

The Story of Phil the file at Club CrossRef

Or

CrossRef system hardware
architecture and queue management.

About Phil

- Dressed in XML
- Matching tags to get by Bouncers.
- UTF-8 encoded accessories (Nice Umlauts)
- A fit 1.80 Megabytes. (Works out to keep under 2.00 Megabytes.)

Phil's Night out

- Likes to mingle with the Processors at Club CrossRef.
- Once he gets there, he has to wait in the Queue!
- Sometimes, he gets bounced.
- Some Processors won't give him the time of day.

Club CrossRef

- Address: <http://doi.crossref.org>
- ID's are required to get in.
- Operated in a RedHat Linux operated Dell PowerEdge Building.
- Club Run by CrossRef with Caucho Resin and Oracle backers.
- Always open. Failover locations for emergencies



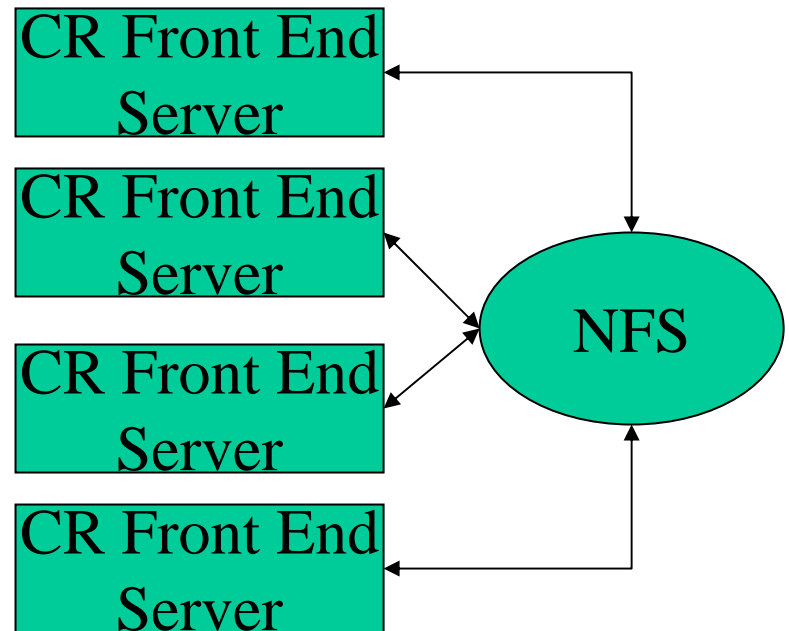
The Building

- Built buy Dell Corporation.
- Poweredge
2650/2850/1750/1850
- All structures are 3 years old or newer.
- Each has 2 CPUs and power/ethernet failover
- Each can take the primary role if required.
- Each interlinked to replicate System software and access Queue



Building Architecture

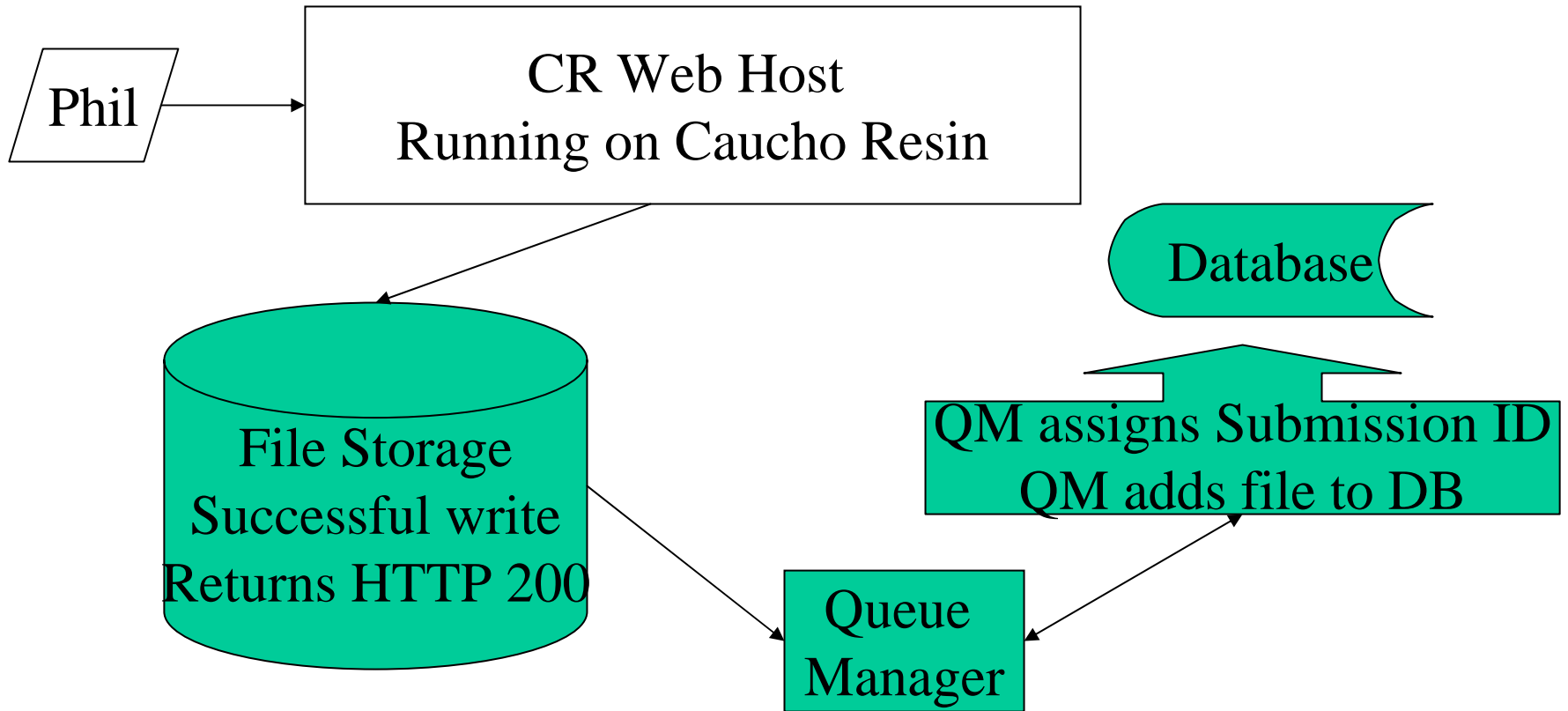
- Each machine shares the System software via NFS.
- Each has a local copy to act as Web Host.
- Each accesses the System Queue files via NFS



