Hewlett Packard
Enterprise

OLAF

CRYOSPARC TUTORIAL

HPE Cray

April, 2020

WHAT IS CRYOSPARC?

- CryoSPARC (**Cryo**-EM **S**ingle **P**article **A**b-Initio **R**econstruction and **C**lassification) is a state of the art HPC software solution for complete processing of single-particle cryo-electron microscopy (cryo-EM) data.
- https://cryosparc.com/

INSTALL AND SETUP

- Export your License ID as an environment variable
  - export LICENSE_ID="<license_id>"
- Download the two files into tarball archives
  - curl -L https://get.cryosparc.com/download/master-latest/$LICENSE_ID -o cryosparc_master.tar.gz
  - curl -L https://get.cryosparc.com/download/worker-latest/$LICENSE_ID -o cryosparc_worker.tar.gz
- Extract the downloaded files
  - tar -xf cryosparc_master.tar.gz cryosparc_master
  - tar -xf cryosparc_worker.tar.gz cryosparc_worker

INSTALL CRYSPARC MASTER PACKAGE

- Install the cryosparc_master package
  - cd cryosparc_master
  - ./install.sh --license $LICENSE_ID --hostname <master_hostname> --dbpath <db_path> --port <port_number> [--insecure] [--allowroot] [--yes]
  - Example
    - ./install.sh --license $LICENSE_ID --hostname olaf1 --port 39200 --dbpath /opt/cryoem/cryosparc-center/3.0.1/cryosparc2_database
- Start cryoSPARC
  - ./bin/cryosparcm start
- Create the first user
  - cryosparcm createuser --email “<user email>” --password “<user password>” --name “<login username>” --firstname “<given name>” --lastname “<surname>”
  - Example
    - cryosparcm createuser --email crayadmin@olaf --password crayadm --username crayadm --firstname edu --lastname edu

INSTALL CRYOSPARC WORKER PACKAGE

- GPU Worker node cryoSPARC installation
  - cd cryosparc_worker
  - ./install.sh --license $LICENSE_ID --cudapath <cuda_path> [--yes]
  - Example
    - ./install.sh --license $LICENSE_ID --cudapath /opt/compiler/nvidia/cuda-10.2

CONNECTION OF CLUSTER WORKLOAD MANAGER

- Connecting a Cluster to cryoSPARC
  - To register the cluster, cluster_info.json and clusters_script.sh need to be provided.
  - cluster_info.json

name : "cluster1"
A unique name for the cluster to be connected (multiple clusters can be connected)
worker_bin_path : "/path/to/cryosparc_worker/bin/cryosparcw"
Path on cluster nodes to the cryosparcw entry point for worker process
cache_path : "/path/to/local/SSD/on/cluster/nodes"
Path on cluster nodes that is a writable location on local SSD on each cluster node. This might be /scratch or similar. This path MUST be the same on all cluster nodes. Note that the installer does not check that this path exists, so make sure it does and is writable. If you plan to use the cluster nodes without SSD, you can leave this blank.
send_cmd_tpl : "ssh loginnode {{ command }}"
Used to send a command to be executed by a cluster node (in case the cryosparc master is not able to directly use cluster commands). If your cryosparc master node is able to directly use cluster commands (like qsub etc) then this string can be just "{{ command }}"
qsub_cmd_tpl : "qsub {{ script_path_abs }}"
The exact command used to submit a job to the cluster, where the job is defined in the cluster script located at {{ script_path_abs }}. This string can also use any of the variables defined below that are available inside the cluster script (num_gpus, num_cpus, etc)
qstat_cmd_tpl : "qstat -as {{ cluster_job_id }}"
Cluster command that will report back the status of cluster job with id {{ cluster_job_id }}.
qdel_cmd_tpl : "qdel {{ cluster_job_id }}"
Cluter command that will kill and remove {{ cluster_job_id }} from the queue.
qinfo_cmd_tpl : "qstat -q"
General cluster information command
transfer_cmd_tpl : "scp {{ src_path }} loginnode:{{ dest_path }}"
Command that can be used to transfer a file {{ src_path }} on the cryosparc master node to {{ dest_path }} on the cluster nodes. This is used when the master node is remotely updating a cluster worker installation. This is optional - if it is incorrect or omitted, you can manually update the cluster worker installation.

CONNECTION OF CLUSTER WORKLOAD MANAGER

- cluster_script.sh

```
{{ script_path_abs }} # the absolute path to the generated submission script
{{ run_cmd }} # the complete command-line string to run the job
{{ num_cpu }} # the number of CPUs needed
{{ num_gpu }} # the number of GPUs needed.
{{ ram_gb }} # the amount of RAM needed in GB
{{ job_dir_abs }} # absolute path to the job directory
{{ project_dir_abs }} # absolute path to the project dir
{{ job_log_path_abs }} # absolute path to the log file for the job
{{ worker_bin_path }} # absolute path to the cryosparc worker command
{{ run_args }} # arguments to be passed to cryosparcw run
{{ project_uid }} # uid of the project
{{ job_uid }} # uid of the job
{{ job_creator }} # name of the user that created the job (may contain spaces)
{{ cryosparc_username }} # cryosparc username of the user that created the job (usually an email)
```

- To actually create or set a configuration for a cluster in cryoSPARC
  - cryosparcm cluster connect # connects new or updates existing cluster configuration, reading cluster_info.json and cluster_script.sh from the current directory, using the name from cluster_info.json
  - Other commands
    - cryosparcm cluster example <cluster_type> # dumps out config and script template files to current working directory. Examples are available for pbs and slurm schedulers but others should be very similar
    - cryosparcm cluster dump <name> # dumps out existing config and script to current working directory
    - cryosparcm cluster remove <name> # removes a cluster configuration from the scheduler

CONNECTION OF CLUSTER WORKLOAD MANAGER

- Example
  - cluster_info.json

```
{
"qdel_cmd_tpl": "scancel {{ cluster_job_id }}",
"worker_bin_path": "/opt/cryoem/cryosparc-center/3.0.1/cryosparc2_worker/bin/cryosparcw",
"title": "cryo_all",
"qinfo_cmd_tpl": "sinfo --format='%.8N %.6D %.10P %.6T %.14C %.5c %.6z %.7m %.7G %.9d %20E'",
"qsub_cmd_tpl": "sbatch {{ script_path_abs }}",
"qstat_cmd_tpl": "squeue -j {{ cluster_job_id }}",
"cache_quota_mb": null,
"cache_reserve_mb": null,
"send_cmd_tpl": "{{ command }}",
"name": "cryo_all"
}
```

