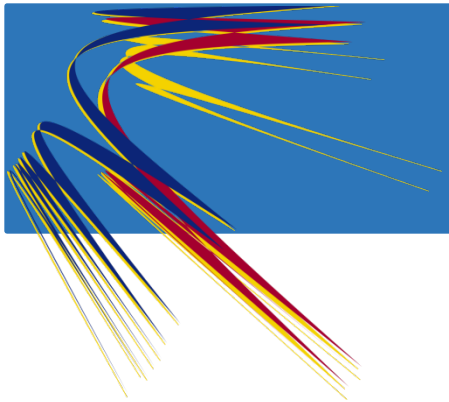


horizon
DIGITAL ECONOMY RESEARCH

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)

The University of Nottingham

UNIVERSITY OF 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

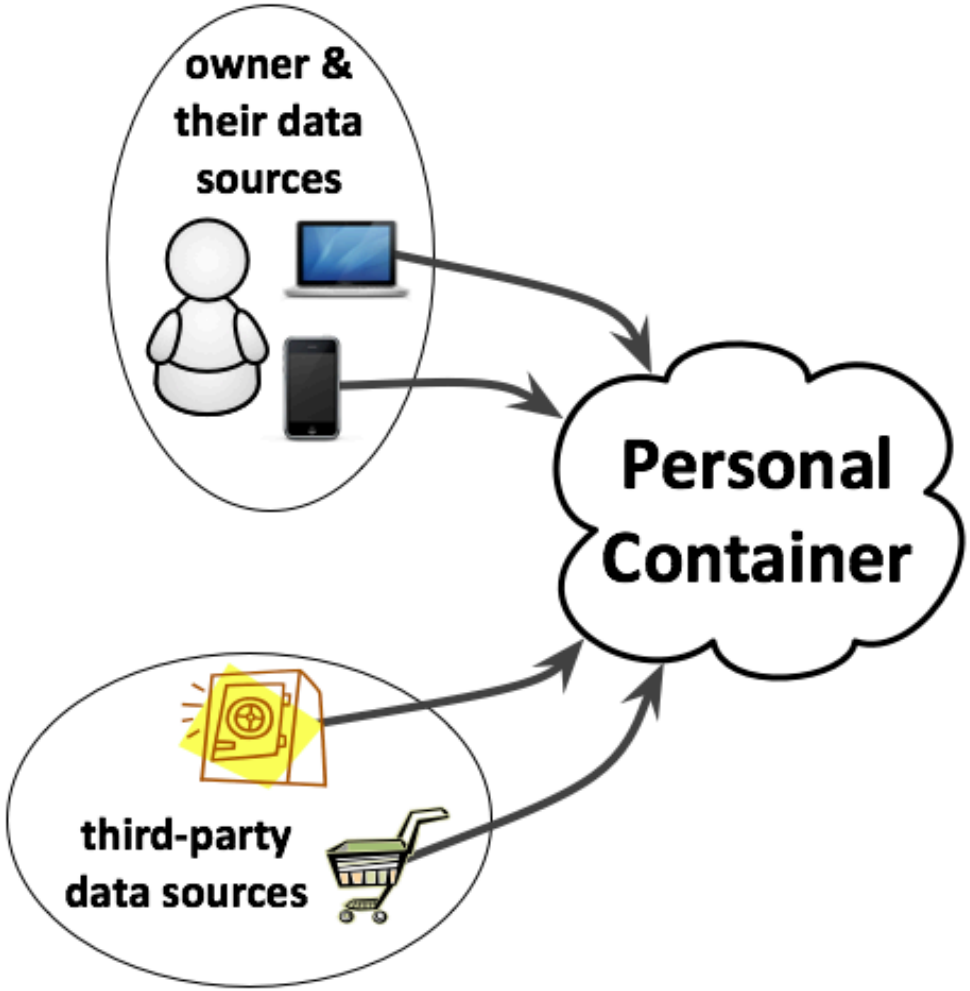
- We **make information about our lives available** online
  - Whether or not we realise it
  - Often in very large, very rich data silos (e.g., Facebook)
- There are more opportunities for **mutually beneficial** exploitation of digital personal data
  - E.g., Shopping basket optimizer;
    Boots medical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

Key Challenge:

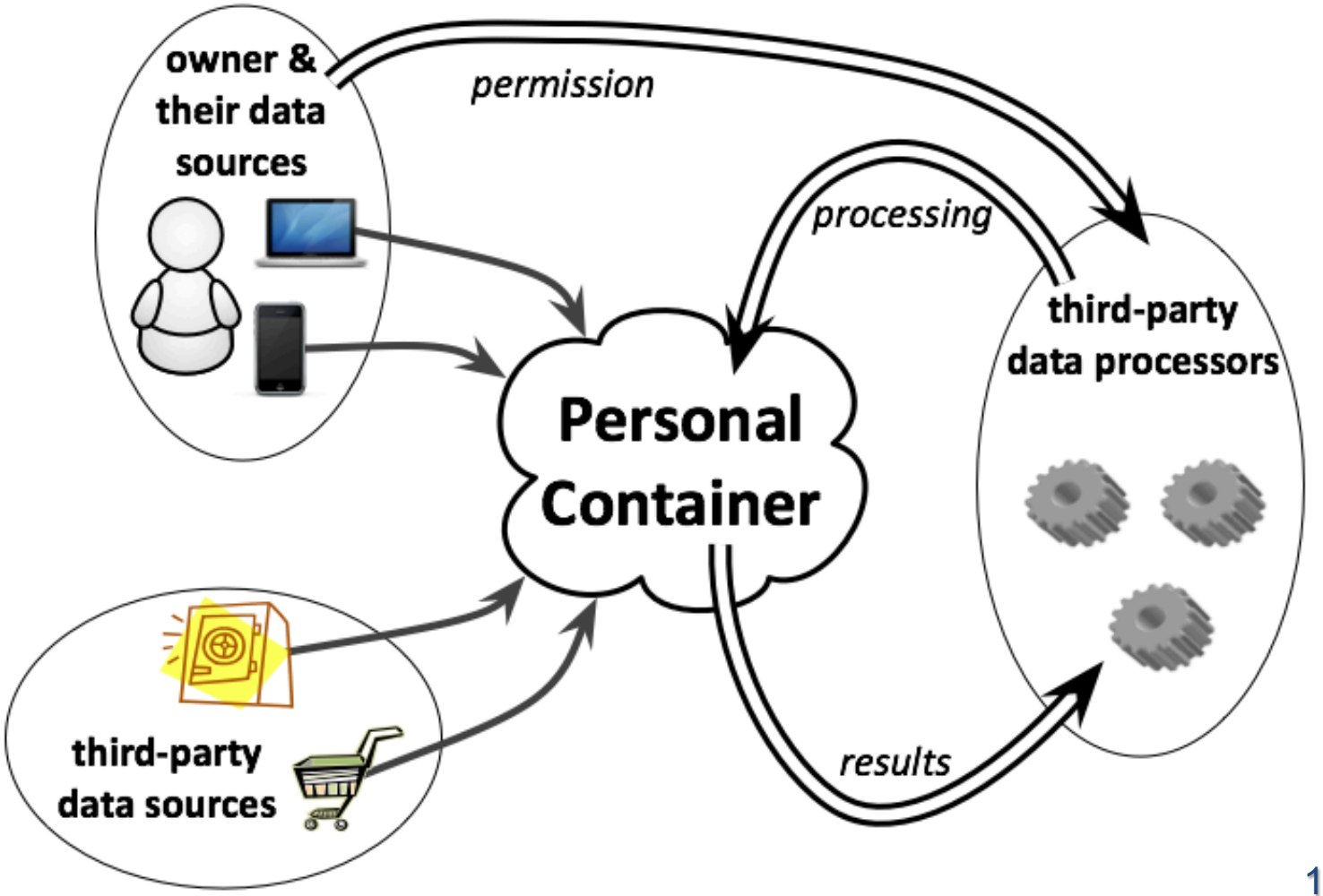
How do we enable individuals to control collection and exploitation of both *their data* and *data about them*?

1) Dataware: nodes

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

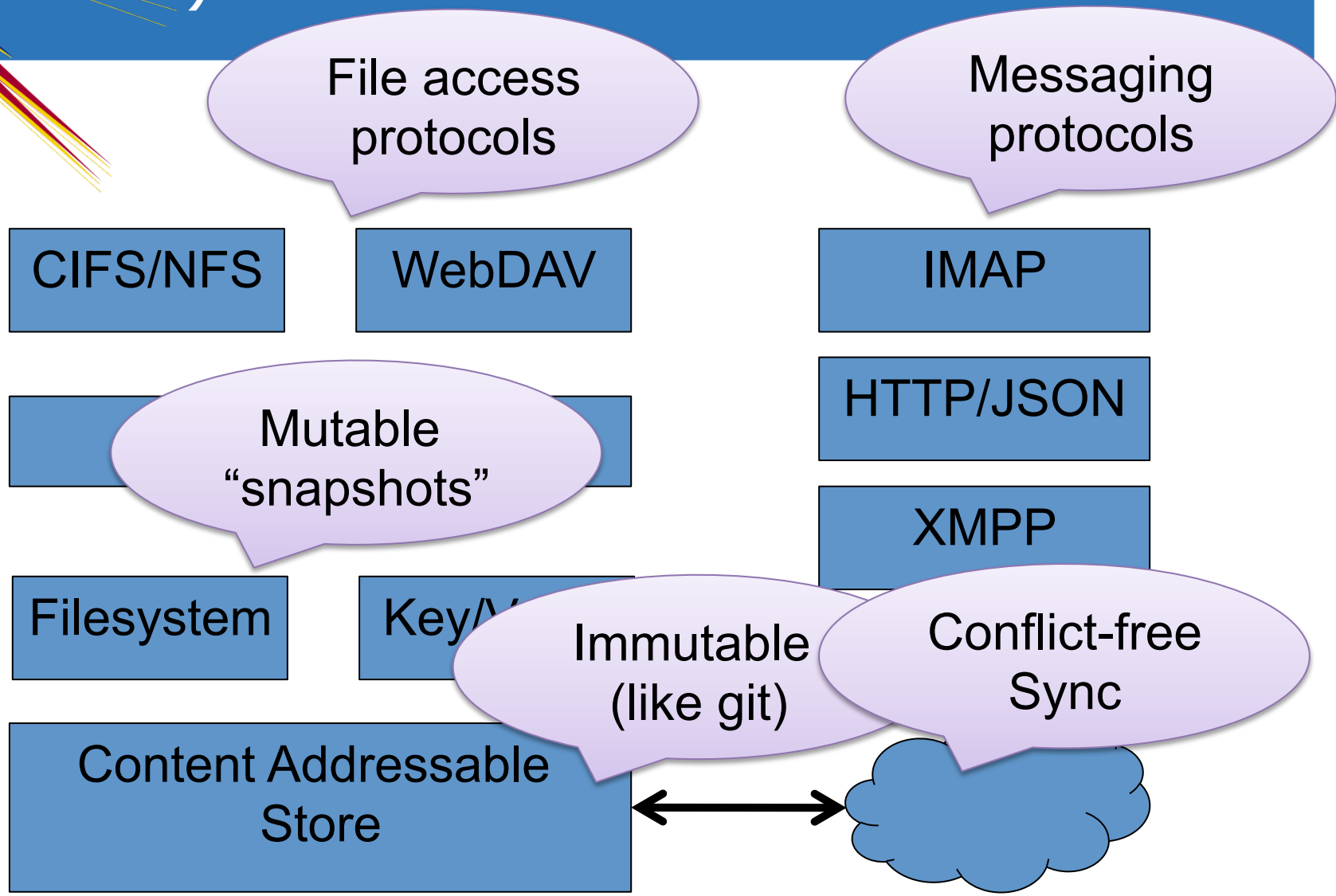
No single “best platform”

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)

