Compute Server– New Features and Enhancements Gurobi 9.0

GUROBI
OPTIMIZATION

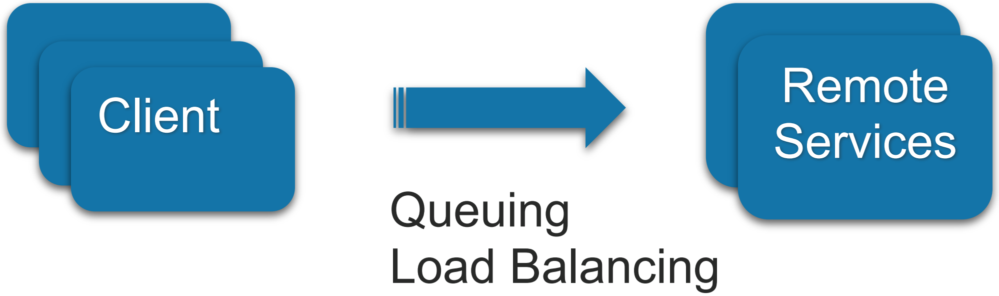
The World's Fastest Solver

Speaker Introduction

- Dr. Michel Jaczynski
  - Senior Architect at Gurobi Optimization
  - He has 20 years of experience building components, frameworks and Cloud-based services for optimization applications
  - Worked at IBM on the decision optimization Cloud services and application frameworks and also at ILOG on integrating optimization with enterprise software and a framework to optimize semiconductor Fabs operations in near real-time.
  - Dr. Jaczynski holds a Ph.D. in Computer Science from the University of Nice Sophia-Antipolis, France.

Copyright © 2019, Gurobi Optimization, LLC

Compute Server Overview



- **Seamlessly offload optimization jobs to one or more servers**
- **Common use cases:**
  - Share resources across users, teams and applications
  - Access more powerful machines
  - Ensure high availability with multiple nodes
  - Build modern applications, service-oriented architecture

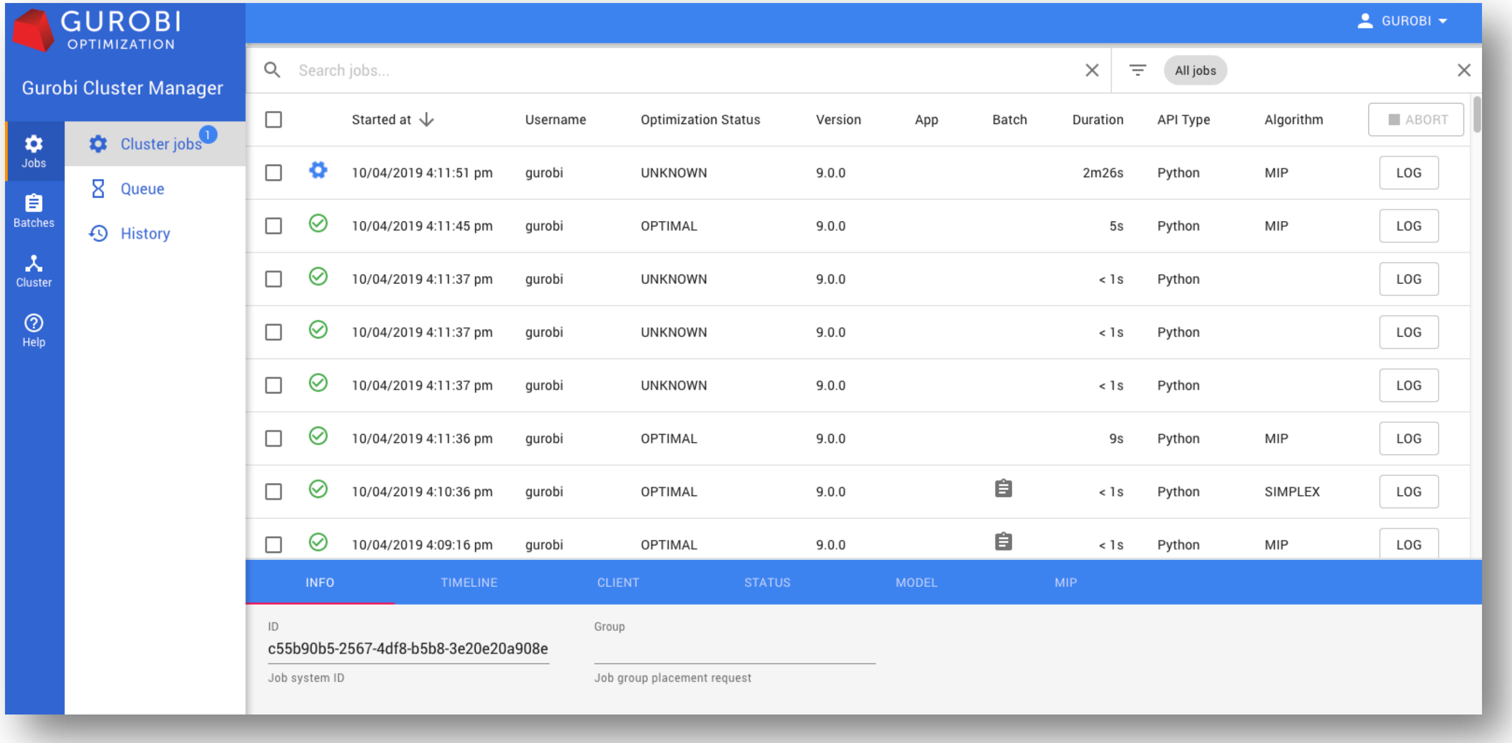
Copyright © 2020, Gurobi Optimization, LLC

Gurobi Compute Server V9.0

- **For IT**
  - Modern and scalable architecture
  - Better security
  - New web UI to manage the cluster

- **For OR Experts**
  - New web UI to monitor jobs
  - New job history
  - New batch optimization

Cluster Manager



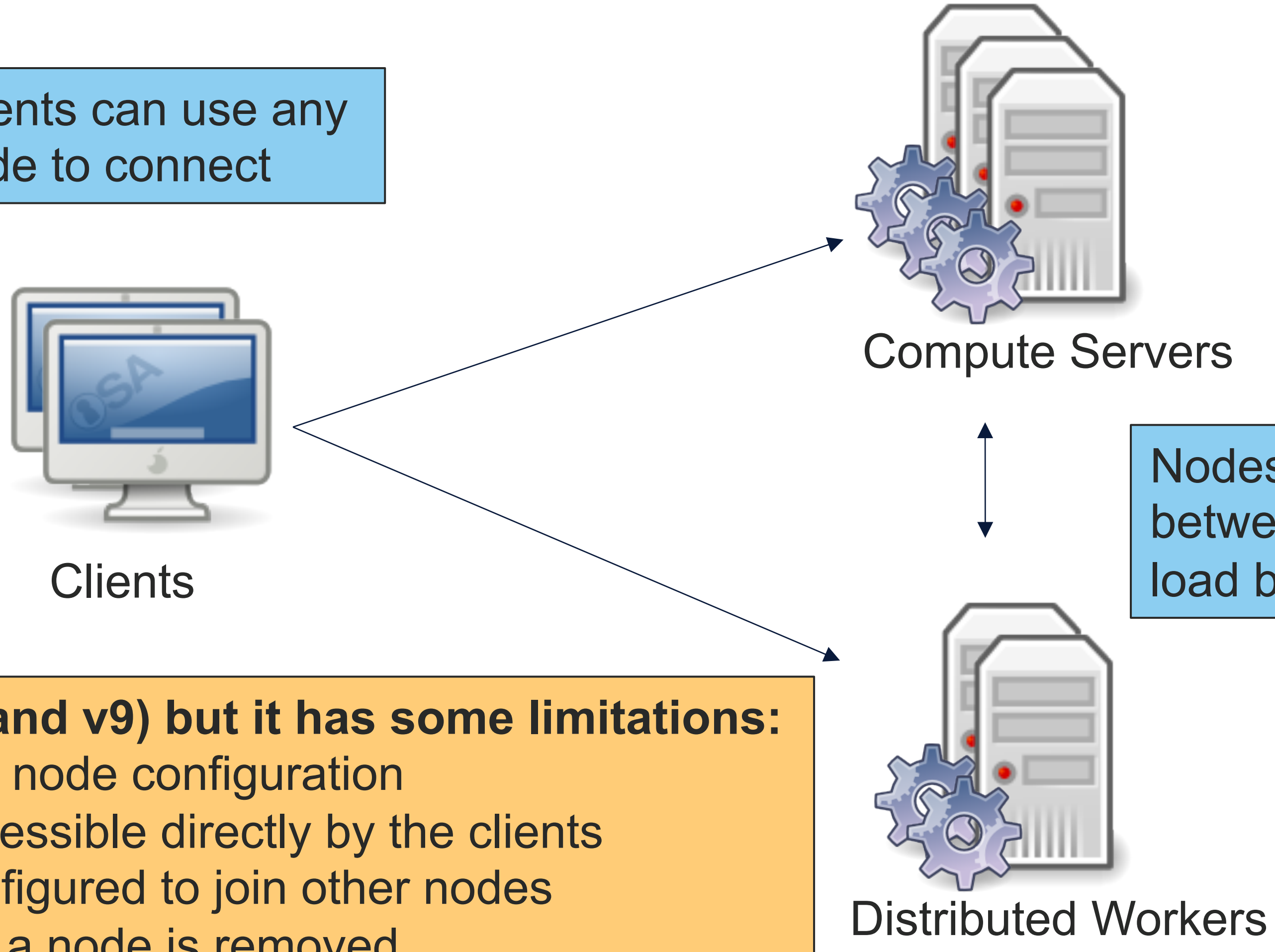
□		Started at ↓	Username	Optimization Status	Version	App	Batch	Duration	API Type	Algorithm	ABORT
□	⚙	10/04/2019 4:11:51 pm	gurobi	UNKNOWN	9.0.0			2m26s	Python	MIP	LOG
□	✓	10/04/2019 4:11:45 pm	gurobi	OPTIMAL	9.0.0			5s	Python	MIP	LOG
□	✓	10/04/2019 4:11:37 pm	gurobi	UNKNOWN	9.0.0			< 1s	Python		LOG
□	✓	10/04/2019 4:11:37 pm	gurobi	UNKNOWN	9.0.0			< 1s	Python		LOG
□	✓	10/04/2019 4:11:37 pm	gurobi	UNKNOWN	9.0.0			< 1s	Python		LOG
□	✓	10/04/2019 4:11:36 pm	gurobi	OPTIMAL	9.0.0			9s	Python	MIP	LOG
□	✓	10/04/2019 4:10:36 pm	gurobi	OPTIMAL	9.0.0		📋	< 1s	Python	SIMPLEX	LOG
□	✓	10/04/2019 4:09:16 pm	gurobi	OPTIMAL	9.0.0		📋	< 1s	Python	MIP	LOG

ID: c55b90b5-2567-4df8-b5b8-3e20e20a908e
Job system ID

Group:
Job group placement request

Compute Server v9 facilitates the deployment and use of optimization services on-premises or on private cloud.

