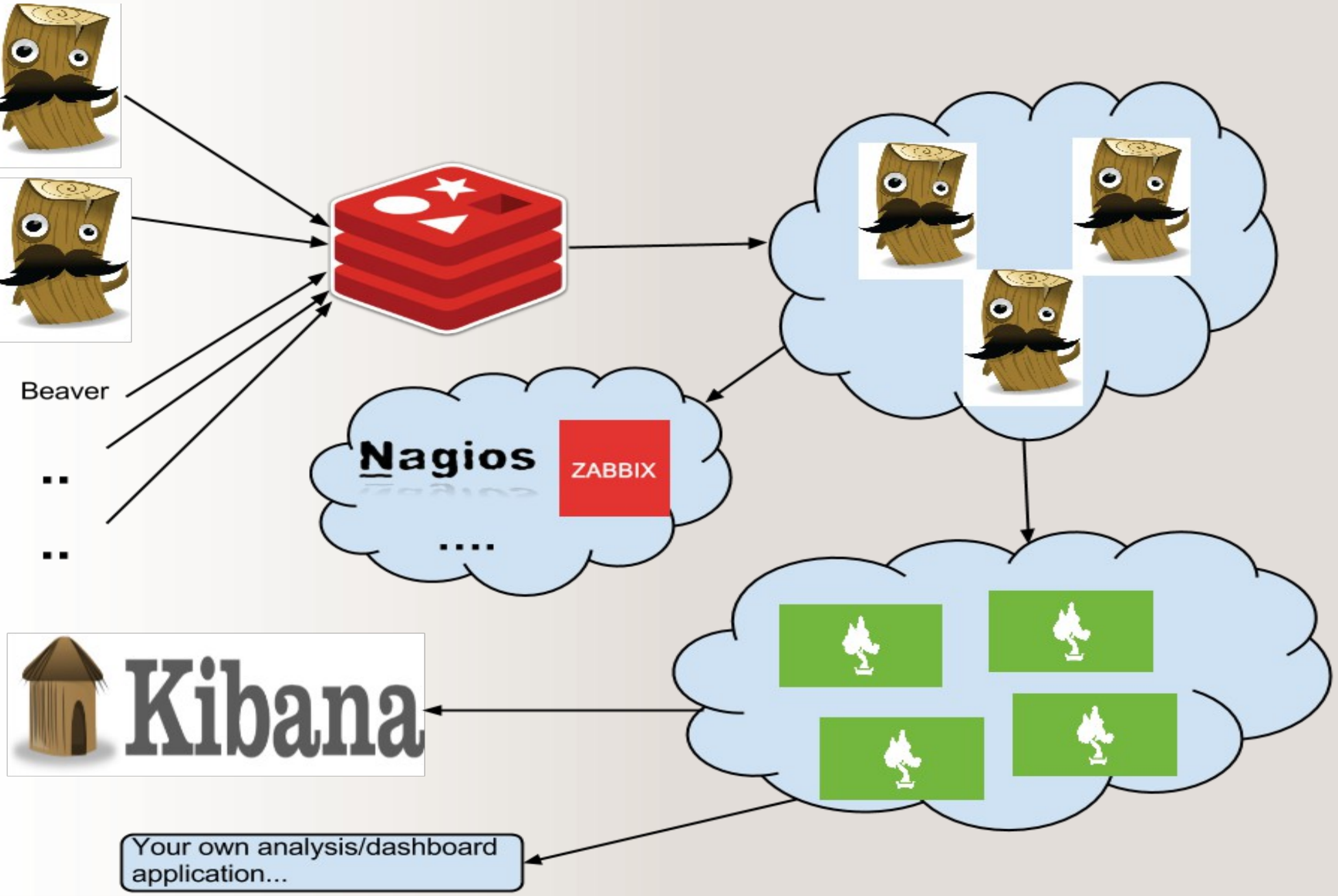


Architecture



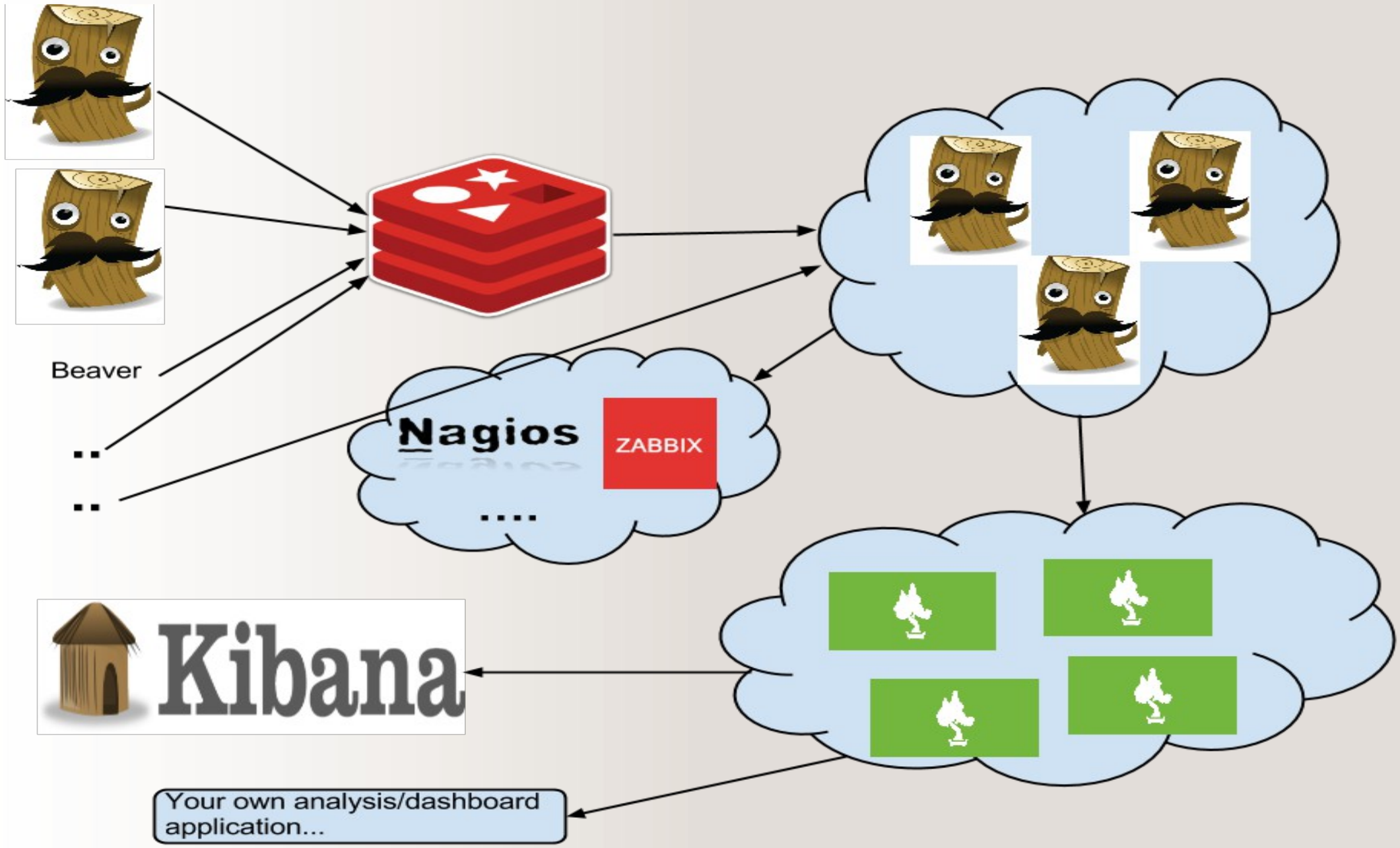
UNINETT

Architecture

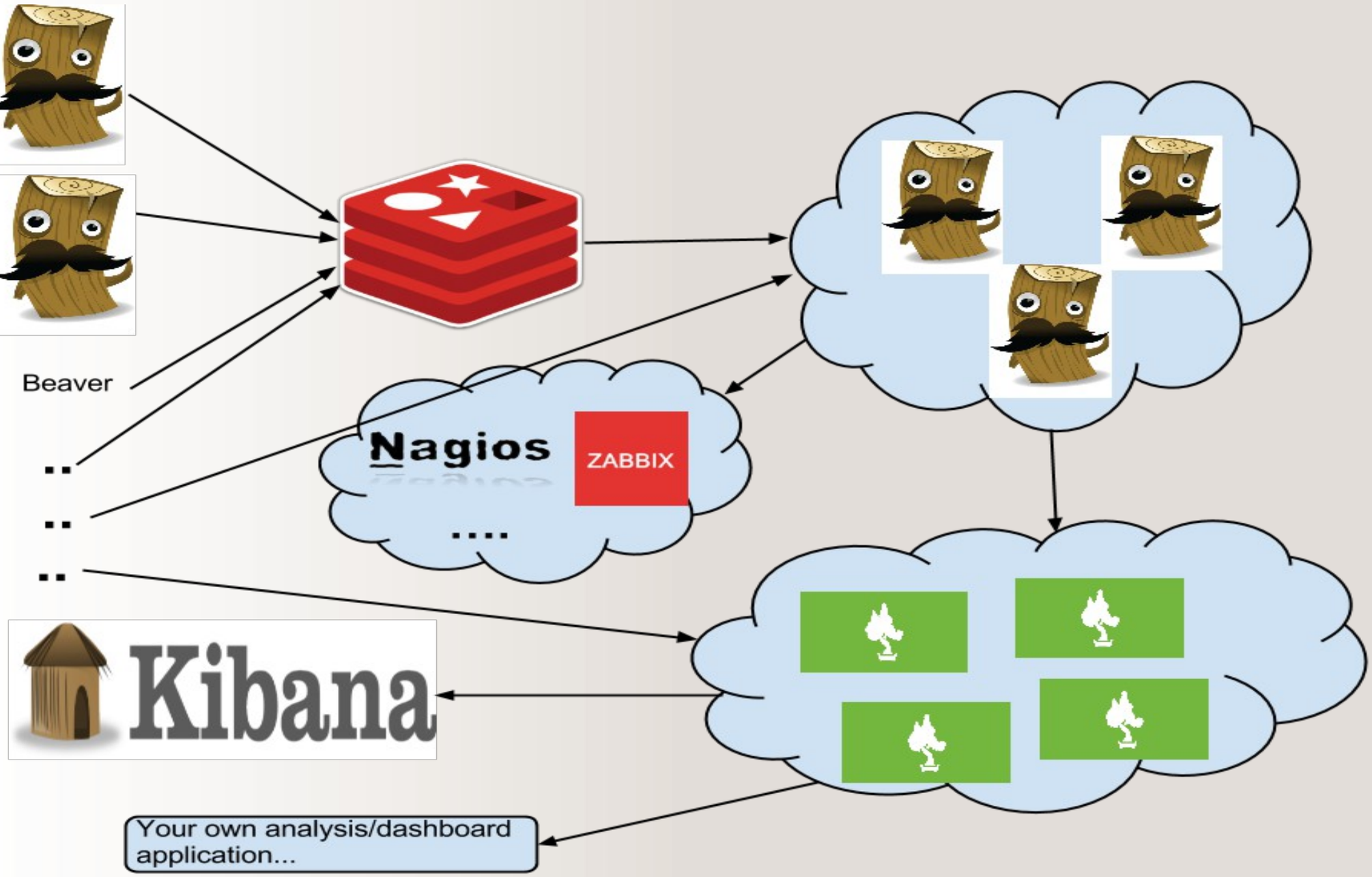


UNINETT

Architecture

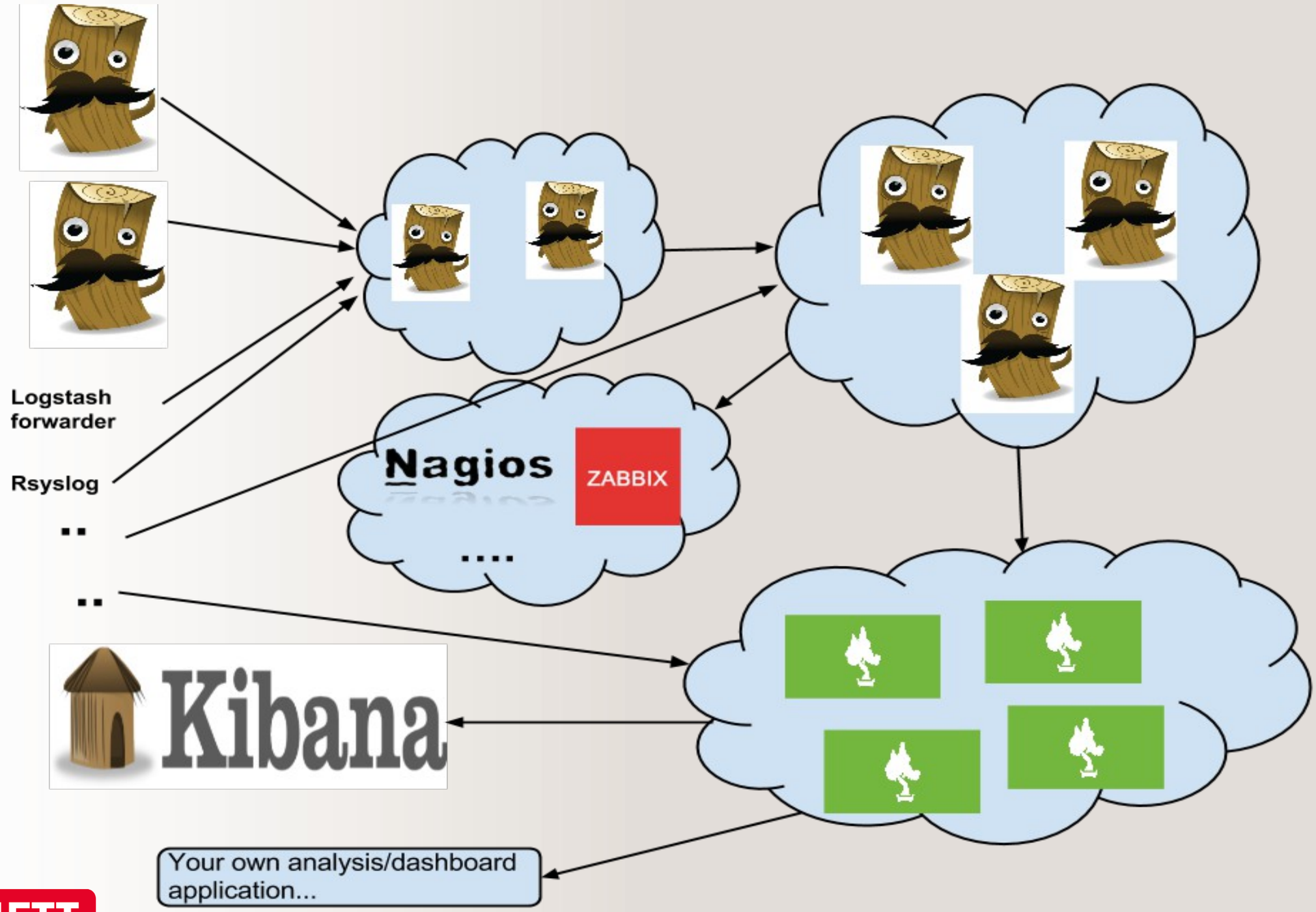


Architecture



UNINETT

Production Architecture



UNINETT

Demo

UNINETT

Possible Input/Processing/Output Options

Experience till yet

- Data Model to make consistent view of log fields across services
  - e.g. Search for user «xyz» in all different services
  - Give me all the logs from host «X» for all the services
- Avoid single point of failure if possible
- Automate setup to help scale easily
- Set minimum number of master nodes in Elasticsearch cluster to avoid split-brain issue
- No users concept in the Elasticsearch or Kibana
- Lack of access & authorization control
- Lack of Hierarichical storage support to exploit fast stroage medium to hot data and slow storage medium for cold data

Questions?