

Dataware Networking

Presented by Anil Madhavapeddy

University of Cambridge

Joint work with Richard Mortier, James Goulding
and Derek McAuley (Nottingham),

Jon Crowcroft, Steve Hand, Malte Schwarzkopf
(Cambridge)



What is Horizon?

- A Digital Economy Research Centre at the University of Nottingham comprising:
 - A Digital Economy Hub
 - £20m from RCUK and university
 - Spokes at Cambridge, Reading, Exeter, Brunel
 - A Doctoral Training Centre
 - £15m from RCUK and university
 - 20 PhD students per year for 5 years
 - Now 120+ partner companies, from 40 in initial bid
 - 3 TEDDI projects
 - ... + future Digital Economy projects

Our Digital Footprints

“Every time we register for a new web service, or upload our photos and videos, we are enlarging our own digital footprints”

- Whether “informed” or not
 - Facebook, Google
- Digital footprint poses **major societal challenges**
- ...but also opportunity for **economic growth**

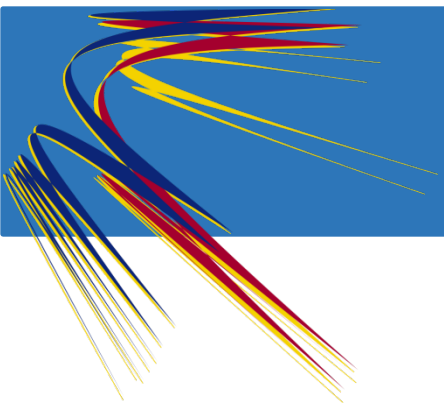


From Digital Nomads...



...to Digital Yurts





Generational Computing





Dataware Research Program

- We **make information about our lives available** online
 - Whether or not we realise it
 - Often in very large, very rich data silos (e.g., Facebook)
- There are more opportunities for **mutually beneficial** exploitation of digital personal data
 - E.g., Shopping basket optimizer;
Boots medical prescription conflict detector

Key Challenge:

How do we enable individuals to control collection and exploitation of both *their data* and *data about them*?