Architecture: Self-Managed Cluster (V8 and V9)



Clients can use any node to connect

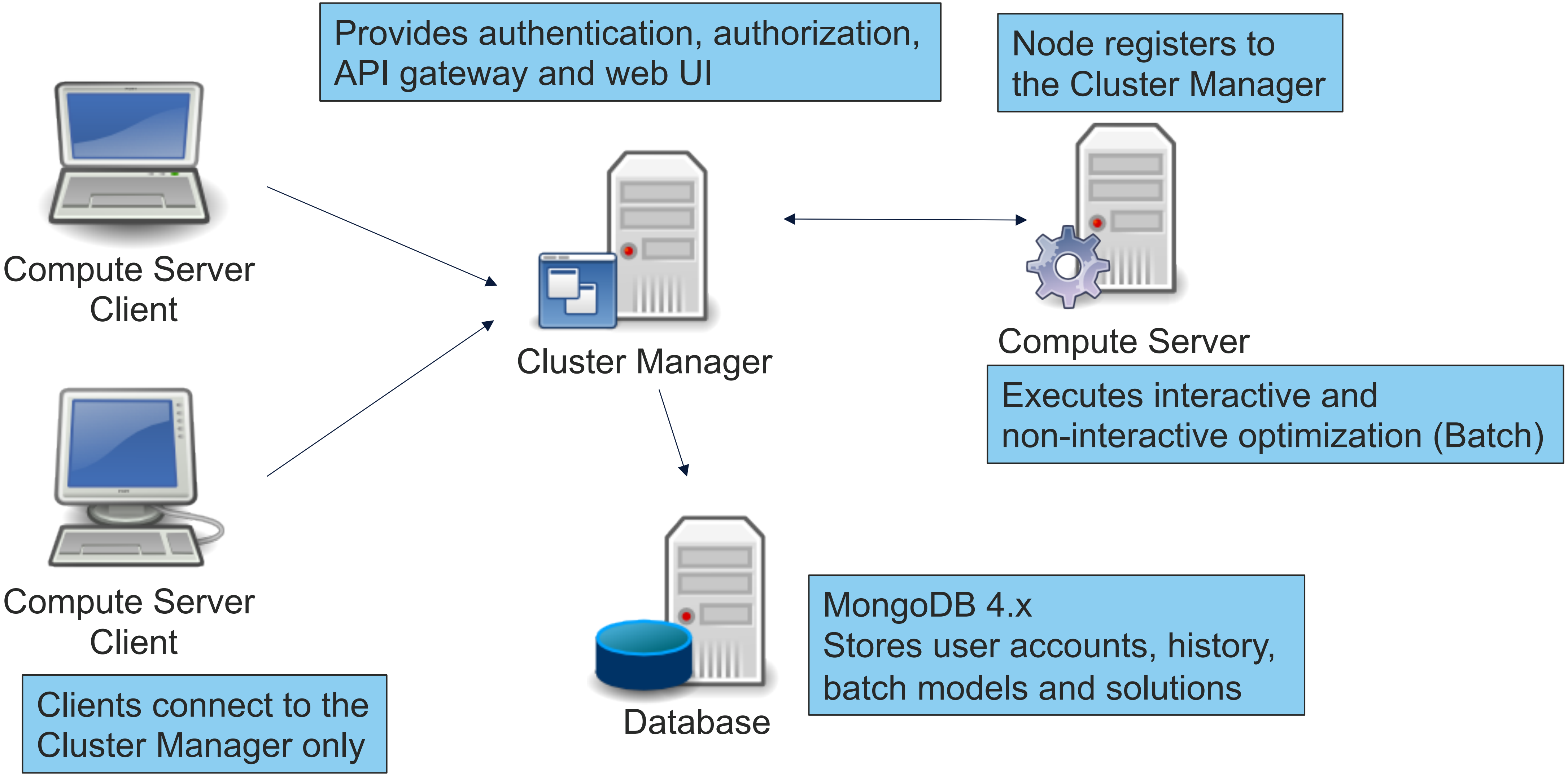
Nodes communicate between each other for load balancing and monitoring

Fully supported (v8 and v9) but it has some limitations:

- Fixed passwords in node configuration
- Nodes must be accessible directly by the clients
- Nodes must be configured to join other nodes
- Job history is lost if a node is removed
- Clients must remain connected during optimization

Copyright © 2020, Gurobi Optimization, LLC

Architecture: Cluster Manager (New in V9)



Provides authentication, authorization, API gateway and web UI

Node registers to the Cluster Manager

Compute Server Client

Cluster Manager

Compute Server

Executes interactive and non-interactive optimization (Batch)

Compute Server Client

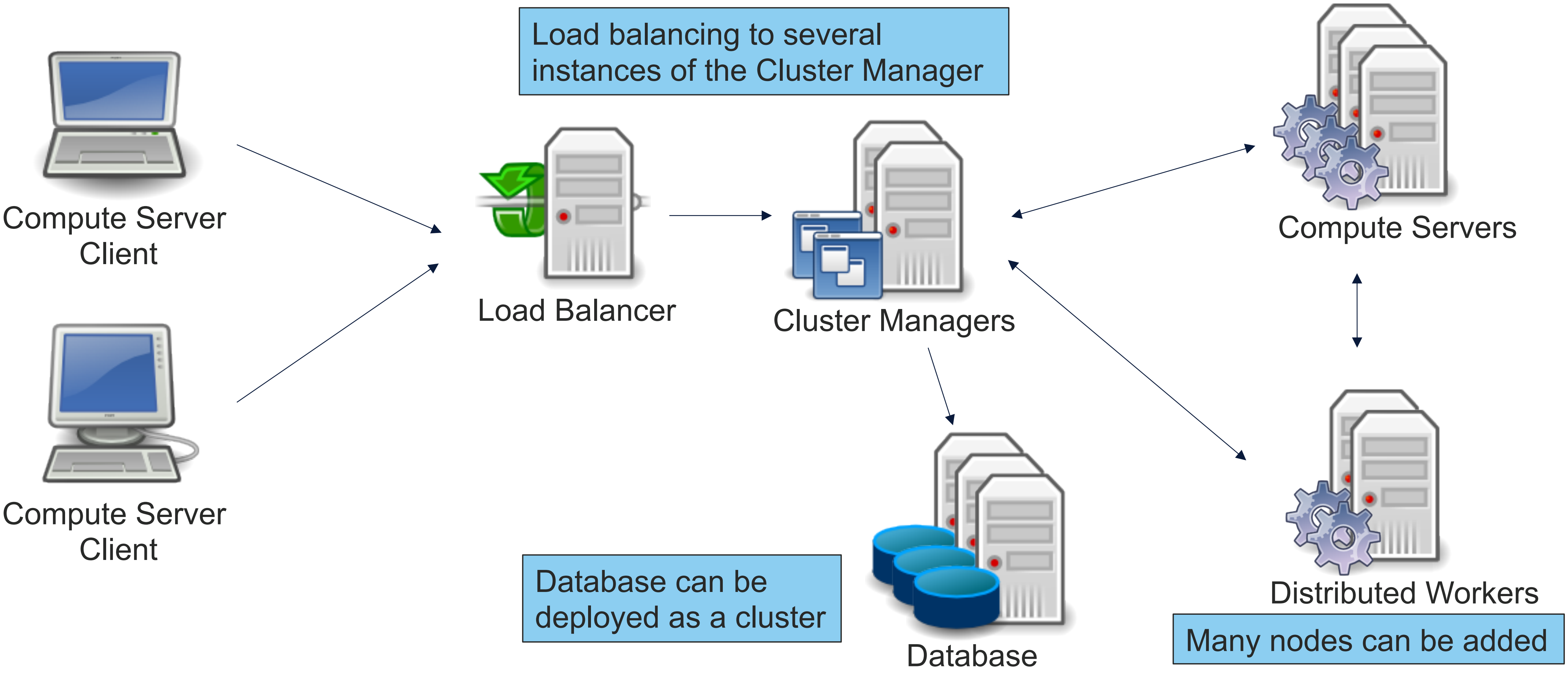
Clients connect to the Cluster Manager only

Database

MongoDB 4.x
Stores user accounts, history, batch models and solutions

Copyright © 2020, Gurobi Optimization, LLC

Architecture: Scalable Cluster Manager (New in V9)



