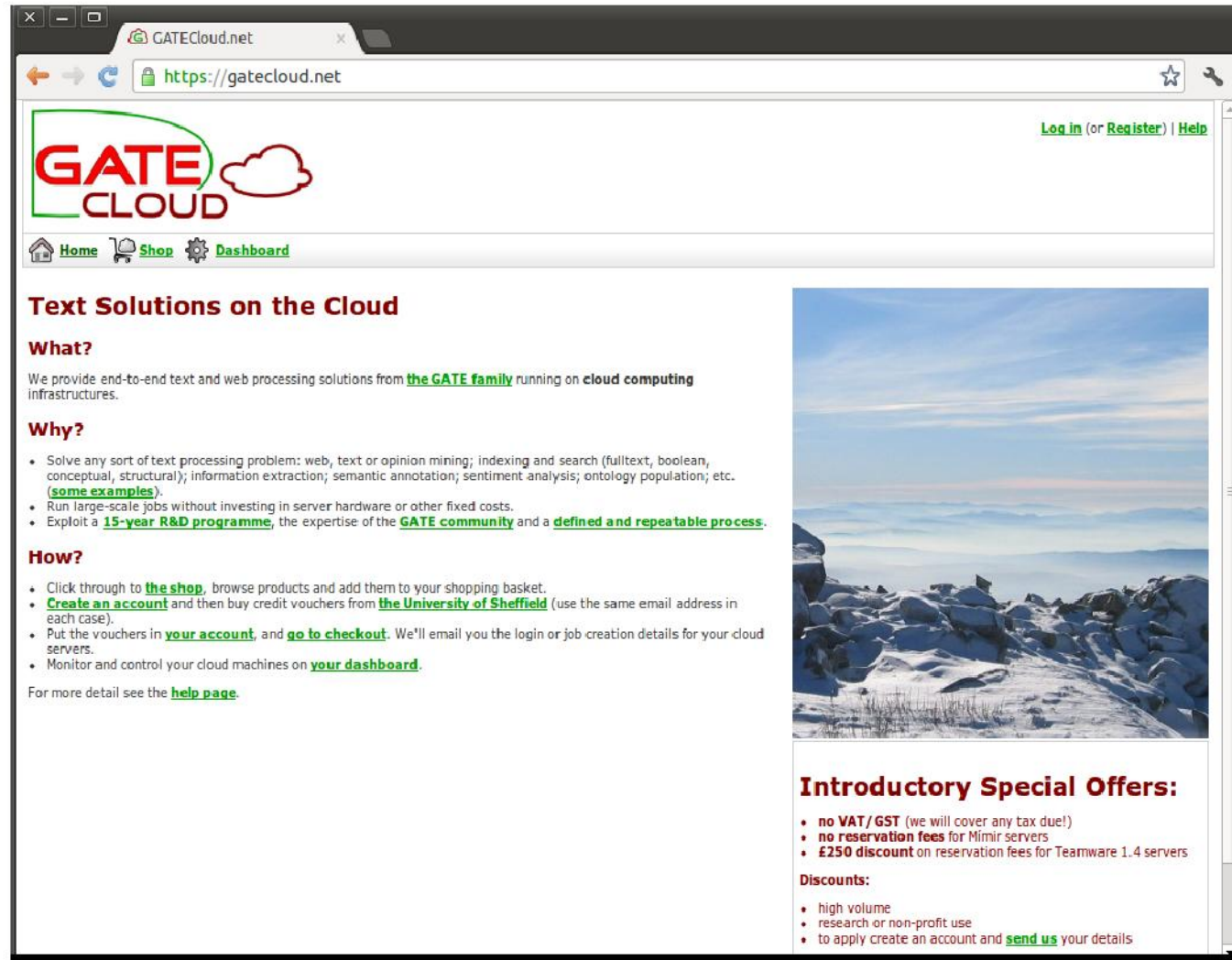


Module 4: Taking GATE to the Cloud

Part II: GATE Cloud

- 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)

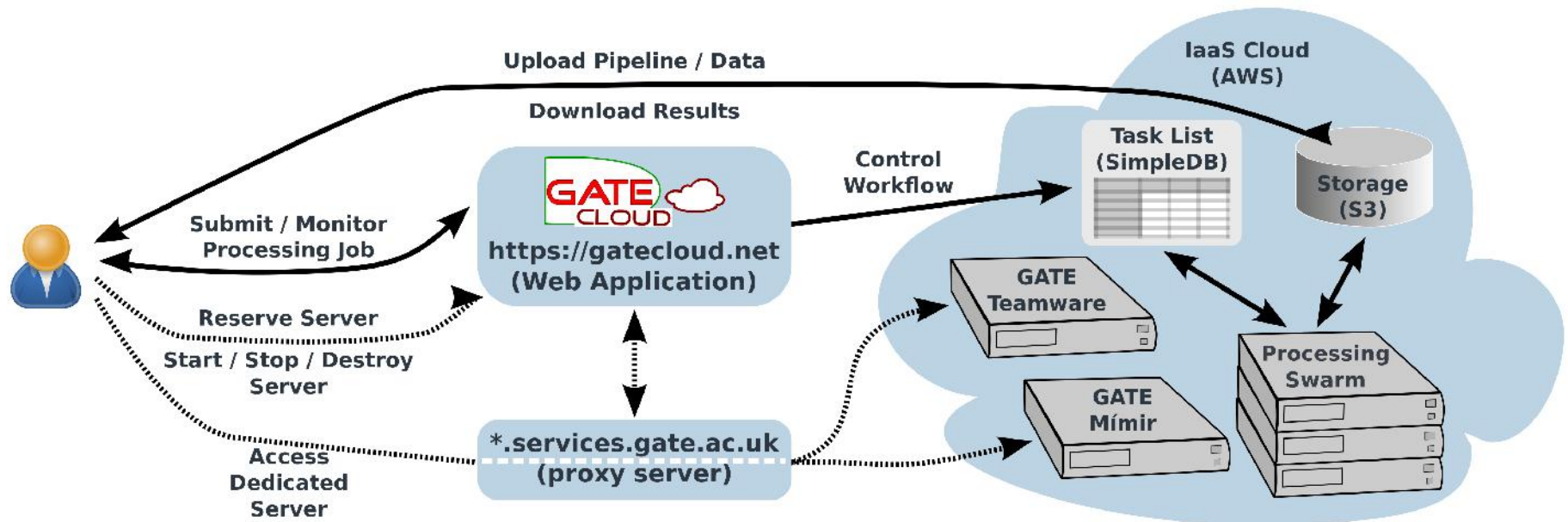
A screenshot of a web browser displaying the GATE Cloud website. The browser's address bar shows "https://gatecloud.net". The website's header features the "GATE CLOUD" logo on the left and "Log in (or Register) | Help" on the right. Below the header is a navigation menu with "Home", "Shop", and "Dashboard" links. The main content area is titled "Text Solutions on the Cloud" and includes sections for "What?", "Why?", and "How?". The "Why?" section lists several benefits of using GATE Cloud, such as solving various text processing problems and exploiting a 15-year R&D programme. The "How?" section provides a step-by-step guide on how to use the service, from browsing products to monitoring cloud machines. On the right side of the page, there is a large image of a snowy mountain landscape and a section titled "Introductory Special Offers:" which lists three key offers: no VAT/GST, no reservation fees for Mimir servers, and a £250 discount on reservation fees for Teamware 1.4 servers. Below this, there is a "Discounts:" section listing high volume, research or non-profit use, and a link to send details.

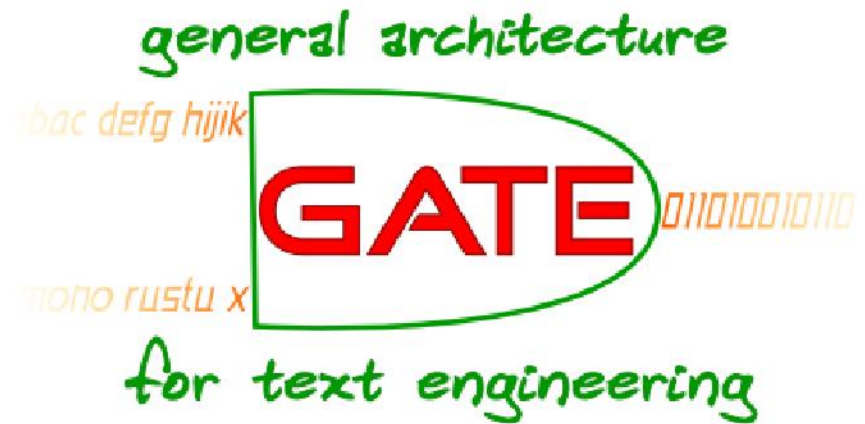
- On demand document processing (a.k.a. *Annotation Jobs*)
 - Massive parallelism
 - On-line job definition tool
 - Many output formats, including Mimir
- On demand servers
 - Teamware
 - Mimir

1. Bring your own documents
2. Bring your own pipeline (or use one of ours)
3. 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 Mimir to power multi-paradigm search