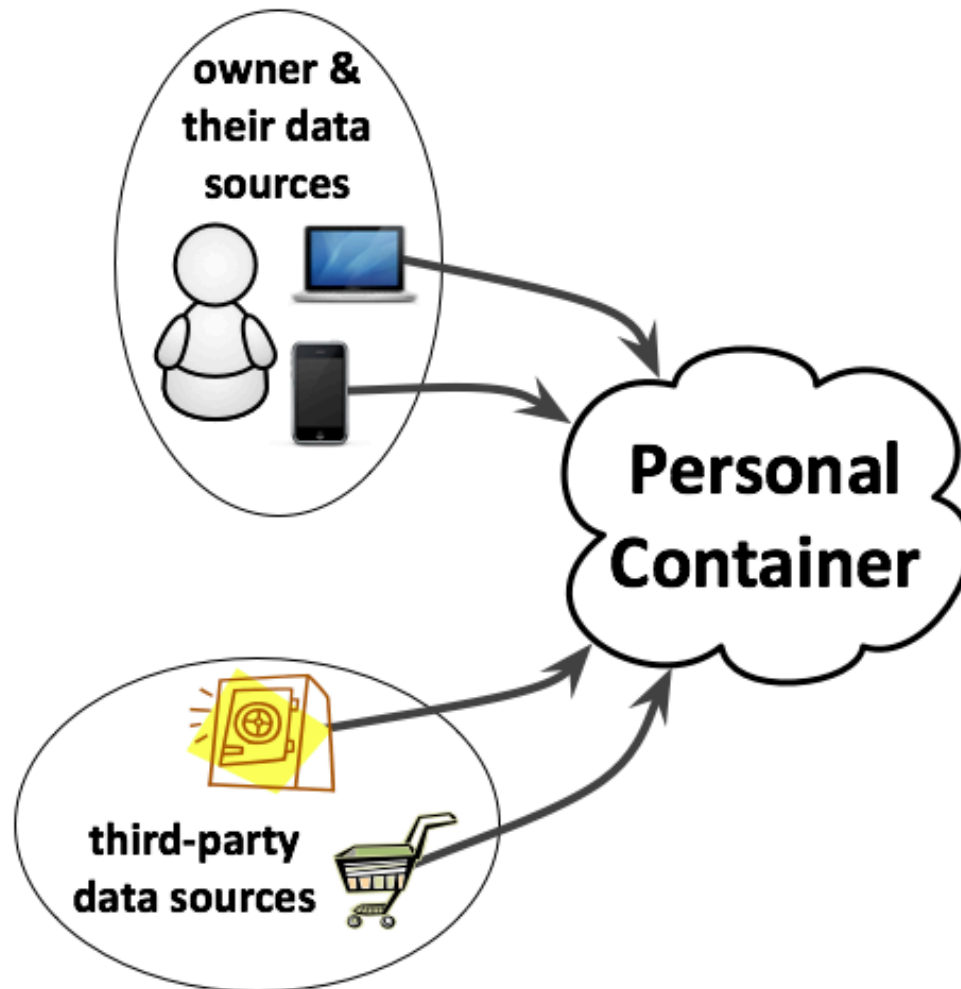


1) Dataware: nodes

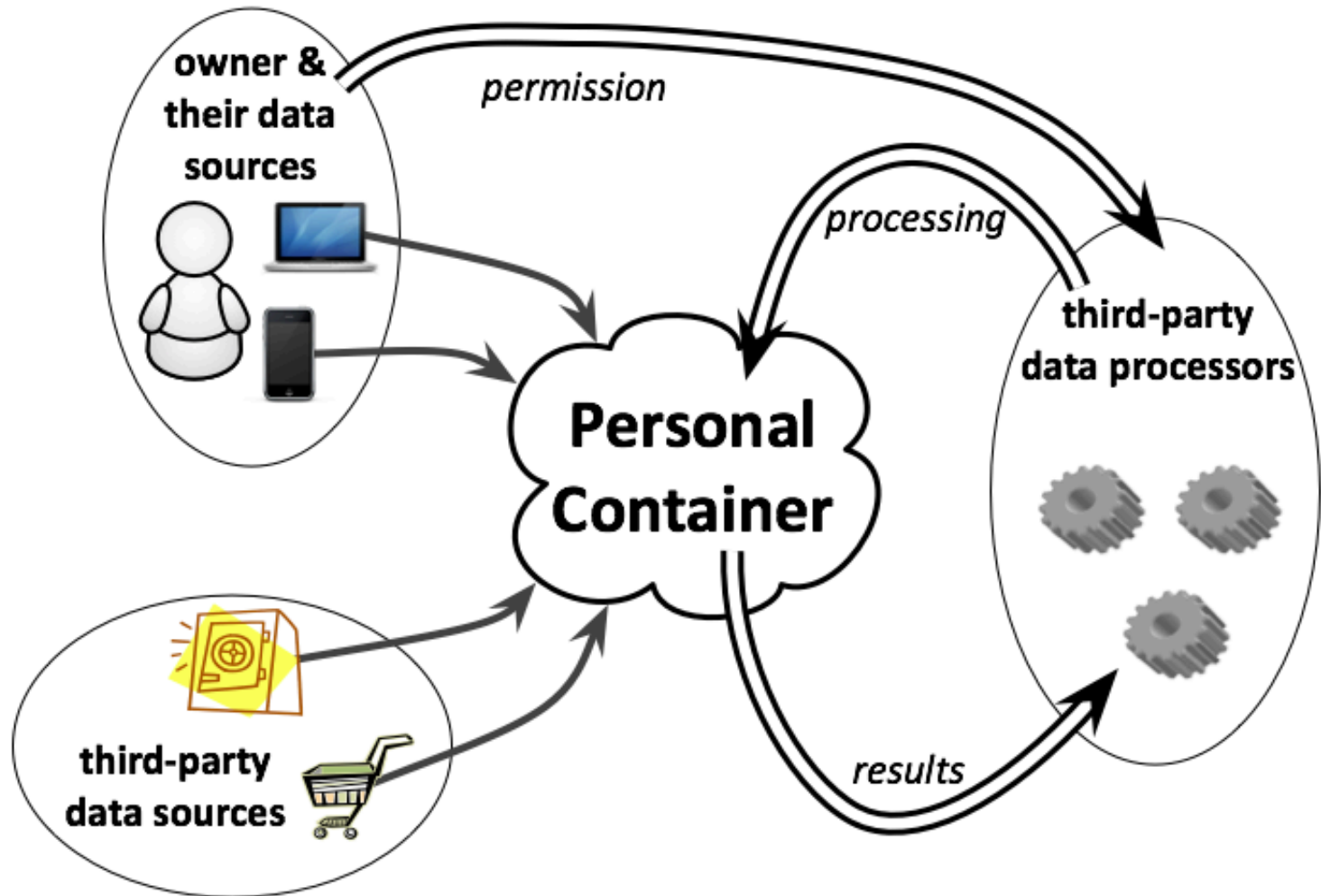


**Personal
Container**

2) Dataware: network



3) Dataware: computation





Evolution or Revolution?

- One way to put this in context is via Van Jacobson's *content centric networking*:
 - Telephones care about building paths (not calls)
 - Internet cared about connections (not data)
 - Content centric networking cares about data (not results)
- **Dataware cares about results**
 - Computation is as mobile as data, if not more so
 - cf. datacentre computing, map-reduce and CIEL



1) Nodes: placement

No single “best platform”

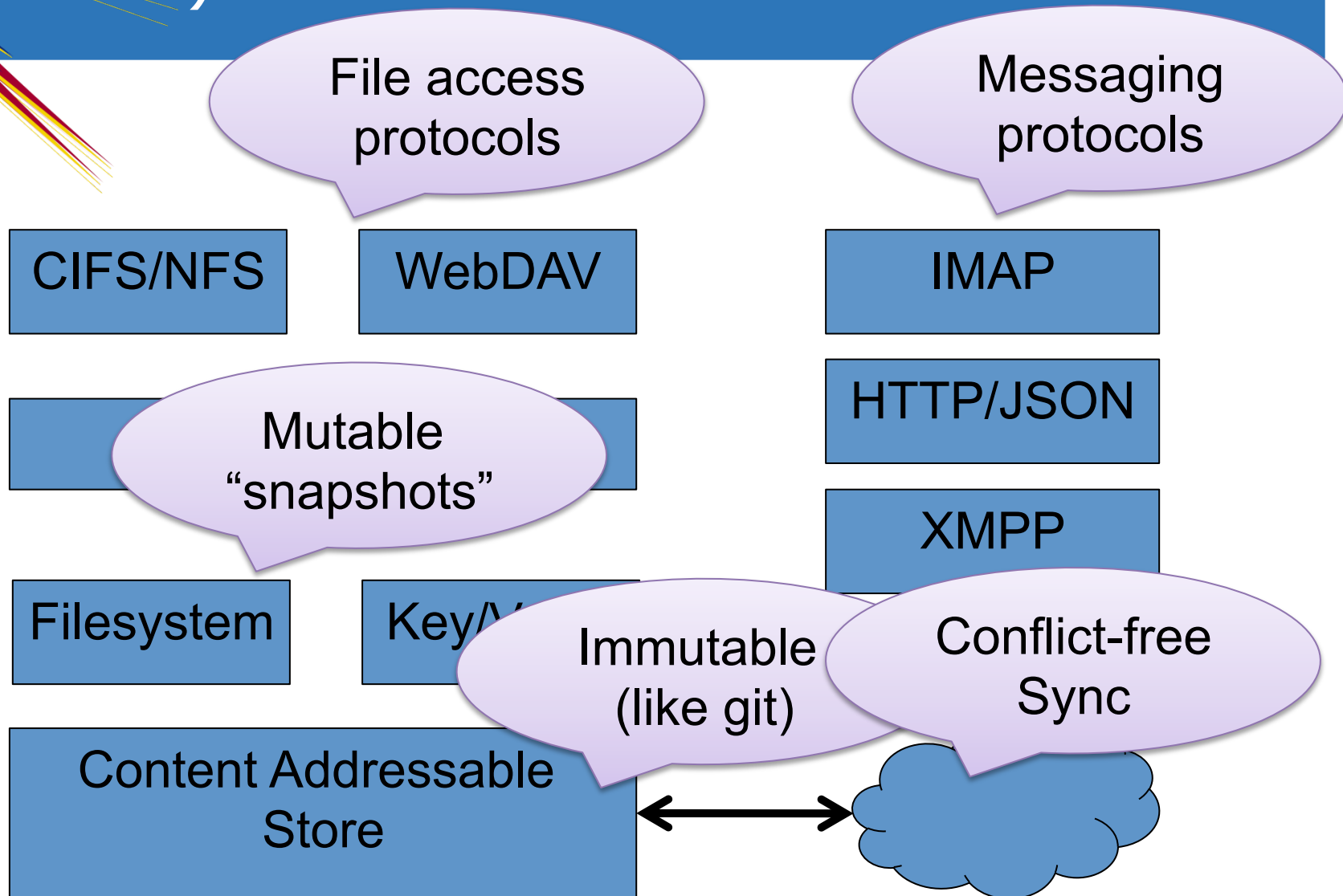
Platform	Google AppEngine	Amazon EC2	Home Computer	Mobile Phone
Storage	Moderate	Moderate	High	Low
Bandwidth	High	High	Limited	Low
Accessibility	Always on	Always on	Variable	Poor
Computation	Limited	Flexible, Plentiful	Flexible, Limited	Limited
Cost	Free	Expensive	Cheap	Cheap
Relability	Medium	High	Medium (failure)	Low (loss)



