CONTINENTAL SCALE CLOUD STORAGE

A lightning fast talk on our experiences
David Jericho, Solutions Architect, AARNet
2017-10-16T13:30:00-08:00
INTRODUCTION
This isn’t a deep dive

I’m going to meander a little

Ask questions along the way

We want to collaborate
BRIEFLY ABOUT ME
Solutions Architect
20 years experience
13 years at AARNet
...and still not bored

Mad motorcyclist
Ice hockey player
Was to be a pilot
Started coding at 5

Parent of 3 children
Likes his dark office
Laziness is a virtue
Still owns SGI kit

DAVID JERICHO
WHO ARE AARNET
WHO ARE AARNET

• Australia’s NREN
  • Founder of Internet in Australia
  • First connection in 1989
  • Own and run our own fibre network

• Owned by 38 universities and CSIRO
  • Not for profit company
  • Acting on the requests of our shareholders

• Built to support research and education
  • High capacity nationally and internationally
  • Connect schools, libraries, galleries, museums, archives, and research institutions
WE’RE GEOGRAPHICALLY LARGE

- **Australia is a huge continent**
  - 45ms from Sydney to Perth
  - 26ms from Brisbane to Melbourne
  - 90ms from Darwin to Perth
  - Bandwidth is never the problem, latency is

- **Fun facts**
  - 25 million people
  - 116 years old as a federation
  - Queen Elizabeth II is our head of state
SPREAD VERY THIN

- 5 major population centres
  - 92% within these 5 areas
  - 95% live within 100 miles of the coastline
  - Nearly all those inland lights are fires

- Large amount of collaboration
  - Often from extreme ends of the network
  - A lot of primary industry

- Interesting data use cases
  - Internet of things
  - Heavy use of video conferencing
  - Square Kilometre Array (SKA-low)

NASA Earth Observatory image by Robert Simmon, 2012
TEXAS COMPARED TO AUSTRALIA

• We do everything with distance in mind
  • Similar population to Texas
  • Similar GDP/GSP to Texas
  • Can fit Texas entirely in Australia’s two largest states

• We always consider latency
  • Everything is tested with latency in mind
  • We haven’t violated causality yet
WHAT IS AARNET’S CLOUDSTOR
CLOUDSTOR IS A PLATFORM FOR STORAGE
(AND HEAVY LIFTING FOR USERS)
CLOUDSTOR’S OWN CLOUD
CLOUDSTOR SITES

• 3 major sites
  • Brisbane, Melbourne, Perth
  • Brisbane to Melbourne is 27ms
  • Melbourne to Perth is 35ms

• At least two geographically disperse replicas
  • Continental scale metadata database
  • Hierarchical namespace consistent in all sites

• BGP Anycast IP ensures we handle the TCP heavy lifting
  • Helps with the campus network problem
  • Wifi doesn’t help
  • We can manage the long haul far better than users usually can
USERS TRAVEL EXTENSIVELY

• 24 hours of CloudStor access
  • Large number of foreign students
  • Nearly all those inland lights are fires

• Large amount of collaboration
  • Mangrove research between Israel, the Philippines, Rockhampton, and Geraldton
  • Bison genomics research between Siberia and Adelaide
  • Even users in Antarctica
GROWTH ALMOST ENTIRELY ORGANIC

- Mostly word of mouth take up until recently
  - Many champions within institutions
  - Institutions own us
  - Data is held within Australia by an Australian company

- Institutions accepting of service
  - Users will use what they can get away with easily
  - If they don’t use us, shadow IT occurs
  - Institutions own us

- Much more focused marketing and deployment in 2017
  - Realisations from our advisory committee and the CIOs
  - We’ll probably see far larger growth again
BEER THE ORIGIN OF CLOUDSTOR
• Playing with ideas on a Friday afternoon
  • ...with a beer in hand
  • ...four days before I was getting married

• We’d been looking to solve the problem
  • ...and I’d been saying no repeatedly
  • ...as there were no tools that really suited
  • ...and even the commercial products didn’t live up to the promise
  • ...so we tried
    • Okay, maybe that was hubris