System Architecture and Work-flow

GATE





Annotation Jobs

Large scale document annotation

- 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

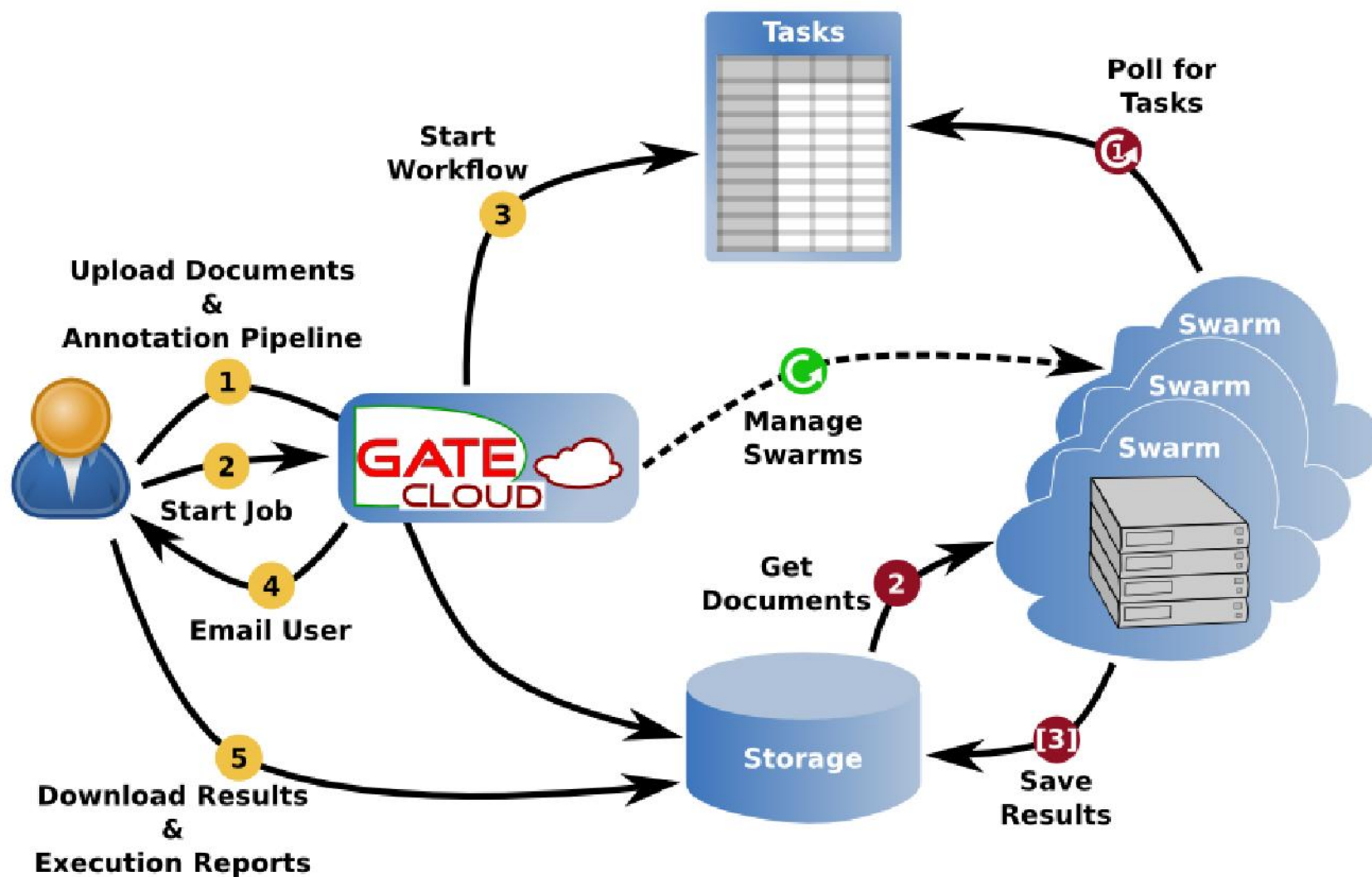
Annotation Jobs: Execution Environment



-
- Ubuntu Linux
 - 64 bit
 - SUN (Oracle) Java 1.6.0
 - 2 GB RAM / thread (on average)
 - GATE Embedded version 7.0
 - GATE Cloud Paralleliser version 2.2
 - Secure sandbox

Running an Annotation Job

GATE



Pick, Checkout, Buy

GATE

ANNNIE plus Measur... x

https://gatecloud.net/shopfront/displayItem/8

BETA
GATE
CLOUD

Valentin Tablan's account | Log out | Help

Home Shop Dashboard Basket (empty) Checkout

ANNNIE plus Measurements and Numbers (annotation service)

Upload your documents and run our pre-packaged named entity annotator with the Numbers and Measurements add-ons £0.00
(plus £1.75 per hour)

Run ANNNIE named entity annotation with the **Numbers and Measurements add-on, on the GATE Cloud**

Annotate your documents on our servers using **ANNNIE** plus the **Numbers** and **Measurements** add-ons.

Features:

- **2GB RAM** per execution thread (on average)
- use the mature ANNNIE named entity recogniser
 - producing the typical ANNNIE annotations, such as *Person, Location, Organization, Date, Money, Address, etc.*
- also includes the **Numbers** add-on
 - produces Number annotations with normalised values
- also includes the **Measurements** add-on
 - produces Measurement annotations for scalar and interval measurements, covering **hundreds of measurement units**. Each measurement is annotated with a **normalised value and unit**.
- bring **your own documents** as .zip or .arc archives
- process any document format that GATE supports, including .txt, .html, .xml, .pdf.
- mix and match output formats, including GATE XML, inline XML, XCES
- send annotated documents directly to your Mimir server for indexing

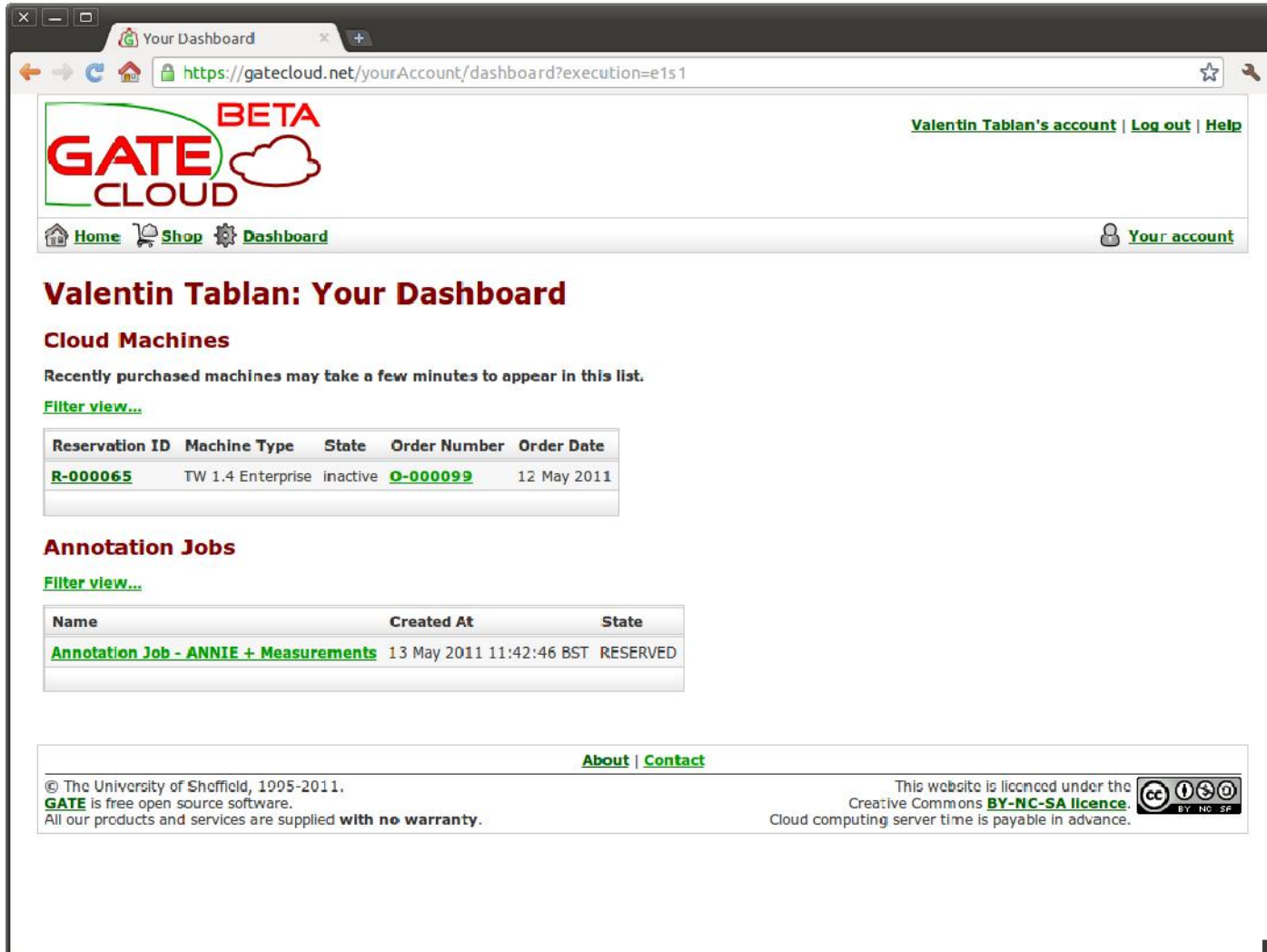
Usage is charged by the hour.

GATE Applications

- ANNNIE
- Language Resources
- Processing Resources
 - ANNNIE Orthomatcher
 - ANNNIE NE Transducer
 - ANNNIE POS Tagger
 - ANNNIE Sentence Splitter
 - ANNNIE Gazetteer
 - ANNNIE English Tokeniser
 - Document Reset PR
 - Datastores

Add to basket

Your Dashboard



The screenshot shows a web browser window with the URL <https://gatecloud.net/yourAccount/dashboard?execution=e1s1>. The page features the GATE Cloud logo with a cloud icon and the word "BETA". The user's account name, "Valentin Tablan's account", is displayed in the top right corner, along with links for "Log out" and "Help". A navigation bar includes "Home", "Shop", and "Dashboard" (the active page), and a "Your account" link with a user icon.

Valentin Tablan: Your Dashboard

Cloud Machines

Recently purchased machines may take a few minutes to appear in this list.

[Filter view...](#)

Reservation ID	Machine Type	State	Order Number	Order Date
R-000065	TW 1.4 Enterprise	inactive	O-000099	12 May 2011

Annotation Jobs


[Filter view...](#)

Name	Created At	State
Annotation Job - ANNIE + Measurements	13 May 2011 11:42:46 BST	RESERVED