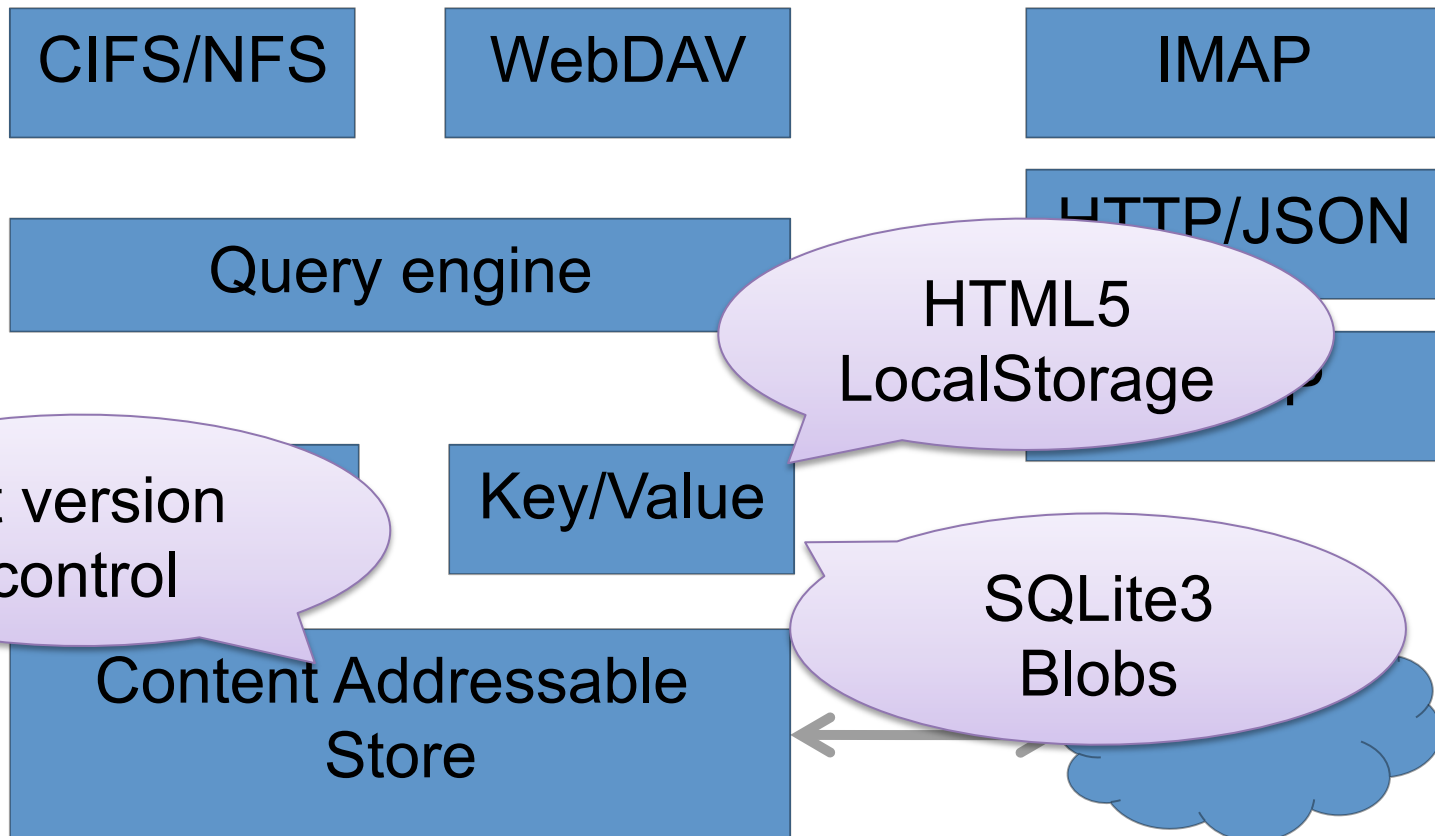
1) Nodes: design principles

- Nodes run in **several locations** (cloud, home, phone)
- No **expert management** available
- **Minimize external dependencies** on third parties.
- Support **incremental** deployment