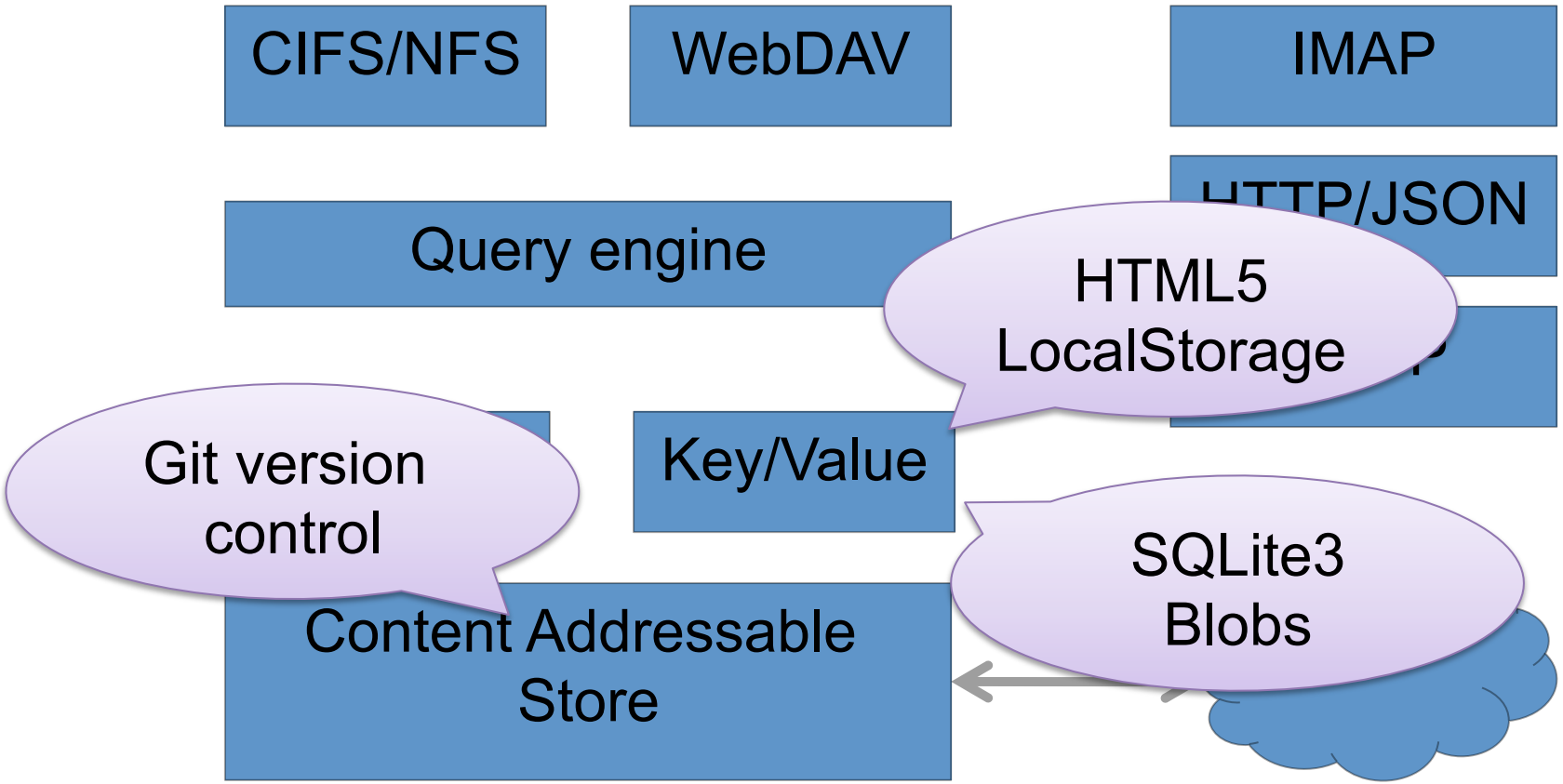
1) Nodes: design principles

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

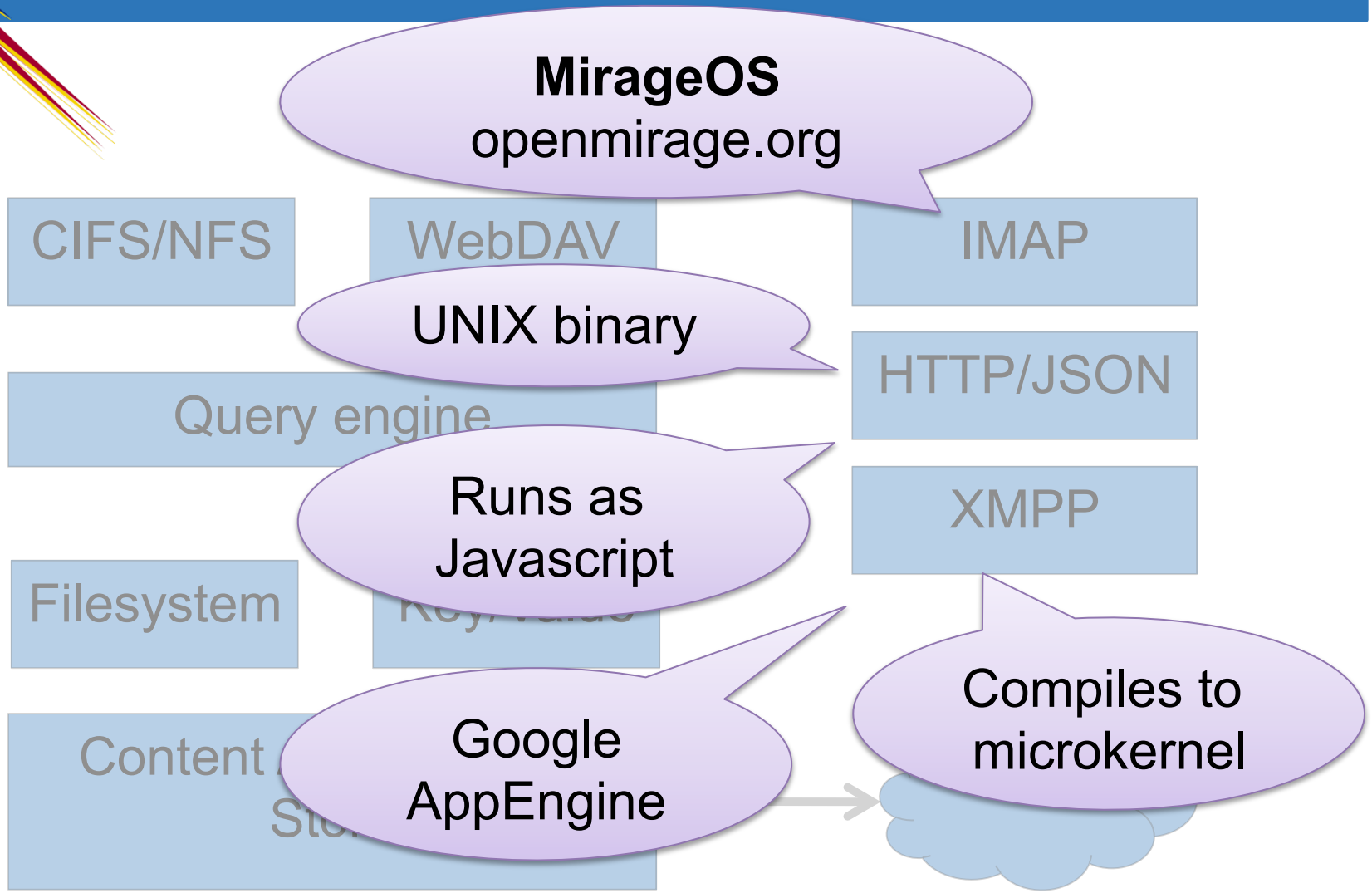
1) Nodes: architecture



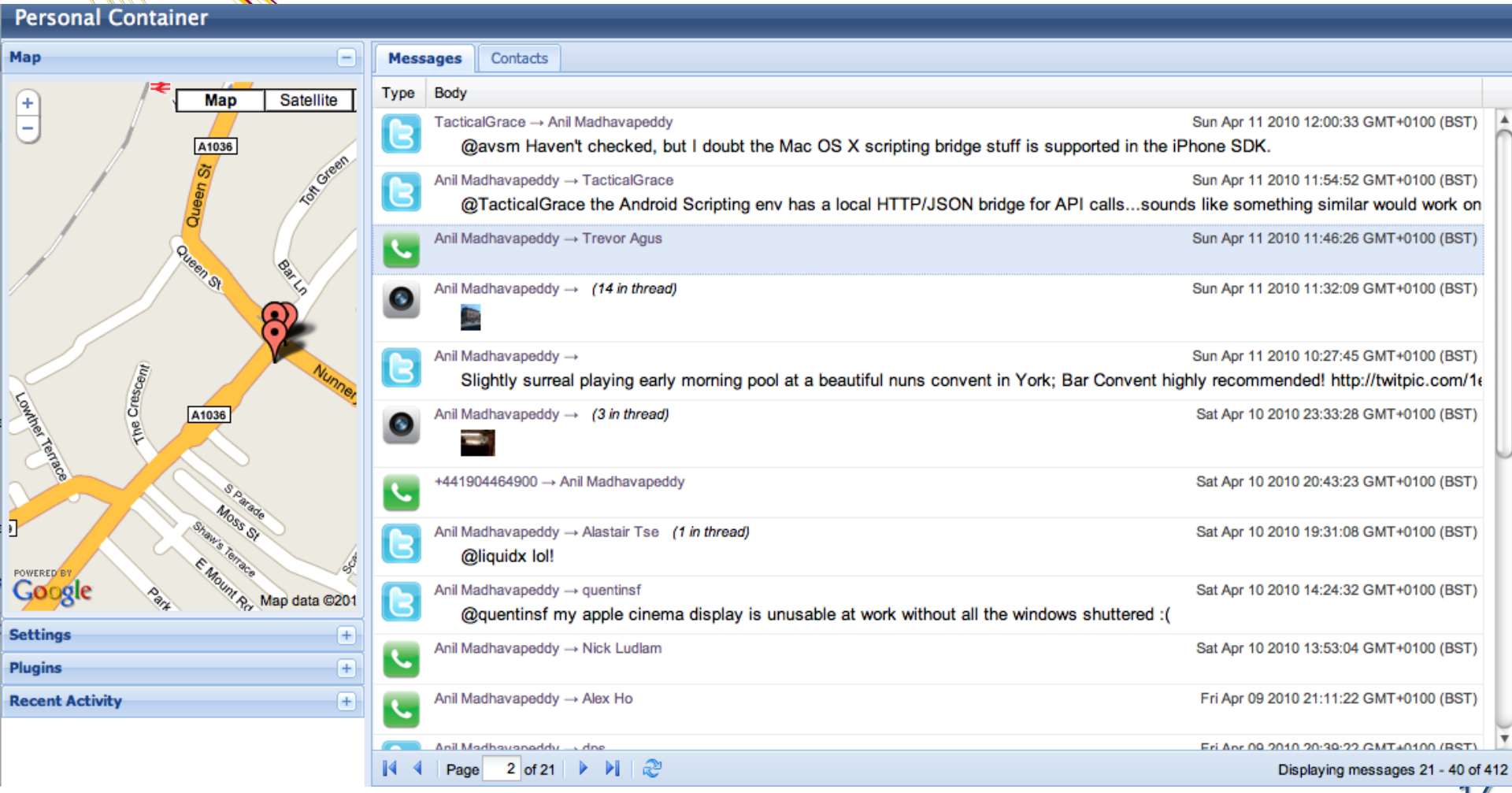