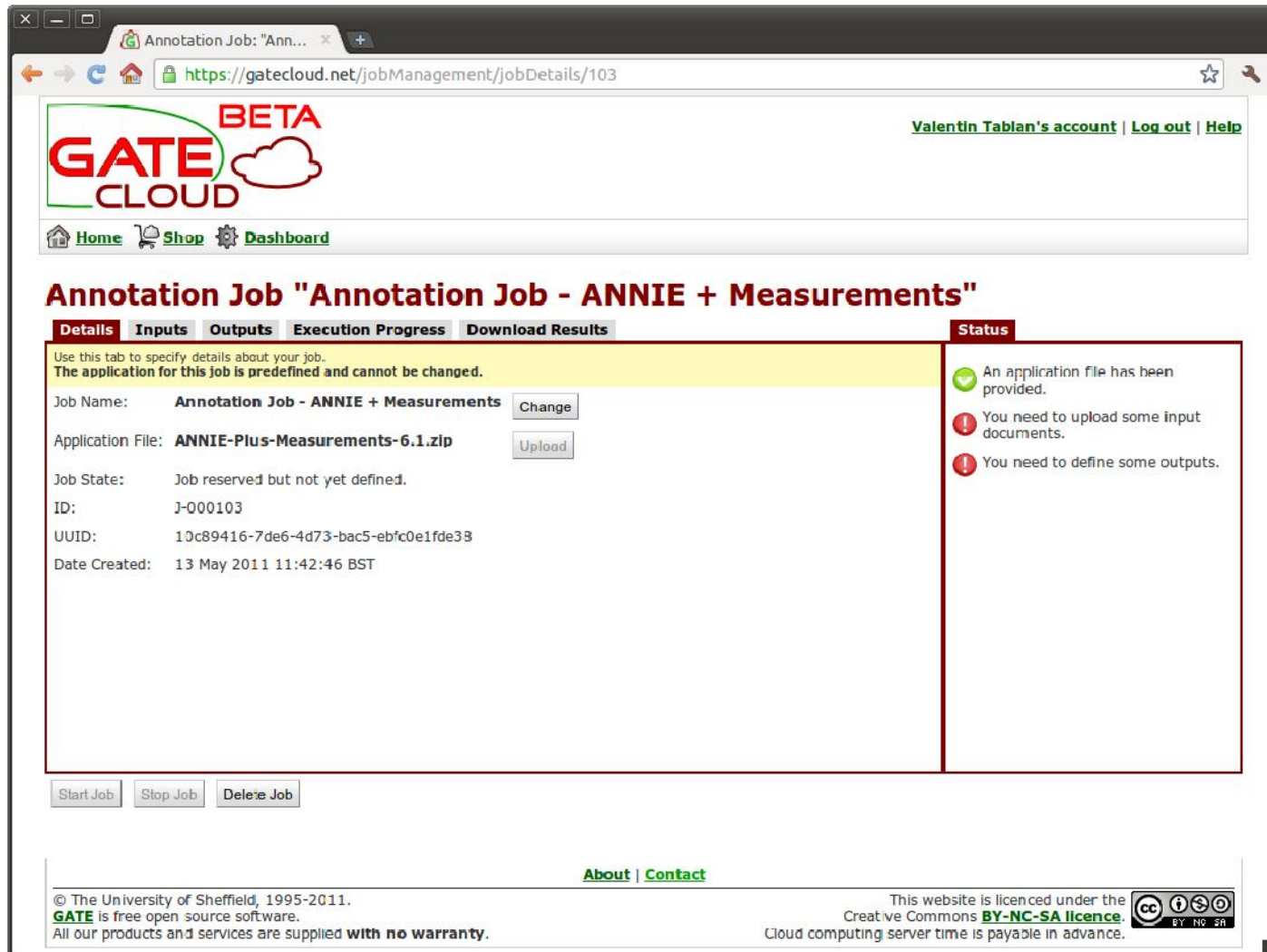
[About](#) | [Contact](#)

© The University of Sheffield, 1995-2011.
GATE is free open source software.
All our products and services are supplied **with no warranty**.

This website is licenced under the
Creative Commons [BY-NC-SA licence](#).
Cloud computing server time is payable in advance.



Job Editor: Details



The screenshot shows a web browser window with the URL <https://gatecloud.net/jobManagement/jobDetails/103>. The page header includes the GATE Cloud logo with a 'BETA' tag and a navigation menu with 'Home', 'Shop', and 'Dashboard'. The user is logged in as 'Valentin Tablan's account' with links for 'Log out' and 'Help'.

Annotation Job "Annotation Job - ANNIE + Measurements"

Details | Inputs | Outputs | Execution Progress | Download Results | **Status**

Use this tab to specify details about your job.
The application for this job is predefined and cannot be changed.

Job Name: **Annotation Job - ANNIE + Measurements**

Application File: **ANNIE-Plus-Measurements-6.1.zip**

Job State: Job reserved but not yet defined.

ID: J-000103

UUID: 10c89416-7de6-4d73-bac5-ebfc0e1fde38

Date Created: 13 May 2011 11:42:46 BST


Status

- An application file has been provided.
- You need to upload some input documents.
- You need to define some outputs.

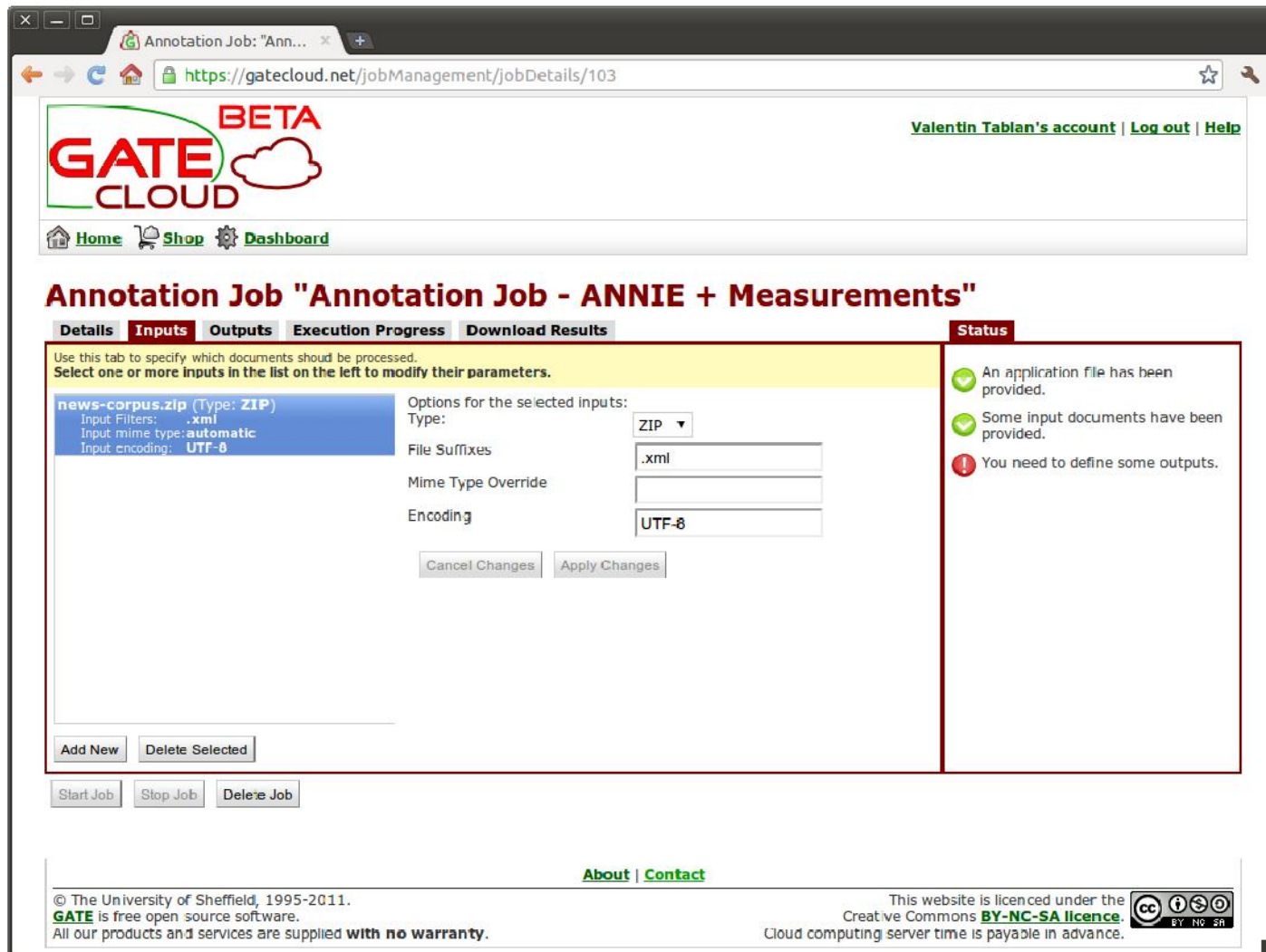
[About](#) | [Contact](#)

© The University of Sheffield, 1995-2011.
GATE is free open source software.
All our products and services are supplied with no warranty.

This website is licenced under the Creative Commons **BY-NC-SA licence**.
Cloud computing server time is payable in advance.



Job Editor: Inputs



The screenshot shows a web browser window with the URL <https://gatecloud.net/jobManagement/jobDetails/103>. The page header includes the GATE Cloud logo with a 'BETA' badge and the user 'Valentin Tablan's account' with links for 'Log out' and 'Help'. Navigation links for 'Home', 'Shop', and 'Dashboard' are present.

Annotation Job "Annotation Job - ANNIE + Measurements"

Details | **Inputs** | Outputs | Execution Progress | Download Results | **Status**

Use this tab to specify which documents should be processed.
Select one or more inputs in the list on the left to modify their parameters.

Input Name	Type	Options
news-corpus.zip	ZIP	Type: ZIP, File Suffixes: .xml, Mime Type Override: , Encoding: UTF-8

Buttons: Cancel Changes, Apply Changes

Buttons: Add New, Delete Selected


Buttons: Start Job, Stop Job, Delete Job

Status

- ✓ An application file has been provided.
- ✓ Some input documents have been provided.
- ! You need to define some outputs.

© The University of Sheffield, 1995-2011.
GATE is free open source software.
All our products and services are supplied with no warranty.

This website is licenced under the Creative Commons **BY-NC-SA licence**.
Cloud computing server time is payable in advance.



Job Editor: Outputs

GATE

The screenshot shows the GATE Cloud web interface for editing an annotation job. The browser address bar shows <https://gatecloud.net/jobManagement/jobDetails/103>. The page title is "Annotation Job 'Annotation Job - ANNIE + Measurements'". The user is logged in as "Valentin Tablan's account".

The interface has a navigation bar with "Home", "Shop", and "Dashboard" links. Below the navigation bar, there are tabs for "Details", "Inputs", "Outputs", "Execution Progress", and "Download Results". The "Outputs" tab is currently selected.

On this tab, you can specify what should happen to the output of your annotation job. Select one of the outputs on the left to change its parameters.