• This wasn’t a managed scoped project
  • It actually worked!
  • And we told people about it.
REQUIREMENTS TO CONTINUE

- Be too easy to use
  - Nobody wants to interact with their data transfer
  - Well, nobody who’s not a nerd
  - All major operating systems must be supported

- Geographic redundancy of data
  - Faults occur, as do natural disasters
  - Campus networks and user device choices effect experience
  - It has to be fast (enough)

- Users aren’t librarians or archivists
  - User data can be highly ephemeral
  - User data is very small
    - 96% of our held data is < 10MB
    - 92% of our held data is < 1MB
  - Institutional policies could kill this even before it got airborne
COMPRESSED HISTORY
WE THOUGHT THIS WOULD WORK JUST FINE

- ownCloud 4.5 and FileSender
  - On top of NFS
    - On top of HDFS
      - On top of Hadoop
  - Sitting next to MariaDB with Galera
  - With Apache prefork using mod_php
  - ...
  - ...
  - I wouldn’t call this design

Source: https://giphy.com/gifs/agconti-friday-raft-l378hZUdgE1qArI7S
REALLY NOT UNDERSTANDING THE IO

- Dedicated hardware ordered Dec 2012
  - 3 identical high spec high disk density machines
  - Benchmarking in Jan 2013
    - We thought we had this covered
    - Hindsight is perfect, we had no idea of what user data really looked like
  - 96% < 10MB, 92% < 1MB
  - We knew we had to grow the hardware, but this was a production proof of concept

- Shingled media disks aren’t what we knew
  - High density disks do not work the same way for writes as common server disks in 2012
EVERYTHING LOOKS FINE

• Life is great, making rapid strides
• Move one machine to Perth in Mar 2013
  • 65ms away
  • Hail Marys are always the best first play

• 300 users in Mar 2013
  • 9 days later, Hadoop stopped behaving
    • Very long story short, Hadoop’s internal buffers were exceeded by the amount of incoming data, and the cluster would collapse when replicating

Source: Gunshow by K.C. Green http://gunshowcomic.com/648
GIVING UP ON HADOOP

• Used a commercial Hadoop offering
  • June 2013, we gave up with the product
  • Engineers at vendor had never considered the use case

• I’ll talk about storage more later
  • Single master with scheduled replication
  • Load was low enough one machine was okay
  • Remote machines doing heavy lifting for user
  • We start looking for alternatives

Source: https://giphy.com/gifs/hcQE1lxTbdvsk
EVERYTHING IS ON HOLD

- Third child born in Jul 2013
  - I do this for my family 😊
- Nobody else was focused on this service
  - Few hundred users
  - Big BETA logo
- No real progress for 2 months
  - And then 6 months of part time work

Source: David Jericho
OWNCLOUD 5.0

- Upgraded to ownCloud 5.0 Dec 2013
  - Even with weeks of preparation work, the upgrade took 12 hours
  - ownCloud made assumptions about databases
  - API changes in ownCloud

- Becoming very evident this a lot of work
  - Email support load
  - Users starting to depend on a beta service
  - Skills needed very wide and varied

Source: http://www.warnerbros.com/madam-satan
WE BROKE A NETWORK

• The nature of traffic was highly irregular
  • TCAM issues on the 7604 platforms
  • Caused excessive jitter on syncing traffic
  • Very slow average throughput

• Using YeAH-TCP unbroke the service
  • Potentially broke other services
  • Fast mode YeAH-TCP was quite aggressive

• Obvious solution is move it
  • Moved to AARNet 4, Oct 2014
  • Hidden note – Upgraded to ownCloud 7
    • Surprisingly painless
    • Thinking we’re onto something here

Source: cisco.com
NEW ZEALAND JOINS IN

• Federation with Tuakiri Feb 15
  • Little risk, as NZ’s population < Sydney
  • Realisation we must manage our own metadata
  • Accept institutional rules will sovereignty issues
  • Highlights lack of user validity checking
    • Trust someone we have no direct relationship with
    • No way to validate a user is still a member

• Wait up, we hadn’t dealt with T&Cs
  • This wasn’t a scoped project
    • We go ahead anyway
    • ...starts a 1 ½ year exercise in sorting out legal issues