1) Nodes: multi-scale portability



1) Nodes: clean slate protocols



1) Node: screen shot

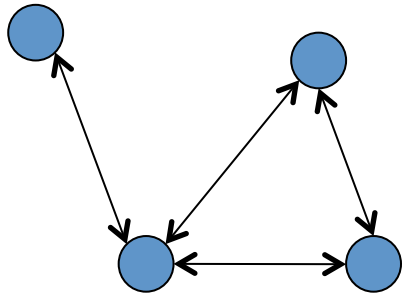


1) Nodes: formal methods

- 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

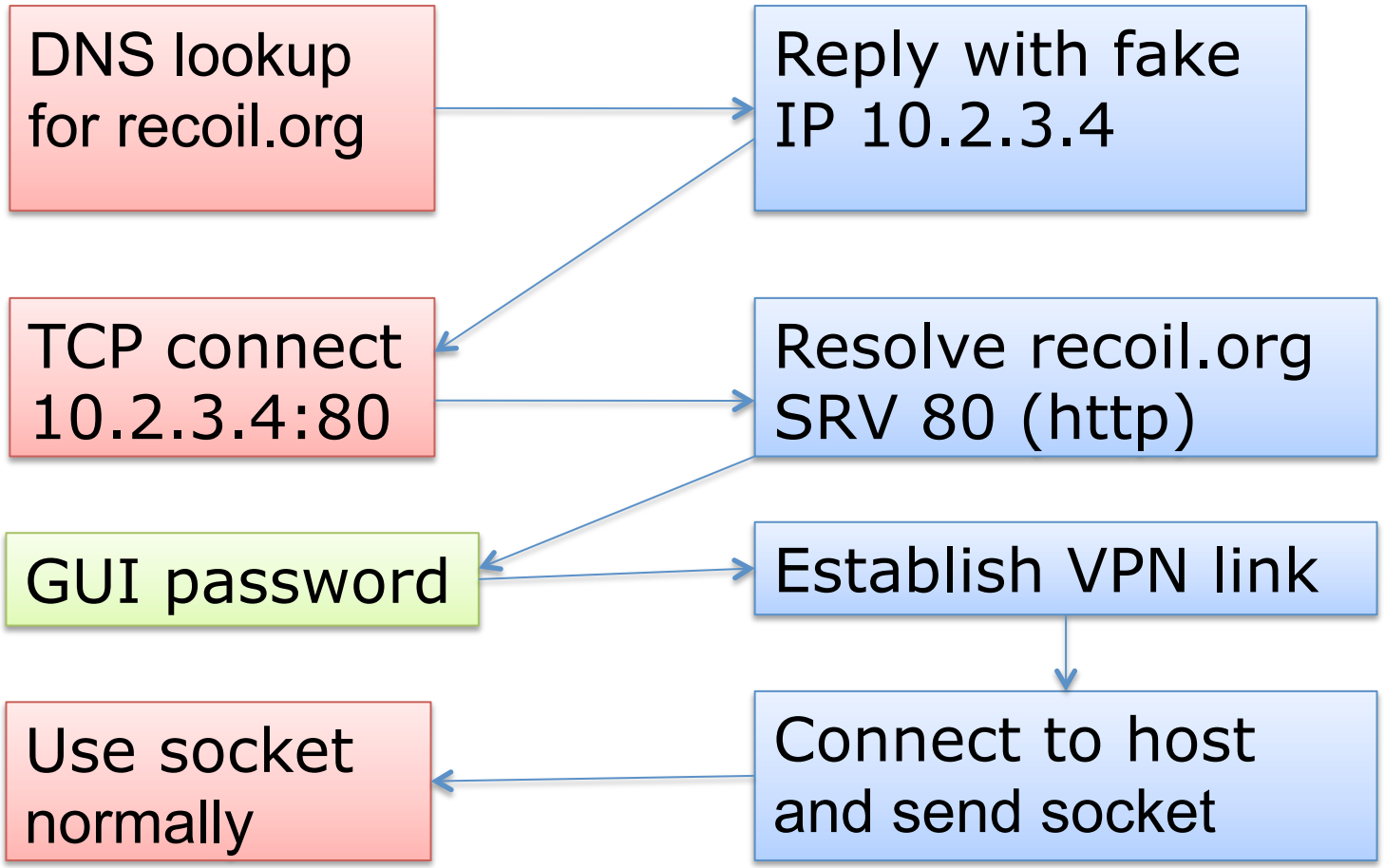


- Need to connect to each of these seamlessly across different connectivity types
- Host support for **route setup** is very inflexible.
  - SSH tunnels or IPSec VPNs
  - Kerberos, Active Directory
  - HTTP proxys, website auth

