Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους

δίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμ για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους



Αλέξιος Πυργιώτης Εθνικό Μετσόβιο Πολιστέχνο October 15, 2013

- 1. Καλημέρα σας, ονομάζομαι Αλέξιος Πυργιώτης Θα σας παρουσιάσω τη διπλωματική μου με τίτλο:...
- 2. Ακούγεται κάπως περίεργο στα ελληνικά... αυτό που πραγματεύται είναι την δημιουργία ενός caching μηχανισμού για το Archipelago, ένα distibuted, storage layer
- 3. Συγκεκριμένα, στην παρουσίαση αυτή θα μιλήσουμε για τον cached, δηλαδή τον caching μηχανισμό μας, αλλά και για το synapsed, ένα συμπληρωματικό εργαλείο που στόχος του είναι να δώσει στον cached δικτυακές δυνατότητες

Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους



Αλέξιος Πυργιώτης

Εθνικό Μετσόβιο Πολυτεχνείο

October 15, 2013

Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους -Contents



- 1. Ο κορμός της παρουσίασης είναι ο εξής:
 - Αρχικά, παρουσιάζουμε κάποια εισαγωγικά που αφορούν το background της εργασίας μας. Αναφέρουμε τι είναι το Synnefo, τι είναι η υπηρεσια okeanos και τι είναι το Αρχιπέλαγο
 - Έπειτα, δείχνουμε τον τρόπο με τον οποίο η υποδομή μας χειριζεται αιτήματα δεδομένων από ένα VM.

Contents

Introduction

Request handling

Caching

Cachedi design



Table of Contents

Introduction

Table of Contents

Introduction

Request handlin

Caching

Cachedi desig



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους Introduction —Synnefo Ας ξεκινήσουμε με την παρούσα κατάσταση.

Storage Service
 Image Service
 Identity Service

Synnefo

sunnefo

Το software που τα ξεκίνησε όλα είναι το Synnefo

..by GRNET -> Και φυσικά τα παιδιά που βλεπετε εδώ

- Compute service, είναι η υπηρεσία η οποία προμηθεύει τους χρήστες με VMs και επιτρέπει το χειρισμό τους
- Network service, είναι η υπηρεσία η οποία δίνει τη δυνατότητα στους χρήστες να δημιουργήσουν ιδιωτικά δίκτυα και να συνδέσουν τα VMs τους σε αυτά.
- Storage service, που κοινώς αποθηκεύει τα αρχεία των χρηστών.
 Στην περίπτωση του Synnefo όμως, έχουμε ένα κοινό σημείο για τα πάντα: είτε είναι αρχεία, είτε δίσκοι των VMS, είτε images
- Image Service, υπεύθυνο για το deployment ενός VM από ένα image. Επίσης, κάνει και παραμετροποιήσεις (παράδειγμα ssh κλειδιά)

Introduction

Synnefo

synnefo

Open source, production-ready, cloud software. Designed since 2010 by GRNET.

Synnefo, as most cloud software, has the following services:

- Compute Service
- Network Service
- Storage Service
- Image Service
- Identity Service



4/24

-okeanos

okeanos

SKECINOS

- IsaS service

- Targeted at the Greek Academic and Research Community

Designed by GRINET

- In production since 2011

- laaS είναι πρακτικά η παροχή εικονικής υποδομής σε χρήστες (δηλαδή πάρε υπολογιστή (VM), δίκτυα, αρχεία κτλ)
- Δωρεάν για τους Ακαδημαϊκους σκοπούς, ήδη γίνονται εργαστήρια στο ΕΜΡ και απ' αυτό το εξάμηνο σε άλλες σχολές

Introduction

okeanos



- laaS service
- Targeted at the Greek Academic and Research Community
- Designed by GRNET
- In production since 2011



