

Contrail: Science in the Cloud Challenges & Opportunities

Jens Jensen <jens.jensen * stfc.ac.uk>
Science and Technology Facilities Council (STFC)
Rutherford Appleton Laboratory (RAL)



contrail is co-funded by the EC
7th Framework Programme

Contrail project in 5 lines or less

- Federation front end to cloud resources
- Single account (and accounting)
- SLAs for resource selection
- IaaS, PaaS (MapReduce, Bag-of-Tasks, DBs), global filesystem (GAFS, based on XtremFS)
- <http://contrail-project.eu/>

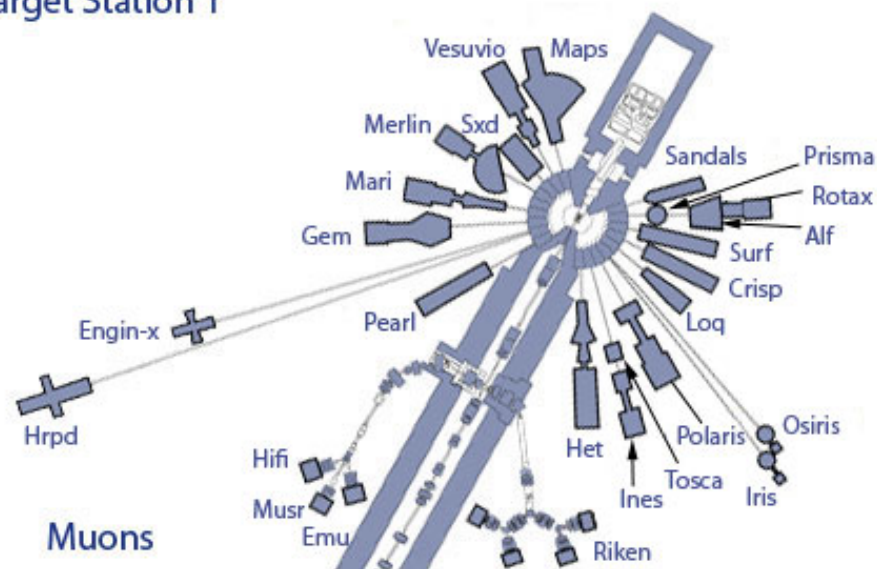
Scientific Application Areas

- Neutron Source: STFC's ISIS
 - <http://www.isis.rl.ac.uk/>
- Earth Observation Data
 - <http://proj.badc.rl.ac.uk/exarch> (CMIP5)
- Genomics
 - <http://www.constellationtechnologies.com/>