Load balancing to several instances of the Cluster Manager

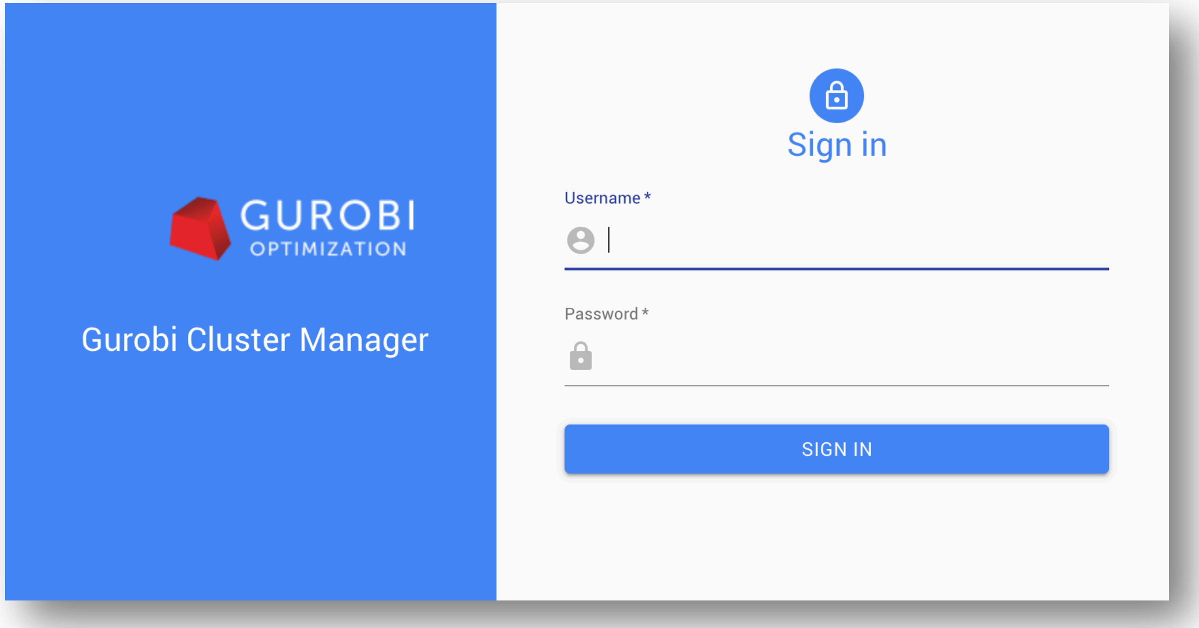
Database can be deployed as a cluster

Many nodes can be added

Copyright © 2020, Gurobi Optimization, LLC

Better Security: User Management

- **User Accounts**
  - Sysadmin creates and manages accounts
  - Users can provide their own passwords
  - Users must login to access the web UI
- **User Roles**
  - Standard users
  - Admin users (kill jobs, access licenses)
  - Sysadmin users
- **Default users (to delete before real deployment)**
  - gurobi/pass
  - admin/admin
  - sysadmin/cluster

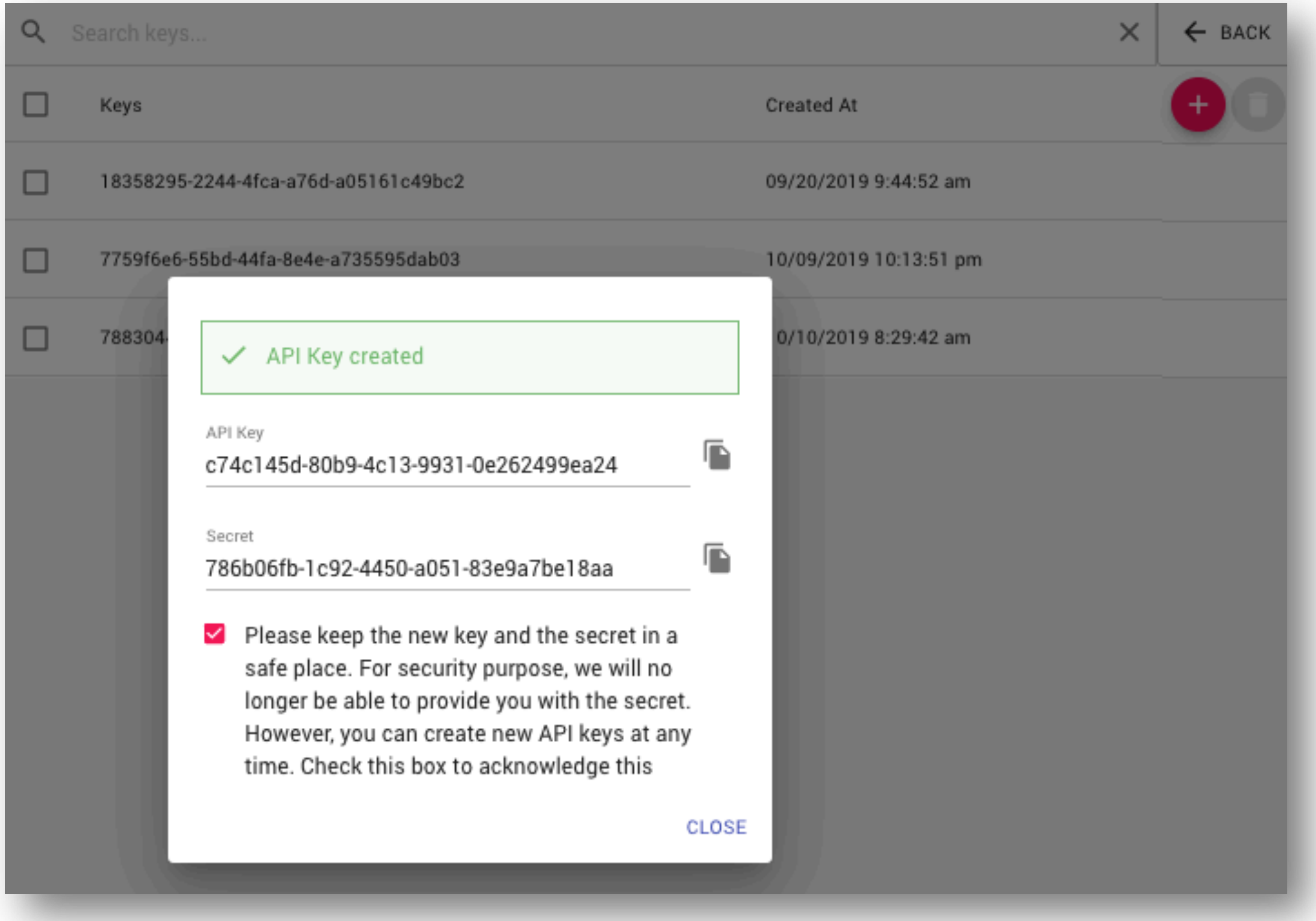


Copyright © 2020, Gurobi Optimization, LLC

Better Security: API Keys and Encryption

- **API keys for applications**
  - Composed of an authentication token and secret key
  - Each user can manage their own API keys
  - Can be embedded into applications

- **Encryption**
  - Support of HTTPS over the wire
  - End-to-end encryption or termination of TLS by the Cluster Manager



□	Keys	Created At
□	18358295-2244-4fca-a76d-a05161c49bc2	09/20/2019 9:44:52 am
□	7759f6e6-55bd-44fa-8e4e-a735595dab03	10/09/2019 10:13:51 pm
□	788304	0/10/2019 8:29:42 am

✓ API Key created

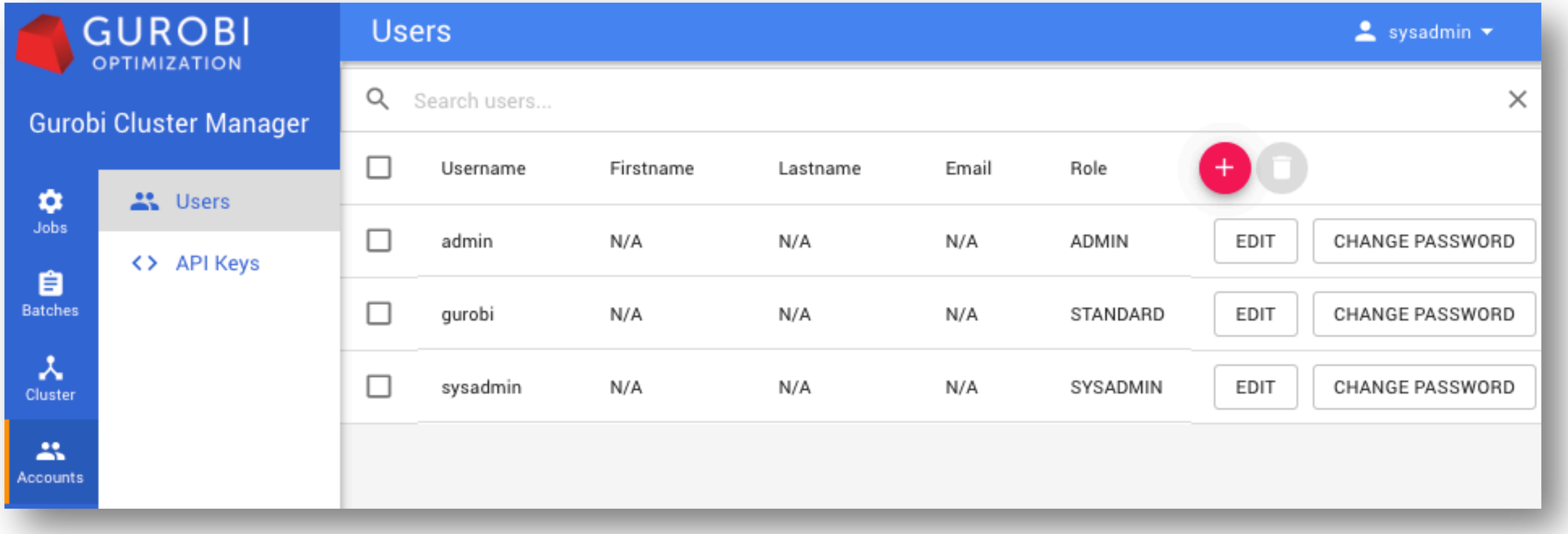
API Key: c74c145d-80b9-4c13-9931-0e262499ea24

Secret: 786b06fb-1c92-4450-a051-83e9a7be18aa

☑ Please keep the new key and the secret in a safe place. For security purpose, we will no longer be able to provide you with the secret. However, you can create new API keys at any time. Check this box to acknowledge this