The Front Door

- All ID's are checked via login and password
- Servlet running on Caucho Resin validates and admits Phil
- 200 Success means Phil is inside. Nothing more.
- Queue Manager gives Phil a Submission ID
- Queue Manager copies Phil to the Database

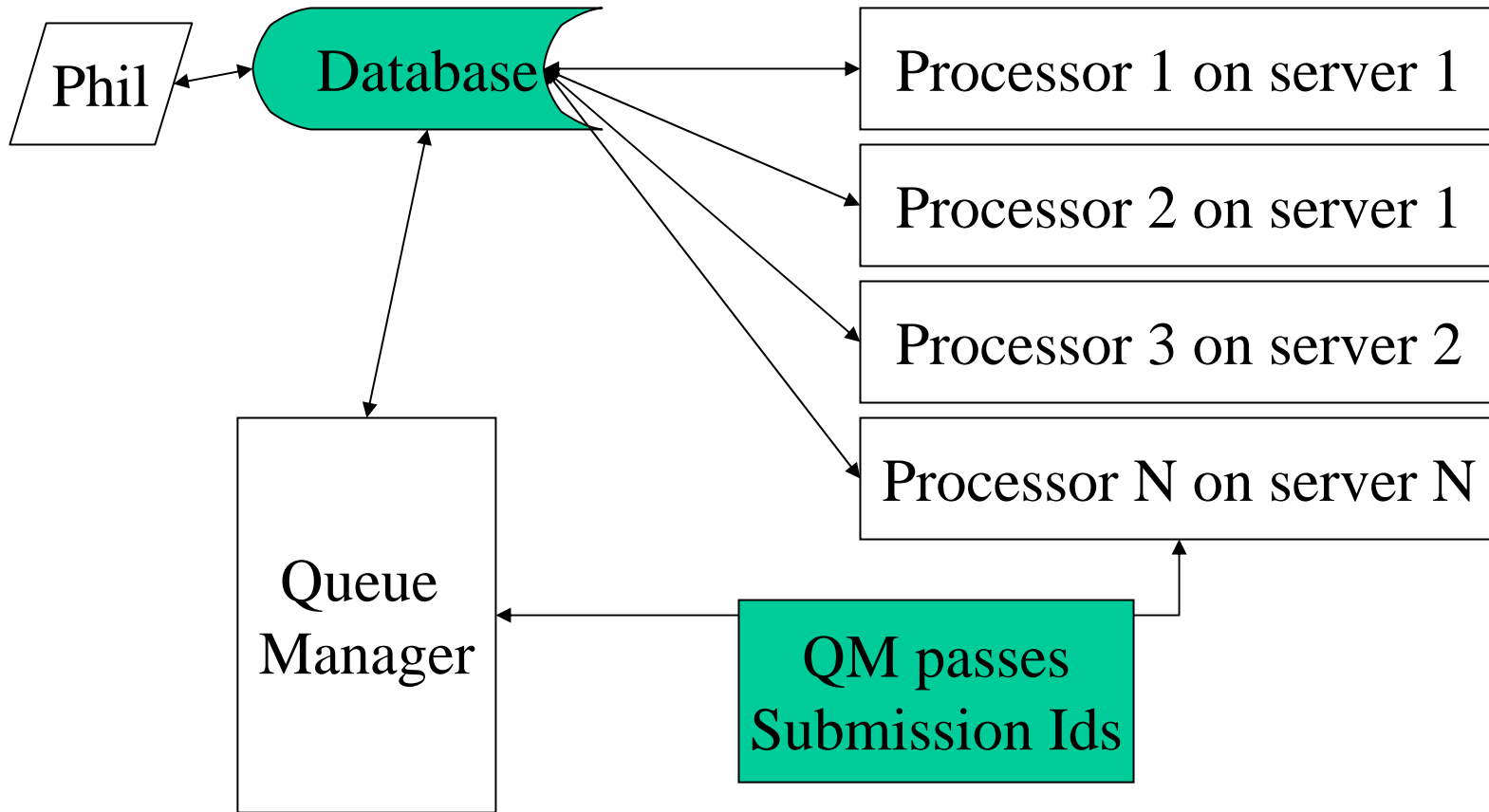
The Front Door



Waiting in the Queue

- Phil waits his turn with a Processor
- Each Submission Processor (SP) can be limited by size, submission type, and submitter.
- Some take only small files and keep the line moving along.
- Larger files, like Phil, take longer. ~ .5sec per DOI.
- SPs ask Queue manager for Submission ID that meets their specifications.
- A full Queue? Add more Submission Processors!

Building Architecture



User Based Filtering

- New ability to limit based on User (login).
- Good for making Deposits with References.
- References take ~.5 sec per DOI and ~1 sec per structured citation (a query).
- Please tell us if you plan to make a large deposit of DOIs with citations.
- A Processor for just your files can be setup and prevent Queue backups.

Queue Admin View

Submission administration - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Home Users **Submissions** Queries Reports Metadata Admin Help

Administration Upload **Show System Queue** Download Metadata

Queue Status: **Active** [Suspend Queue](#)
[Stored Query Monitor](#)
[Refresh queues](#)

6 submissions being processed

ID	Submitter	Name	Progress	Type	Received	started	Size [bytes]	SP ID
294096178	springer	scirebdata_2006-10-25-22-06-10.xml	80% done	XML metadata	25-Oct-06 16:47:52	25-Oct-06 16:47:53	254792	cr6.crossref.org-spm2
294031394	science_retro	10250690.xml	13% done	XML metadata	25-Oct-06 11:51:13	25-Oct-06 16:50:12	1694888	cr6.crossref.org-spm4sr