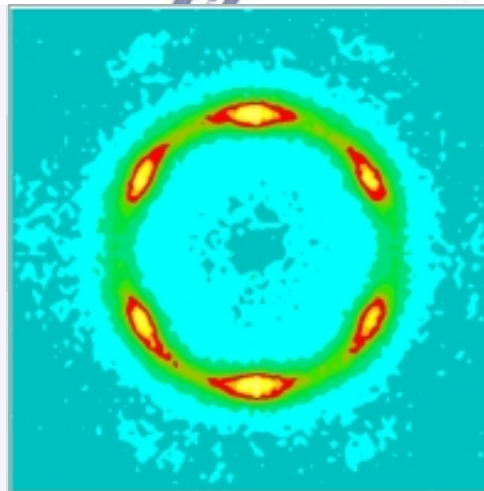
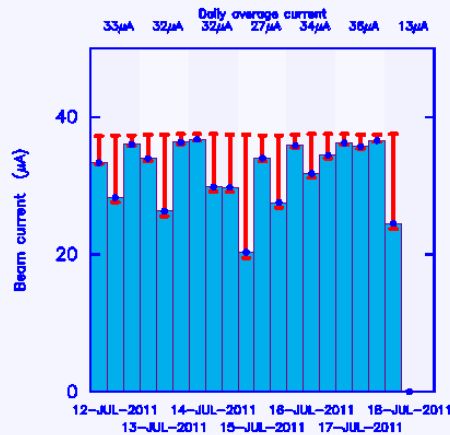
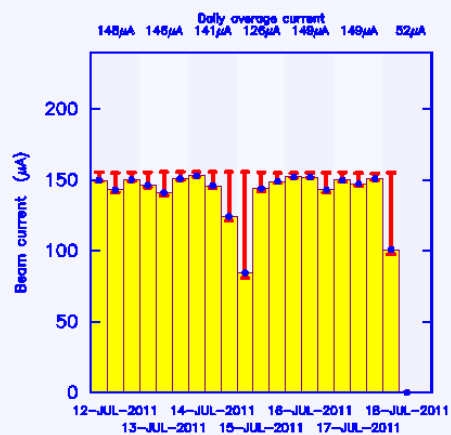
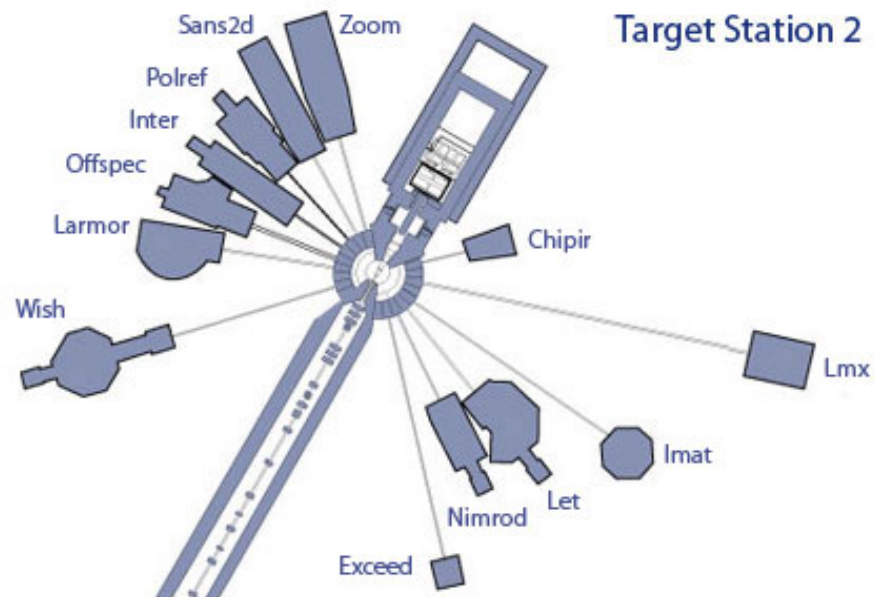
(there are other non-science applications:
geospatial referencing, media streaming)

Neutron Source – SANS

Target Station 1



Target Station 2



Neutron Source

- Small angle neutron scattering
- Locate “interesting” parameters (10^{50} parameter space)
- Perform experiment
- Feed back into model – χ^2 -fit

Model: **Ellipsoid** (change)

<input type="checkbox"/> scale	<input type="text" value="1"/>	
<input checked="" type="checkbox"/> radius_b [Å]	<input type="text" value="48.7"/>	± 0.27
<input checked="" type="checkbox"/> radius_a [Å]	<input type="text" value="8.02"/>	± 0.057
<input checked="" type="checkbox"/> background [1/cm]	<input type="text" value="0.0176"/>	± 0.044
<input type="checkbox"/> contrast [1/Å ²]	<input type="text" value="3.2e-06"/>	

