

Module 4: Taking GATE to the Cloud

Part II: GATE Cloud



Text Mining Life Cycle



Define Task

Source Data Process Data **Exploit** Results

Teamware

Web Crawling Corporate Archive Academic Papers Patents GATE Developer GATE Embedded

GATECloud.net

Reports Data mining

...

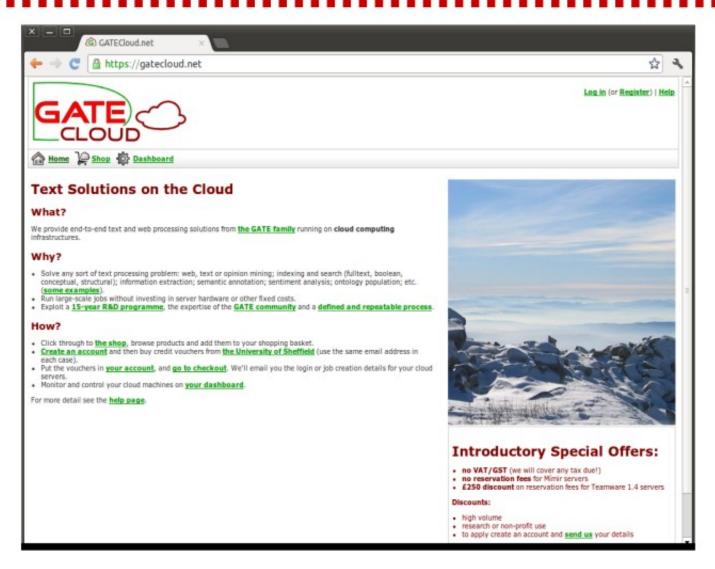
GATECloud.net



- A cloud based service from the GATE team
- Usual cloud benefits:
 - Pay-as-you-go, no upfront hardware costs
 - No sysadmin work
 - Web-based management tools
 - Always latest version, maintained by us
- Not-so-usual cloud benefits
 - Based on open-source software
 - Bring your own pipeline
 - No API required (all interaction is browser based)

https://gatecloud.net





Current Offering



- On demand document processing (a.k.a. Annotation Jobs)
 - Massive parallelism
 - On-line job definition tool
 - Many output formats, including Mimir
- On demand servers
 - Teamware
 - Mímir

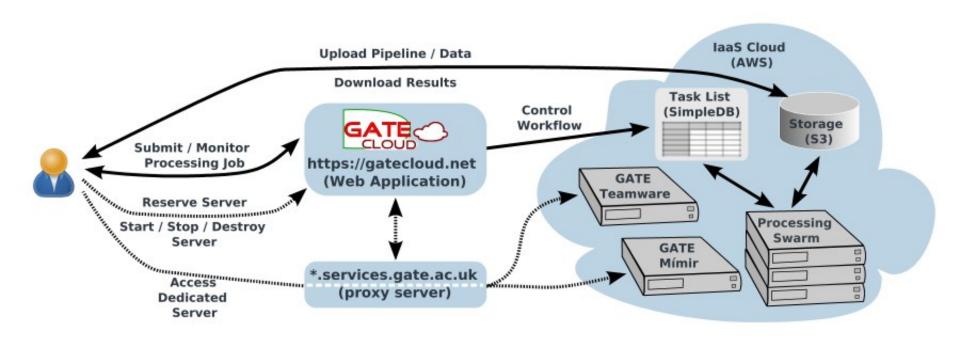
End-to-End Solution

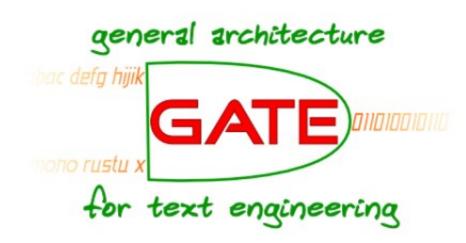


- Bring your own documents
- 2.Bring your own pipeline (or use one of ours)
- Automatically annotate the documents on a massively parallel infrastructure
 - Talk to us if you have special requirements
- 4.Use Teamware for pipeline fine-tuning and quality assurance
- 5.Use Mímir to power multi-paradigm search