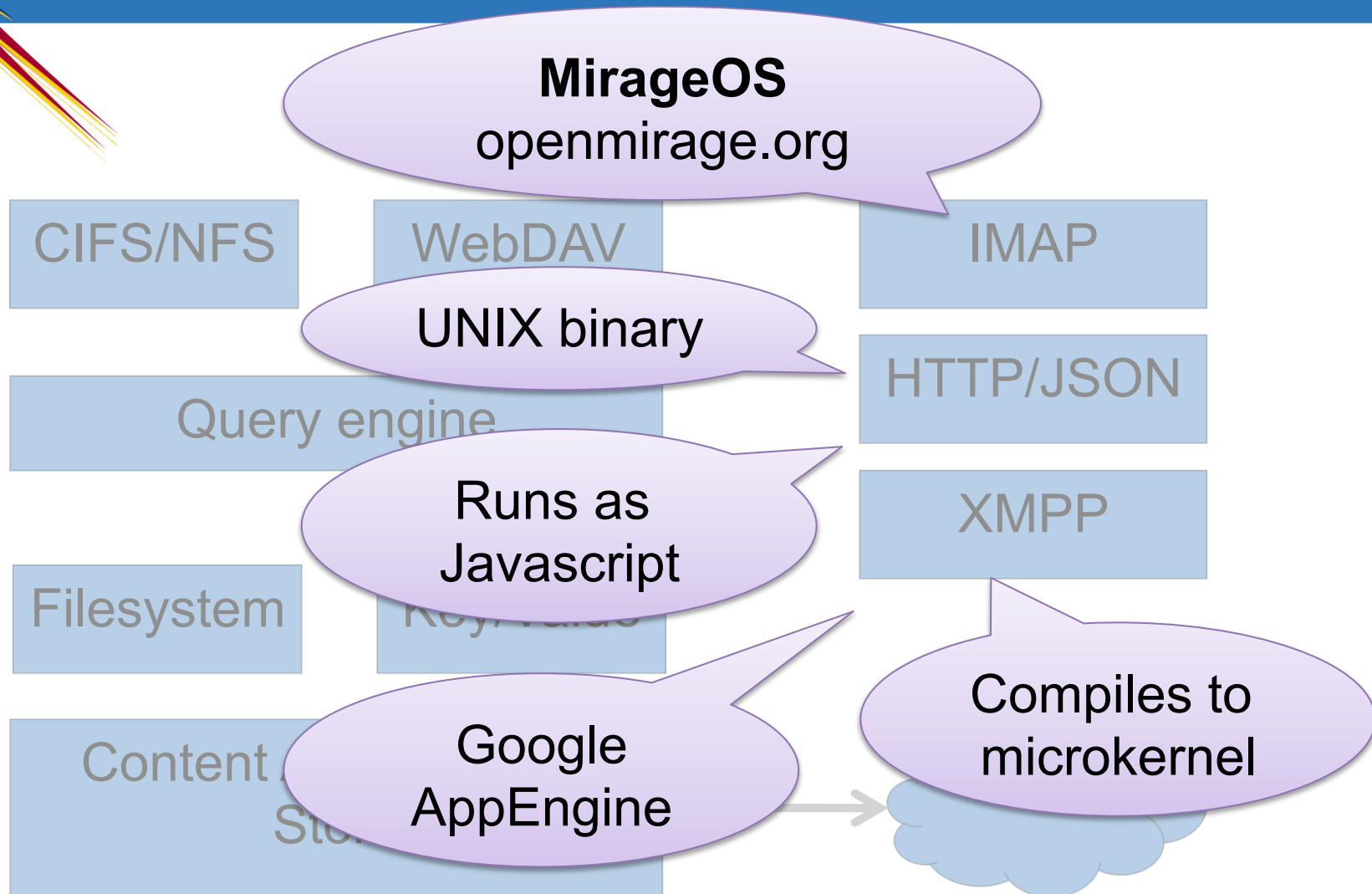
1) Nodes: architecture



1) Nodes: multi-scale portability



1) Nodes: clean slate protocols



1) Node: screen shot

Personal Container

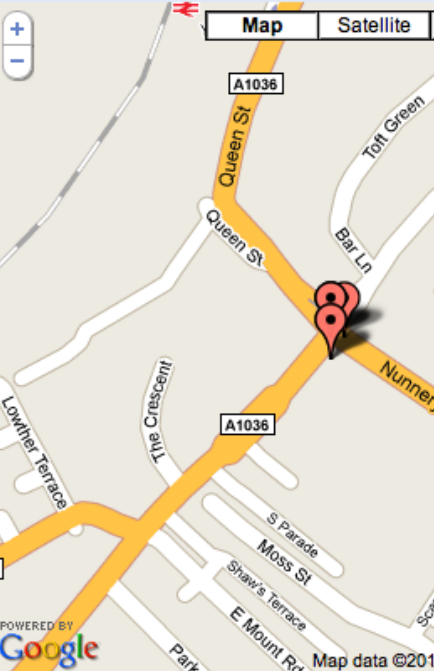
Map

+

-

Map

Satellite















Settings

Plugins

Recent Activity


Messages

Contacts

Type	Body	
	TacticalGrace → Anil Madhavapeddy @avsm Haven't checked, but I doubt the Mac OS X scripting bridge stuff is supported in the iPhone SDK.	Sun Apr 11 2010 12:00:33 GMT+0100 (BST)
	Anil Madhavapeddy → TacticalGrace @TacticalGrace the Android Scripting env has a local HTTP/JSON bridge for API calls...sounds like something similar would work on	Sun Apr 11 2010 11:54:52 GMT+0100 (BST)
	Anil Madhavapeddy → Trevor Agus	Sun Apr 11 2010 11:46:26 GMT+0100 (BST)
	Anil Madhavapeddy → (14 in thread)	Sun Apr 11 2010 11:32:09 GMT+0100 (BST)
	Anil Madhavapeddy → Slightly surreal playing early morning pool at a beautiful nuns convent in York; Bar Convent highly recommended! http://twitpic.com/16	Sun Apr 11 2010 10:27:45 GMT+0100 (BST)
	Anil Madhavapeddy → (3 in thread)	Sat Apr 10 2010 23:33:28 GMT+0100 (BST)
	+441904464900 → Anil Madhavapeddy	Sat Apr 10 2010 20:43:23 GMT+0100 (BST)
	Anil Madhavapeddy → Alastair Tse (1 in thread) @liquidx lol!	Sat Apr 10 2010 19:31:08 GMT+0100 (BST)
	Anil Madhavapeddy → quentinsf @quentinsf my apple cinema display is unusable at work without all the windows shuttered :(Sat Apr 10 2010 14:24:32 GMT+0100 (BST)
	Anil Madhavapeddy → Nick Ludlam	Sat Apr 10 2010 13:53:04 GMT+0100 (BST)
	Anil Madhavapeddy → Alex Ho	Fri Apr 09 2010 21:11:22 GMT+0100 (BST)
	Anil Madhavapeddy → dpc	Fri Apr 09 2010 20:30:22 GMT+0100 (BST)