The left sidebar shows a list of outputs:

- GATE_XML: \${filename}-gate.xml
Annotation Set Annotation Type
[Default] Person
[Default] Location
[Default] Organization
[Default] Measurement
- MIMIR: http://ip-10-230-31-156.eu-west
Annotation Set Annotation Type
[Default] Person
[Default] Location
[Default] Organization
[Default] Measurement
[Default] Money
[Default] Date
[Default] Ratio
[Default] Sentence
- XCES: \${filename}-measurements.xml
Annotation Set Annotation Type
[Default] Measurement

The "MIMIR" output is selected. The "Details for the selected output:" section shows:

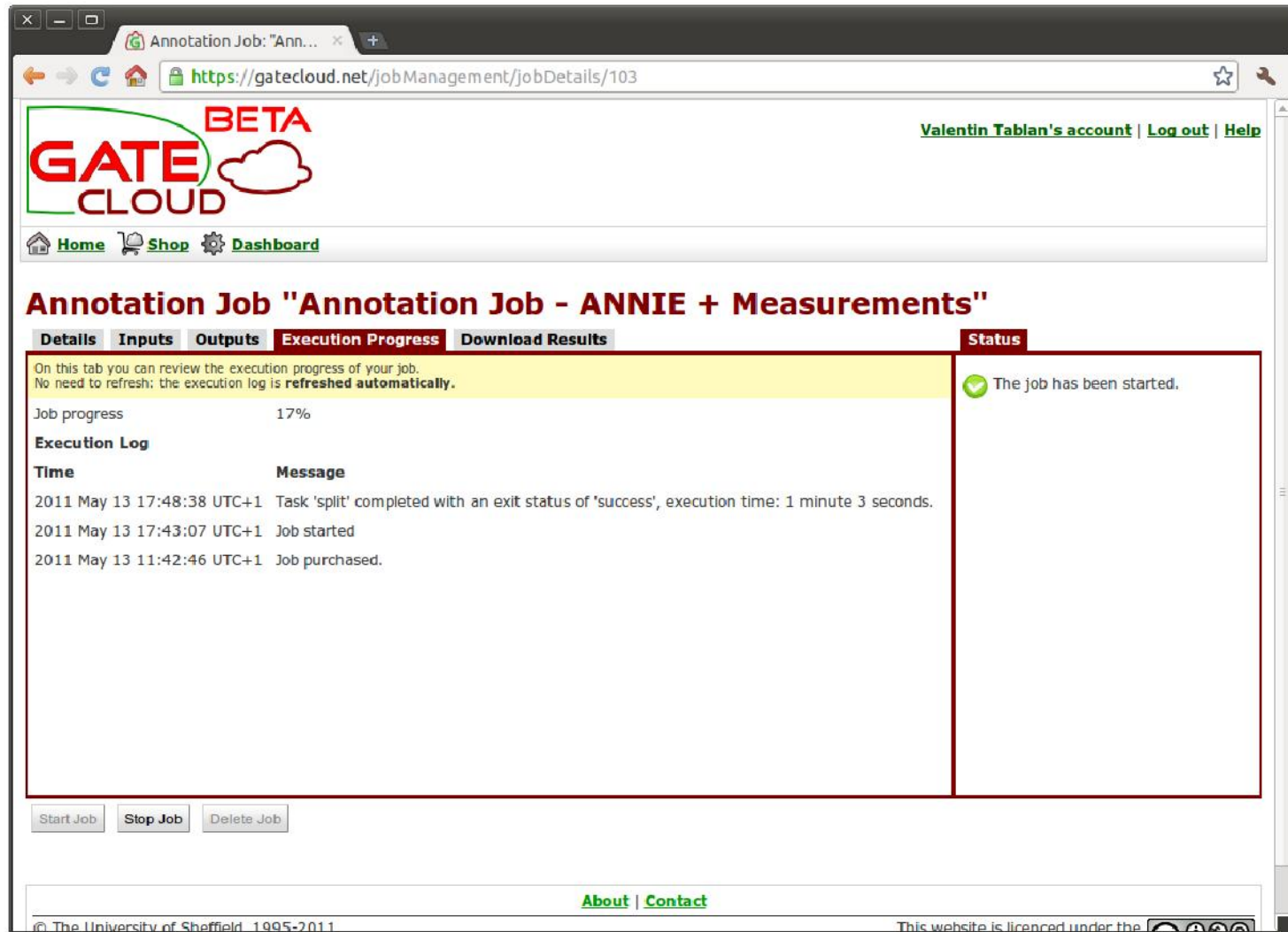
- Type: MIMIR
- Index URL: http://ip-10-230-31-156.eu-
- Username: [Empty field]
- Password: [Empty field]

The "Output filters (which annotations should be saved):" section shows a table of filters:

Annotation Set:	Annotation Type:	
[Default]	Person	Delete
[Default]	Location	Delete
[Default]	Organization	Delete
[Default]	Measurement	Delete
[Default]	Money	Delete
[Default]	Date	Delete

At the bottom of the page, there are buttons for "Start Job", "Stop Job", and "Delete Job". The footer contains "About | Contact" and "© The University of Sheffield, 1995-2011. This website is licenced under the [Creative Commons License]".

Job Editor: Progress

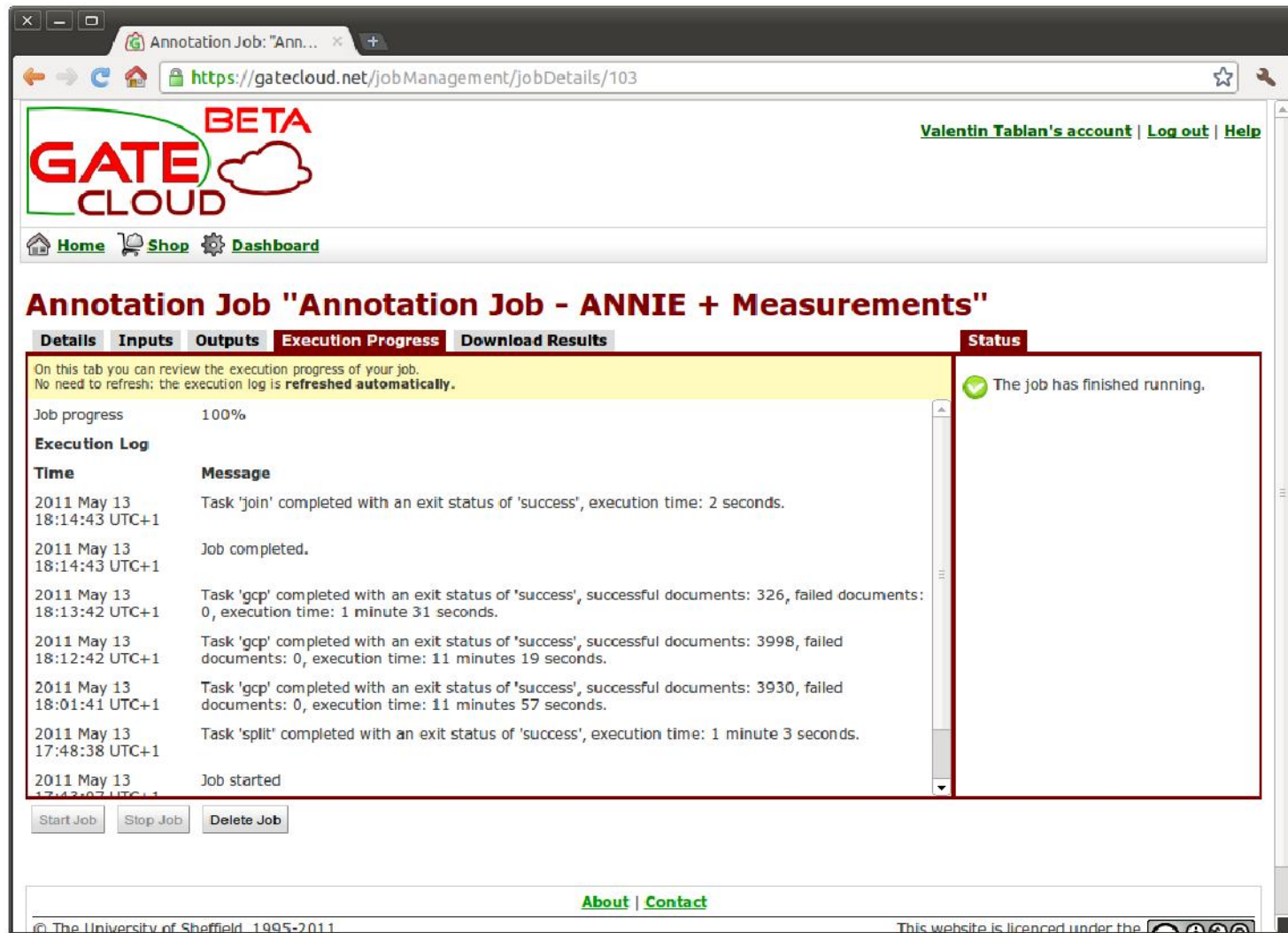


The screenshot shows a web browser window with the URL <https://gatecloud.net/jobManagement/jobDetails/103>. The page features the GATE Cloud logo with a 'BETA' tag and a cloud icon. The user is logged in as 'Valentin Tablan' and has options to 'Log out' or 'Help'. The main heading is 'Annotation Job "Annotation Job - ANNIE + Measurements"'. Below this, there are tabs for 'Details', 'Inputs', 'Outputs', 'Execution Progress', 'Download Results', and 'Status'. The 'Execution Progress' tab is active, showing a progress bar at 17% and a message: 'On this tab you can review the execution progress of your job. No need to refresh: the execution log is refreshed automatically.' The 'Status' tab shows a green checkmark and the message 'The job has been started.' The 'Execution Log' table contains the following entries:

Time	Message
2011 May 13 17:48:38 UTC+1	Task 'split' completed with an exit status of 'success', execution time: 1 minute 3 seconds.
2011 May 13 17:43:07 UTC+1	Job started
2011 May 13 11:42:46 UTC+1	Job purchased.

At the bottom of the page, there are buttons for 'Start Job', 'Stop Job', and 'Delete Job', and a footer with '© The University of Sheffield 1995-2011' and 'This website is licenced under the'.

Job Editor: Done!