  - Cluster_script.sh

```
#!/bin/bash
#SBATCH --job-name=cryosparc_center_{{ project_uid }}_{{ job_uid }}
#SBATCH --partition=cryoem
#SBATCH --output={{ job_log_path_abs }}
#SBATCH --error={{ job_log_path_abs }}
#SBATCH --gpus={{ num_gpu }}
#SBATCH --gres=gpu:{{ num_gpu }}
#SBATCH --ntasks={{ num_cpu }}
#SBATCH --cpus-per-task=1

source $MODULESHOME/init/bash
module load cuda/10.2
module load cryosparc-center/3.0.1-worker

{{ run_cmd }}
```

UPDATES

- Run cryosparcm update command to check for cryoSPARC updates
  - cryosparcm update –check
  - Example

    $ cryosparcm update --check
    CryoSPARC current version v2.15.0
    update starting on Wed Mar 18 12:09:52 EDT 2020
    current version v2.15.0
    new version v3.0.0
    Update available!
  - cryosparcm update --list
  - Example

    $ cryosparcm update --list
    CryoSPARC current version v2.14.0
    update starting on Wed Mar 18 12:11:42 EDT 2020
    Available versions:
    v2.0.18
    v2.0.20
    v2.0.23
    ...
    v2.14.2
    v2.15.0
    v3.0.0
    To install a specific version, use
    $ cryosparcm update --version=vXX.YY.ZZ[-branchname]

UPDATES

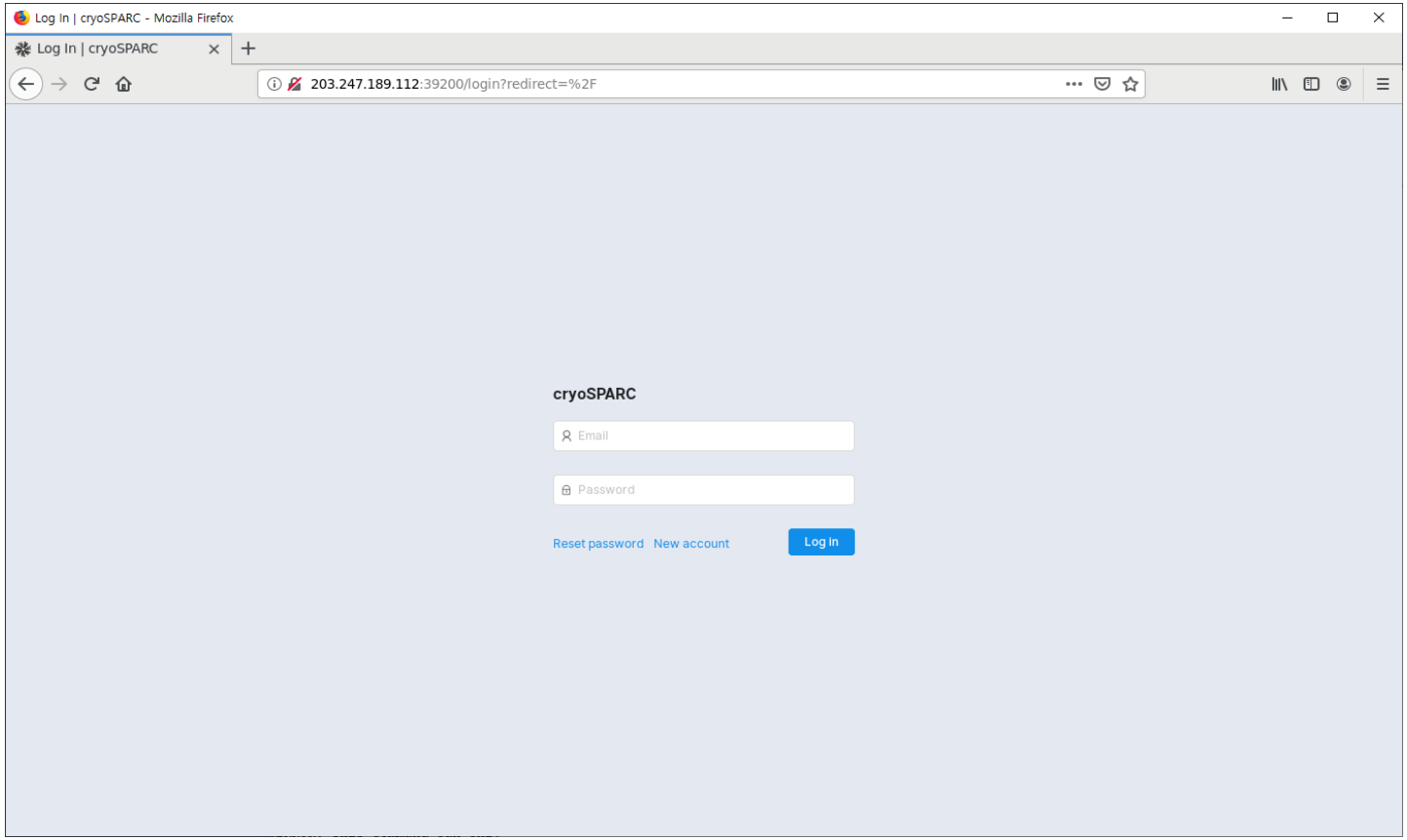
- Master update
  - cryosparcm update or cryosparcm update –version=vXX.YY.ZZ
- Worker update
  - Copy cryosparc2_worker.tar.gz that it is downloaded when updating master into worker installation directory.
  - bin/cryosparcm update

CRYO-EM DATA PROCESSING IN CRYOSPARC: INTRODUCTORY TUTORIAL

- https://guide.cryosparc.com/processing-data/cryo-em-data-processing-in-cryosparc-introductory-tutorial
- A subset of 20 movies from the EMPIAR-10025 T20S Proteasome dataset
- Launch firefox or web browser on your laptop
  - module load cryosparc-center/3.0.1-master
  - firefox