Copyright © 2020, Gurobi Optimization, LLC

Better Security: Sysadmin



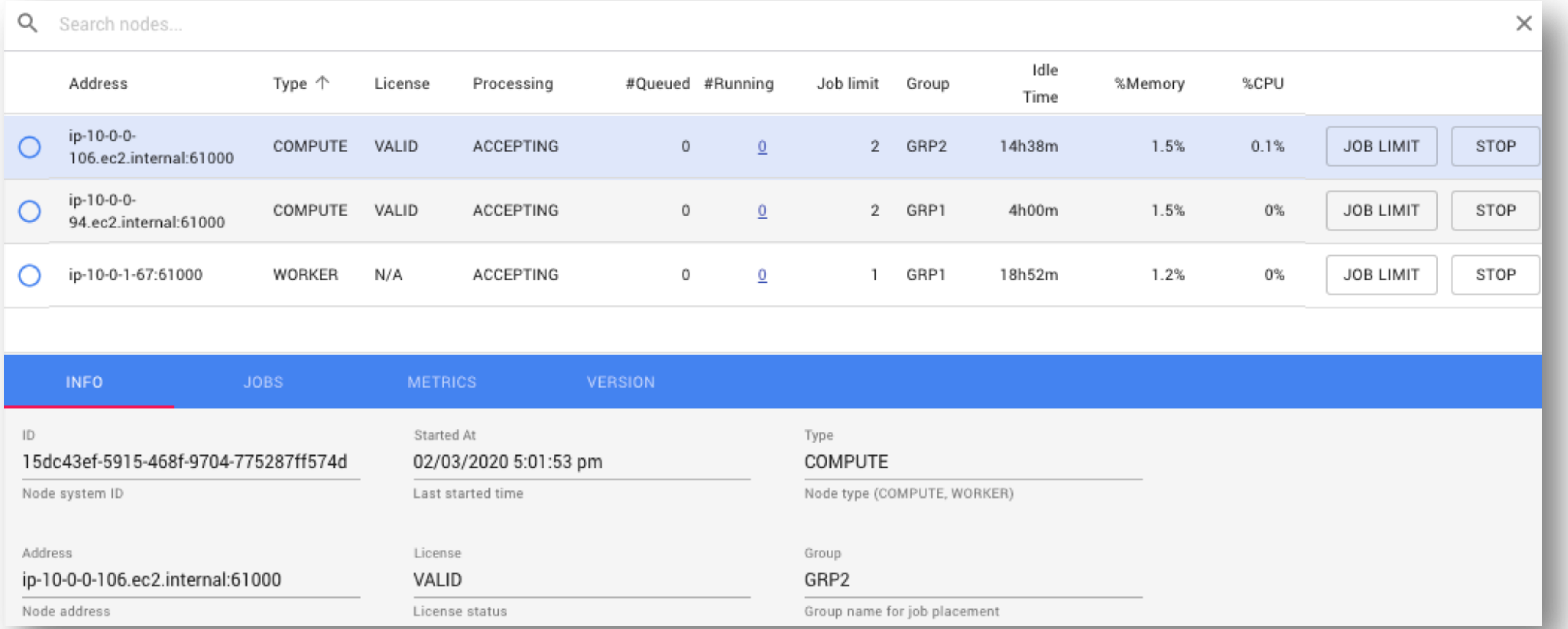
□	Username	Firstname	Lastname	Email	Role		
□	admin	N/A	N/A	N/A	ADMIN	EDIT	CHANGE PASSWORD
□	gurobi	N/A	N/A	N/A	STANDARD	EDIT	CHANGE PASSWORD
□	sysadmin	N/A	N/A	N/A	SYSADMIN	EDIT	CHANGE PASSWORD

- Sysadmin can create, delete and change passwords of all users
- Sysadmin can list and delete any API key

Copyright © 2020, Gurobi Optimization, LLC

Managing the Cluster: Nodes

- List and filter nodes
- Change job limit
- Start/stop nodes



Search nodes...

	Address	Type ↑	License	Processing	#Queued	#Running	Job limit	Group	Idle Time	%Memory	%CPU		
○	ip-10-0-0-106.ec2.internal:61000	COMPUTE	VALID	ACCEPTING	0	0	2	GRP2	14h38m	1.5%	0.1%	JOB LIMIT	STOP
○	ip-10-0-0-94.ec2.internal:61000	COMPUTE	VALID	ACCEPTING	0	0	2	GRP1	4h00m	1.5%	0%	JOB LIMIT	STOP
○	ip-10-0-1-67:61000	WORKER	N/A	ACCEPTING	0	0	1	GRP1	18h52m	1.2%	0%	JOB LIMIT	STOP

ID
15dc43ef-5915-468f-9704-775287ff574d
Node system ID

Started At
02/03/2020 5:01:53 pm
Last started time

Type
COMPUTE
Node type (COMPUTE, WORKER)

Address
ip-10-0-0-106.ec2.internal:61000
Node address

License
VALID
License status

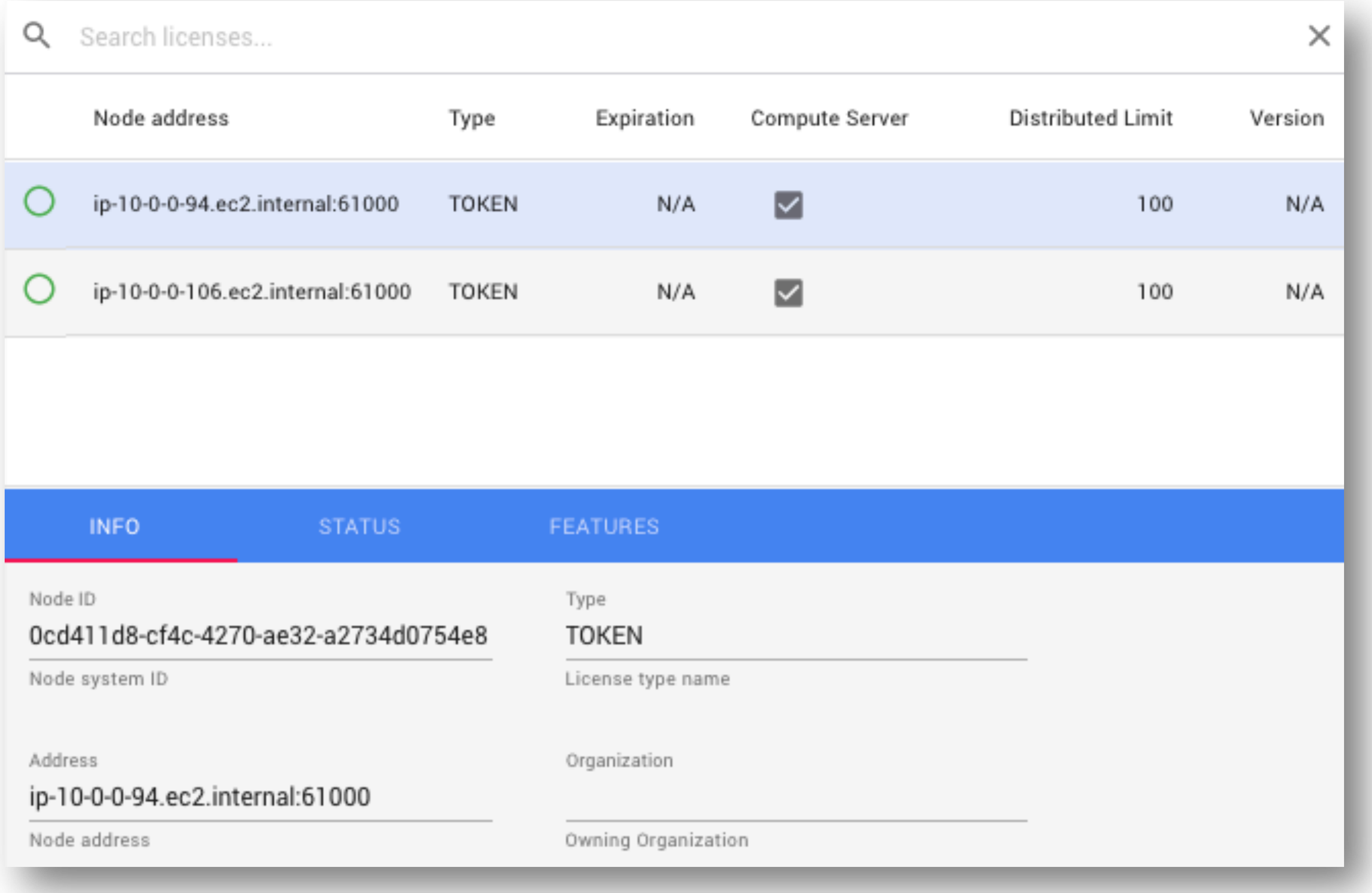
Group
GRP2
Group name for job placement

Copyright © 2020, Gurobi Optimization, LLC

Managing the Cluster: Licenses

- List node licenses
- Check expiration

- V9 adds support of token server licensing



Search licenses...

	Node address	Type	Expiration	Compute Server	Distributed Limit	Version
○	ip-10-0-0-94.ec2.internal:61000	TOKEN	N/A	☑	100	N/A
○	ip-10-0-0-106.ec2.internal:61000	TOKEN	N/A	☑	100	N/A

Node ID
0cd411d8-cf4c-4270-ae32-a2734d0754e8
Node system ID

Type
TOKEN
License type name

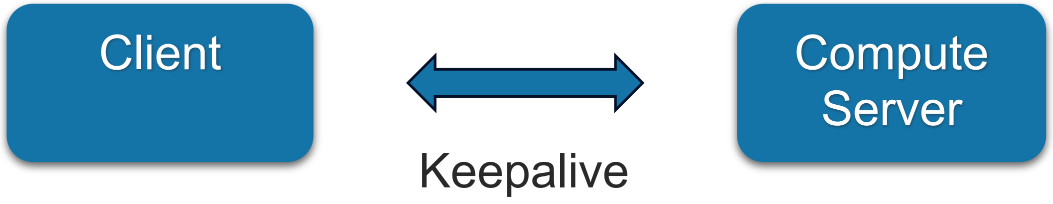
Address
ip-10-0-0-94.ec2.internal:61000
Node address

Organization

Owning Organization

Copyright © 2020, Gurobi Optimization, LLC

Submitting Jobs: Interactive Optimization (V8 and V9)



Execute optimization commands

- **Interactive optimization tasks (Standard Jobs)**
  - Client always connected to server
  - Client remotely controls all optimization steps (model building, optimization)
  - Callbacks are supported

Copyright © 2020, Gurobi Optimization, LLC

Submitting Jobs: Client Configuration for Cluster Manager

- Client license file has new properties
  - **CSMANAGER**: URL to the manager
  - **CSAPIACCESSID** and **CSAPISECRET**: API key and secret
  - **USERNAME** and **PASSWORD**: username/password (not recommended)
  - **CSAUTHTOKEN**: time limited token after using the login command
- New login command to update the client license file