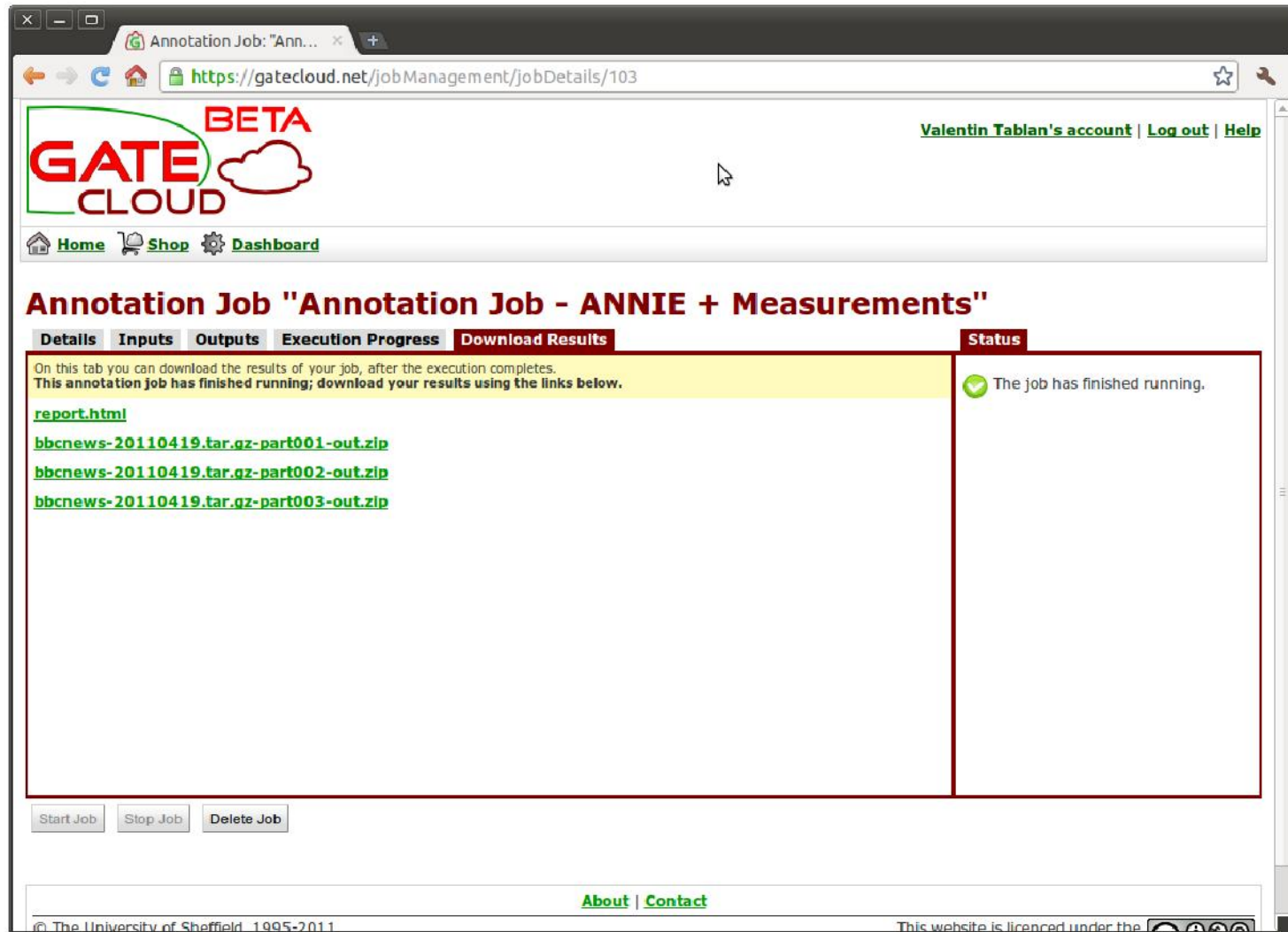


The screenshot shows a web browser window with the URL <https://gatecloud.net/jobManagement/jobDetails/103>. The page header includes the GATE Cloud logo with a 'BETA' tag and navigation links for Home, Shop, and Dashboard. The main content area is titled 'Annotation Job "Annotation Job - ANNIE + Measurements"' and features a 'Status' tab. A yellow notification bar states: 'On this tab you can review the execution progress of your job. No need to refresh: the execution log is refreshed automatically.' The job progress is shown as 100%. The 'Status' panel on the right indicates 'The job has finished running.' with a green checkmark. Below this, an 'Execution Log' table provides a detailed history of tasks and their completion times.

Time	Message
2011 May 13 18:14:43 UTC+1	Task 'join' completed with an exit status of 'success', execution time: 2 seconds.
2011 May 13 18:14:43 UTC+1	Job completed.
2011 May 13 18:13:42 UTC+1	Task 'gcp' completed with an exit status of 'success', successful documents: 326, failed documents: 0, execution time: 1 minute 31 seconds.
2011 May 13 18:12:42 UTC+1	Task 'gcp' completed with an exit status of 'success', successful documents: 3998, failed documents: 0, execution time: 11 minutes 19 seconds.
2011 May 13 18:01:41 UTC+1	Task 'gcp' completed with an exit status of 'success', successful documents: 3930, failed documents: 0, execution time: 11 minutes 57 seconds.
2011 May 13 17:48:38 UTC+1	Task 'split' completed with an exit status of 'success', execution time: 1 minute 3 seconds.
2011 May 13 17:43:07 UTC+1	Job started

At the bottom of the job details section, there are three buttons: 'Start Job', 'Stop Job', and 'Delete Job'. The footer of the page contains 'About | Contact' links and copyright information: '© The University of Sheffield, 1995-2011' and 'This website is licenced under the...'

Job Editor: Results



The screenshot shows a web browser window with the URL <https://gatecloud.net/jobManagement/jobDetails/103>. The page header includes the GATE Cloud logo with a 'BETA' tag and a user account menu for 'Valentin Tablan' with options for 'Log out' and 'Help'. A navigation bar contains 'Home', 'Shop', and 'Dashboard' links.

Annotation Job "Annotation Job - ANNIE + Measurements"

Navigation tabs: Details | Inputs | Outputs | Execution Progress | **Download Results** | Status

Download Results (Left Panel):

On this tab you can download the results of your job, after the execution completes.
This annotation job has finished running; download your results using the links below.

- [report.html](#)
- [bbcnews-20110419.tar.gz-part001-out.zip](#)
- [bbcnews-20110419.tar.gz-part002-out.zip](#)
- [bbcnews-20110419.tar.gz-part003-out.zip](#)

Status (Right Panel):

✓ The job has finished running.

Buttons at the bottom: Start Job | Stop Job | Delete Job

Footer: [About](#) | [Contact](#)
© The University of Sheffield, 1995-2011 | This website is licenced under the [Creative Commons License]

Hands on: Access your own account



- Navigate to <http://gatecloud.net>
- Log in to the account you were given
 - Change your email address
 - Change the password

Hands on: Run the ANNIE pipeline

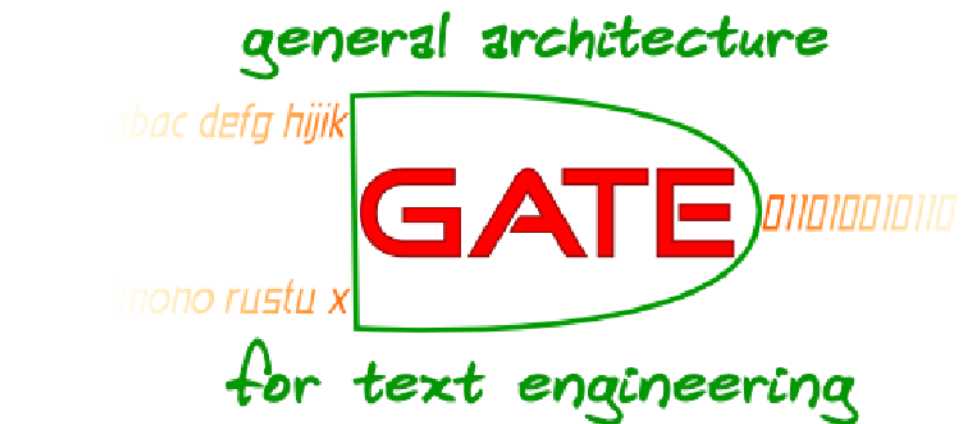


- Choose the pre-packaged ANNIE pipeline
- Use `news-corpus-small.zip` as input
- Output to GATE XML format
 - from the default annotation set
 - with all annotation types included
- Download the annotated documents, and inspect them using GATE Developer.
- Look at the execution reports.

Hands on: Extra Credit

The logo for GATE (Global Access to Text Engineering) is displayed in the top right corner. It consists of the word "GATE" in a bold, red, sans-serif font, enclosed within a green, rounded rectangular border.

-
- Provide your own pipeline (e.g. export ANNIE from Developer)
 - Produce output in XCES, selecting only Person, Location, Organization, Date annotation types.



Severs (Machines)

Rent ready-made GATE-based virtual servers.

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 server 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 overwrites the current state of the data volume.

Destroy a Machine Reservation

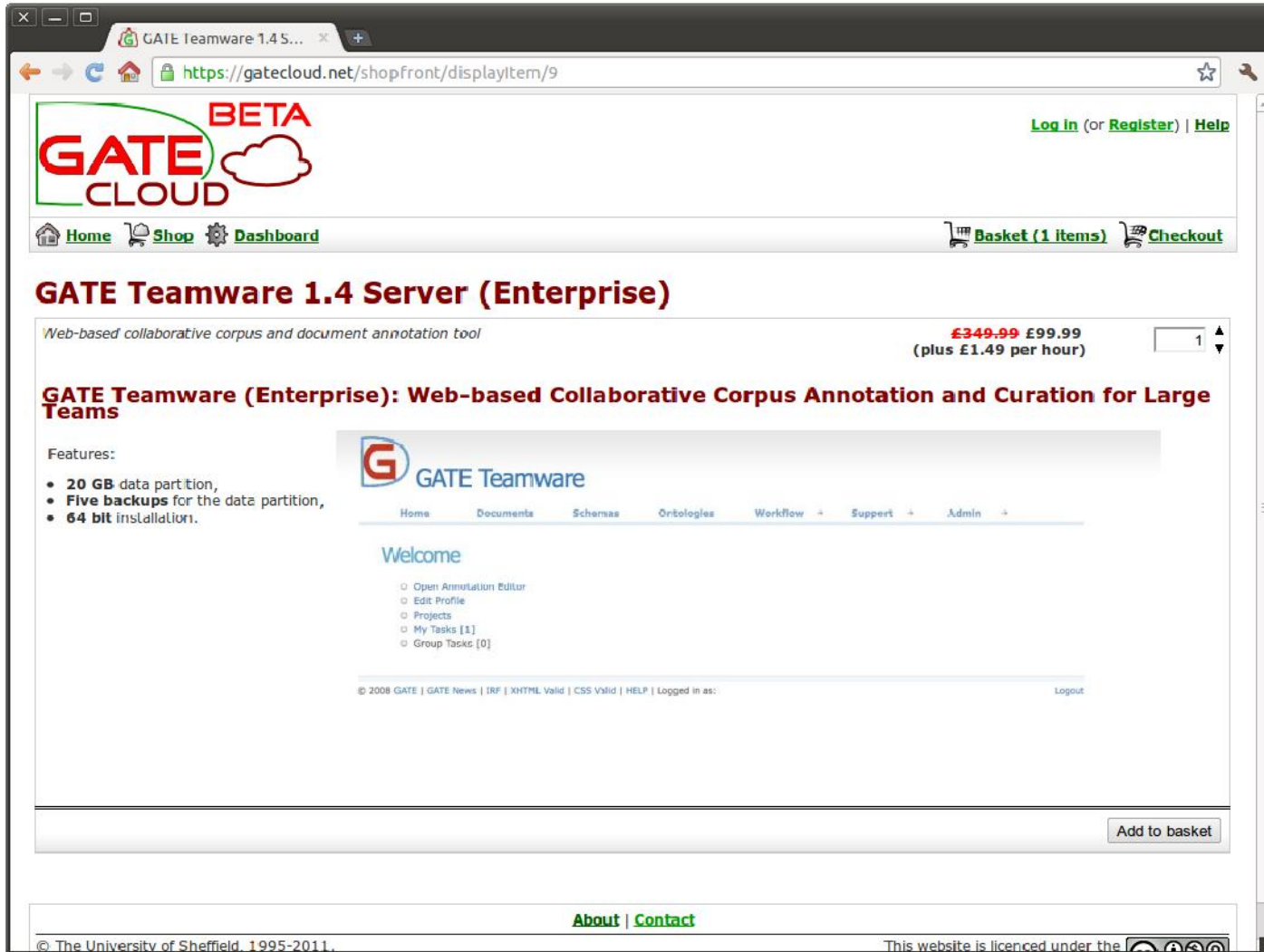


Destroy all data associated with the reservation.