System Architecture and Work-flow







Annotation Jobs

Large scale document annotation

Annotation Jobs



- Parallel and distributed annotation of documents using GATE applications.
- Upload your own document archives (.zip, .tar.gz, .tar.bz2, .arc)
- Upload your own GATE pipeline, ...or use a pre-packaged one
- Download the outputs as .zip files,
 ...or send the annotated documents directly to Mimir for indexing

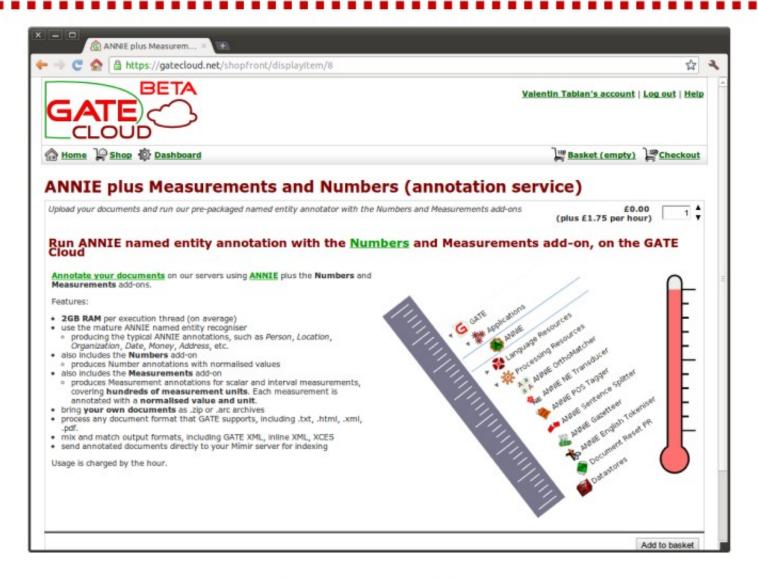
Annotation Jobs: Execution Environment



- Ubuntu Linux
- 64 bit
- SUN (Oracle) Java 1.6.0
- 2 GB RAM / thread (on average)
- GATE Embedded version 7.1
- GATE Cloud Paralleliser version 2.3
- Secure sandbox