Page

2 of 21



Displaying messages 21 - 40 of 412

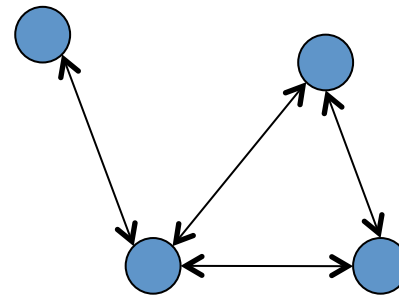


1) Nodes: formal methods

- Since there are few expert users, may be **years** between software updates.
 - **Secure language**: MirageOS (openmirage.org)
 - **Privilege separation**: Capsicum (USENIX Security 2010), TESLA (under submission)
 - **Hardware/software** co-design: CHERI MIPS processor (DARPA CRASH)
 - **Single user**: one server, one person. Simple authentication.

2) Naming: personal cloud

- Many devices per user:
 - Laptop
 - Desktop
 - Phone
 - Home Devices
- Need to connect to each of these seamlessly across different connectivity types
- Host support for **route setup** is very inflexible.
 - SSH tunnels or IPSec VPNs
 - Kerberos, Active Directory
 - HTTP proxys, website auth

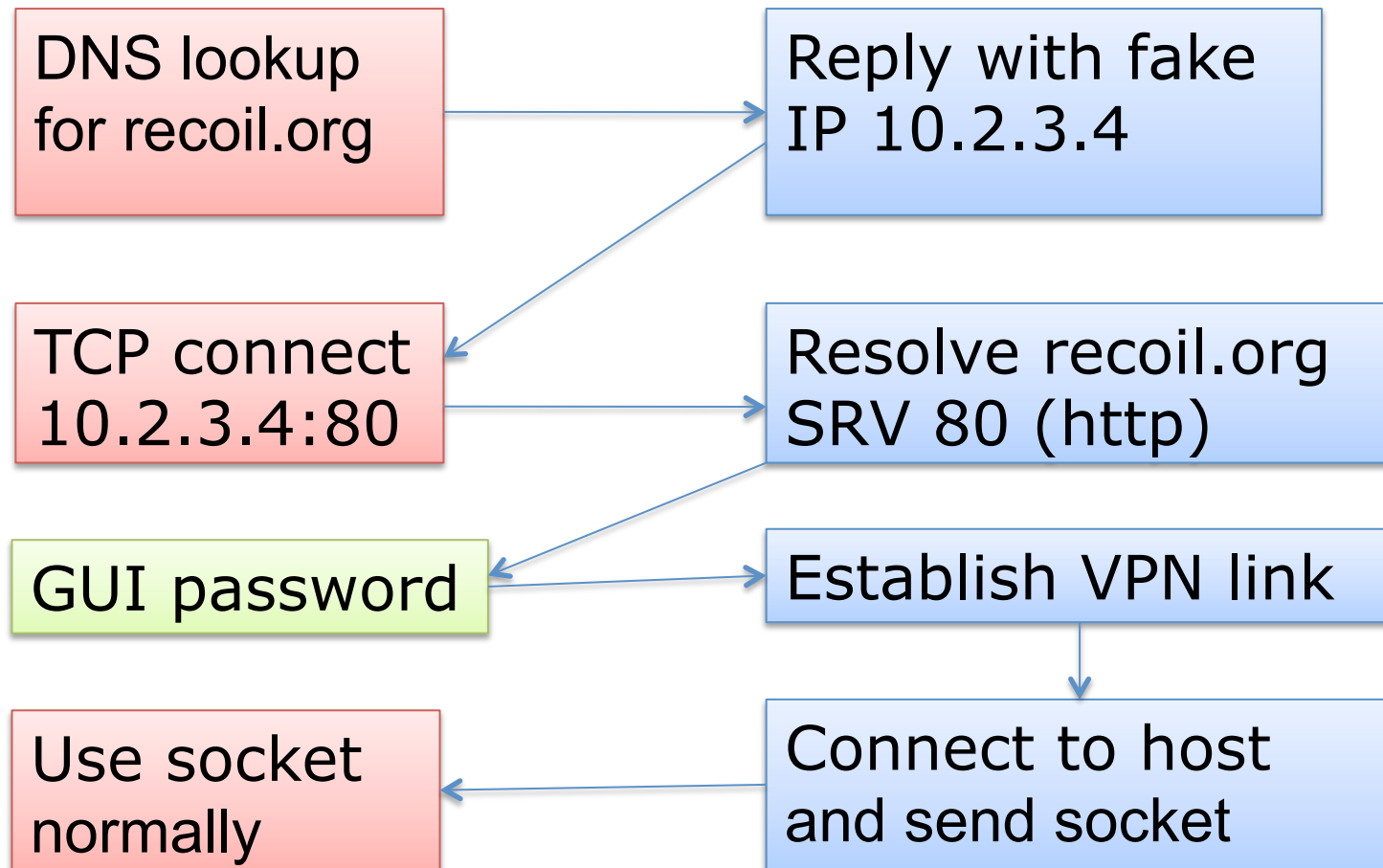




2) Naming: approach

- **Goal:** secure auto-configuration of devices to a user's personal cloud, including route management
 - Use **DNSSEC** to give every user a global ID
 - Similar to XMPP XID (e.g. anil.recoil.org)
 - Extend OS to support **user-level application flows**
 - **Context-aware scheduler** to figure out user activities
 - MarcoPolo: <http://www.symonds.id.au/marcopolo/>
 - Shelf: <http://code.movieos.org/shelf/>