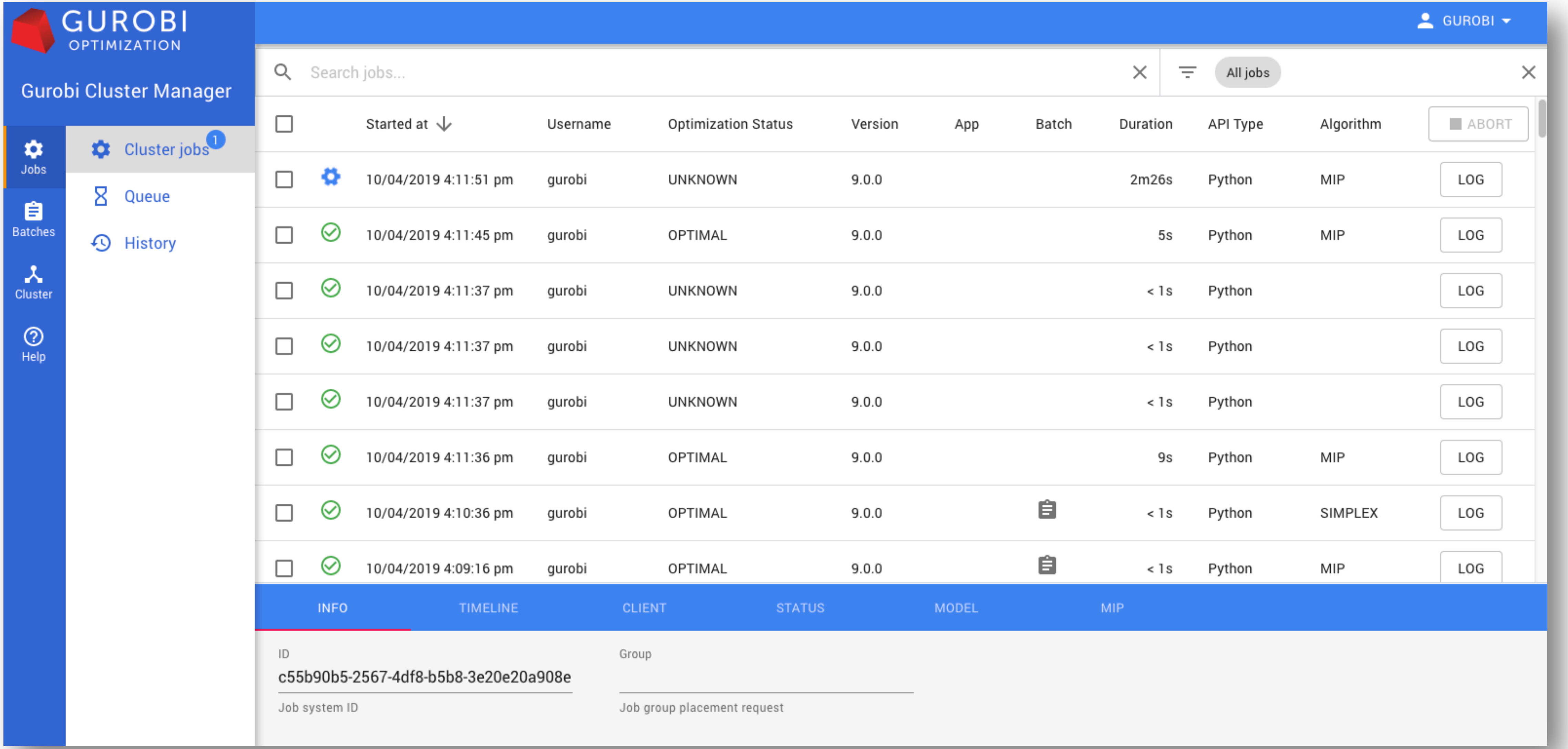
```
grbcluster login --manager=http://localhost:61080 --username=gurobi --password=pass

gurobi_cl glass4.mps
```

- New environment parameters

```
with gp.Env(empty=True) as env:
        env.setParam('CSManager', 'http://localhost:61080')
        env.setParam('CSAPIAccessID', '0e8c35d5-ff20-4e5d-a639-10105e56b264')
        env.setParam('CSAPISecret', 'd588f010-ad47-4310-933e-1902057661c9')
        env.start()
```

Copyright © 2020, Gurobi Optimization, LLC

Submitting Jobs: Monitoring and History



☐		Started at ↓	Username	Optimization Status	Version	App	Batch	Duration	API Type	Algorithm	ABORT
☐	⚙	10/04/2019 4:11:51 pm	gurobi	UNKNOWN	9.0.0			2m26s	Python	MIP	LOG
☐	✓	10/04/2019 4:11:45 pm	gurobi	OPTIMAL	9.0.0			5s	Python	MIP	LOG
☐	✓	10/04/2019 4:11:37 pm	gurobi	UNKNOWN	9.0.0			< 1s	Python		LOG
☐	✓	10/04/2019 4:11:37 pm	gurobi	UNKNOWN	9.0.0			< 1s	Python		LOG
☐	✓	10/04/2019 4:11:37 pm	gurobi	UNKNOWN	9.0.0			< 1s	Python		LOG
☐	✓	10/04/2019 4:11:36 pm	gurobi	OPTIMAL	9.0.0			9s	Python	MIP	LOG
☐	✓	10/04/2019 4:10:36 pm	gurobi	OPTIMAL	9.0.0		📋	< 1s	Python	SIMPLEX	LOG
☐	✓	10/04/2019 4:09:16 pm	gurobi	OPTIMAL	9.0.0		📋	< 1s	Python	MIP	LOG