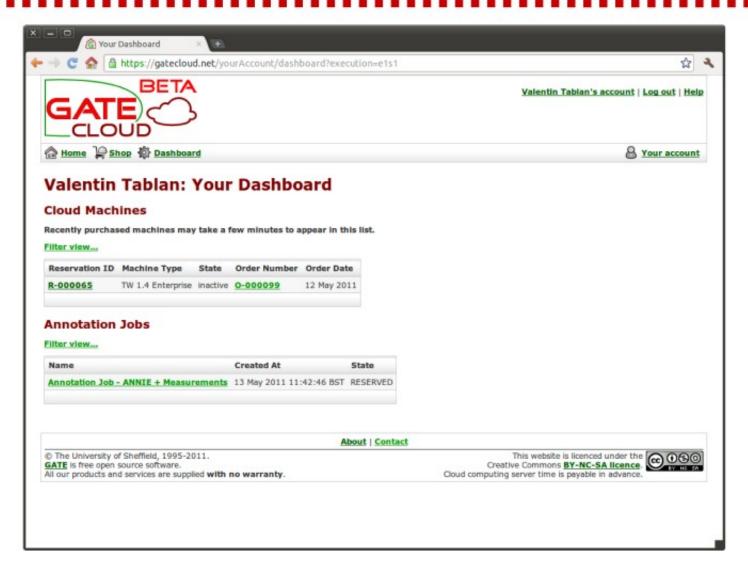
Pick, Checkout, Buy





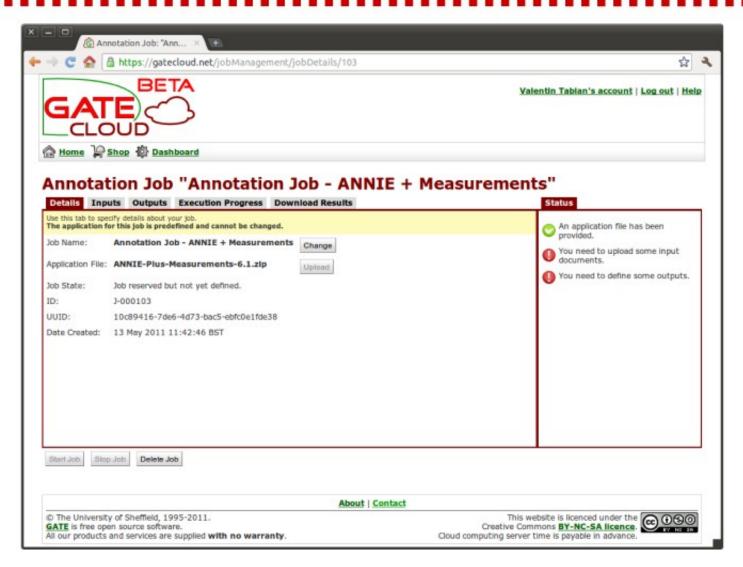
Your Dashboard





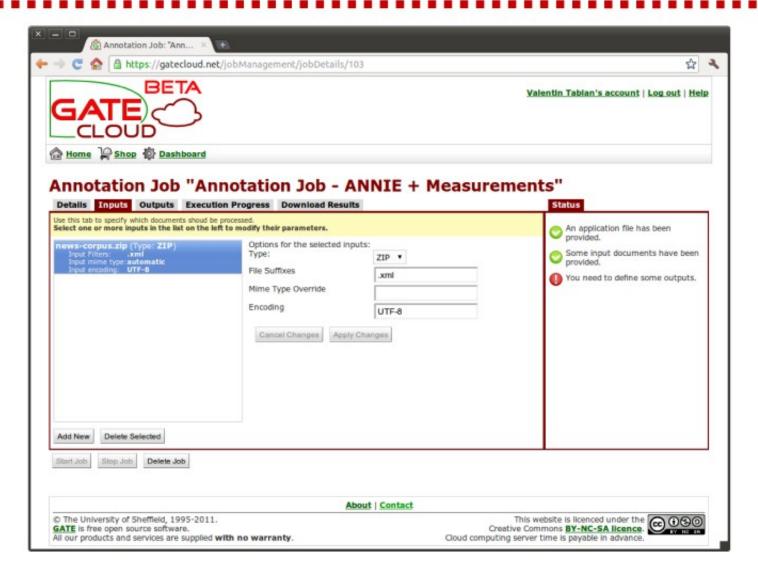
Job Editor: Details