2) Naming: DNS cookies (WIP)

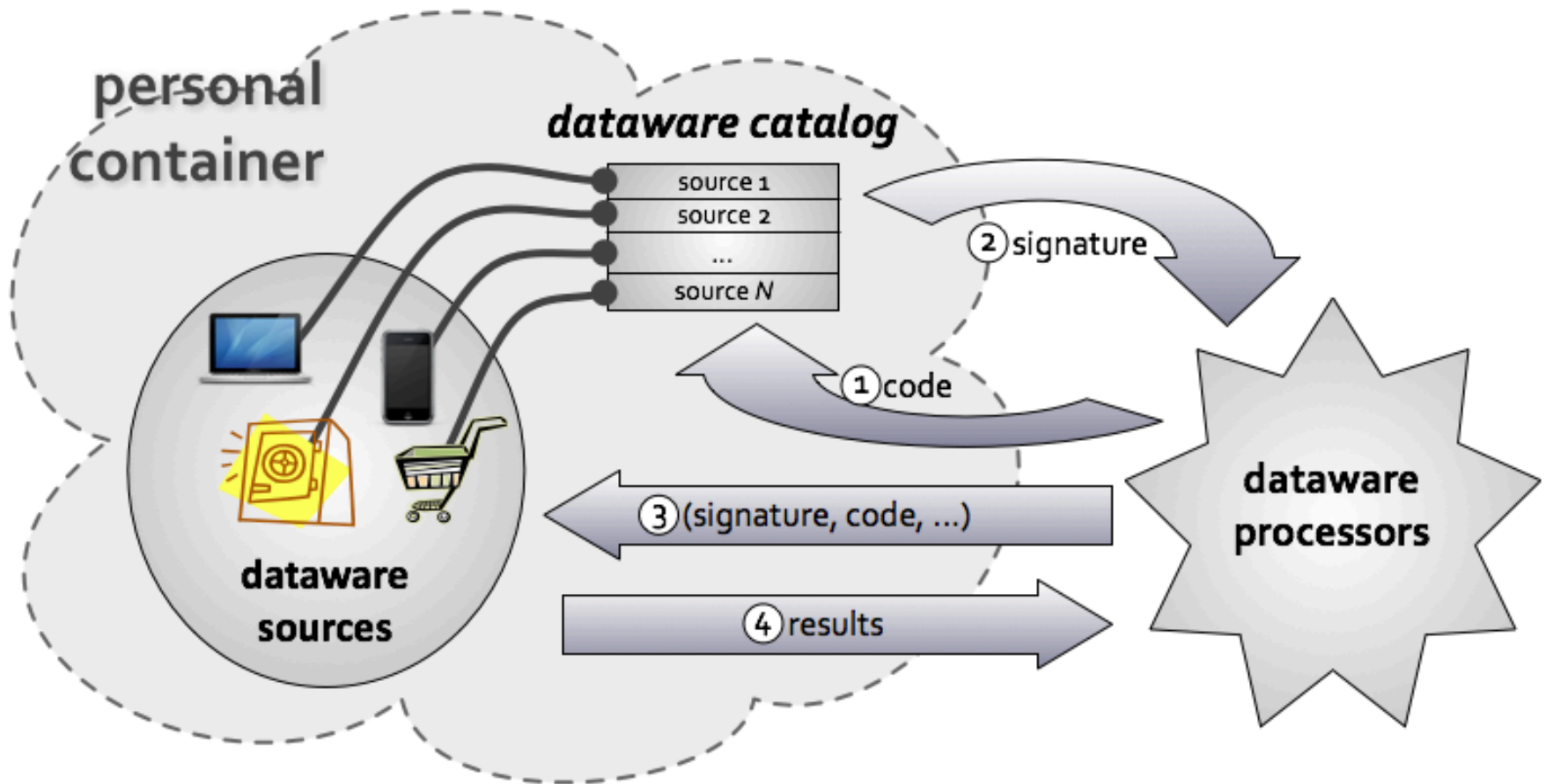




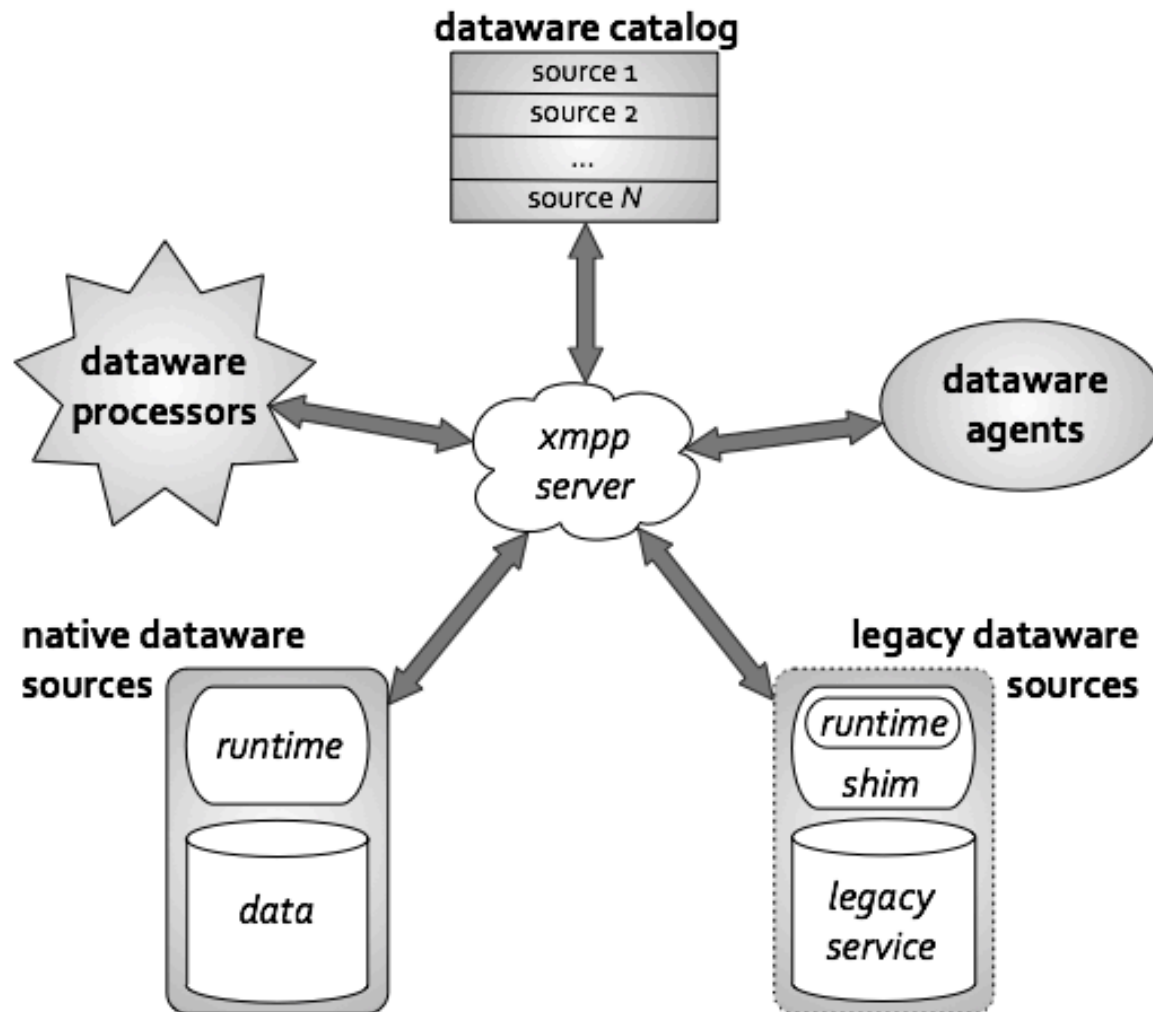
2) Naming: discussion

- **DNS cookies** combine name resolution + sockets into **flows** (“host OpenFlow”)
- **Delayed resolution** lets kernel establish VPN based on a named service
- Interacts well with real-world (e.g. internal DNS at workplace) that currently fails.
- Still several issues (socket handoff) that are very platform-kernel specific.

3) Computation: registration



3) Computation: routing requests



Prototypes: Shim Login



my data sphere

Log in | Facebook - Mozilla Firefox

Facebook, Inc. (US) https://ssl.facebook.com/login.php?api_key=1089099891703838&skip_ap

Facebook login

Log in to use your Facebook account with data-chant.