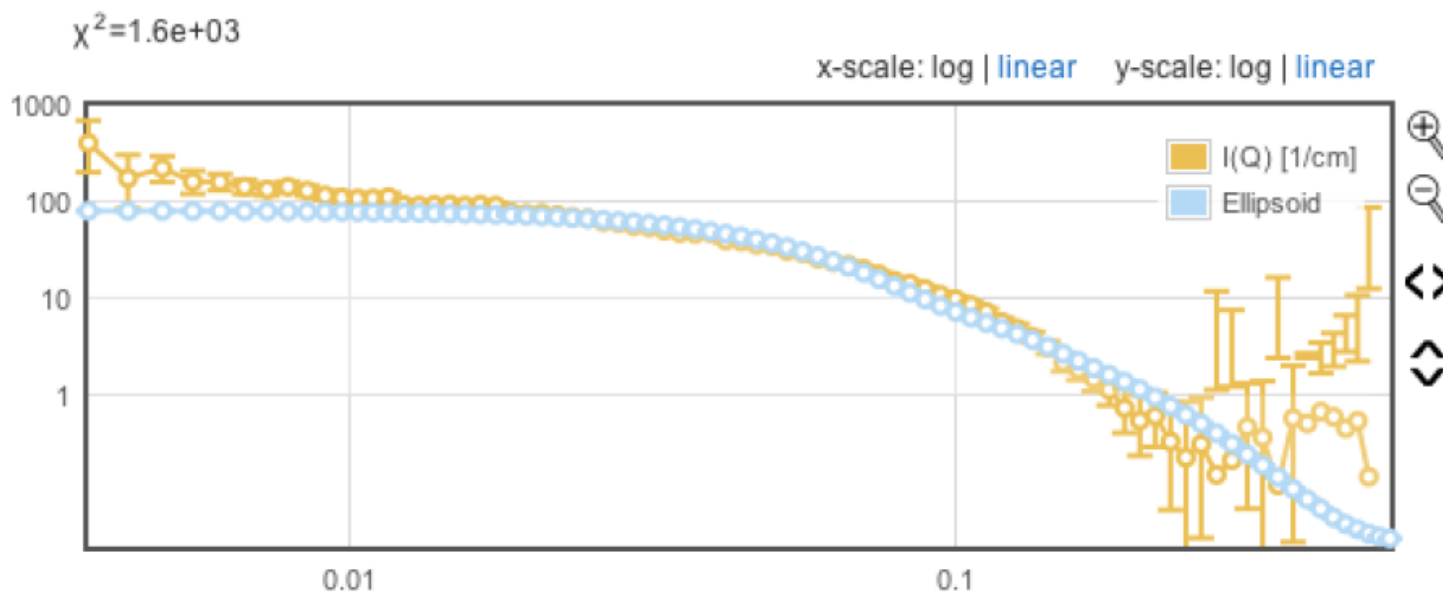
Current analysis setup

Specify the Q range to be used in the fit [optional]:

Q min [1/Å] Q max [1/Å]

Specify your smearing parameters [optional]:

None Data default Point smearing Slit smearing

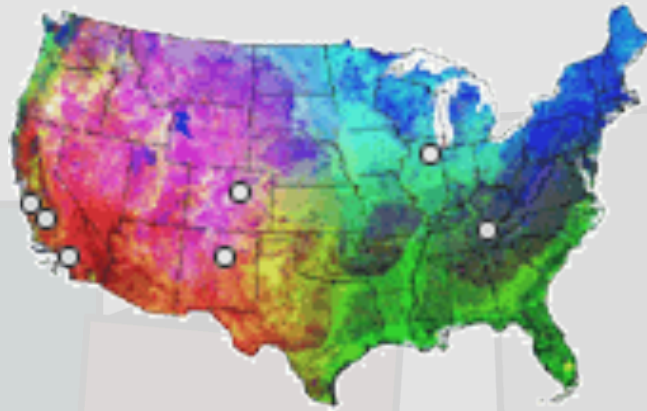


ExArch – “Exascale climate model data”

- BADC (STFC, (NERC), UK), IPSL (FR), DKRZ (DE)
- CMIP5: Coupled Model Intercomparison Project, Phase 5 (<http://cmip-pcmdi.llnl.gov/cmip5/>)
- Input to Fifth Intergovernmental Panel on Climate Change Assessment Report

Climate Data Processing

- OGC Web Processing Service (WPS): async jobs based on Climate Data Operators (CDO)
- Pull data from OPeNDAP services (www.opendap.org)
- ESGF federated access (OpenID+PKI, SAML)



Genomics

- Challenge – NGS
 - Target customers: life sciences, academia
 - Resources not available to most companies
- Services front end based on Ensembl and PlasMapper (www.ensembl.org, <http://wishart.biology.ualberta.ca/PlasMapper/>)
- Open genome data, confidential processing

Genomics

- Genome/Transcriptome: in-silico experiments to link genotype to phenotype
 - Assembly, alignment, SNP detection,
 - Access to commercial tools?
- Need Azure? (SAS70 II certified cloud data centre)
- WS-Federation?