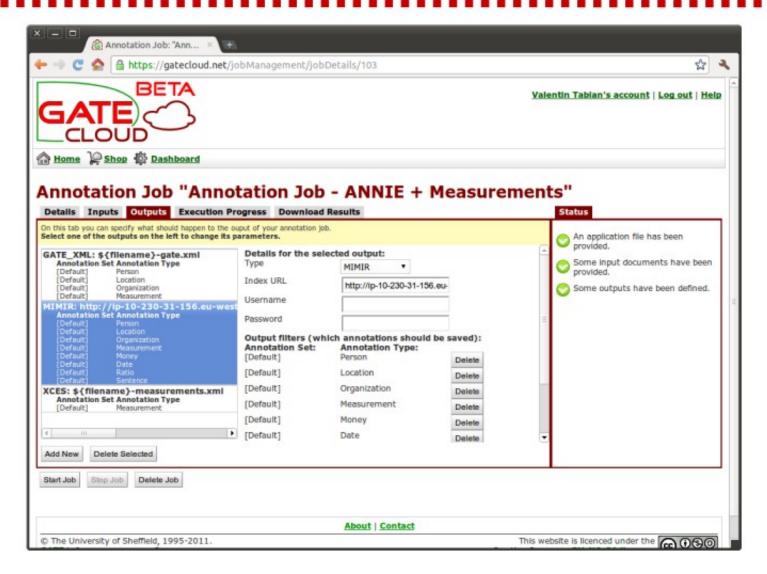
Job Editor: Inputs





Job Editor: Outputs





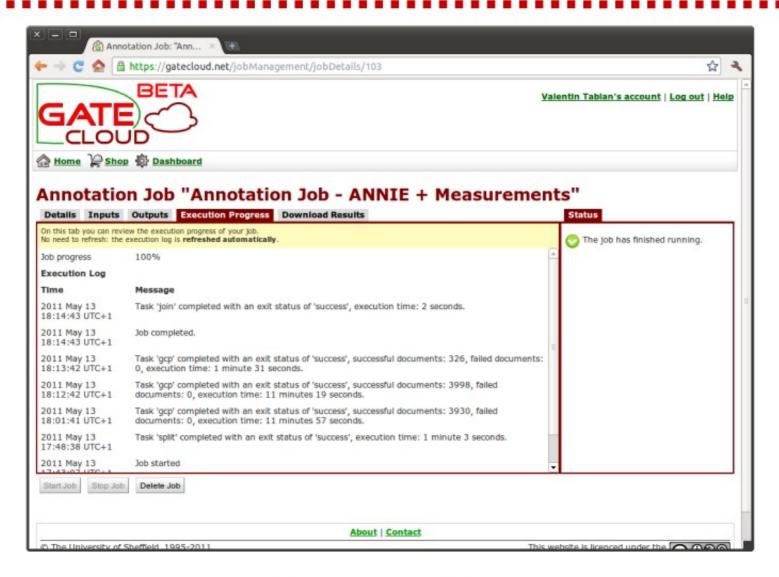
Job Editor: Progress





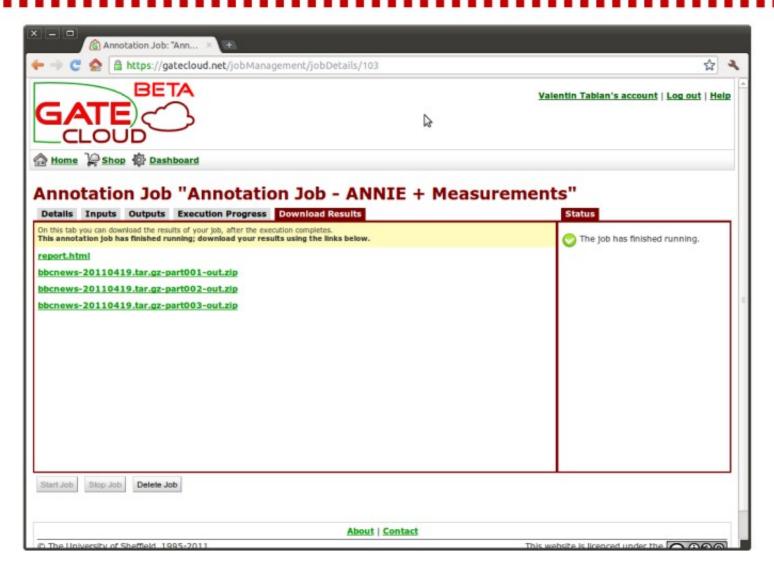
Job Editor: Done!