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/

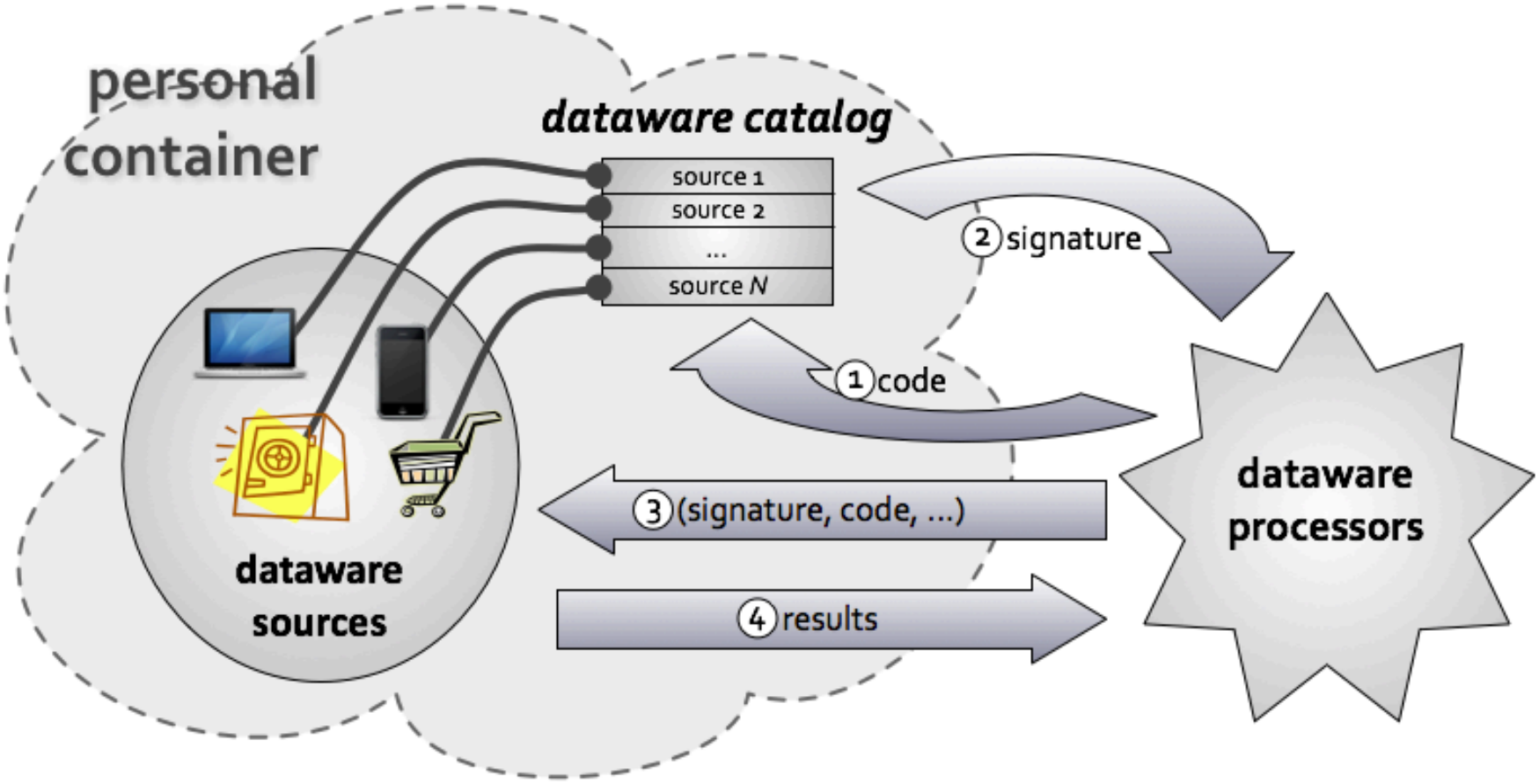
2) Naming: DNS cookies (WIP)



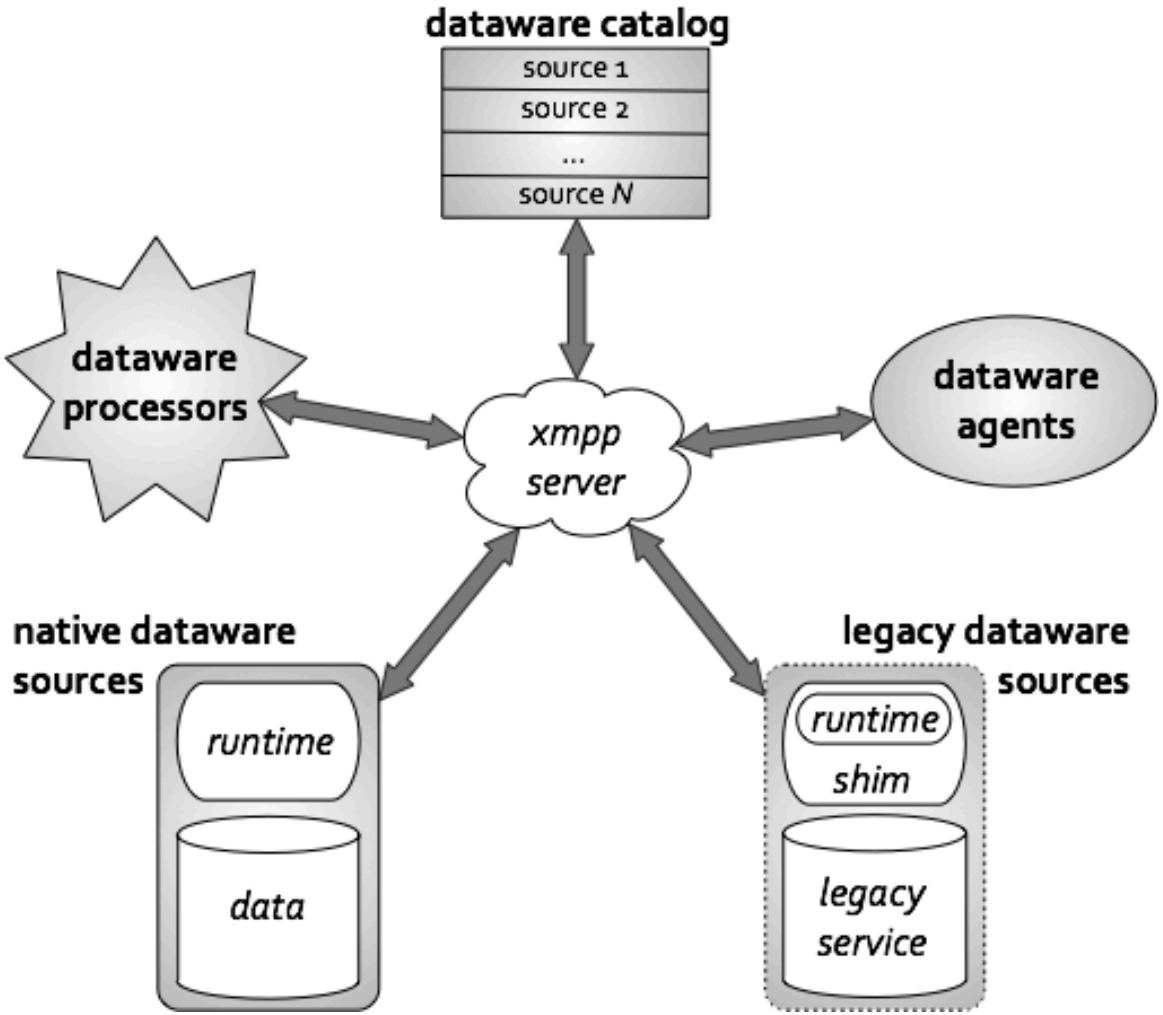
2) Naming: discussion

- **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.

