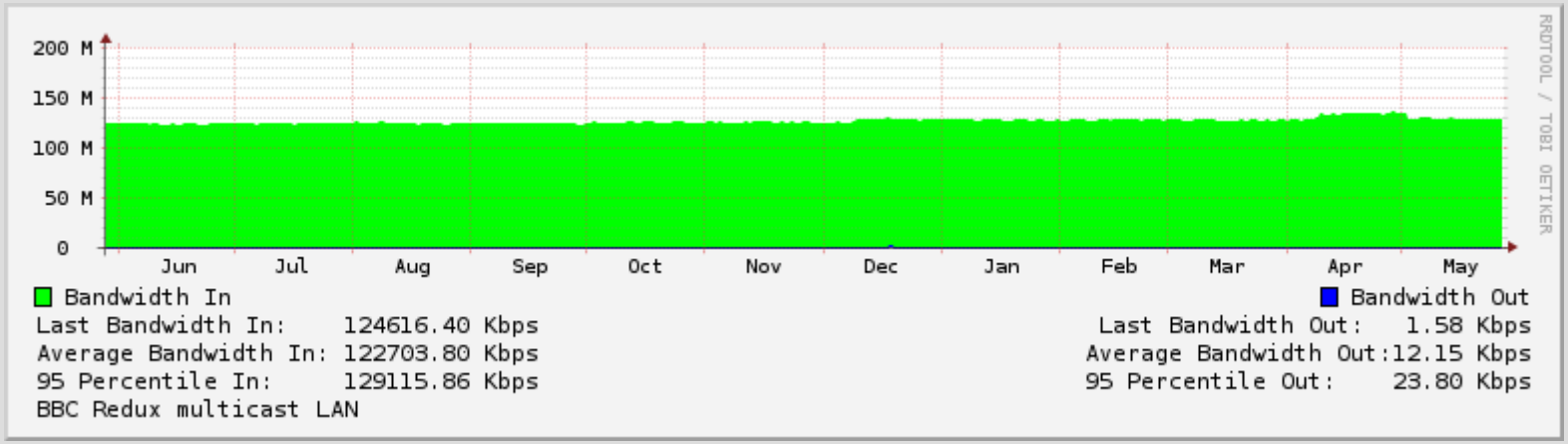
BBC Redux

Tom Bird, Portfast Ltd / BBC R&D
tom@portfast.co.uk / tomb@rd.bbc.co.uk

The problem:

- Store all output for a major broadcaster in the UK



The stats:

- TV and Radio programmes
- 323,344 of them
- 2 years of BBC output
- 12,771 days of content
- >100Mbit/sec of new data, constantly

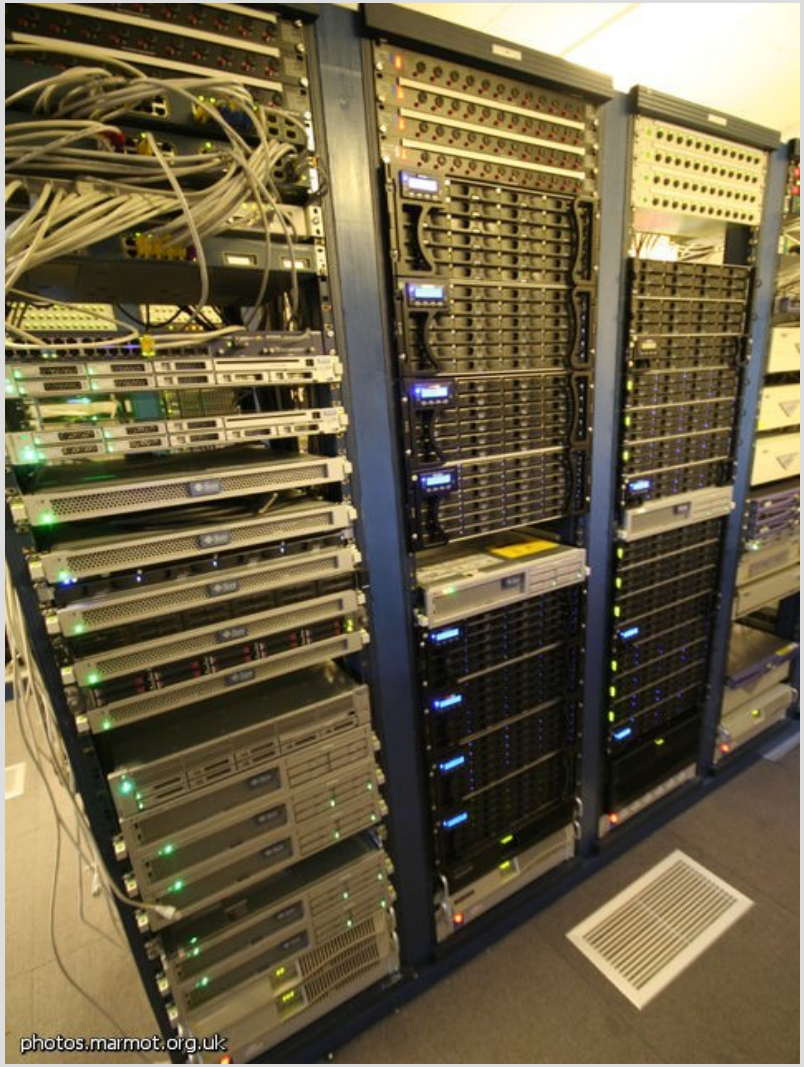
The hardware:

- Sun T1000 for acquisition
- Sun T2000 for storage
- X86-64 for transcoding
- X86-64 for database

Content acquisition

- All BBC terrestrial content is being multicasted from R&D (AS31459), 233.122.227.0/24
- Bit streams taken directly from the air through a standard aerial, DVB card and Linux, and turned into UDP
- No transcode, it's already MPEG-2
- Metadata is already available in the streams
- ... so acquisition can be automatic!

A nice picture of some racks



Hardware

- Sun T2000 servers
- Commodity SATA drives (500G – 2T)
- RAID boxes present a RAID system to the host as a single SCSI LUN
- SAS / Ultra 320 to the host
- 10 gigabit ethernet to storage nodes

File system

- Has to work on Solaris
- Tried UFS, too slow, died under load
- UFS is a traditional inode based unix file system, uses fsck and so on
- This isn't a bad idea but there are other ways....

ZFS

- No FSCK
- Instant snap shots
- Object based
- Copy on write (this bit is important)
- Is present in Solaris 10
- Use OpenSolaris

ZFS + crashing disk box =

Sun fixed it

- “Uberblock” and superblock got corrupted
- Uberblock is for finding the superblock
- Due to “copy on write”, previous copies of the superblock are still on disk
- Sun rolled it back to a previous copy, by using `dd` to edit bytes on disk (!!)

It broke again.

Sun fixed it again

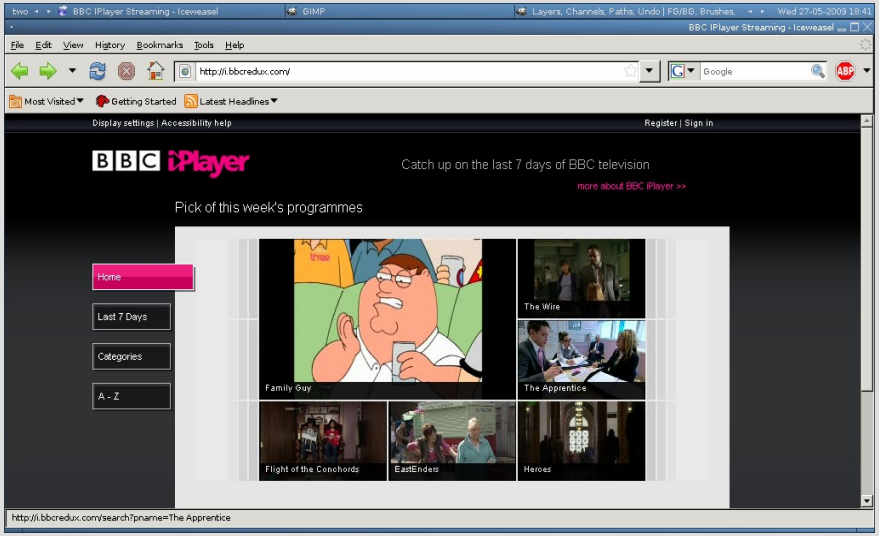
- This time they had better utilities
- Fixed with less pain
- Still not something the end user can do
- However, Stuff was online again!

Content coordination

- Master server has a database of all content and where it is
- Client makes a request to the main DB and gets redirected to the real location
- Same story if the client needs to upload
- All mod_perl except for stuff that needs to be really quick, which is done in C

But why does it exist?

- Initially as a proof of concept for a flash iplayer
- Due to modular transcoding architecture, it is now used to feed content to the real iplayer.



What else?

- Broadcast compliance checking
- Subtitle extraction
- Test bed for anything else we can think of

That's all, folks

Questions?