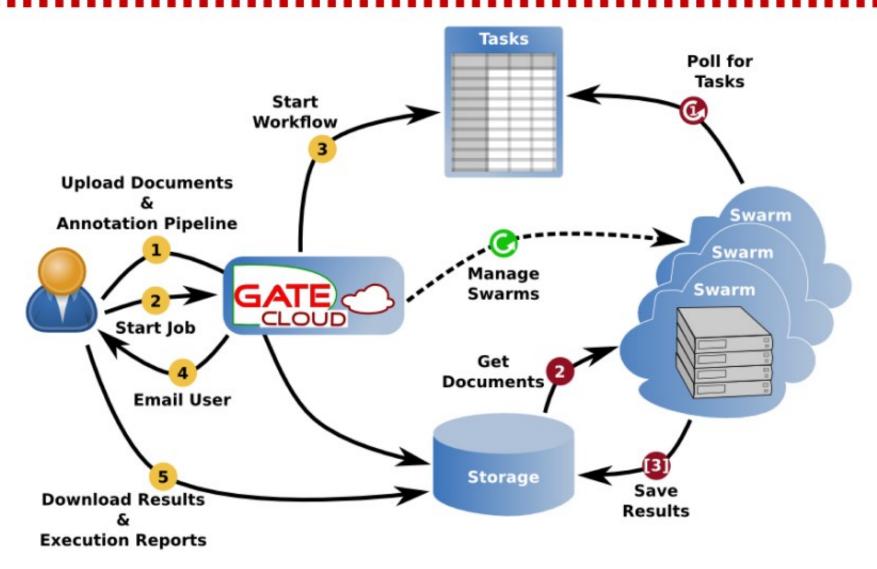
Job Editor: Results





Running an Annotation Job





Hands on: Create your own account



- Navigate to http://gatecloud.net
- Create a new account
 - Make sure to use a valid email address that you can access!
- Use the voucher you received to add some funds to the account.
- Visit your account page and check your balance.

Hands on: Run the ANNIE pipeline

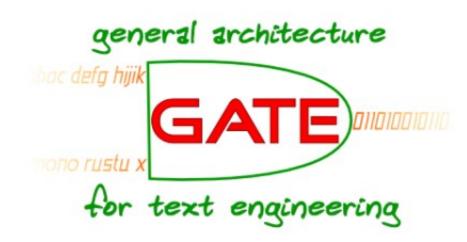


- Choose the pre-packaged ANNIE pipeline
- Use news-corpus-small.zip as input
- Output to GATE XML format
 - from the default annotation set
 - with all annotation types included
- Download the annotated documents, and inspect them using GATE Developer.
- Look at the execution reports.

Extra Hands on: Run your custom pipeline



- In GATE Developer:
 - modify the ANNIE pipeline to also annotate 'payments', e.g. {Person} ({Token}) [0,5] {Money}
 - Export the modified pipeline for GATECloud.net
- On GATECloud.net
 - Create a new custom annotation job
 - Upload news-corpus-small.zip as input
 - Upload the modified pipeline
 - Output to XCES format, from the default annotation set, including only {Person}, {Money}, {Payment}
- Download the annotated documents, and inspect them using a text editor.
- Look at the execution reports.



Severs (Machines)

Rent ready-made GATE-based virtual servers.

Reserve a Server



Create a new Machine Reservation

- A new data volume is created and reserved for you.
- A new machine appears on your dashboard.
- You pay the one-off price.
- The reservation is in the 'inactive' state.

Start an Instance



Start-up a virtual server (a.k.a. Instance) based on a reservation.

- The reserved server is started and you can use it like any other server.
- You pay the hourly price for as long as the sever is running.
- Your reservation is now 'active'

Stop a Running Instance



Stop the running instance

- The started server is shut down.
- You stop paying hourly charges.
- Your reservation is once again 'inactive'.

Create a backup



Make a copy of your data volume.

- You can only do that while the reservation is 'inactive'.
- You can restore this backup back onto your data volume whenever you want.
 - Doing so <u>overwrites</u> the current state of the data volume.

Destroy a Machine Reservation

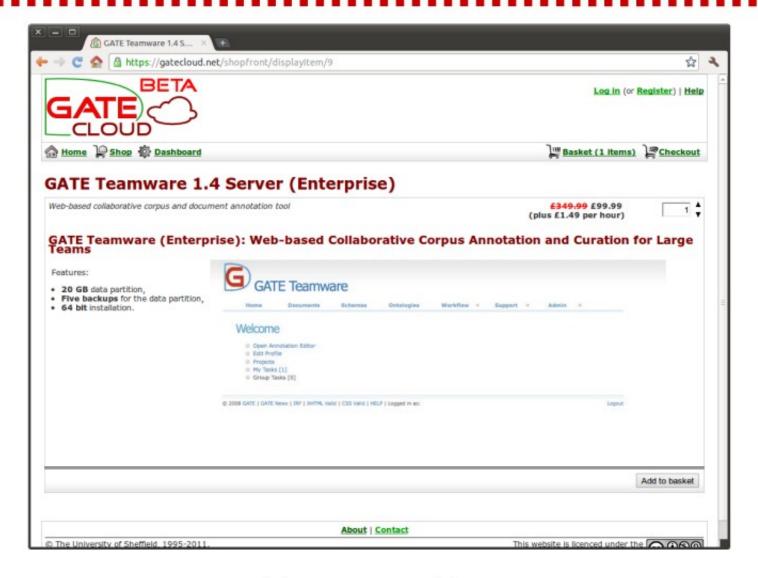


Destroy all data associated with the reservation.