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους Introduction

-okeanos

okeanos

OKECINOS

LaaS sarvice

Targeted at the Greek Academic and Research Community
Obergined by GRINET

- laaS είναι πρακτικά η παροχή εικονικής υποδομής σε χρήστες (δηλαδή πάρε υπολογιστή (VM), δίκτυα, αρχεία κτλ)
- Δωρεάν για τους Ακαδημαϊκους σκοπούς, ήδη γίνονται εργαστήρια στο ΕΜΡ και απ' αυτό το εξάμηνο σε άλλες σχολές

Introduction

okeanos

ବkeanos

- laaS service
- Targeted at the Greek Academic and Research Community
- Designed by GRNET
- In production since 2011
- ...and of course powered by Synnefo.



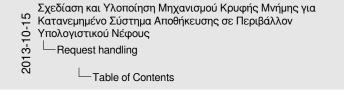


Table of Contents

Request handing

Request handling

Table of Contents

Introduction

Request handling

Caching

Cachedi desig



Τι είναι η διαχείριση των αιτημάτων ενός VM? Είναι η εφαρμογή πολιτικών και επεξεργασία των αιτημάτων σε όλη την πορεία τους μέχρι το να φτάσουν στο storage.

Δηλαδή έχουμε ένα εικονικό μηχάνημα <κλικ> ... το storage μας <κλικ> και πρέπει με κάποιο τρόπο τα δεδομένα του μηχανήματος να φτάσουν σε εμάς <κλικ> Ένας απλός τρόπος θα ήταν να τα συνδέσουμε. Άλλωστε όταν τρέχει VM, ο hypervisor κοιτάει block device. Θα μπορούσε να ήταν κομμάτι του storage Είναι αυτό αρκετό; <κλικ> Όχι, χρειαζόμαστε επίσης **FIXME**:

Request handling

What is request handling?



R
0-15
Τ.
ď
Ξ
201
.,

Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους

at is request handling?				

What is request handling?

-Request handling

Τι είναι η διαχείριση των αιτημάτων ενός VM? Είναι η εφαρμογή πολιτικών και επεξεργασία των αιτημάτων σε όλη την πορεία τους μέχρι το να φτάσουν στο storage.

Δηλαδή έχουμε ένα εικονικό μηχάνημα <κλικ> ... το storage μας <κλικ> και πρέπει με κάποιο τρόπο τα δεδομένα του μηχανήματος να φτάσουν σε εμάς <κλικ> Ένας απλός τρόπος θα ήταν να τα συνδέσουμε. Άλλωστε όταν τρέχει VM, ο hypervisor κοιτάει block device. Θα μπορούσε να ήταν κομμάτι του storage Είναι αυτό αρκετό; <κλικ> Όχι, χρειαζόμαστε επίσης **FIXME**:

Request handling





Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους — Request handling



What is request handling?

Τι είναι η διαχείριση των αιτημάτων ενός VM? Είναι η εφαρμογή πολιτικών και επεξεργασία των αιτημάτων σε όλη την πορεία τους μέχρι το να φτάσουν στο storage.

Δηλαδή έχουμε ένα εικονικό μηχάνημα <κλικ> ... το storage μας <κλικ> και πρέπει με κάποιο τρόπο τα δεδομένα του μηχανήματος να φτάσουν σε εμάς <κλικ> Ένας απλός τρόπος θα ήταν να τα συνδέσουμε. Άλλωστε όταν τρέχει VM, ο hypervisor κοιτάει block device. Θα μπορούσε να ήταν κομμάτι του storage Είναι αυτό αρκετό; <κλικ> Όχι, χρειαζόμαστε επίσης **FIXME**:

Request handling







Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους — Request handling

What is request handling?

What is request handling?

Τι είναι η διαχείριση των αιτημάτων ενός VM? Είναι η εφαρμογή πολιτικών και επεξεργασία των αιτημάτων σε όλη την πορεία τους μέχρι το να φτάσουν στο storage.