Source: Kiwi!, Dony Permedi 2006
LAUNCHING GROUP SERVICES

- Easy to do using role accounts Mar 15
  - Worked in the test
  - We thought we were great
  - Oops...
  - Lack of testing at scale hid major issues from us
  - Revealed many sync client issues

- Scaling issues again
  - Exponential increase in SQL queries
  - Introduce MaxScale to help spread SQL queries around

- Meanwhile regular support load increasing
  - 5000 users by this point

Source: http://i.imgur.com/wGTM8lw.gifv
STILL EXPLORING STORAGE

- Engage devs to work on Swift May 15
  - OpenStack Swift is coming out ahead in all tests
  - Engage a number of developers to help
  - Large object support an issue
    - Telling a user they can’t store > 5GB?
  - Totally dependent upon the OC metadata
  - Lack of a POSIX like FS is an issue
    - Object stores are another presentation

Source: Ren and Stimpy, The Boy Who Cried Rat 1991
Hiring a dedicated developer Jun 15
- We’ve realised this is devops
- The traditional network mindset doesn’t work:
  - Scope
  - Design
  - Build
  - Deploy
  - Operate
- Oops...

Used the skilling up period to build new network
- Moved switching to its own Layer 3 VPN
- Changed rules on how we can operate the switching

Start testing a global EOS storage system with CERN
- ASGC from Taiwan join in
- L. Mascetti, D. Jericho, A. Hsu, “Global EOS: exploring the 300-ms-latency region” CHEP’16

Source: A break from a sequence of animated gifs
AARNET’S FIRST PETABYTE

• Network company – not HPC Jan 16
  • We’ve decided on CERN’s EOS at this point
    • CERN’s experiences with Geneva <-> Wigner at 22ms
    • Our experiences with the Global EOS test
    • Our experiences replacing a troublesome mirror
    • POSIX like file system
    • It worked at 65ms!

• Actually 2 petabytes
  • CERN have 200 petabytes on disk and a team of developers
  • They were very keen for us to contribute

Source: David Jericho
BRANDED CLIENT AND MIGRATING USERS

- **Branded client Apr 16**
  - CloudStor becomes a consistent brand to the user
    - We can stop asking the user which version they’re running
  - Client for free on the App Store and Google Play

- **Moving users to the EOS environment May 16**
  - 20,000 users
    - 15,000 new users in 12 months
    - 40 million objects
    - Always more work than first thought
  - EOS rollout was itself trouble free

Source: https://cloudstor.aarnet.edu.au/
THE SECOND DEDICATED DEVELOPER

- Another developer hired Jun 16
  - Found a developer at a university
    - He’d managed to break CloudStor repeatedly
    - Figured he could spend his time unbreaking
  - Our rate of work is increasing
    - The nature of outages is changing
      - From first order to second and third order
- The Michaels won’t let me take a photo
  - They remind me of The Bobs
  - I’ll let them argue which is which

Source: Office Space (1999)
OWNCLOUD 8.2

• We skipped 8.1, going to 8.2 Aug 16
  • Why did I do this to myself again?
    • More regressions
    • Far more subtle each time
  • 12 hours to perform
    • Once again, database migrations

• Started using containers for the app stack
  • A learning curve
    • Rollback is now a matter of seconds
    • We now consider ourselves crazy for not doing it sooner

Source: Star Trek: The Next Generation
OWNCLOUD 9.0

• We estimated 2 weeks of work Nov 16
  • It’s still not done in Oct 17
    • More regressions
    • Far more subtle each time
  • 12 hours to perform
    • Once again, database migrations

• We’ll be skipping straight to ownCloud 10
  • Self managing security issues
    • Massive regressions
    • Technical debt is fun

Source: Star Trek: The Next Generation
ENGAGING A SECURITY AUDITOR

• Engaged a security auditor Jan 17
  • Increasing requests from users for audit
    • ISO2700x?
    • Legal departments requesting