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

Making a Reservation:

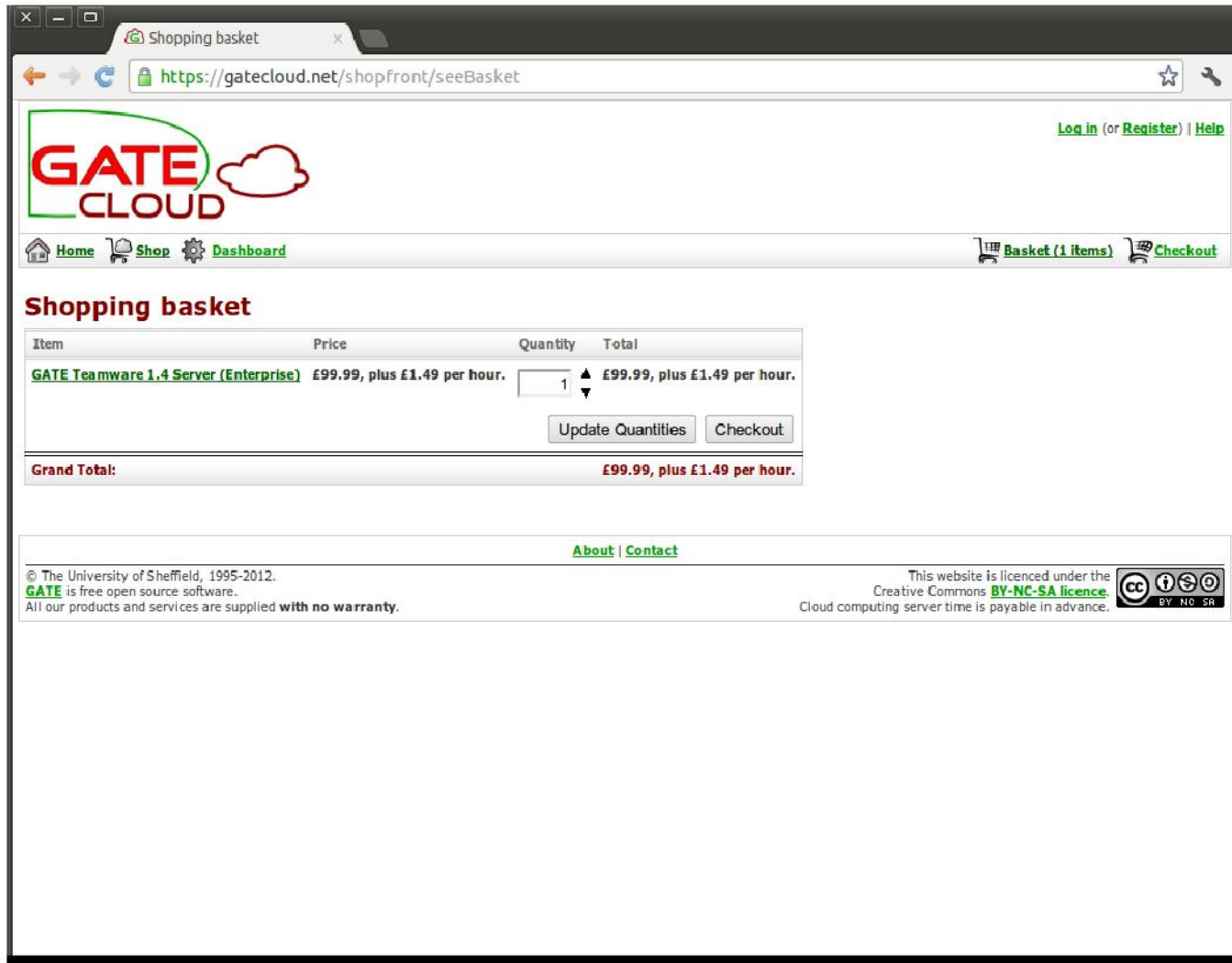
1) Find Required Server



The screenshot shows a web browser window with the URL <https://gatecloud.net/shopfront/displayItem/9>. The page features the GATE Cloud logo with a "BETA" tag and navigation links for Home, Shop, and Dashboard. A shopping basket icon shows "1 item" and a Checkout button. The main content area displays the product "GATE Teamware 1.4 Server (Enterprise)" with a price of £349.99 (reduced from £99.99) plus £1.49 per hour. Below the product name is a description: "Web-based collaborative corpus and document annotation tool". A list of features includes 20 GB data partition, five backups, and 64 bit installation. A preview of the GATE Teamware interface is shown, including a navigation menu and a "Welcome" message with links to Open Annotation Editor, Edit Profile, Projects, My Tasks [1], and Group Tasks [0]. The page footer contains copyright information for The University of Sheffield (1995-2011) and a license notice.

Making a Reservation:

2) Add to Basket



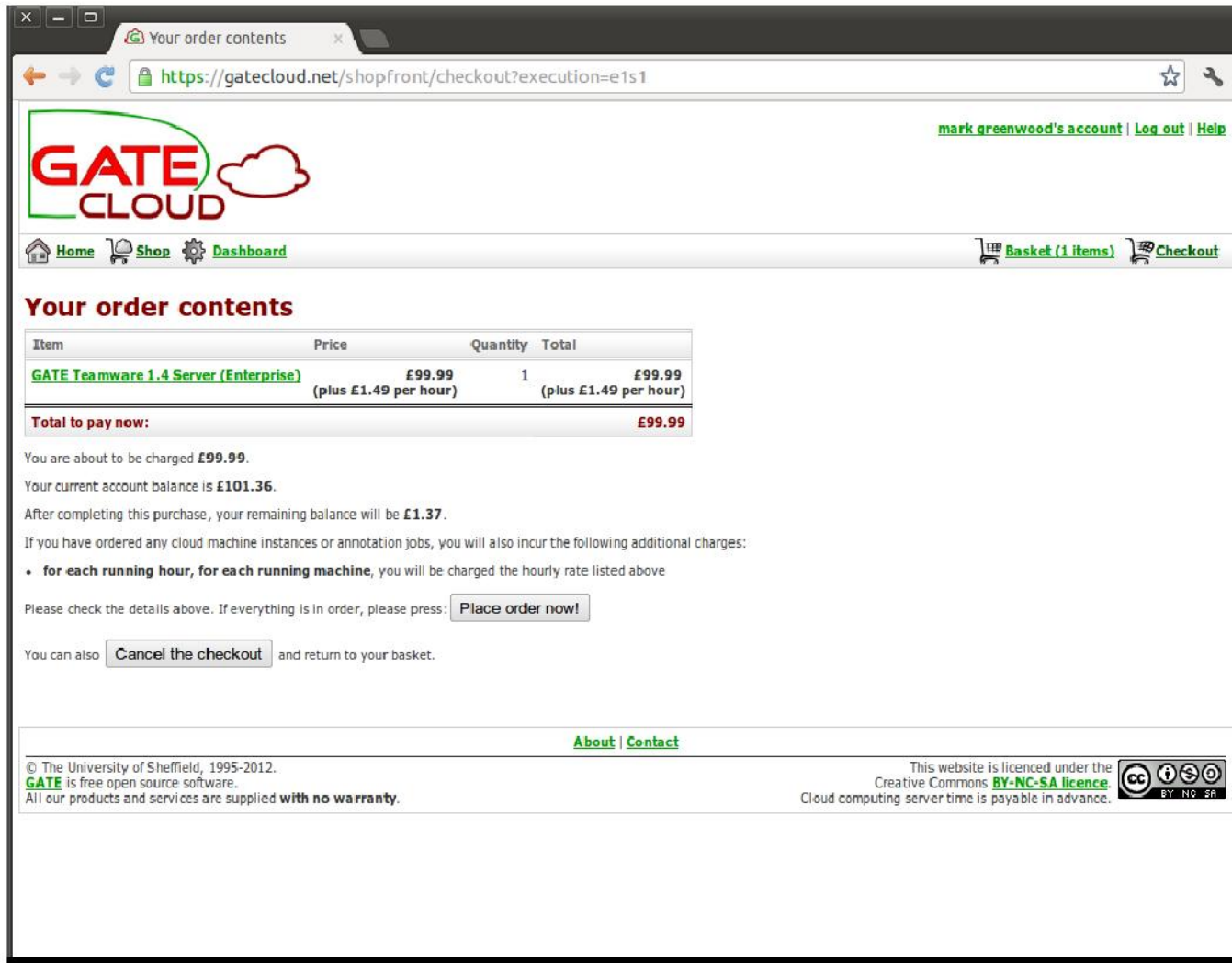
The screenshot shows a web browser window with the URL <https://gatecloud.net/shopfront/seeBasket>. The page features the GATE CLOUD logo and navigation links for Home, Shop, and Dashboard. A shopping basket icon indicates 1 item, and a Checkout button is visible. The main content area is titled "Shopping basket" and contains a table with the following data:

Item	Price	Quantity	Total
GATE Teaaware 1.4 Server (Enterprise)	£99.99, plus £1.49 per hour.	<input type="text" value="1"/>	£99.99, plus £1.49 per hour.

Below the table are buttons for "Update Quantities" and "Checkout". A "Grand Total:" section shows the total price as "£99.99, plus £1.49 per hour." The footer includes copyright information for The University of Sheffield (1995-2012), a Creative Commons BY-NC-SA license, and a disclaimer that all products and services are supplied with no warranty.