Genomics

- Underlying tools: BLAST, BLAT
- Sequence annotations databases

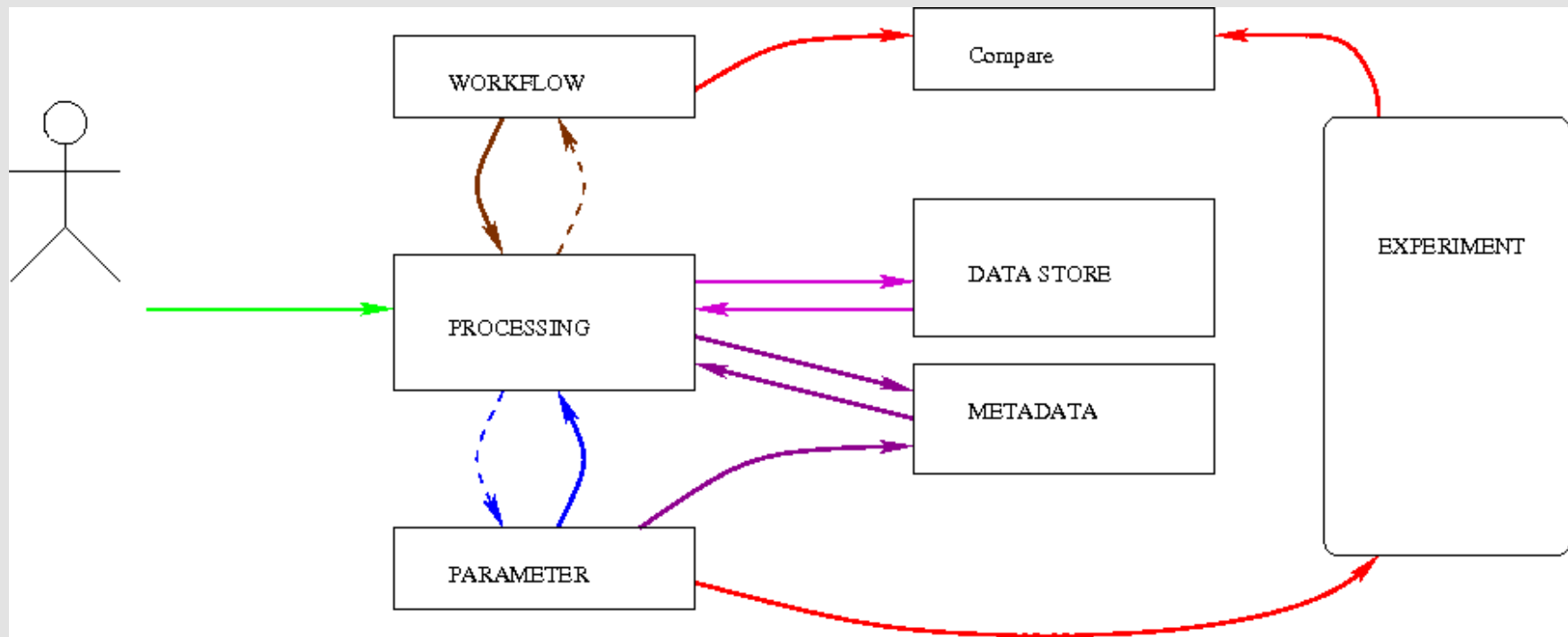
Why use *clouds* for science?

- The conference case
- Responsiveness: more resources as needed
- PaaS: code to higher level services
 - Improve reuse, quicker dev't
- Service provider: increase small-user occupancy
 - Compare the grid model

Why use *Contrail* for science?

- Federated: single account, single sign-on
- Collaboration attributes
- Use of existing A(A(A))I
- Easier cloudbursting
- Single, shared, accessible storage: GAFS
- SLA-based service selection, including QoP
- Protected services for *omics

Generic(ish) science processing



Issues

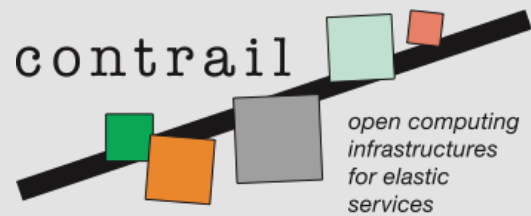
- Coding to PaaS
 - Suitability of PaaS (vs building VM imgs)
 - Adapting existing code/building new code
 - Testing environment
- Less (coding, support) effort for more science
- Who does the coding (workflows)?

Issues

- Workflow – at which level?
- Data! (Interfacing to)
- User experience: non-techie + techie
- Managing risks and user expectations (as in any project)
- Cost effectiveness of service provision/
consumption

Standards (bodies) – SDOs

- Role of OGF
 - Federations (→DCI), accounting (→UR)
 - Security (→cloudsec)
 - OCCI
 - Interoperation?
- OGC (WPS), SNIA (CDMI?), DMTF (OVF)
- OASIS, W3C, IETF



contrail is co-funded by the
EC 7th Framework Programme

Funded under: FP7 (Seventh Framework Programme)

Area: Internet of Services, Software & virtualization
(ICT-2009.1.2)

Project reference: 257438

Total cost: 11,29 million euro

EU contribution: 8,3 million euro

Execution: From 2010-10-01 till 2013-09-30

Duration: 36 months

Contract type: Collaborative project (generic)