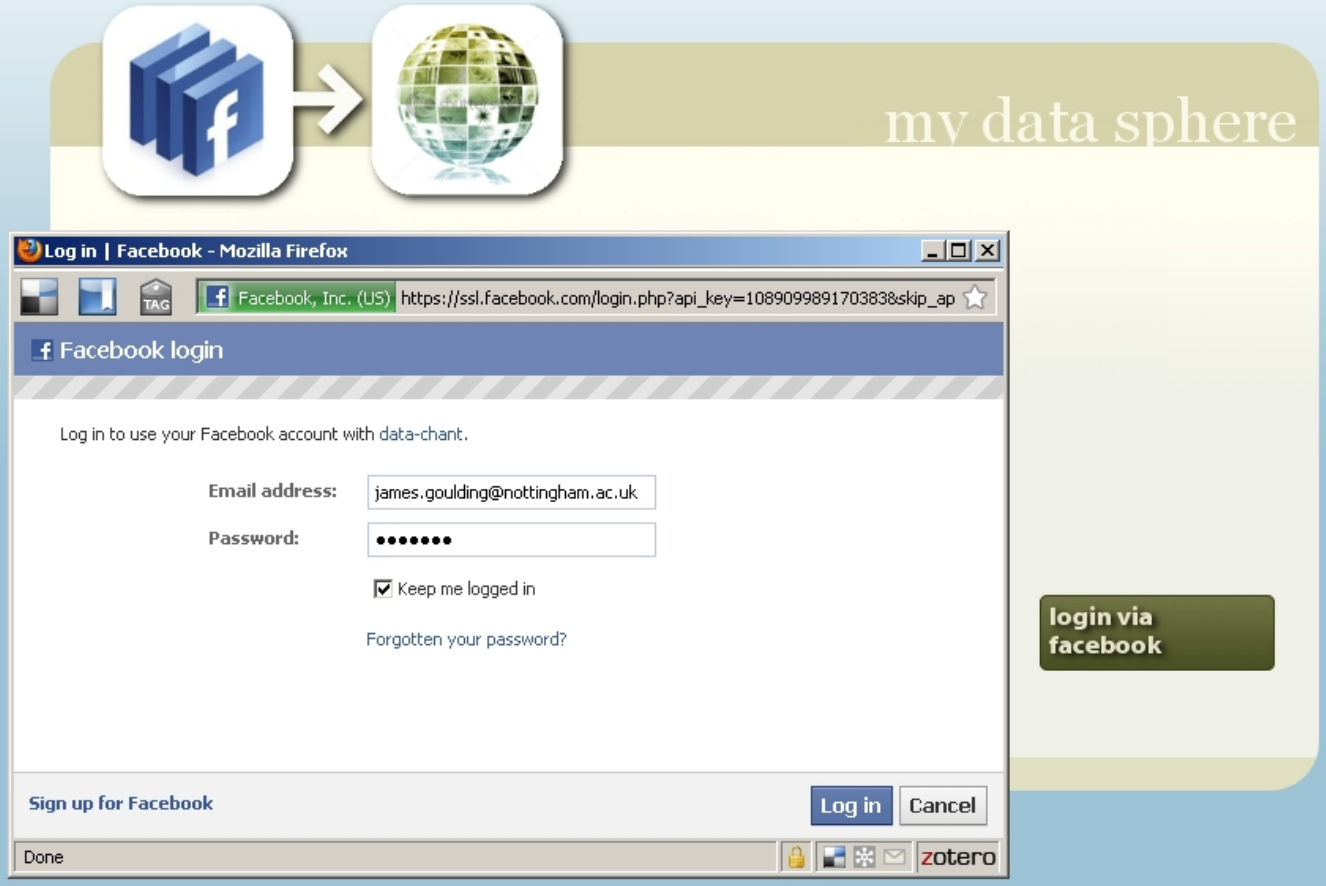
3) Computation: registration



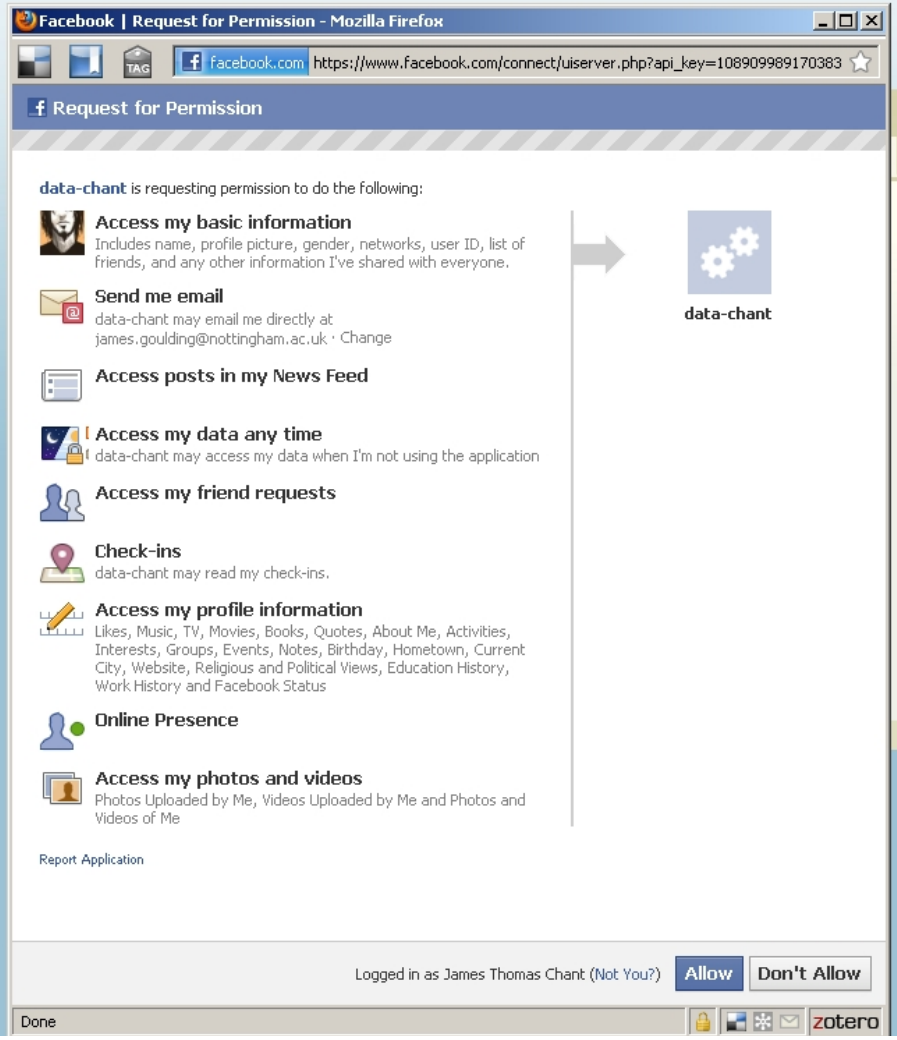
3) Computation: routing requests

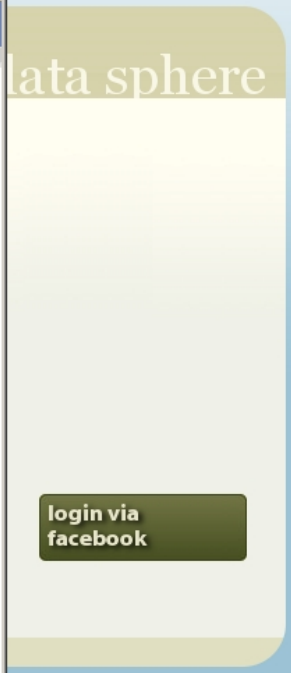


Prototypes: Shim Login

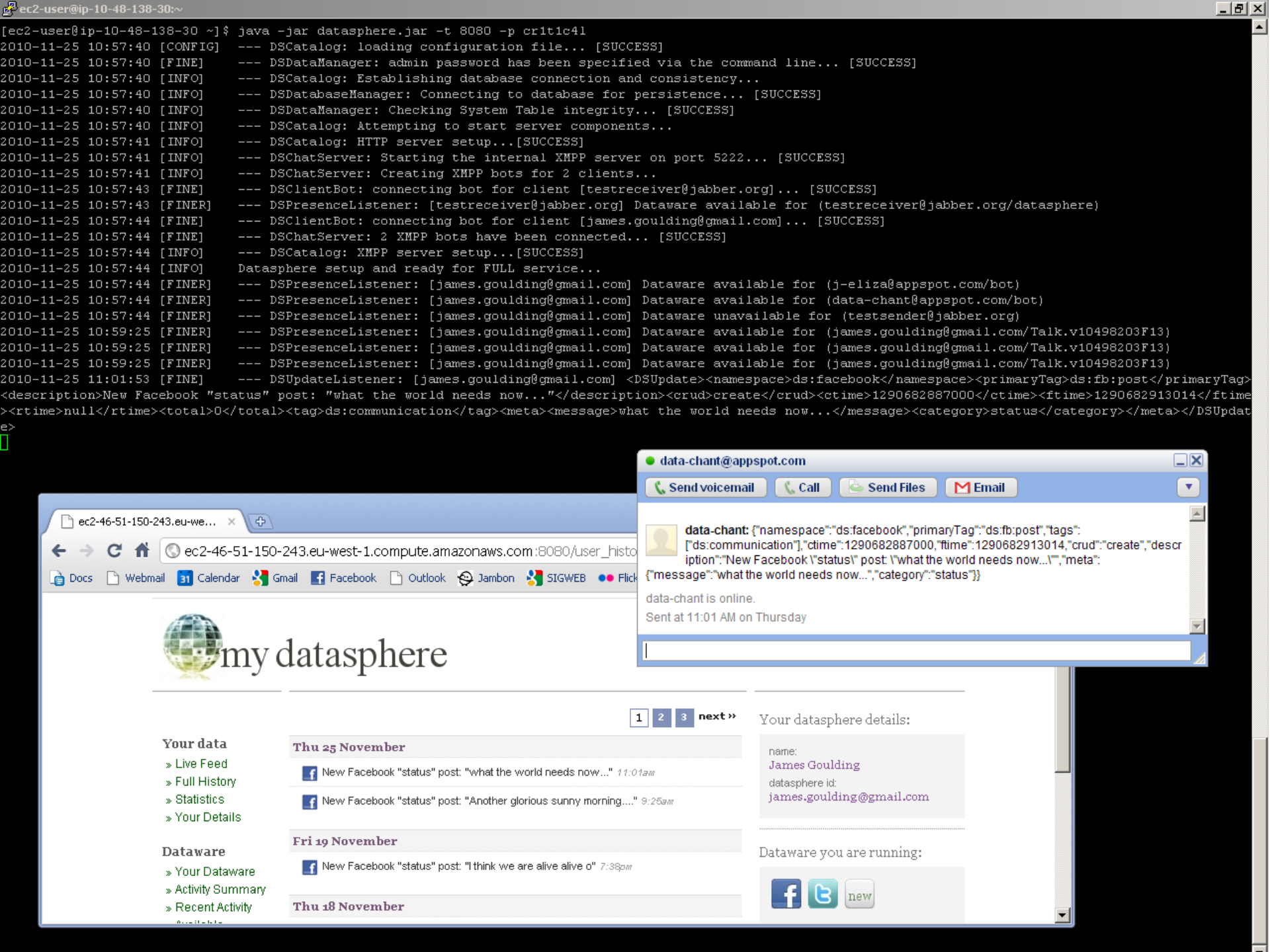


Prototypes: Shim Permission





Prototypes: Web Interface


ec2-user@ip-10-48-138-30:~
[ec2-user@ip-10-48-138-30 ~]$ java -jar datasphere.jar -t 8080 -p cr1t1c41
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></DSUpdat
e>
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","descr
iption":"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
ec2-46-51-150-243.eu-we...
ec2-46-51-150-243.eu-west-1.compute.amazonaws.com:8080/user_histo
Docs
Webmail
Calendar
Gmail
Facebook
Outlook
Jambon
SIGWEB
my datasphere
1
2
3
next »
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
Your datasphere details:
name:
James Goulding
datasphere id:
james.goulding@gmail.com
Dataware you are running:
new

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!)