Making a Reservation:

3) Checkout



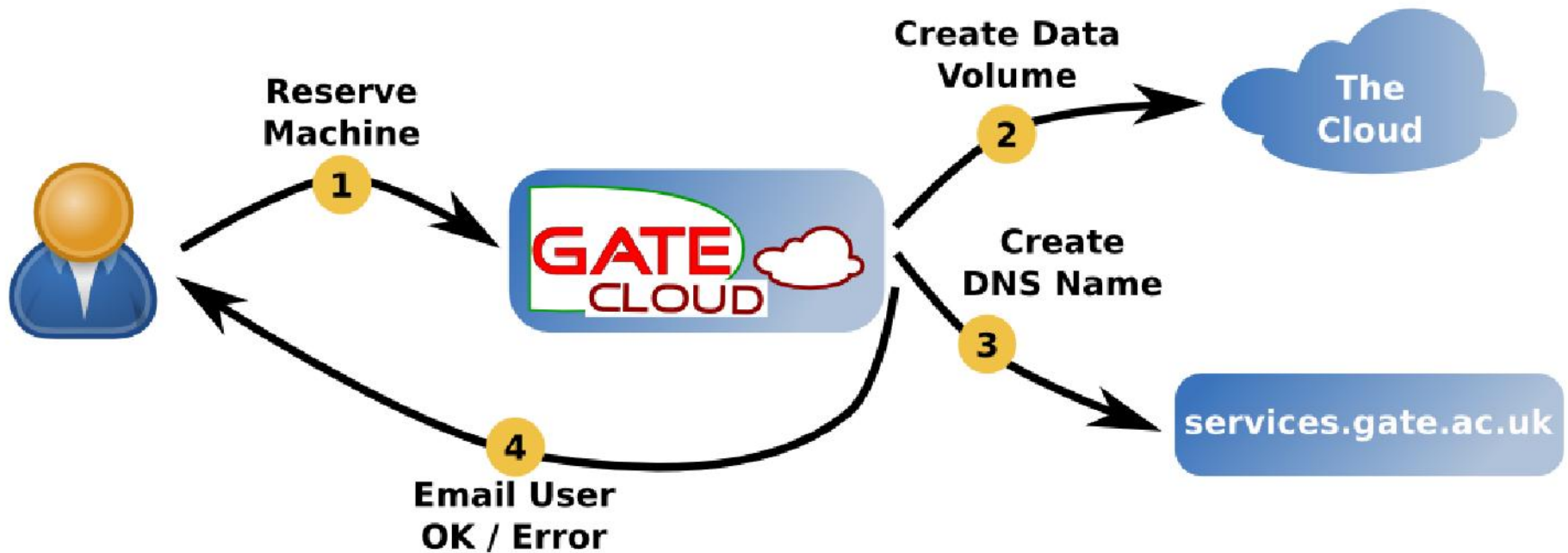
The screenshot shows a web browser window with the URL <https://gatecloud.net/shopfront/checkout?execution=e1s1>. The page header includes the GATE CLOUD logo, navigation links for Home, Shop, and Dashboard, and a shopping cart icon showing 'Basket (1 items)' and a 'Checkout' button. The main content area is titled 'Your order contents' and features a table with the following data:

Item	Price	Quantity	Total
GATE Teamware 1.4 Server (Enterprise)	£99.99 (plus £1.49 per hour)	1	£99.99 (plus £1.49 per hour)
Total to pay now:			£99.99

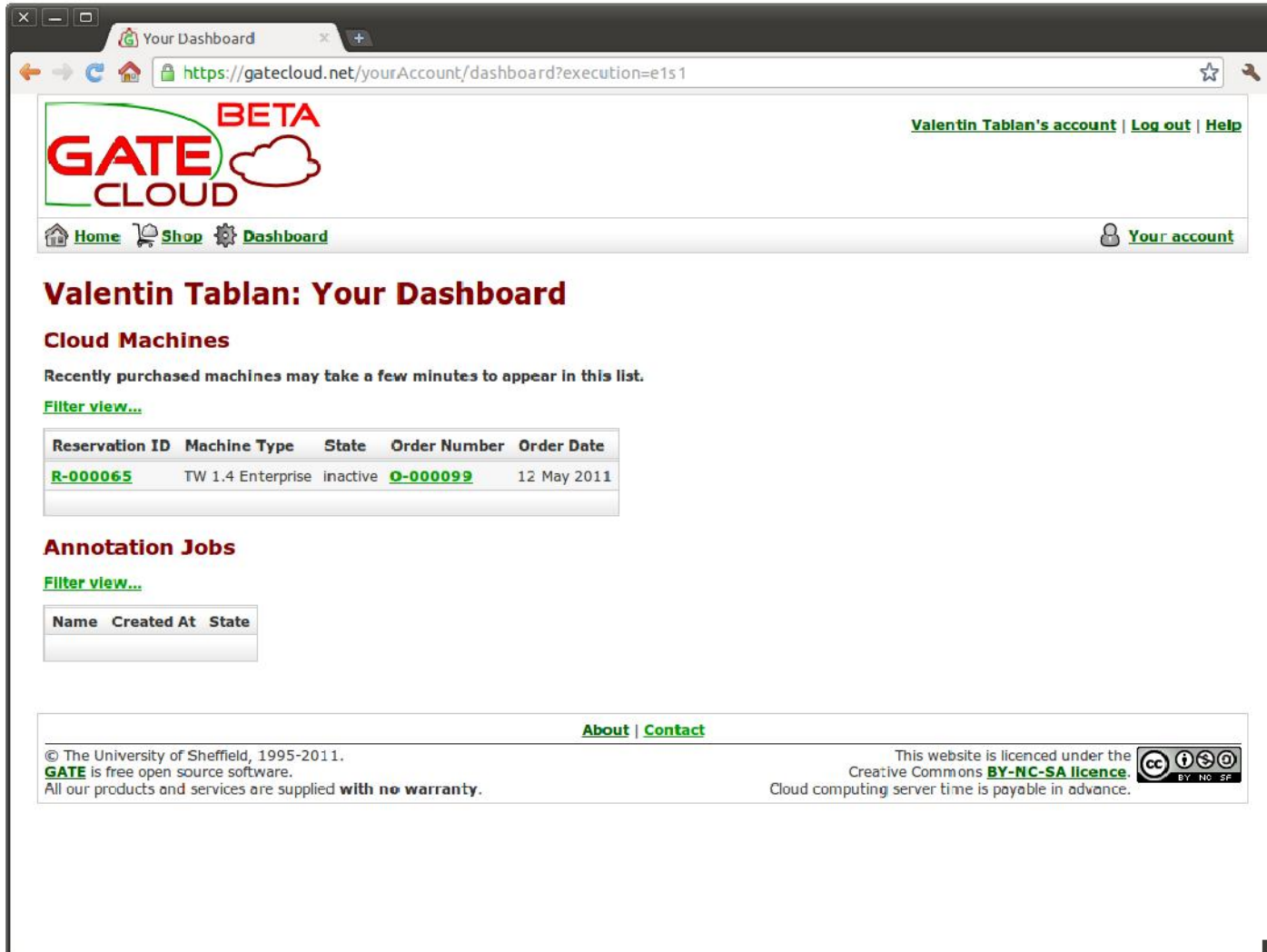
Below the table, the user is informed: 'You are about to be charged £99.99. Your current account balance is £101.36. After completing this purchase, your remaining balance will be £1.37. If you have ordered any cloud machine instances or annotation jobs, you will also incur the following additional charges: for each running hour, for each running machine, you will be charged the hourly rate listed above.' A 'Place order now!' button is provided, along with a 'Cancel the checkout' button to return to the basket. The footer contains copyright information for The University of Sheffield (1995-2012), a Creative Commons BY-NC-SA license, and a disclaimer: 'All our products and services are supplied with no warranty. Cloud computing server time is payable in advance.'

Reserving a Server

GATE



Your Dashboard



The screenshot shows a web browser window with the URL <https://gatecloud.net/yourAccount/dashboard?execution=e1s1>. The page features the GATE Cloud logo with a 'BETA' badge and a cloud icon. The user is logged in as 'Valentin Tablan's account' with links for 'Log out' and 'Help'. A navigation bar includes 'Home', 'Shop', and 'Dashboard' (selected), along with a 'Your account' link. The main content area is titled 'Valentin Tablan: Your Dashboard' and contains two sections: 'Cloud Machines' and 'Annotation Jobs'. The 'Cloud Machines' section includes a table with one entry: Reservation ID R-000065, Machine Type TW 1.4 Enterprise, State inactive, Order Number O-000099, and Order Date 12 May 2011. The 'Annotation Jobs' section is currently empty. The footer contains copyright information for The University of Sheffield (1995-2011), a Creative Commons BY-NC-SA license logo, and a disclaimer that all products and services are supplied with no warranty.

Valentin Tablan: Your Dashboard

Cloud Machines

Recently purchased machines may take a few minutes to appear in this list.

[Filter view...](#)

Reservation ID	Machine Type	State	Order Number	Order Date
R-000065	TW 1.4 Enterprise	inactive	O-000099	12 May 2011

Annotation Jobs


[Filter view...](#)

Name	Created At	State
------	------------	-------

[About](#) | [Contact](#)

© The University of Sheffield, 1995-2011.
GATE is free open source software.
All our products and services are supplied **with no warranty**.

This website is licenced under the Creative Commons **BY-NC-SA licence**.
Cloud computing server time is payable in advance.



Your Machine Reservation

GATE

The screenshot shows a web browser window with the URL <https://gatecloud.net/yourAccount/machineReservationDetails/65>. The page title is "Machine Reservation R-000065". The navigation bar includes "Home", "Shop", "Dashboard", and "Your account".