• We completed with only minor issues
  • The majority were fixed in hours
    • SHA1 signed certificates and so on
  • We now have a document we can present
  • Has changed the mindset of the team
    • Security issues now forefront and are fixed instantly

Source: Office Space (1999)
IT’S CONTAINERS ALL THE WAY DOWN

• **CloudStor is 100% containerised Mar 17**
  • Storage, databases, orchestration, app stack
  • Orchestrated by Rancher
    • Power to the developers
  • 4 ½ years after starting this I no longer understand it all
    • Improving development processes mean it’s not opaque

• **CI and CD works**
  • QED

Source: AARNet’s Slack
THE NEED FOR SPEED

• Deploy S3 gateways Apr 17
  • De facto object store interface
    • Set and forget uploading of terabytes

• Soft launch the Rocket uploader
  • Custom written tool talking xroot directly
    • Batches and unbundles small files into larger chunks
  • The common user is now getting a gigabit from their desktop or laptop
  • Data fully visible via the S3 or ownCloud interfaces

Source: Veronica Jericho
TAPE SILOS, WHUT?

- Deploying backup onto tape silos Jun 17
  - To protect us from ourselves
    - Value add being able to backup other services
    - Realised as part of this just how many pies CloudStor has put fingers into
  - Using EOS’ modify time propagation to keep job sizes manageable
    - Sub-second discovery of all changes for a user since a timestamp
    - 2 seconds to find all changes in the last 24 hours across 200 million objects
  - Data being highly ephemeral and in use
    - Many small deltas hidden deep in directory structures

- Future plans are to move to CERN’s EOS CTA
  - Integrated HSM like behaviour with archival and backup policy capabilities
  - Be able to properly support archives, as opposed to backup

- Brought a dedicated storage developer into the team Jun 16

I AM COMPLETELY OPERATIONAL AND ALL MY CIRCUITS ARE FUNCTIONING PERFECTLY.
THE NEED FOR SPEED ELSEWHERE

- Heavy instrumentation has been ongoing
  - 99.99% of our queries are serviced under a second
    - It’s those outliers that cause issue
    - Users panic (understandably)
      - Research reveals that other services fail continuously
      - Their failures are hidden by the client
      - User experience engineering!

- Attacking the SQL layer Oct 17
  - ProxySQL statistics
    - Found SQL improvements with 3 orders of magnitude

Source: David Jericho
THE STORAGE STORY
HADOOP FAILED US EARLY

- XtreemFS
  - Was great with large files, didn’t work with small files
- Ceph
  - Paxos doesn’t work with latency beyond 4ms
    - At 12ms, 4 GByte/s turned to 40 MByte/s
- GlusterFS
  - Let’s just say I’m not a fan
- pNFS
  - We ended up using NFS continentally for over 2 years
- OrangeFS
  - Designed with HPC use cases in mind, not cloud storage
- CERN’s EOS
  - AND WE HAVE A WINNER!

- The issue was never bandwidth
  - Latency always hurt us with everything we tried
  - Data being highly ephemeral and in use
  - 2012 looks very different from 2017 in this space
Running Beryl-Aquamarine v0.3.x
- Have travelled the journey from v0.3.100 to v0.3.270
- Australia-Europe time zone differences worked to our advantage
  - Bug found -> report -> fix -> test -> deploy

Upgrade to Citrine v4.x
- The focus is stability and availability

Source: L. Mascetti (CERN/IT-ST), D. Jericho (AARNet), A. Hsu (ASGC)
WHERE WE’RE GOING
CloudStor Rocket at 6.9Gbps to production
THINGS WE’VE LEARNED AS A NETWORK DOING SOFTWARE
IT’S ALL ABOUT THE ORCHESTRATION

- The focus is a product
  - Not projects

- Write down what we’re doing
  - Reducing induced error
  - Reducing monolithic blocks

- Metrics and monitoring are worth more than the time it takes
  - Plan to spend as much time on metrics and monitoring as on code or systems work
  - It returns when things actually go bad

Source: CCO licensed
Tasks always take 5x longer than thought (often longer)

Products don’t have an end point; just an iteration

You don’t have to go it alone; work finding partners pays off

You’ll always have some unhappy users; roll with the punches