- The equivalent to destroying an actual server.
- Your data volume and all the backups are deleted.
- This cannot be undone.

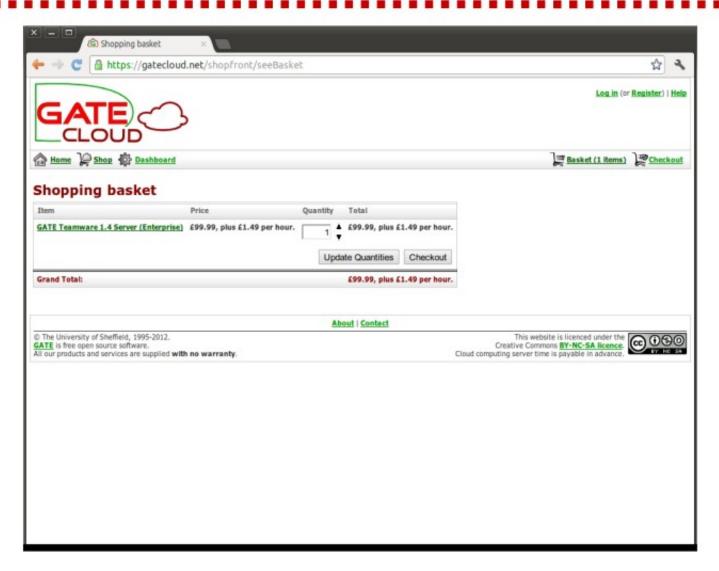
Making a Reservation: 1) Find Required Server