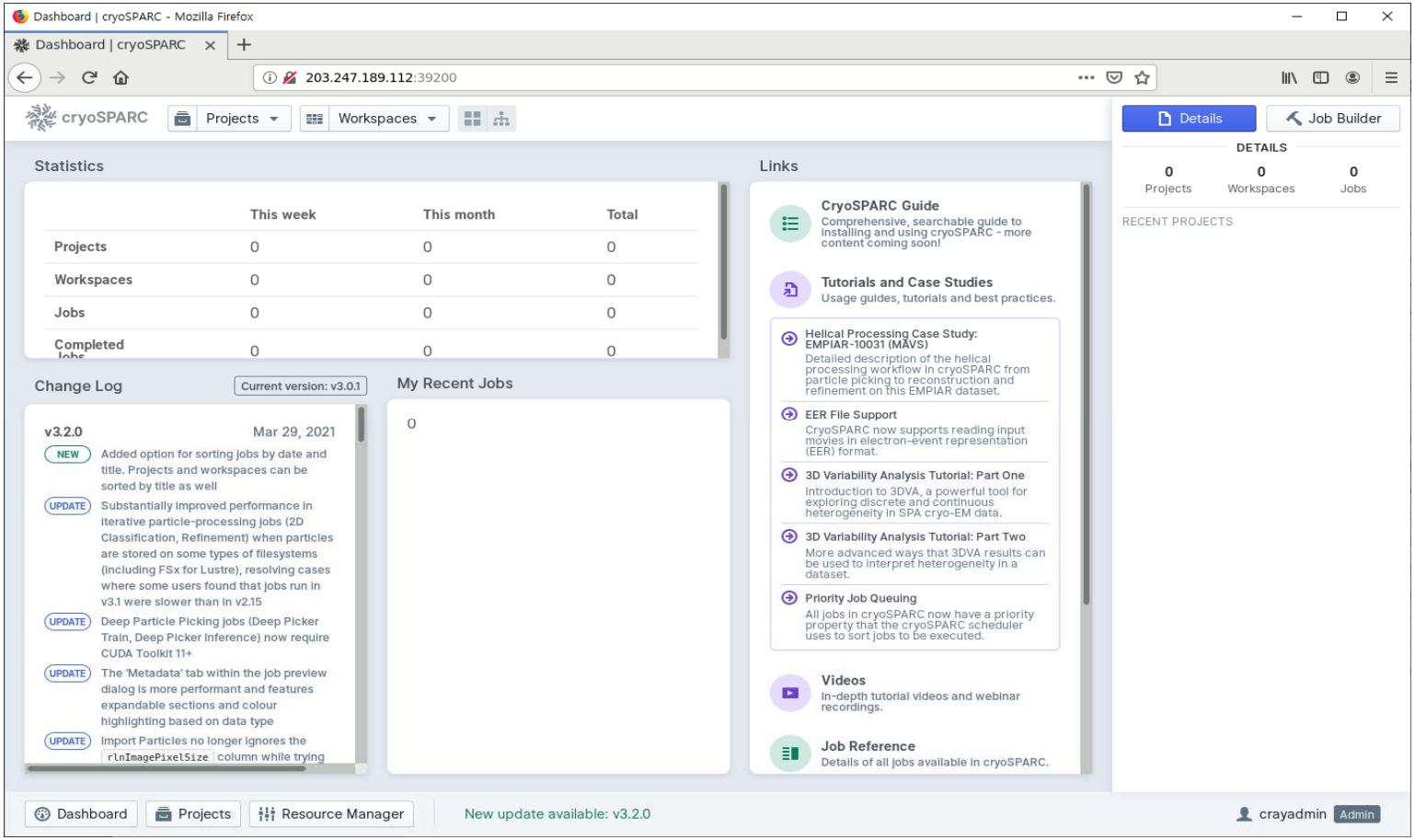
CONNECT TO MASTER

- **http://203.247.189.112:39200**



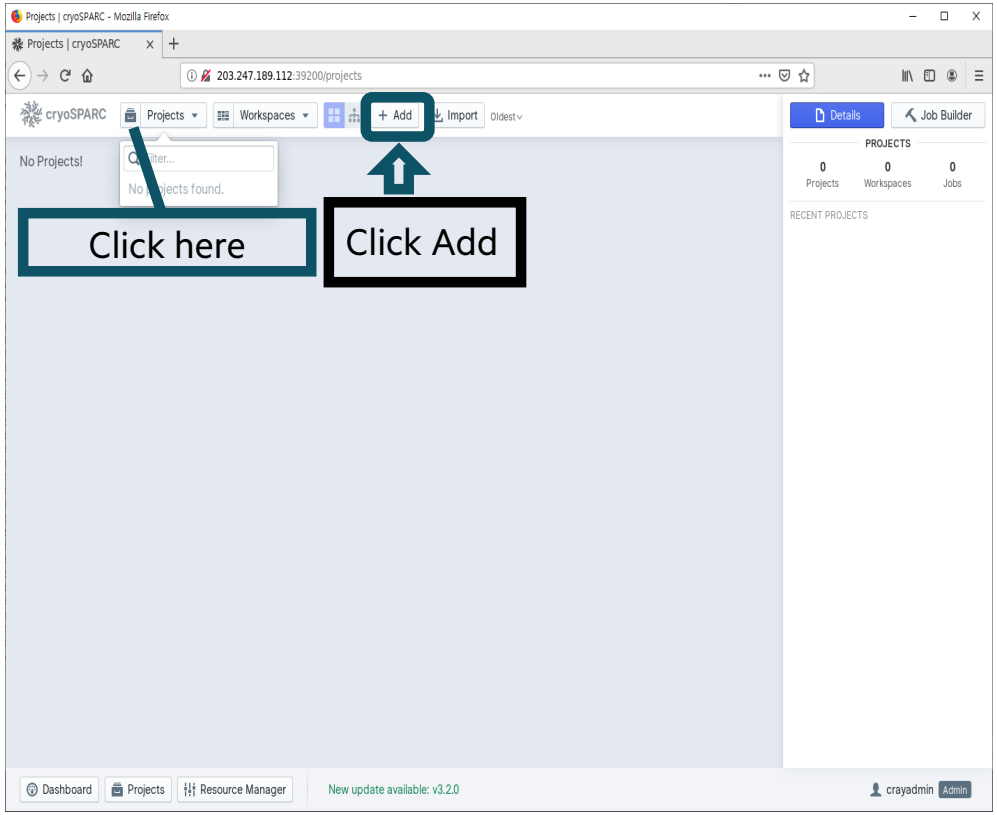
INTRODUCTION: DASHBOARD, PROJECTS, WORKSPACES AND JOBS

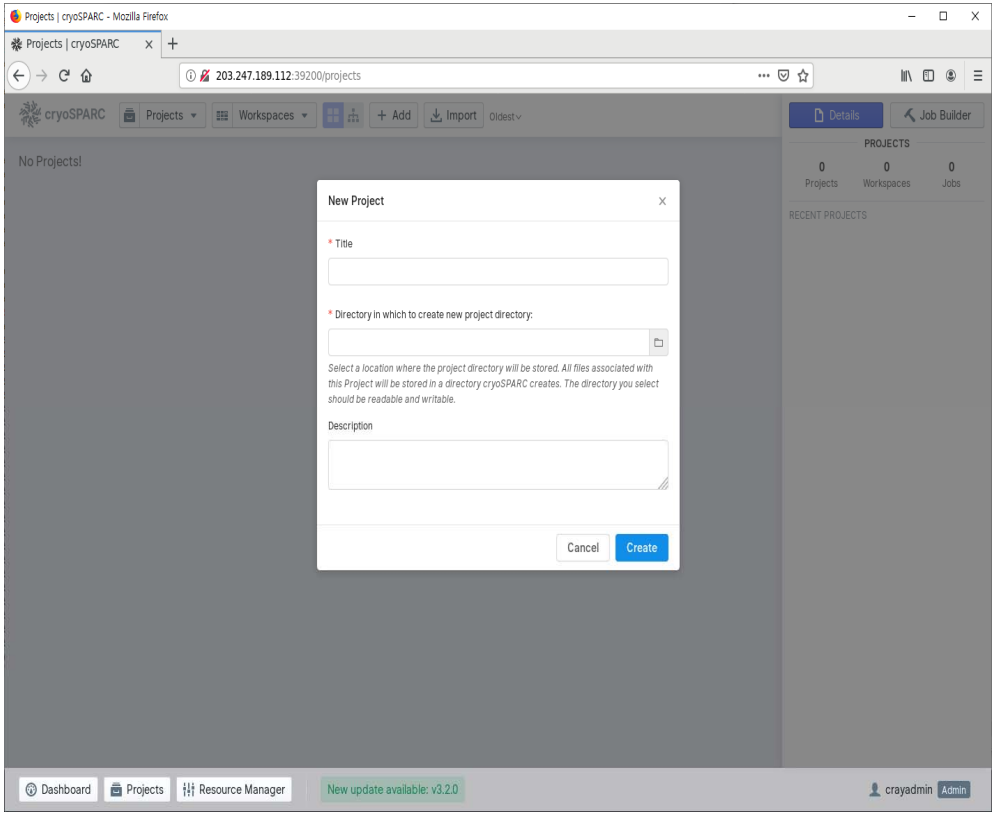
- Dashboard
  - At-a-glance information on your Projects, Workspaces and status of jobs
- Workflow is organized by Project, e.g. P1, P2, etc
- Projects contain one or more Workspaces, which in turn house jobs.