294090957	springer2	references.xml	95% done	pipelined query	25-Oct-06 16:14:42	25-Oct-06 16:40:31	308992	cr6.crossref.org-spq1
294029127	science_retro	10250688.xml	98% done	XML metadata	25-Oct-06 11:44:58	25-Oct-06 16:25:02	1598752	cr8.crossref.org-spm3sr
294030353	science_retro	10250689.xml	59% done	XML metadata	25-Oct-06 11:48:05	25-Oct-06 16:34:57	1750000	cr3.crossref.org-spm2
294093354	elsevier	mrw2006-10-25-16-22-00-1016001.xml	93% done	XML metadata				cr6.crossref.org-spm1

Submission ID

Submission Processor ID

37 submissions in metadata queue

ID	Submitter	Name	Received	Size [bytes]	Live
<input type="checkbox"/> 294032396	science_retro	10250691.xml	25-Oct-06 11:54:31	1846520	X Live
<input type="checkbox"/> 294033260	science_retro	10250692.xml	25-Oct-06 11:57:14	1168376	X Live
<input type="checkbox"/> 294034360	science_retro	10250693.xml	25-Oct-06 12:00:48	1470120	X Live

Find: Find Next Find Previous Highlight all Match case

Done

Fasterfox Adblock

When to Call about your Data

- If you have a large amount of data you plan to submit in a short time frame. (Really Big)
- You have a tight window (real time?) to turn around the deposit and make data available.
- You have many references to add.
- If you have questions about making deposits.

Processor Dislikes

Submission Processors parse files based on
Schema and Encoding rules

- Unencoded entities!
- Non-printing characters at the start of a file.
- More than one prefix in a file.
- Tags out of order.

What happened to Phil? and Handling Rejection

- Submission Administration is your best friend. <http://doi.crossref.org/servlet/useragent?func=adminSubmission>
- Find Phil there.
- Find reasons for rejection.
- No email from the system means the SP could not parse the file, even to extract the email address!