Δηλαδή έχουμε ένα εικονικό μηχάνημα <κλικ> ... το storage μας <κλικ> και πρέπει με κάποιο τρόπο τα δεδομένα του μηχανήματος να φτάσουν σε εμάς <κλικ> Ένας απλός τρόπος θα ήταν να τα συνδέσουμε. Άλλωστε όταν τρέχει VM, ο hypervisor κοιτάει block device. Θα μπορούσε να ήταν κομμάτι του storage Είναι αυτό αρκετό; <κλικ> Όχι, χρειαζόμαστε επίσης **FIXME**:

Request handling







Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους — Request handling

Policy enforcement?

Storage agnosticity?

Τι είναι η διαχείριση των αιτημάτων ενός VM? Είναι η εφαρμογή πολιτικών και επεξεργασία των αιτημάτων σε όλη την πορεία τους μέχρι το να φτάσουν στο storage.

Δηλαδή έχουμε ένα εικονικό μηχάνημα <κλικ> ... το storage μας <κλικ> και πρέπει με κάποιο τρόπο τα δεδομένα του μηχανήματος να φτάσουν σε εμάς <κλικ> Ένας απλός τρόπος θα ήταν να τα συνδέσουμε. Άλλωστε όταν τρέχει VM, ο hypervisor κοιτάει block device. Θα μπορούσε να ήταν κομμάτι του storage Είναι αυτό αρκετό; <κλικ> Όχι, χρειαζόμαστε επίσης **FIXME**:

Request handling





- Policy enforcement?
- Storage agnosticity?



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους — Request handling



Η λύση που χρησιμοποιήσαμε είναι το Archipelago

- Software-defined: αν και είναι ένα όρος μαρκετινγκ, εμείς κανονικά. Σημαίνει με το software OPIZEIΣ το storage (εφαρμογή policy, αλλαγή πορείας του request)
- τρέχει σε πολλούς κόμβους

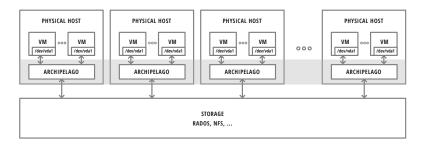
Our solution

- αποτελείται από διακριτά κομμάτια
- κάνει CoW (εξήγησε ότι τα images είναι λίγα, τα VMs πολλά, όπως όταν ένα process κάνει fork)
- μπορούμε χρησιμοποιήσουμε ότι θέλουμε

Request handling

Our solution

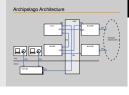
Archipelago



Key features: 1) Software-defined 2) Distributed) Modular Copy-On-Write Storage agnostic

Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους — Request handling

-Archipelago Architecture

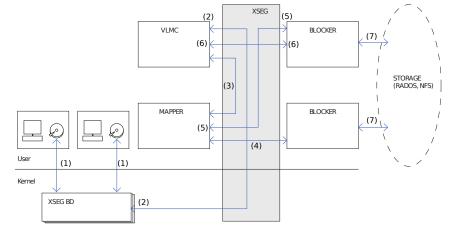


 Το VM στέλνει αίτημα στο δίσκο του, ο δίσκος είναι εικονικός, θα το δει ο hypervisor (εξήγησε τι είναι ο hypervisor) και θα το στείλει στον δίσκο που το έχουμε πει. (xsegbd)

• 2) **FIXME**:

Request handling

Archipelago Architecture





2	Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους
2	Request handling
3	└─RADOS

RADOS
The object store component of Ceph Bierystem.
Key features:
Papication
Faul tolerance
Self-management
Scalability

Request handling

RADOS

The object store component of Ceph filesystem.

Key features:

- Replication
- Fault tolerance
- Self-management
- Scalability



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους —Request handling

2013-10

RADOS
The object store component of Cuph Resystem.
Key feature:

* Replication

* Replication

* Replication

* Saul biolarance

* Scalabilly

Sport Sease:

Will will appea coache * SMMBs. < Ims

Will will appea coache * SMMBs. < Ims

Will will appea coache * SMMBs. Tims

Will will appea coache *

Request handling

RADOS

The object store component of Ceph filesystem.