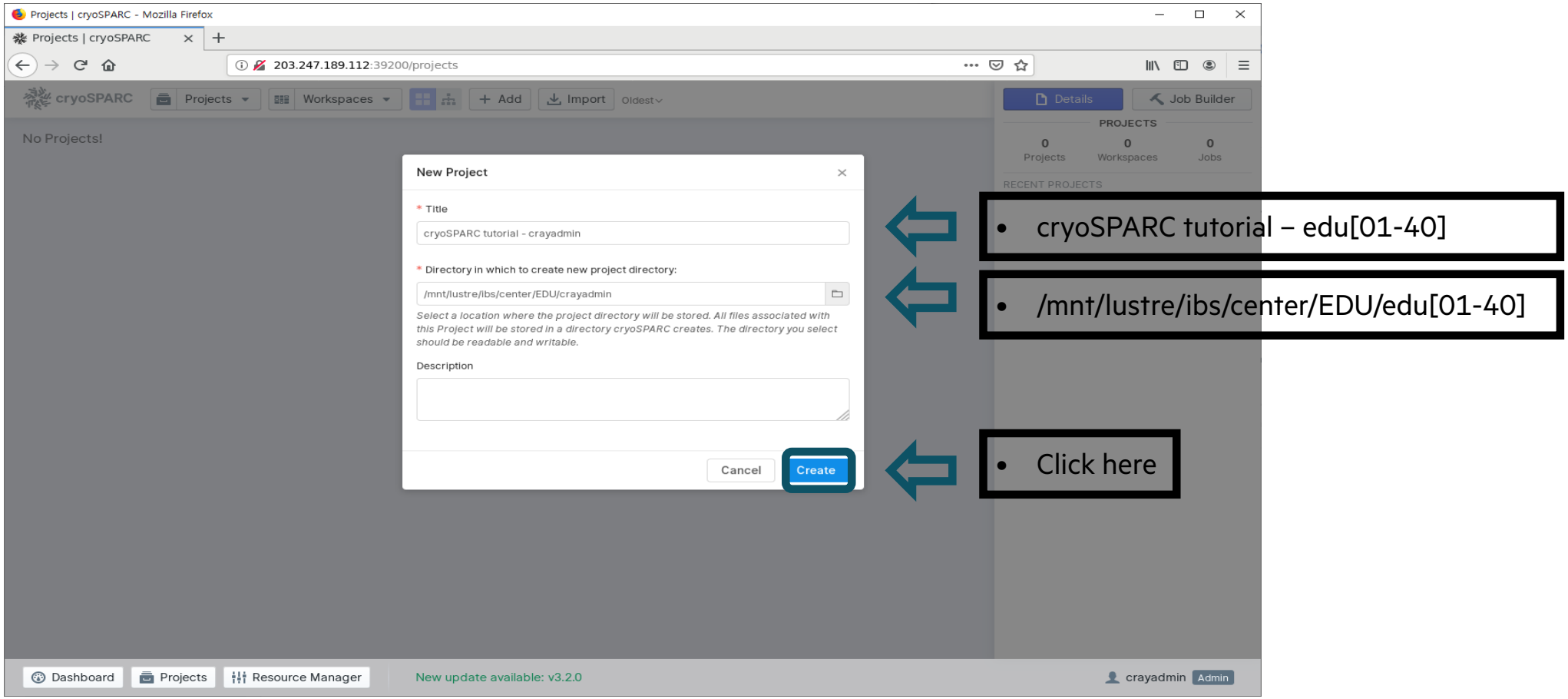
STEP 1

- Create a Project





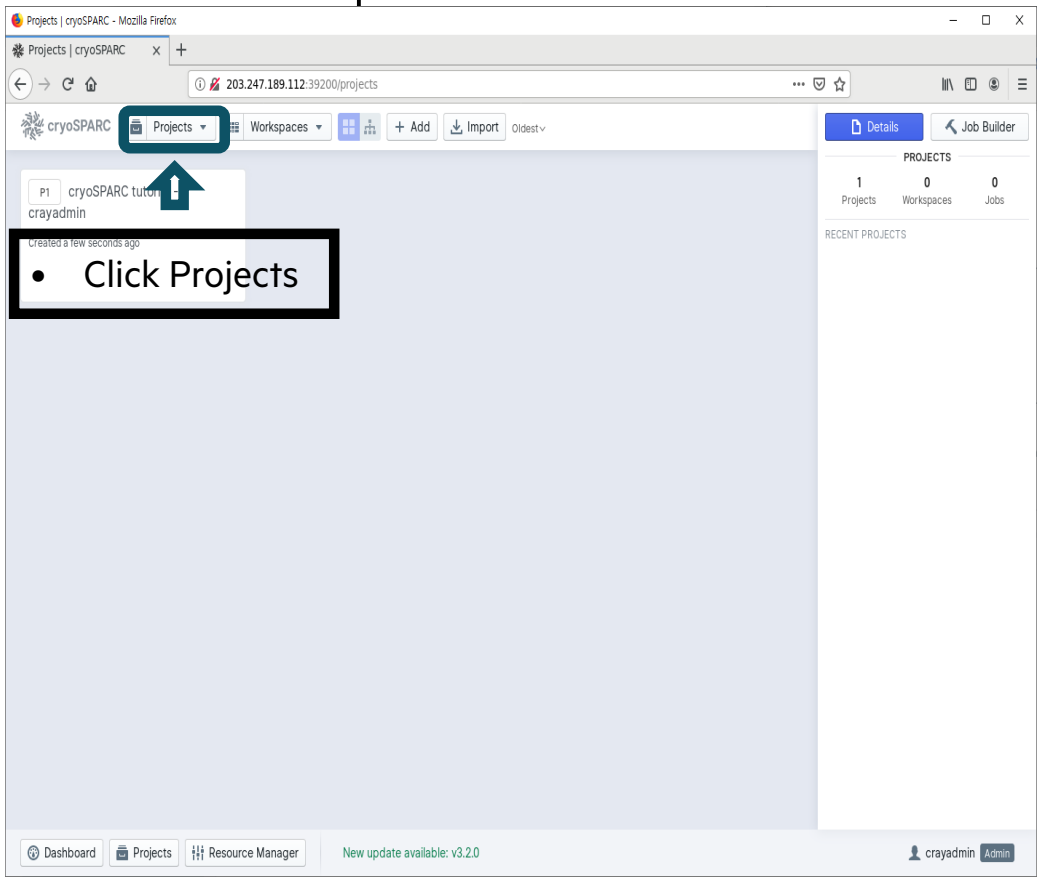
STEP 1