ID
c55b90b5-2567-4df8-b5b8-3e20e20a908e
Job system ID

Group

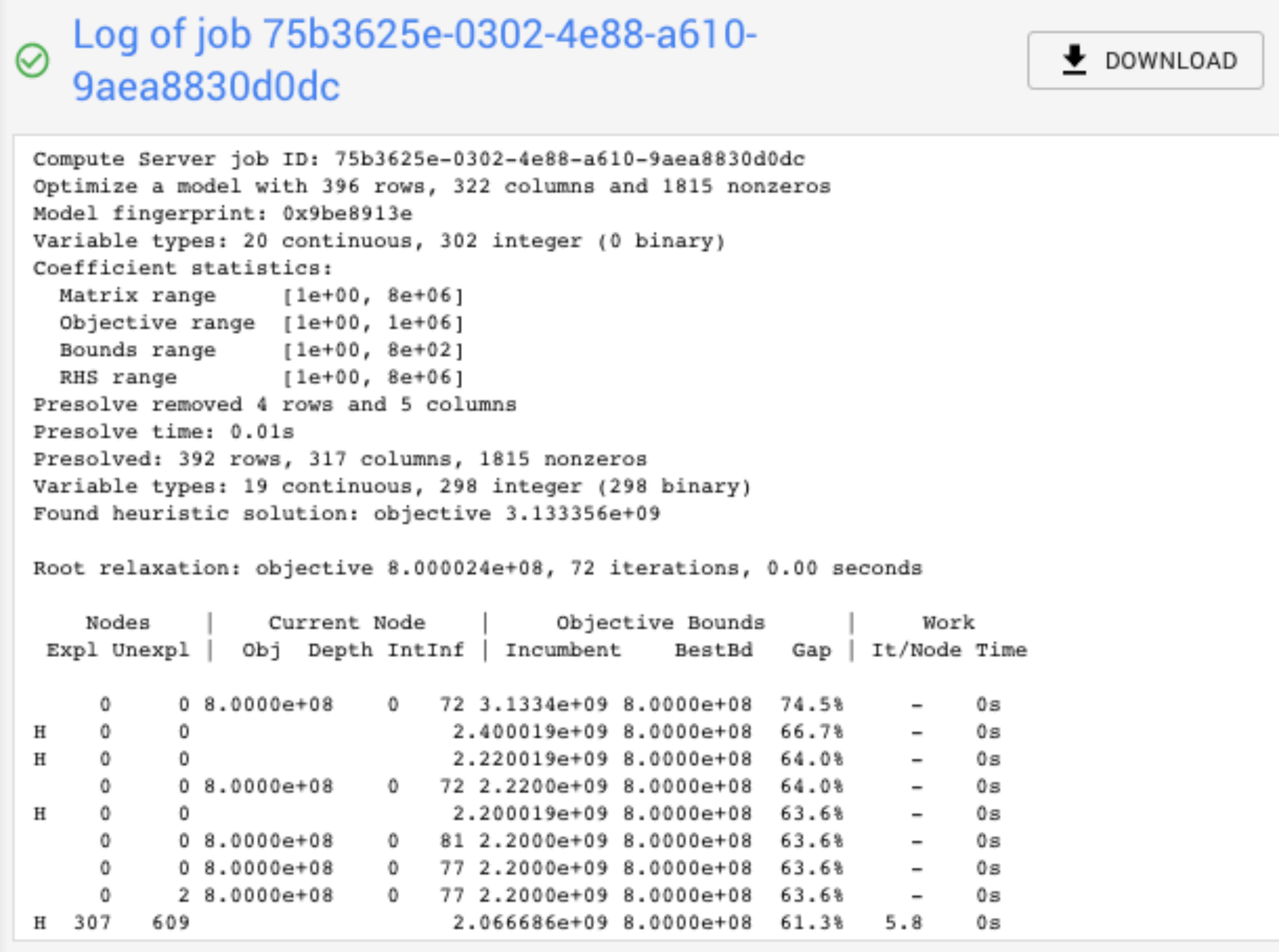
Job group placement request

Copyright © 2020, Gurobi Optimization, LLC

Submitting Jobs: Logs

- Logs can be accessed live when the job is running

- Logs are also archived in the history

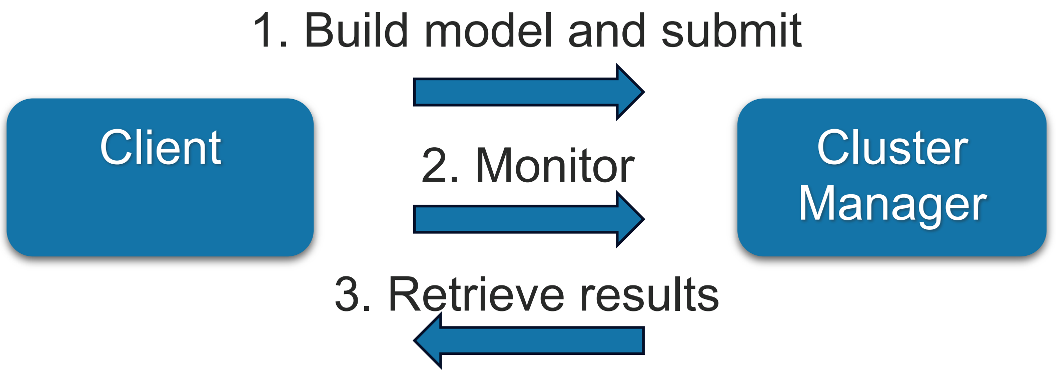


Log of job 75b3625e-0302-4e88-a610-9aea8830d0dc

```
Compute Server job ID: 75b3625e-0302-4e88-a610-9aea8830d0dc
Optimize a model with 396 rows, 322 columns and 1815 nonzeros
Model fingerprint: 0x9be8913e
Variable types: 20 continuous, 302 integer (0 binary)
Coefficient statistics:
  Matrix range     [1e+00, 8e+06]
  Objective range  [1e+00, 1e+06]
  Bounds range     [1e+00, 8e+02]
  RHS range        [1e+00, 8e+06]
Presolve removed 4 rows and 5 columns
Presolve time: 0.01s
Presolved: 392 rows, 317 columns, 1815 nonzeros
Variable types: 19 continuous, 298 integer (298 binary)
Found heuristic solution: objective 3.133356e+09

Root relaxation: objective 8.000024e+08, 72 iterations, 0.00 seconds

    Nodes    |    Current Node    |     Objective Bounds      |     Work
 Expl Unexpl |  Obj  Depth IntInf | Incumbent    BestBd   Gap | It/Node Time

     0     0 8.0000e+08    0   72 3.1334e+09 8.0000e+08  74.5%     -    0s
H    0     0                    2.400019e+09 8.0000e+08  66.7%     -    0s
H    0     0                    2.220019e+09 8.0000e+08  64.0%     -    0s
     0     0 8.0000e+08    0   72 2.2200e+09 8.0000e+08  64.0%     -    0s
H    0     0                    2.200019e+09 8.0000e+08  63.6%     -    0s
     0     0 8.0000e+08    0   81 2.2000e+09 8.0000e+08  63.6%     -    0s
     0     0 8.0000e+08    0   77 2.2000e+09 8.0000e+08  63.6%     -    0s
     0     2 8.0000e+08    0   77 2.2000e+09 8.0000e+08  63.6%     -    0s
H  307   609                    2.066686e+09 8.0000e+08  61.3%   5.8    0s
```

Copyright © 2020, Gurobi Optimization, LLC

Batch Optimization: Non-Interactive Tasks (New in V9)

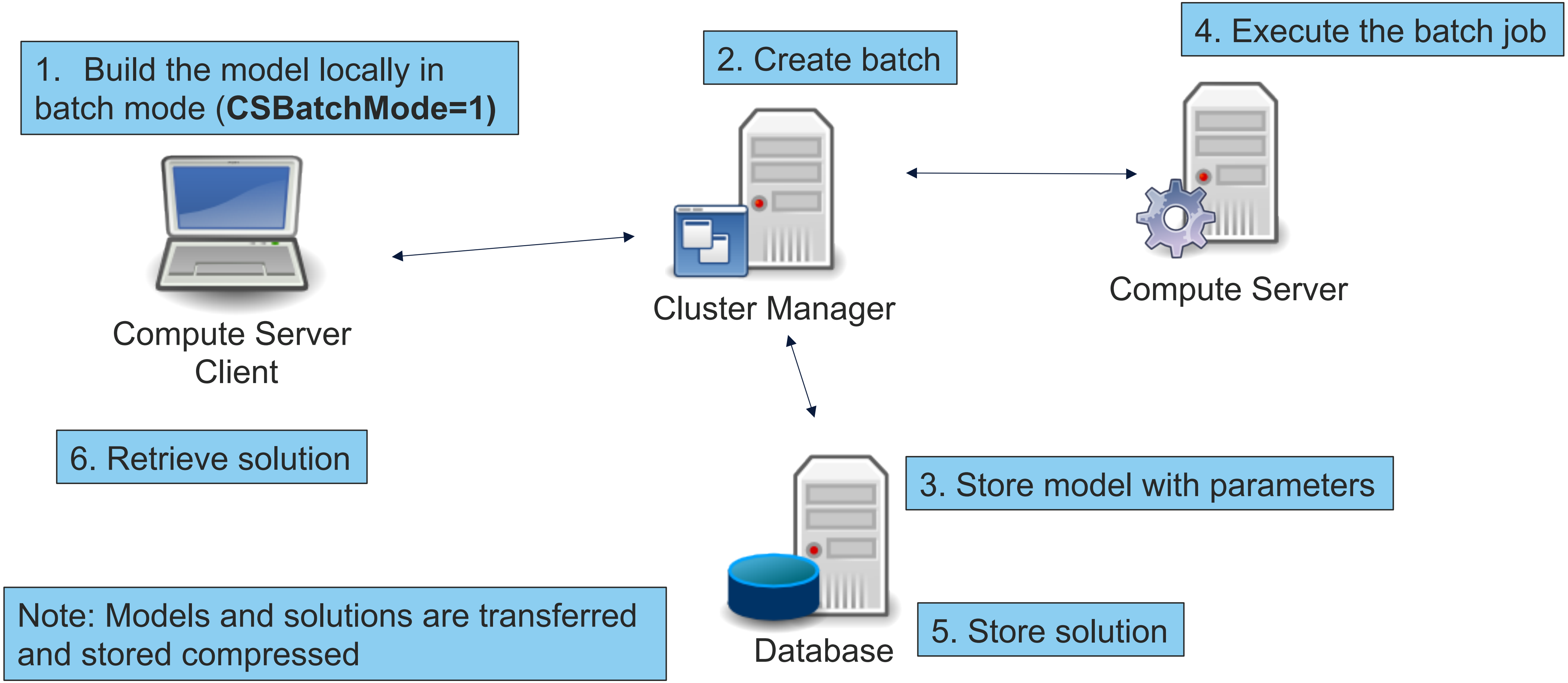


Batch processing of optimization

- **Three steps:**
  - Build locally and submit the model, a batch ID is returned
  - Monitor using the batch ID
  - Retrieve the results as a JSON file
- **Benefits:**
  - Clients can disconnect in between steps
  - No client resources are used once the model has been submitted
  - Results can be retrieved by a different service of the client application

Copyright © 2020, Gurobi Optimization, LLC

Batch Optimization: Architecture



1. Build the model locally in batch mode (**CSBatchMode=1**)

2. Create batch

3. Store model with parameters

4. Execute the batch job

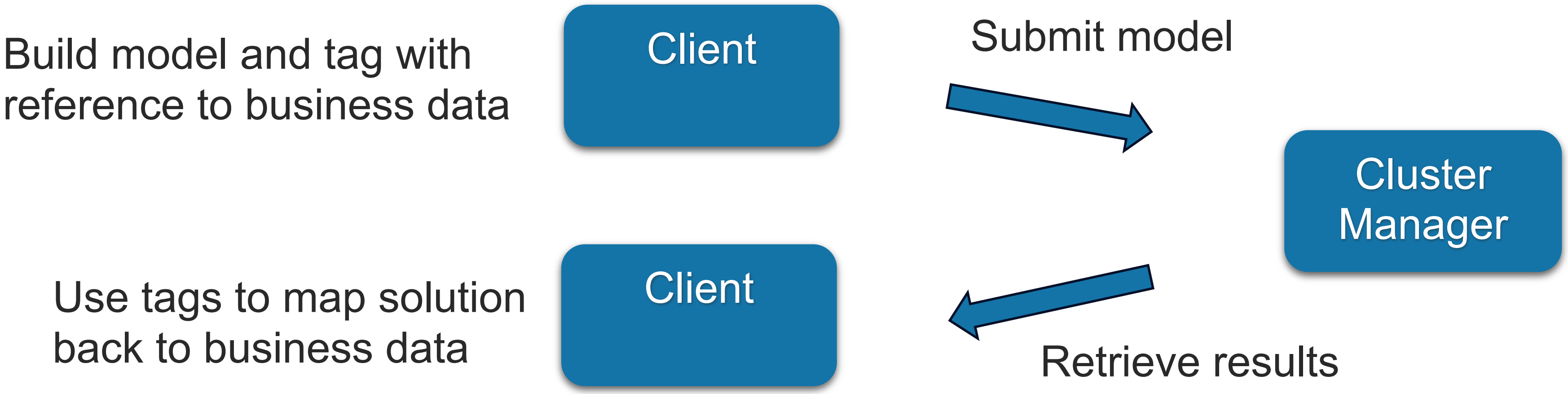
5. Store solution

6. Retrieve solution

Note: Models and solutions are transferred and stored compressed

Copyright © 2020, Gurobi Optimization, LLC

Batch Optimization: Variable and Constraint Tagging


Build model and tag with reference to business data
Client
Submit model
Cluster Manager
Use tags to map solution back to business data
Client
Retrieve results

- **Benefits of tagging**
  - Solution only contains the tagged variables and constraints
  - Solution can be interpreted standalone, without the submitted model context

Copyright © 2020, Gurobi Optimization, LLC

Batch Optimization: Model Building Example

- Create an environment connected to a Cluster Manager
- Set CSBatchMode=1
- Tag relevant variables and constraints with custom data
- Call model.optimizeBatch()
- Batch ID is returned

```
with gp.Model("assignment", env=env) as m:
    # Assignment variables: x[w,s] == 1 if worker w is assigned to shift s.
    x = m.addVars(availability, vtype=GRB.BINARY, name="x")

    # Set tags encoding the assignments for later retrieval of the schedule.
    # Each tag is a JSON string of the format
    #   {
    #     "Worker": "<Name of the worker>",
    #     "Shift":  "String representation of the shift"
    #   }
    #
    for k, v in x.items():
        name, timeslot = k
        d = {"Worker": name, "Shift": shiftname[timeslot]}
        v.VTag = json.dumps(d)
```