Key features:

- Replication
- Fault tolerance
- Self-management
- Scalability

Speed issues:

VM with page-cache: > 90MB/s, < 1ms VM without page-cache: < 7MB/s, 10ms



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους
- Request handling
- RADOS

component of Caph Blasystem.

a nord

cher.> 90MBb; < first
cache: < 7MBb; , 10ms

RADOS

Key features:
 Replication

Fault tolerand

Request handling

RADOS

The object store component of Ceph filesystem.

Key features:

- Replication
- Fault tolerance
- Self-management
- Scalability

Speed issues:

VM with page-cache: > 90MB/s, < 1ms VM without page-cache: < 7MB/s, 10ms

Thesis goal: make this faster.

10 / 24



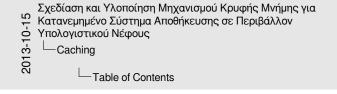


Table of Contents

Cathing

Caching

Table of Contents

Introduction

Request handling

Caching

Cachedi desig



No doubt we must do caching Caching is ... similar to the memory bigrarchy model

Caching

No doubt we must do caching Caching is \dots similar to the memory hierarchy model, \dots



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους — Caching — Block solutions Block solutions

• Boards

• The Control Contr

Caching

Block solutions

- Bcache
- ...

Pros: ... Cons: ...



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους

Caching

Object solutions

Object solutions

• Mericached
• Coorbase

Proc. —

Coors —

Caching

Object solutions

- Memcached
- Couchbase

Pros: ... Cons: ...



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους — Caching — Page-cache Page-cache

What if we used the page-cache?

Proc: ...

Caching

Page-cache

What if we used the page-cache?

Pros: ... Cons: ...



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους

— Caching

— Conclusion

Conclusion

Carnot use those reliably with with Archipelago, may use them later.

Have to create our own solution

Caching

Conclusion

- Cannot use those reliably with with Archipelago, may use them later.
- Have to create our own solution



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους Cachedi design

Table of Contents

Cached design

Cachedi design

Table of Contents

Introduction

Request handling

Caching

Cachedi design



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους

— Cachedi design

— Our decision

Our decision

reate something close understands Archipelago logic neasure the best possible performance we can get

Cachedi design

Our decision

- create something close understands Archipelago logic
- measure the best possible performance we can get



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους

— Cachedi design

— Requirements

Requirements

For air information we pose the following requirements:

- Nativity
- Plaggability
- In marrory
- Law indexing overhead

Cachedi design

Requirements

For our implementation we pose the following requirements:

- Nativity
- Pluggability
- In-memory
- Low indexing overhead



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους

— Cachedi design

— Cached design

Cached design

operations image compount image

Cachedi design

Cached design

operations image component image



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους

— Cachedi design

— Xcache design

Xcache design

show with red the xcache design go to xcache picture Show what xcache does

Cachedi design

Xcache design

show with red the xcache design go to xcache picture Show what xcache does



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους

— Cachedi design

— Xworkq design

Xworkq design
show the workq design
what workq does

Cachedi design

Xworkq design

show the xworkq design what xworkq does



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους

— Cachedi design

— Xwaitq design

Xwaitq design
show washit design,
what sealiq does

Cachedi design

Xwaitq design

show xwaitq design what xwaitq does



Σχεδίαση και Υλοποίηση Μηχανισμού Κρυφής Μνήμης για Κατανεμημένο Σύστημα Αποθήκευσης σε Περιβάλλον Υπολογιστικού Νέφους — Cachedi design — Bucket pool

Bucket pool

Buckets have been designed to do this first

Cachedi design

Bucket pool

Buckets have been designed to do this list