Email address:

Password:

☒ Keep me logged in

[Forgotten your password?](#)

Sign up for Facebook

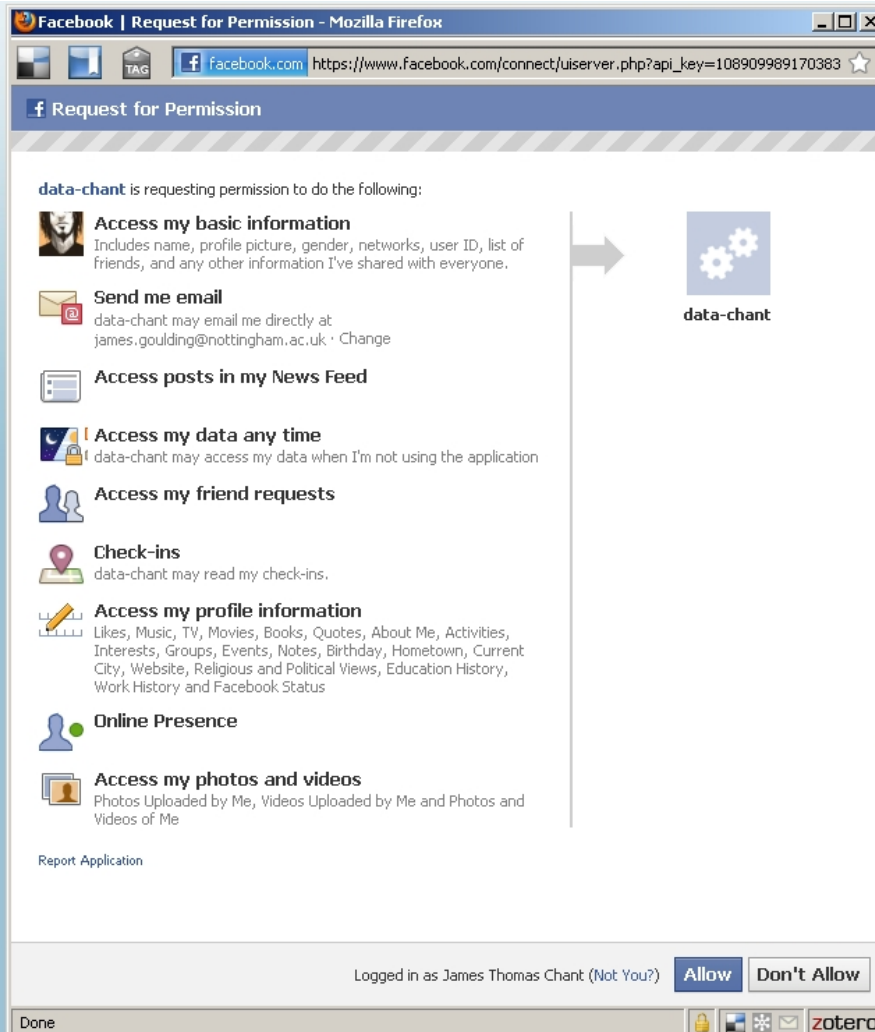
Log in Cancel

Done

zotero

login via
facebook

Prototypes: Shim Permission



data sphere

login via
facebook

Prototypes: Web Interface

[My Data](#) [Dataspheres](#) [History](#) [Statistics](#) [Datablog](#) [Login](#) [Register](#)

 my datasphere

Your data

- » Live Feed
- » Full History
- » Statistics
- » Your Details

Datasphere

- » Your Datasphere
- » Activity Summary
- » Recent Activity
- » Available

Sharing

- » Accounts Summary
- » Your Requests
- » Sharing Stats

Datasphere Source summary:

 common name:
Facebook Datasphere Shim
namespace:
ds:facebook
subscription status:
ACTIVE
[unsubscribe](#)