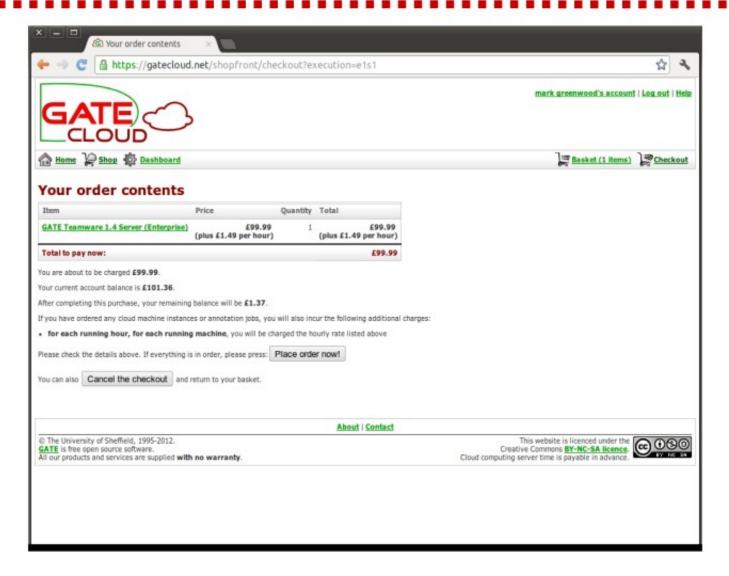
Making a Reservation: 2) Add to Basket





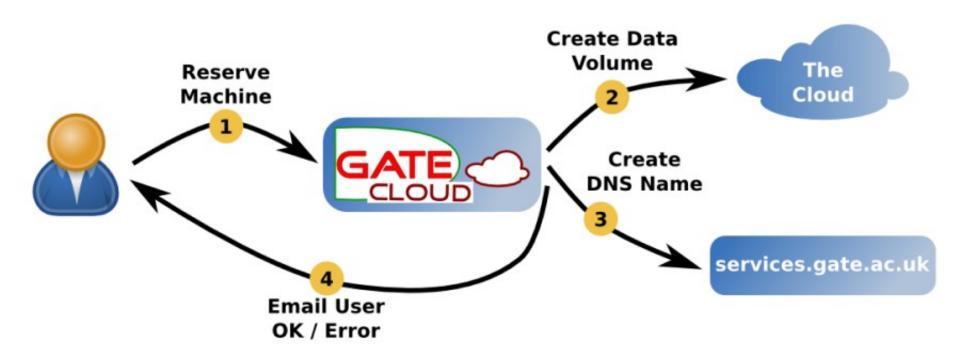
Making a Reservation: 3) Checkout





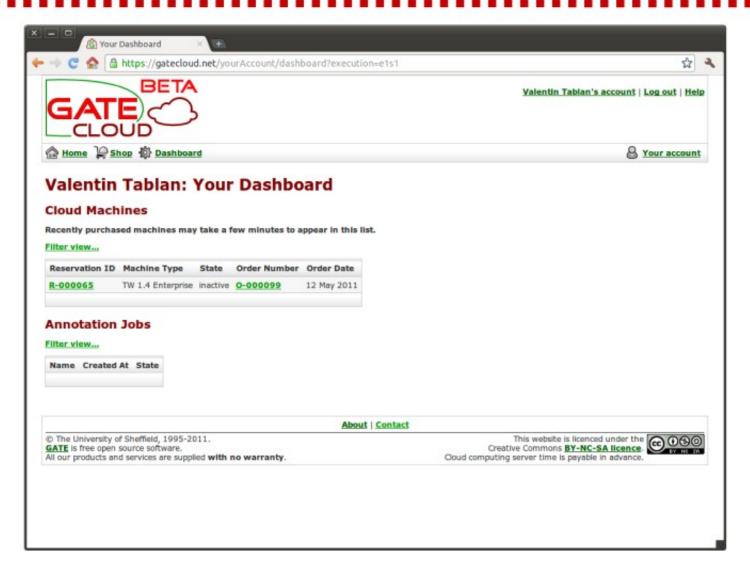
Reserving a Server





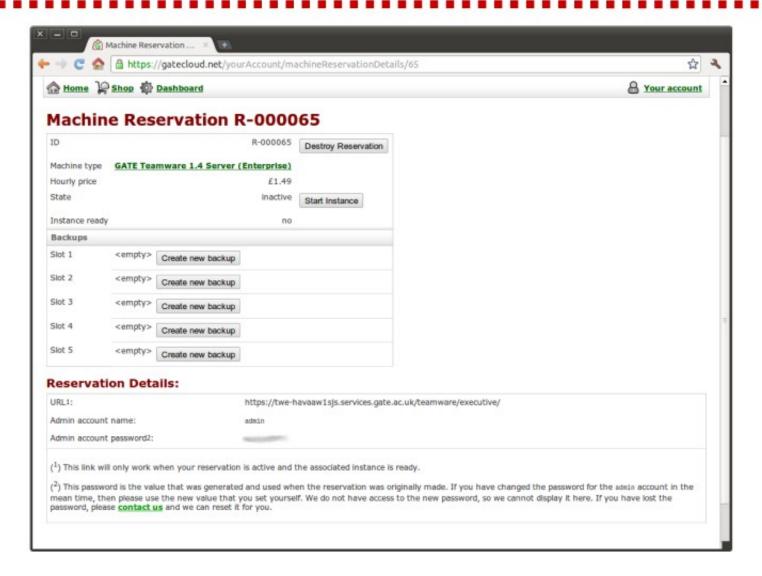
Your Dashboard





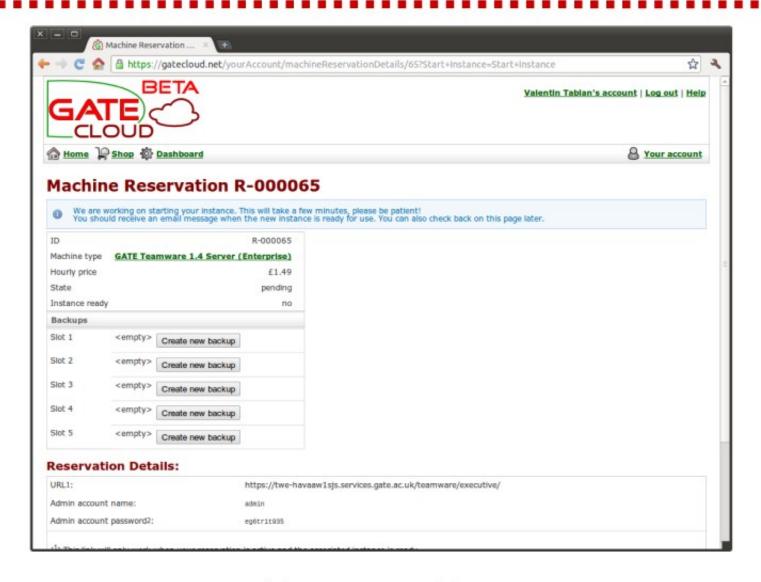