Copyright © 2020, Gurobi Optimization, LLC

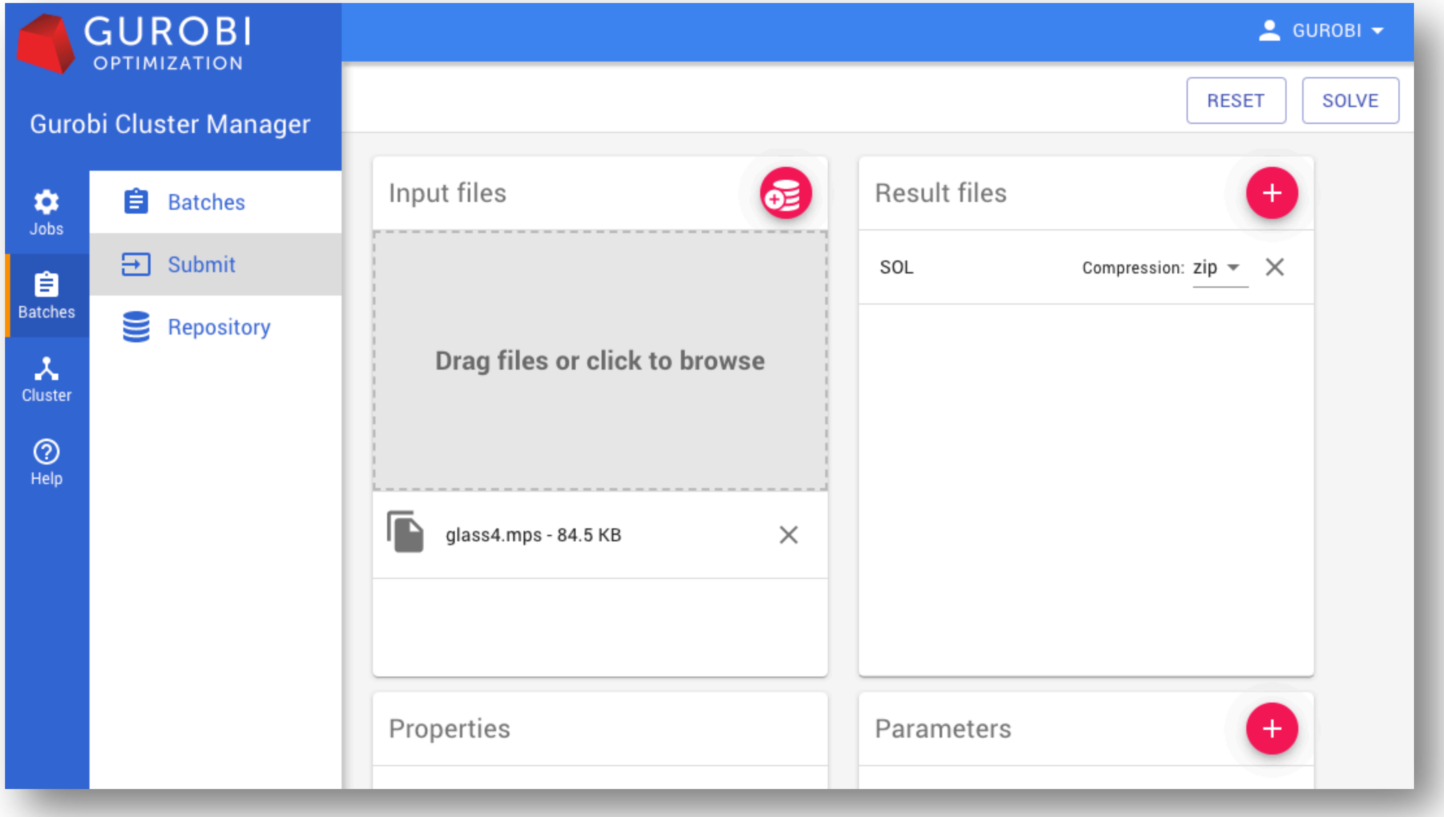
Batch Optimization: Retrieving the Solution

- Monitor the batch status
- When completed, get the solution file
- JSON format for easy integration
- Only tagged elements will be exported
- More details can be exported with JSONSolDetail parameter
- Extract business solution using variable value and related tags

```
{
  "SolutionInfo": {
    …
  },
  "Vars": [
    {
      "VTag": [
        "{\"Shift\": \"Wednesday 8:00\", \"Worker\": \"Amy\"}"
      ],
      "X": "1"
    },
    {
      "VTag": [
        "{\"Shift\": \"Thursday 8:00\", \"Worker\": \"Amy\"}"
      ],
      "X": "1"
    },
…
```

Copyright © 2020, Gurobi Optimization, LLC

Batch Optimization: UI and Command Line



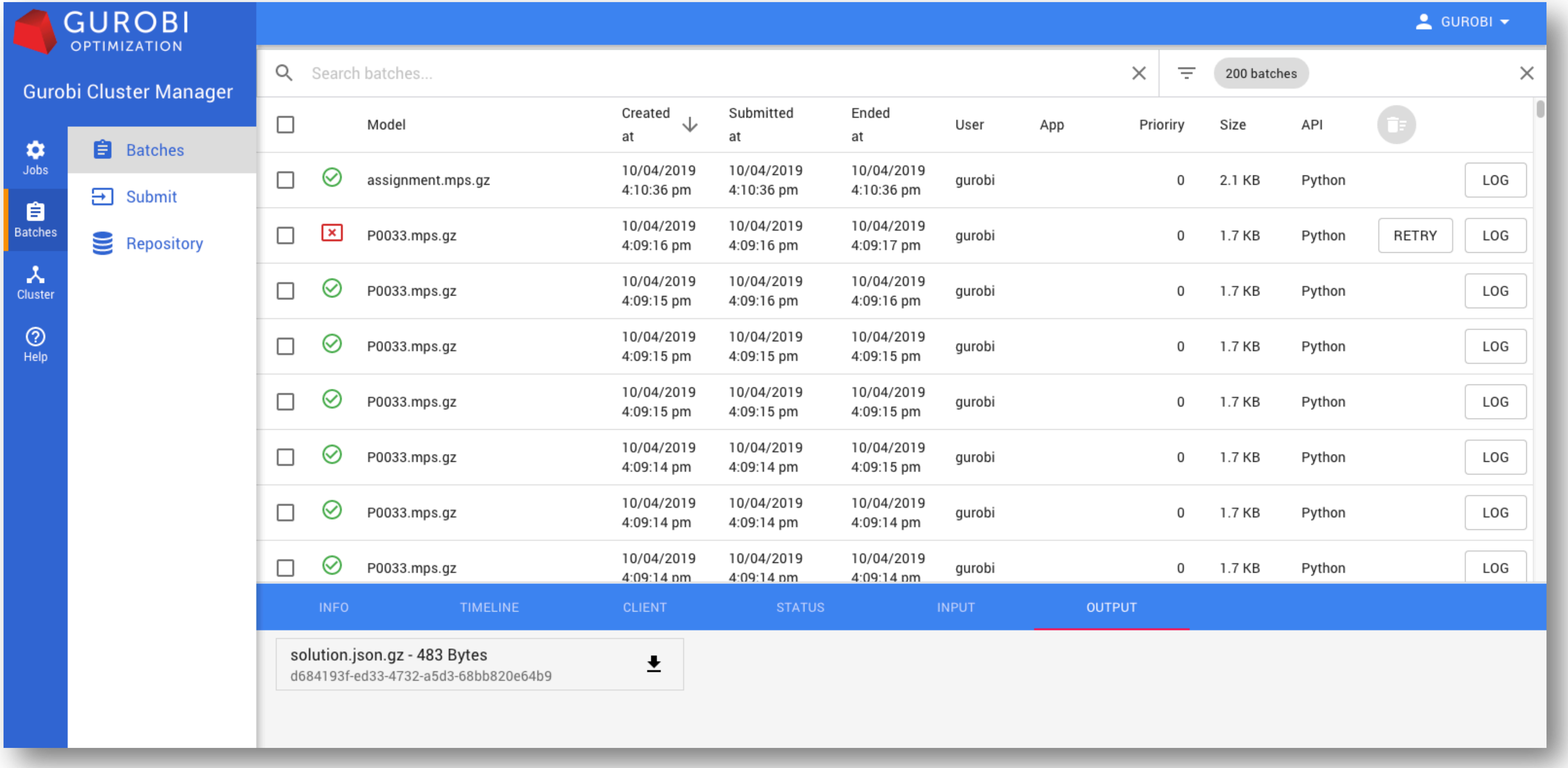
glass4.mps - 84.5 KB

SOL Compression: zip

```
$ grbcluster batch solve glass4.mps ResultFile=glass4.sol
info  : Batch 10230a87-8ab9-4277-8c00-e949627516cb created
info  : Uploading glass4.mps...
info  : Batch 10230a87-8ab9-4277-8c00-e949627516cb submitted…
```