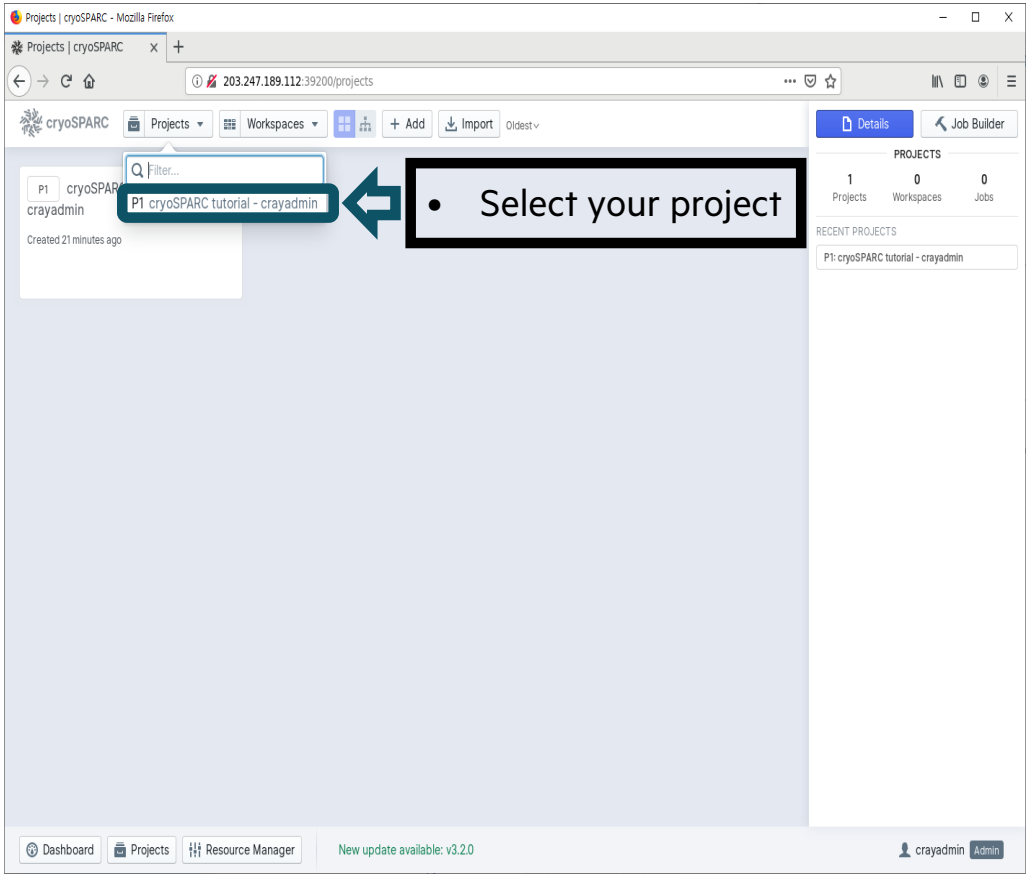


- cryoSPARC tutorial – edu[01-40]
- /mnt/lustre/ibs/center/EDU/edu[01-40]
- Click here