Machine Reservation R-000065

ID	R-000065	Destroy Reservation
Machine type	GATE Teamware 1.4 Server (Enterprise)	
Hourly price	£1.49	
State	inactive	Start Instance
Instance ready	no	

Backups

Slot 1	<empty>	Create new backup
Slot 2	<empty>	Create new backup
Slot 3	<empty>	Create new backup
Slot 4	<empty>	Create new backup
Slot 5	<empty>	Create new backup

Reservation Details:

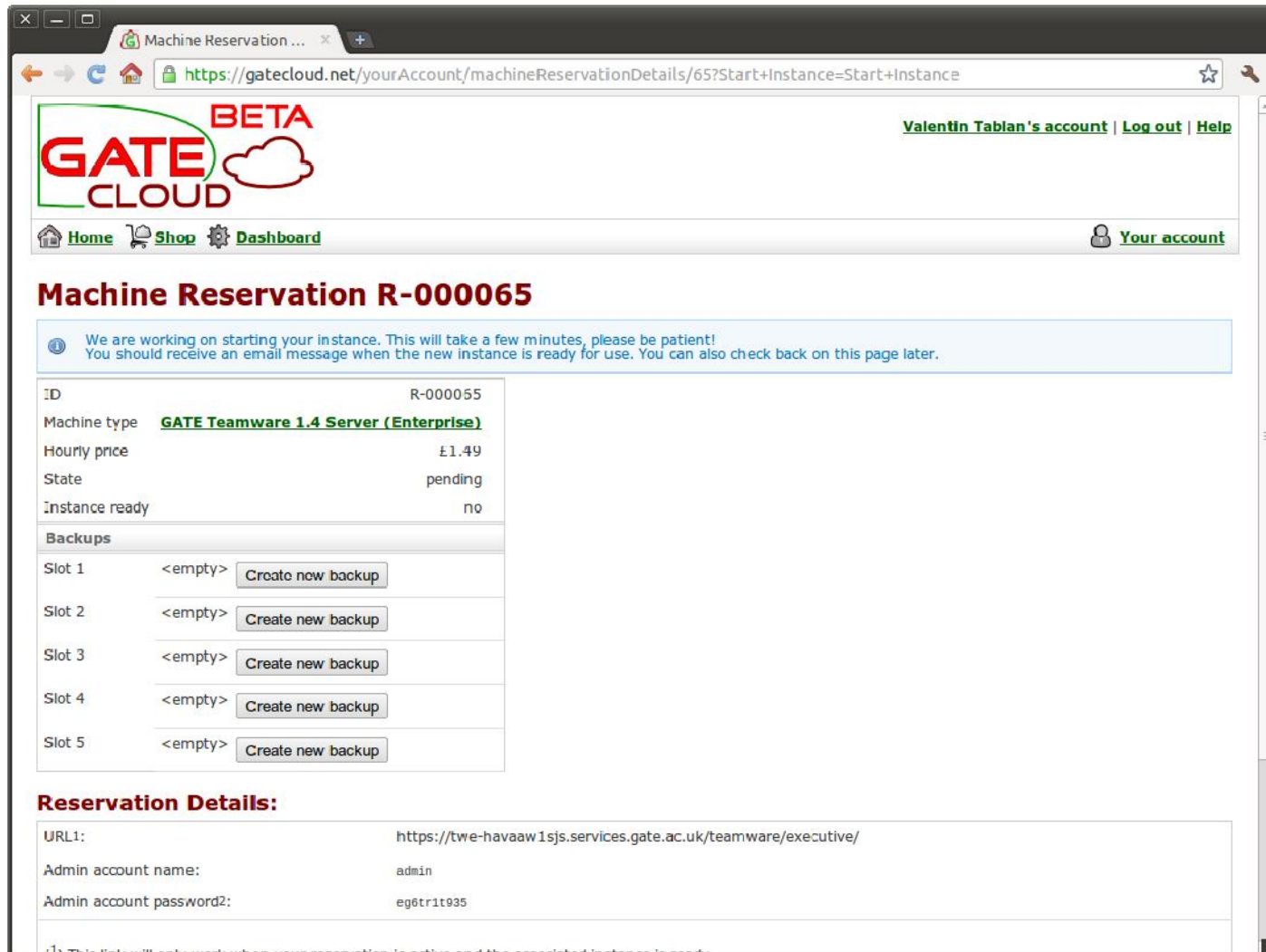
URL1:	https://twe-havaaw1sjs.services.gate.ac.uk/teamware/executive/
Admin account name:	admin
Admin account password2:	██████████

⁽¹⁾ This link will only work when your reservation is active and the associated instance is ready.

⁽²⁾ This password is the value that was generated and used when the reservation was originally made. If you have changed the password for the admin account in the mean time, then please use the new value that you set yourself. We do not have access to the new password, so we cannot display it here. If you have lost the password, please [contact us](#) and we can reset it for you.

Starting a Server:

1) Press Start



The screenshot shows a web browser window with the URL <https://gatecloud.net/yourAccount/machineReservationDetails/65?Start+Instance=Start+Instance>. The page header includes the GATE Cloud logo with a 'BETA' badge, the user's account name 'Valentin Tablan's account', and links for 'Log out' and 'Help'. A navigation bar contains 'Home', 'Shop', and 'Dashboard' links, along with a 'Your account' link.

Machine Reservation R-000065

We are working on starting your instance. This will take a few minutes, please be patient!
You should receive an email message when the new instance is ready for use. You can also check back on this page later.

ID	R-000055
Machine type	GATE Teamware 1.4 Server (Enterprise)
Hourly price	£1.49
State	pending
Instance ready	no

Backups

Slot 1	<empty>	<input type="button" value="Create new backup"/>
Slot 2	<empty>	<input type="button" value="Create new backup"/>
Slot 3	<empty>	<input type="button" value="Create new backup"/>
Slot 4	<empty>	<input type="button" value="Create new backup"/>
Slot 5	<empty>	<input type="button" value="Create new backup"/>

Reservation Details:

URL1:	https://twe-havaaw1sjs.services.gate.ac.uk/teamware/executive/
Admin account name:	admin
Admin account password2:	eg@tr1t935

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

GATE

The screenshot shows a web browser window displaying the GATE Cloud interface. The page title is "Machine Reservation R-000065". The main content area shows the reservation details:

ID	R-000065
Machine type	GATE Teamware 1.4 Server (Enterprise)
Hourly price	£1.49
State	active <input type="button" value="Stop Instance"/>
Instance ready	yes

Below the details is a "Backups" section with five slots, each containing "<empty>" and a "Create new backup" button.

Reservation Details:

URL1:	https://twe-havaaw1sis.services.gate.ac.uk/teamware/executive/
Admin account name:	admin
Admin account password2:	██████████

Footnote (1): This link will only work when your reservation is active and the associated instance is ready.
Footnote (2): This password is the value that was generated and used when the reservation was originally made. If you have changed the password for the admin account in the

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

GATE

Machine Reservation ... Login | GATECloud.net ...

https://twe-havaaw1sjs.services.gate.ac.uk/teamware/executive/login.jsp;jsessionid=7CDE764D69D5BF4FBE03FA3322FA474F

G GATECloud.net Teamware 1.4

Login

Login

Username *

admin

Password *

Remember Me

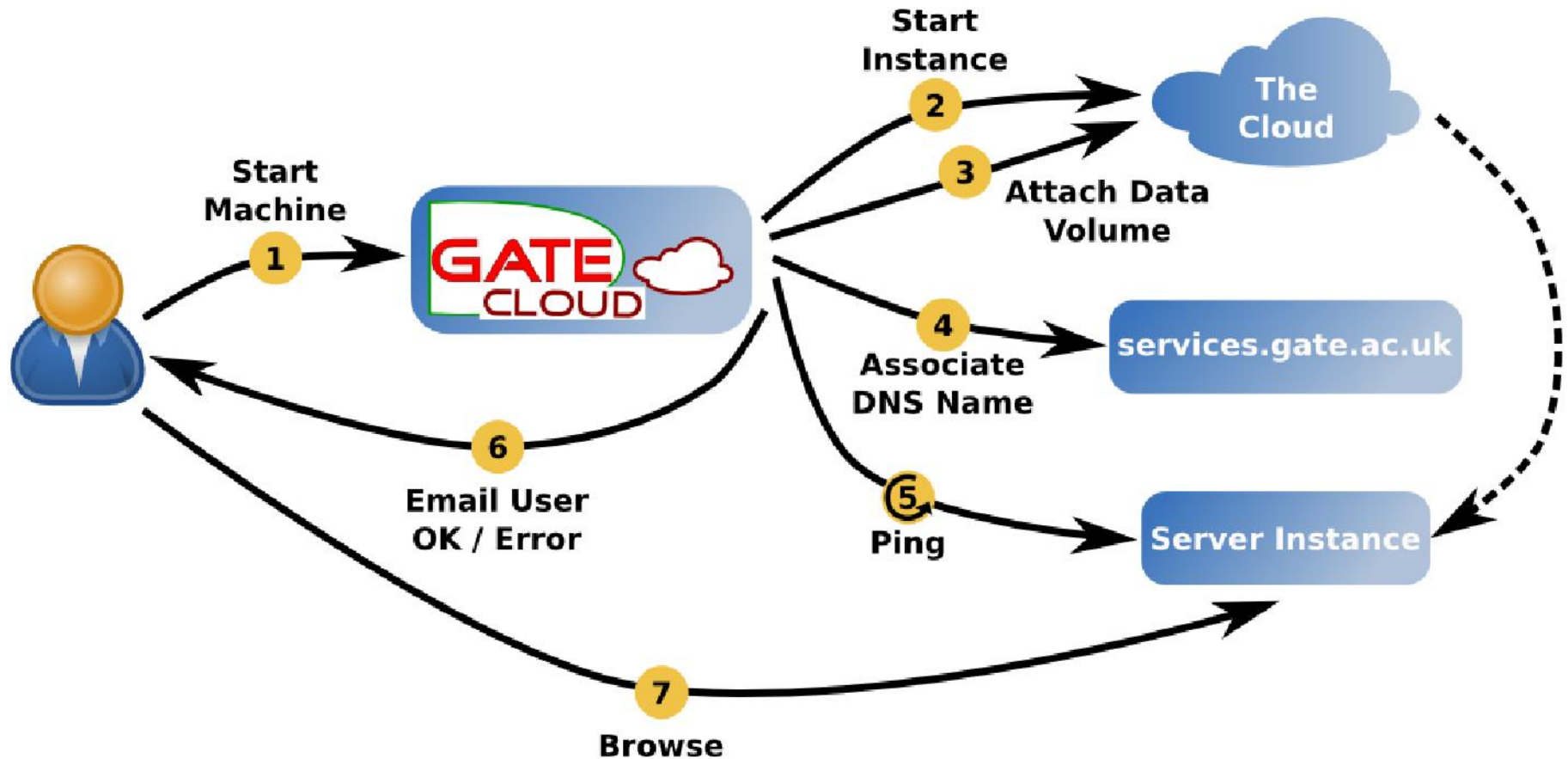
Login

Not a member? [Sign-up](#) for an account..

Forgot your password? Have your [password](#) hint e-mailed to you.

© 2007-2011 GATE team (AGPL licence) | [GATE News](#) | [GATE Teamware](#) | [HELP](#) | Time zone: **UTC**

Starting a Server



Questions?



<http://gate.ac.uk>