Submission Administration Results



Software developed by atypon

Home Users **Submissions** Queries Reports Metadata Admin Help

Administration Upload Show System Queue Download Metadata

Search Criteria: Received >= 2006-10-25

Submission ID

	ID	Filename	Depositor	Area	Type	Received	Started	Finished
E	294095948	240994205-FU_15524922_9999_9999A_9999	wiley	L	XML metadata	25-Oct-06 16:46:31	2006-10-25 16:46:32.0	2006-10-25 16:47:23.0
E	294093411	DOI_Query-Journal_...				25-Oct-06 16:30:15	2006-10-25 17:16:39.0	2006-10-25 17:16:39.0
E	294073006	240962185-FP_02776715_2006_25_23_SIM2	wiley	L	XML metadata	25-Oct-06 14:18:42	2006-10-25 14:22:39.0	2006-10-25 14:23:41.0
E	294072976	240962587-FP_16165187_2006_6_10_MABI8	wiley	L	XML metadata	25-Oct-06 14:18:31	2006-10-25 14:21:36.0	2006-10-25 14:22:19.0
E	294069035	PHCOv26i5-20061025105916.xml	extenza	L	XML metadata	25-Oct-06 13:59:30	2006-10-25 15:15:43.0	2006-10-25 15:16:01.0

Link to Phil!

Click Submission ID for Details

The screenshot shows a web application interface with a navigation bar at the top containing 'Administration', 'Upload', and 'Show System'. Below the navigation bar is a 'Submission Detail' section. A blue callout box labeled 'Submission ID' points to the 'Submission ID: 294095948' field. Another blue callout box labeled 'Link to Phil!' points to the '[Click to View]' link in the 'Content' field. A third blue callout box labeled 'What went wrong' points to the XML message content in the 'Message' field. The XML message shows a 'completed' status for the submission but a 'Failure' status for the record, with a message: 'Out of date version: 20061025161551'. The 'Stack Trace' field at the bottom shows 'null'.

Administration Upload Show System

Submission ID

Submission Detail

Submission ID: 294095948

Version: 0

Depositor: wiley

Name: 240994205-FU_15524922_9999_9999

Size: 17920 [bytes]

Area: Live

Type: XML metadata

Received: 25-Oct-06 16:46:31

Started: 25-Oct-06 16:46:32

Finished: 25-Oct-06 16:47:27

Content: [\[Click to View\]](#)

Has Note:

Modified: 2006-10-25 16:47:24.0

Error Level: Error

Message:

```
<?xml version="1.0" encoding="UTF-8"?>
<doi_batch_diagnostic status="completed"
sp="cr6.crossref.org">
  <submission_id>294095948</submission_id>

  <batch_id>240994205-FU_15524922_9999_9999A_9999_CYT020348.xml
  <record diagnostic status="Failure">
    <doi>10.1002/(ISSN)1552-4930</doi>
    <msg>Out of date version: 20061025161551</msg>
  </record diagnostic>
</doi_batch_diagnostic>
```

Stack Trace: null

Link to Phil!

What went wrong

What about this specific DOI?

- To find a specific DOI's deposit history use the Reports → Misc. Reports!
- See the Deposit history with Handy Links to Submission ID's.

DOI History Search



Software developed by  atypon

Home

Users

Submissions

Queries

Reports

Metadata Admin

Query Report

Metadata Report

Monthly Report

Misc. Reports

Please provide a DOI to view its history

DOI:

DOI History Results



Software developed by atypon

[Home](#) [Users](#) [Submissions](#) [Queries](#) **[Reports](#)** [Metadata Admin](#) [Help](#)

DOI History Report for 10.1049/ip-com:19990279

Submission ID

Click to manage the Conflict

Date	Event	Detail	By Whom
16-May-05 10:26:54	Add Back Year	Submission: 18251491	iee
11-Jul-05 02:27:06	Marked as Primary	Conflict: 194120	Super Admin Group
26-Sep-05 04:39:26	Transfer Ownership	Submission: 213398816	Super Admin Group
13-Jul-06 03:09:55	Update	Submission: 266067783	aip

Legend: =Content =Note =Detail =Save

Conclusions

- Club CrossRef is a fashionable place with solid, scalable infrastructure.
- If Phil is well formed, UTF-8 encoded, and orderly, he will have a swell time with a Processor.
- Its not that the Processor is ignoring Phil, it is waiting for a different kind of Phil or a specific user Phil.
- The Queue is flexible and allows us to shape file traffic to make most use of our servers.

Important things to Remember

- You can find all your submissions through Submission Administration.
- A success on the submission via http does NOT equal successfully parsed and deposited File.
- Try to keep files under 2 MB.
- Got a lot of Data submit? Please call!