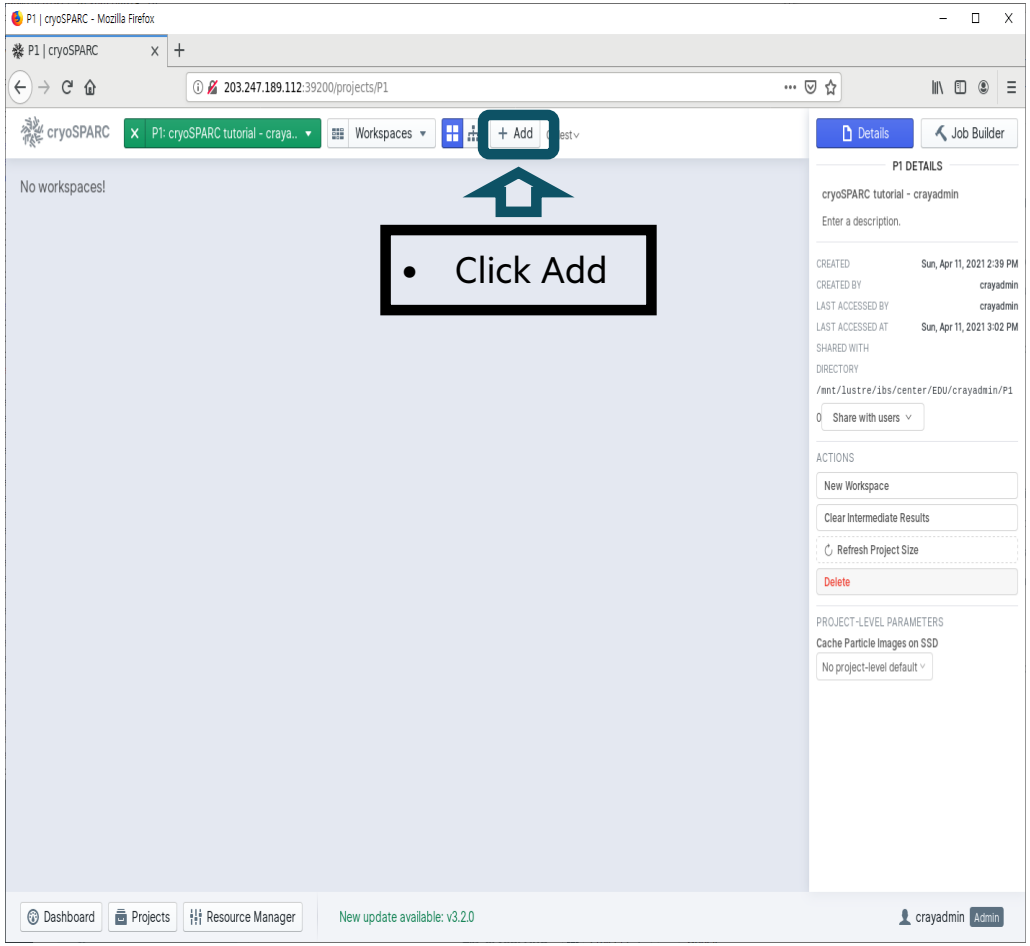
STEP 2

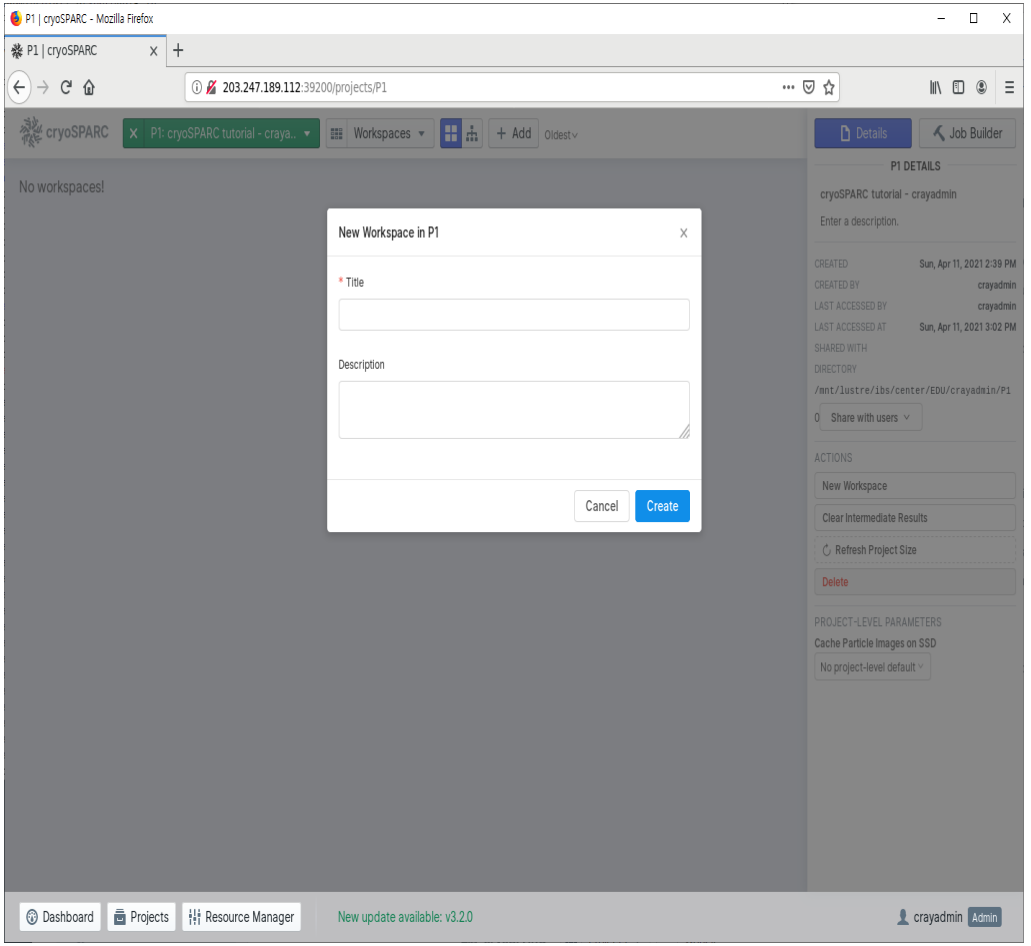
- Create a Workspace



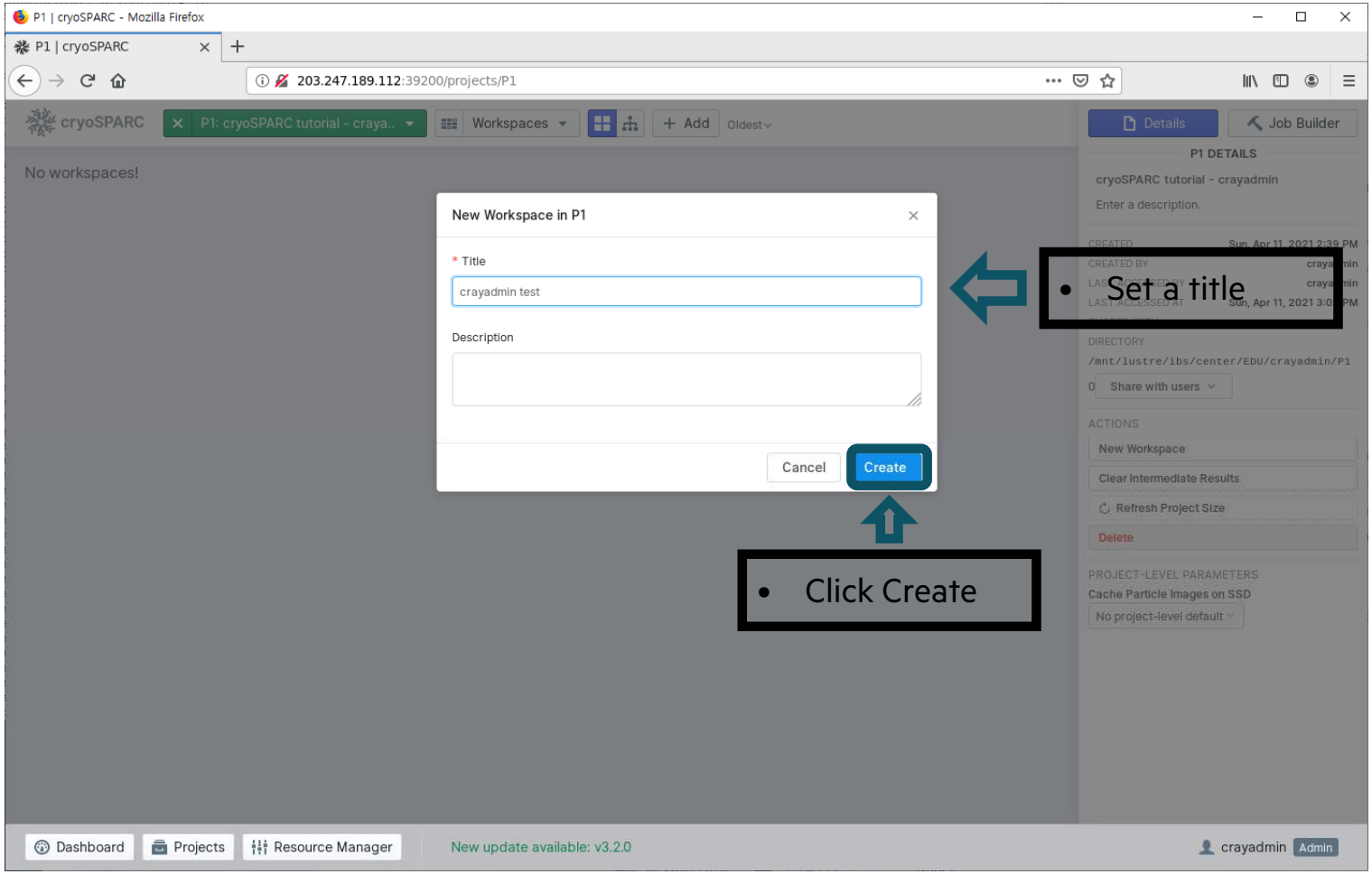


STEP 2


P1 | cryoSPARC - Mozilla Firefox
P1 | cryoSPARC
203.247.189.112:39200/projects/P1
cryoSPARC
P1: cryoSPARC tutorial - craya..
Workspaces
+ Add
Details
Job Builder
No workspaces!
P1 DETAILS
cryoSPARC tutorial - crayadmin
Enter a description.
CREATED Sun, Apr 11, 2021 2:39 PM
CREATED BY crayadmin
LAST ACCESSED BY crayadmin
LAST ACCESSED AT Sun, Apr 11, 2021 3:02 PM
SHARED WITH
DIRECTORY
/mnt/lustre/ibs/center/EDU/crayadmin/P1
Share with users