Copyright © 2020, Gurobi Optimization, LLC

Batch Optimization: Monitoring



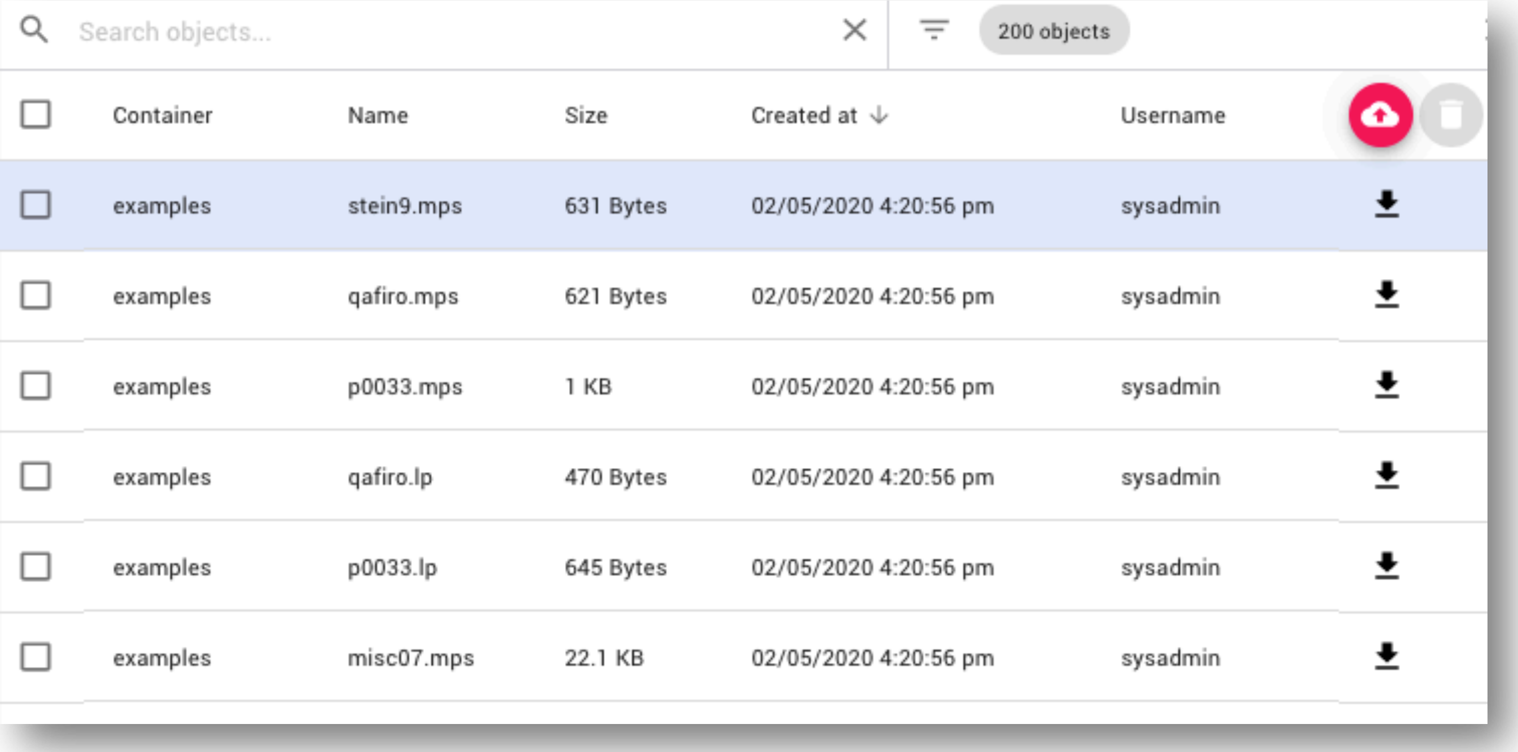
☐		Model	Created at	Submitted at	Ended at	User	App	Priroiry	Size	API		
☐	✓	assignment.mps.gz	10/04/2019 4:10:36 pm	10/04/2019 4:10:36 pm	10/04/2019 4:10:36 pm	gurobi		0	2.1 KB	Python		LOG
☐	☒	P0033.mps.gz	10/04/2019 4:09:16 pm	10/04/2019 4:09:16 pm	10/04/2019 4:09:17 pm	gurobi		0	1.7 KB	Python	RETRY	LOG
☐	✓	P0033.mps.gz	10/04/2019 4:09:15 pm	10/04/2019 4:09:16 pm	10/04/2019 4:09:16 pm	gurobi		0	1.7 KB	Python		LOG
☐	✓	P0033.mps.gz	10/04/2019 4:09:15 pm	10/04/2019 4:09:15 pm	10/04/2019 4:09:15 pm	gurobi		0	1.7 KB	Python		LOG
☐	✓	P0033.mps.gz	10/04/2019 4:09:15 pm	10/04/2019 4:09:15 pm	10/04/2019 4:09:15 pm	gurobi		0	1.7 KB	Python		LOG
☐	✓	P0033.mps.gz	10/04/2019 4:09:14 pm	10/04/2019 4:09:14 pm	10/04/2019 4:09:15 pm	gurobi		0	1.7 KB	Python		LOG
☐	✓	P0033.mps.gz	10/04/2019 4:09:14 pm	10/04/2019 4:09:14 pm	10/04/2019 4:09:14 pm	gurobi		0	1.7 KB	Python		LOG
☐	✓	P0033.mps.gz	10/04/2019 4:09:14 pm	10/04/2019 4:09:14 pm	10/04/2019 4:09:14 pm	gurobi		0	1.7 KB	Python		LOG

solution.json.gz - 483 Bytes
d684193f-ed33-4732-a5d3-68bb820e64b9

Copyright © 2020, Gurobi Optimization, LLC

Batch Optimization: File Repository

- Share models or parameters with your team
- Avoid uploading again the same files



Search objects...

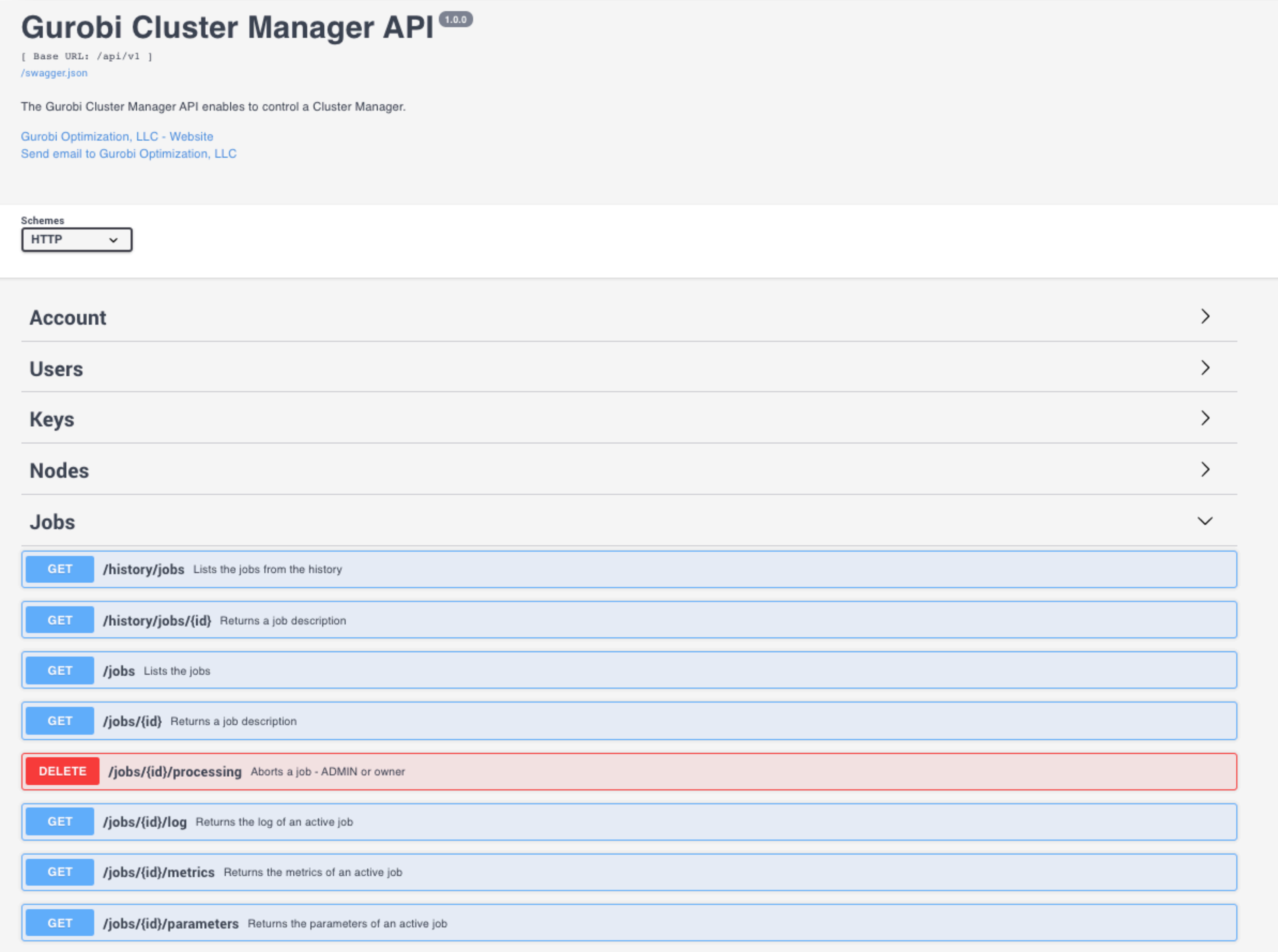
200 objects

□	Container	Name	Size	Created at ↓	Username	
□	examples	stein9.mps	631 Bytes	02/05/2020 4:20:56 pm	sysadmin	⬇
□	examples	qafiro.mps	621 Bytes	02/05/2020 4:20:56 pm	sysadmin	⬇
□	examples	p0033.mps	1 KB	02/05/2020 4:20:56 pm	sysadmin	⬇
□	examples	qafiro.lp	470 Bytes	02/05/2020 4:20:56 pm	sysadmin	⬇
□	examples	p0033.lp	645 Bytes	02/05/2020 4:20:56 pm	sysadmin	⬇
□	examples	misc07.mps	22.1 KB	02/05/2020 4:20:56 pm	sysadmin	⬇

Copyright © 2020, Gurobi Optimization, LLC

REST API

- Can be used to integrate and automate cluster monitoring and management
- Covers all the entities
  - Users
  - API Keys
  - Cluster nodes
  - Jobs and history
  - Model and solution files
  - Batches



Gurobi Cluster Manager API 1.0.0

[Base URL: /api/v1]

/swagger.json

The Gurobi Cluster Manager API enables to control a Cluster Manager.

Gurobi Optimization, LLC - Website
Send email to Gurobi Optimization, LLC

Schemes: HTTP

Account

Users

Keys

Nodes

Jobs

GET /history/jobs Lists the jobs from the history

GET /history/jobs/{id} Returns a job description

GET /jobs Lists the jobs

GET /jobs/{id} Returns a job description

DELETE /jobs/{id}/processing Aborts a job - ADMIN or owner

GET /jobs/{id}/log Returns the log of an active job

GET /jobs/{id}/metrics Returns the metrics of an active job

GET /jobs/{id}/parameters Returns the parameters of an active job

Demo

- **Installation**
  - Start MongoDB and manager
  - Start two local nodes
  - Display nodes and licenses

- **Submitting Jobs**
  - Login as a user on the command line
  - Submit a job
  - Display the job info in the Cluster Manager

- **Submitting Batches**
  - Run workforce batch example
  - Submit a batch from the UI
  - Submit a batch from the command line

Copyright © 2020, Gurobi Optimization, LLC

Thank you

GUROBI OPTIMIZATION

The World's Fastest Solver