Your Machine Reservation





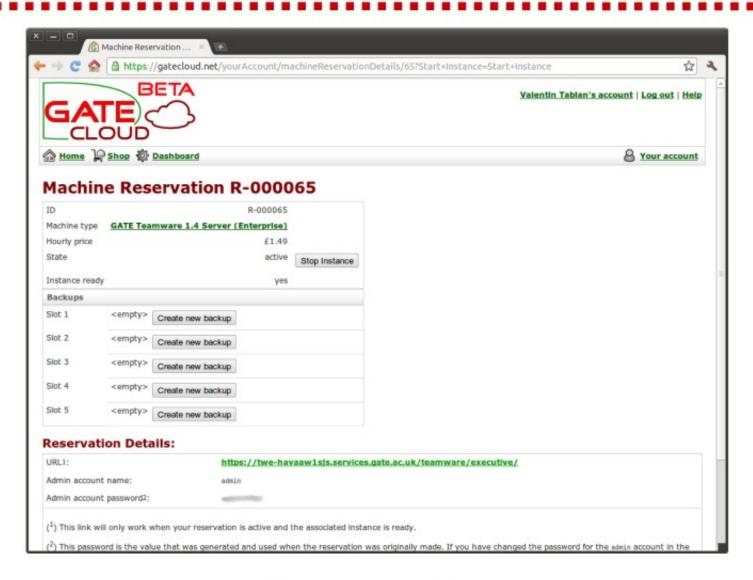
Starting a Server: 1) Press Start





2) Wait for the Email (or refresh the page)





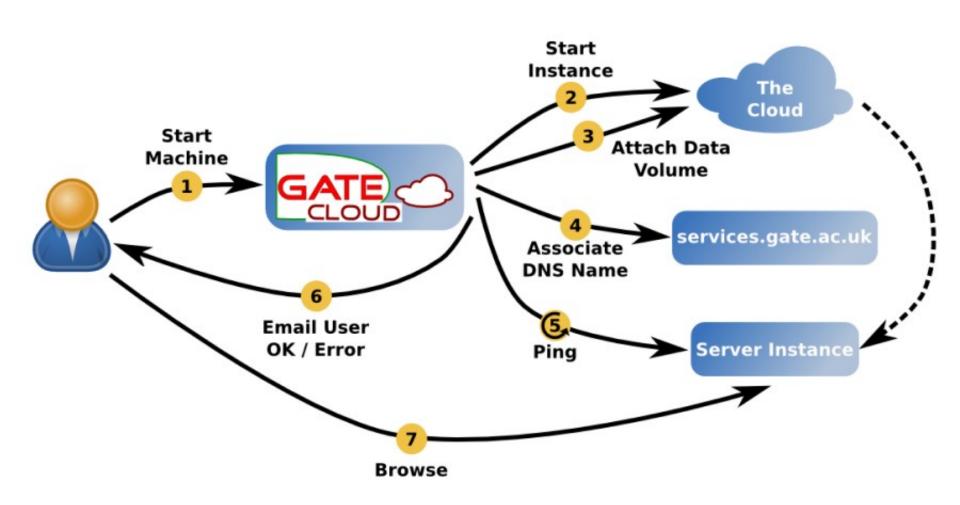
3) There is no "3": Your Server is Ready





Starting a Server





Questions?



http://gate.ac.uk http://gatecloud.net