ACTIONS
New Workspace
Clear Intermediate Results
Refresh Project Size
Delete
PROJECT-LEVEL PARAMETERS
Cache Particle Images on SSD
No project-level default
Dashboard
Projects
Resource Manager
New update available: v3.2.0
crayadmin
Admin
Click Add

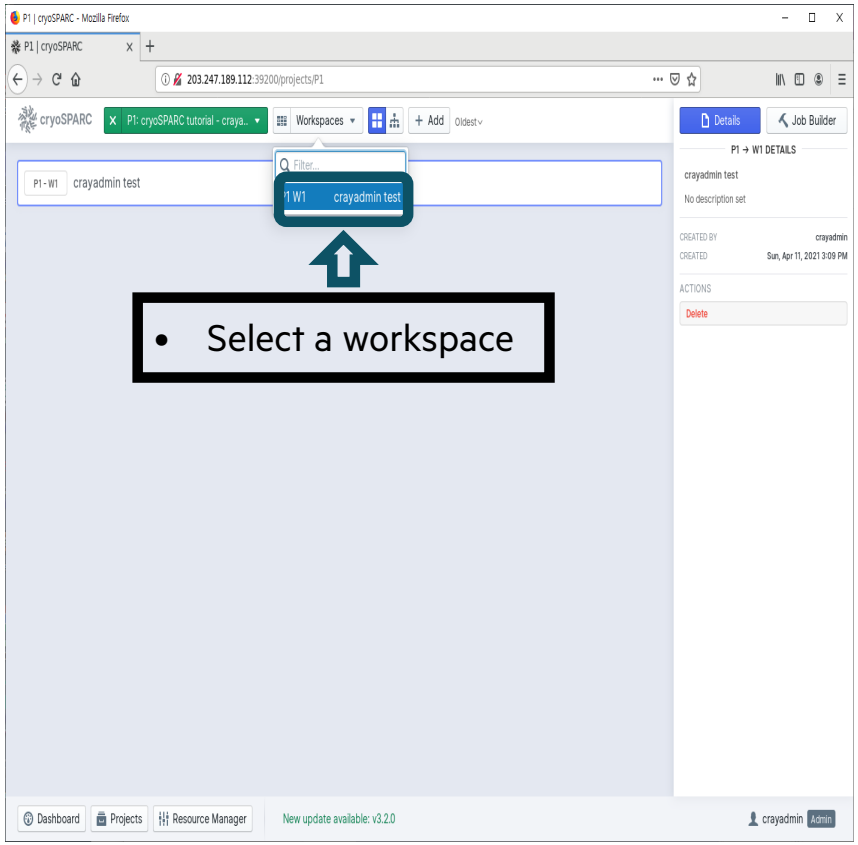

P1 | cryoSPARC - Mozilla Firefox
P1 | cryoSPARC
203.247.189.112:39200/projects/P1
cryoSPARC
P1: cryoSPARC tutorial - craya..
Workspaces
+ Add
Details
Job Builder
No workspaces!
New Workspace in P1
* Title
Description
Cancel
Create
Dashboard
Projects
Resource Manager
New update available: v3.2.0
crayadmin
Admin