1 2 next »

Fri 19 November

 New Facebook "status" post: "I think we are alive alive o" 7:38pm

Thu 18 November

 In your Facebook bio, your "quotes" has changed to "Quotation changed again so there! And again!" 9:47am

 In your Facebook bio, your "birthday" has changed to "03/10/1995" 9:47am

Wed 17 November

 New Facebook "status" post: "mighty booooooshka" 5:05pm

Tue 16 November

 New Facebook "status" post: "test" 2:10pm

 In your Facebook bio, your "bio" has been removed. 2:02pm

 In your Facebook bio, your "hometown" has changed to Olton, Solihull, United Kingdom (107013716004816) 2:01pm

Your datasphere details:

name:
James Goulding
datasphere id:
james.goulding@gmail.com

Datasphere you are running:

Datasphere requests:

Available datasphere:

[see all...](#)

```

ec2-user@ip-10-48-138-30:~$ java -jar datasphere.jar -t 8080 -p critic41
2010-11-25 10:57:40 [CONFIG] --- DSCatalog: loading configuration file... [SUCCESS]
2010-11-25 10:57:40 [FINE] --- DSDataManager: admin password has been specified via the command line... [SUCCESS]
2010-11-25 10:57:40 [INFO] --- DSCatalog: Establishing database connection and consistency...
2010-11-25 10:57:40 [INFO] --- DSDatabaseManager: Connecting to database for persistence... [SUCCESS]
2010-11-25 10:57:40 [INFO] --- DSDataManager: Checking System Table integrity... [SUCCESS]
2010-11-25 10:57:40 [INFO] --- DSCatalog: Attempting to start server components...
2010-11-25 10:57:41 [INFO] --- DSCatalog: HTTP server setup...[SUCCESS]
2010-11-25 10:57:41 [INFO] --- DSChatServer: Starting the internal XMPP server on port 5222... [SUCCESS]
2010-11-25 10:57:41 [INFO] --- DSChatServer: Creating XMPP bots for 2 clients...
2010-11-25 10:57:43 [FINE] --- DSClientBot: connecting bot for client [testreceiver@jabber.org]... [SUCCESS]
2010-11-25 10:57:43 [FINER] --- DSPresenceListener: [testreceiver@jabber.org] Dataware available for (testreceiver@jabber.org/datasphere)
2010-11-25 10:57:44 [FINE] --- DSClientBot: connecting bot for client [james.goulding@gmail.com]... [SUCCESS]
2010-11-25 10:57:44 [FINE] --- DSChatServer: 2 XMPP bots have been connected... [SUCCESS]
2010-11-25 10:57:44 [INFO] --- DSCatalog: XMPP server setup...[SUCCESS]
2010-11-25 10:57:44 [INFO] Datasphere setup and ready for FULL service...
2010-11-25 10:57:44 [FINER] --- DSPresenceListener: [james.goulding@gmail.com] Dataware available for (j-eliza@appspot.com/bot)
2010-11-25 10:57:44 [FINER] --- DSPresenceListener: [james.goulding@gmail.com] Dataware available for (data-chant@appspot.com/bot)
2010-11-25 10:57:44 [FINER] --- DSPresenceListener: [james.goulding@gmail.com] Dataware unavailable for (testsender@jabber.org)
2010-11-25 10:59:25 [FINER] --- DSPresenceListener: [james.goulding@gmail.com] Dataware available for (james.goulding@gmail.com/Talk.v10498203F13)
2010-11-25 10:59:25 [FINER] --- DSPresenceListener: [james.goulding@gmail.com] Dataware available for (james.goulding@gmail.com/Talk.v10498203F13)
2010-11-25 10:59:25 [FINER] --- DSPresenceListener: [james.goulding@gmail.com] Dataware available for (james.goulding@gmail.com/Talk.v10498203F13)
2010-11-25 11:01:53 [FINE] --- DSUpdateListener: [james.goulding@gmail.com] <DSUpdate><namespace>ds:facebook</namespace><primaryTag>ds:fb:post</primaryTag>
<description>New Facebook "status" post: "what the world needs now..."</description><crud>create</crud><ctime>1290682887000</ctime><ftime>1290682913014</ftime>
<rtime>null</rtime><total>0</total><tag>ds:communication</tag><meta><message>what the world needs now..."</message><category>status</category></meta></DSUpdate>
e>

```

ec2-46-51-150-243.eu-west-1.compute.amazonaws.com:8080/user_history

Docs Webmail 31 Calendar Gmail Facebook Outlook Jambon SIGWEB Flickr

my datasphere

Your data

- » Live Feed
- » Full History
- » Statistics
- » Your Details

Dataware

- » Your Dataware
- » Activity Summary
- » Recent Activity

Thu 25 November

- New Facebook "status" post: "what the world needs now..." 11:01am
- New Facebook "status" post: "Another glorious sunny morning..." 9:25am

Fri 19 November

- New Facebook "status" post: "I think we are alive alive o" 7:38pm

Thu 18 November

data-chant@appspot.com

Send voicemail Call Send Files Email

data-chant: { "namespace": "ds:facebook", "primaryTag": "ds:fb:post", "tags": ["ds:communication", "ctime": 1290682887000, "ftime": 1290682913014, "crud": "create", "description": "New Facebook \"status\" post: \"what the world needs now...\", \"meta\": { \"message\": \"what the world needs now...\", \"category\": \"status\" } } }

data-chant is online.

Sent at 11:01 AM on Thursday

Your datasphere details:

name:
James Goulding
datasphere id:
james.goulding@gmail.com

Dataware you are running:





Dataware Summary

- So conceptually, a *dataware application* is:
 - A network of running components,
 - Each processing your personal data,
 - In a manner acceptable to you
- Your *dataware* implements your *personal container*:
 - Defining APIs to your data,
 - Enabling third-parties to process your data,
 - While ensuring *you* retain control,
 - And they *don't* get copies of your data
 - (unless you want them to!)