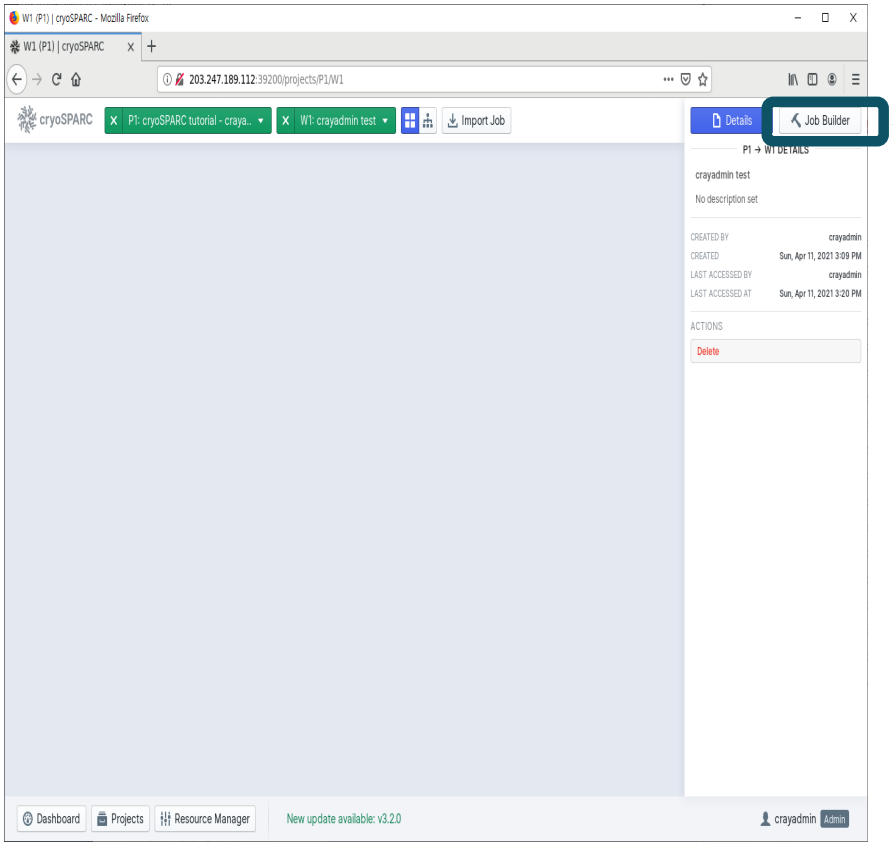
STEP 2


P1 | cryoSPARC - Mozilla Firefox
P1 | cryoSPARC
203.247.189.112:39200/projects/P1
cryoSPARC
P1: cryoSPARC tutorial - craya..
Workspaces
Add
Oldest
No workspaces!
Details
Job Builder
P1 DETAILS
cryoSPARC tutorial - crayadmin
Enter a description.
DIRECTORY
/mnt/lustre/ibs/center/EDU/crayadmin/P1
Share with users
ACTIONS
New Workspace
Clear Intermediate Results
Refresh Project Size
Delete
PROJECT-LEVEL PARAMETERS
Cache Particle Images on SSD
No project-level default
New Workspace in P1
* Title
crayadmin test
Description
Cancel
Create
Set a title
Click Create
Dashboard
Projects
Resource Manager
New update available: v3.2.0
crayadmin
Admin

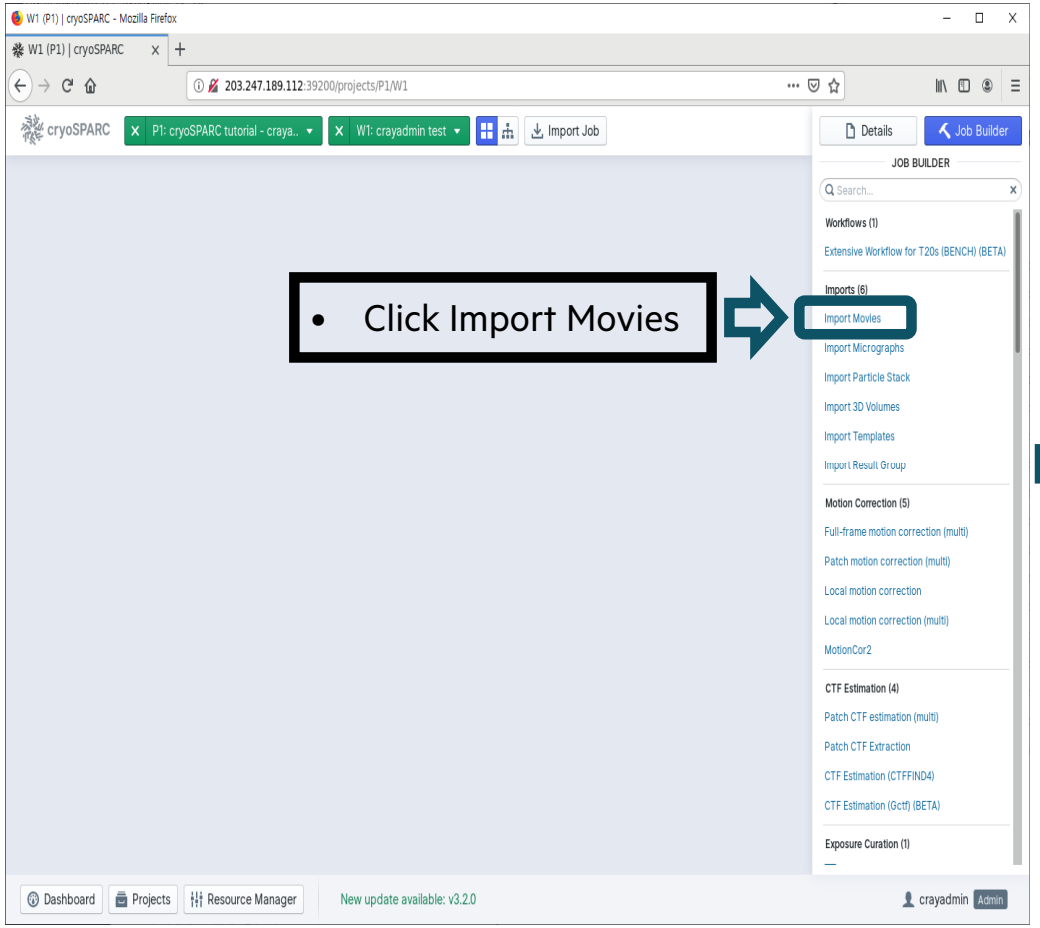
STEP 3

- Import Movies

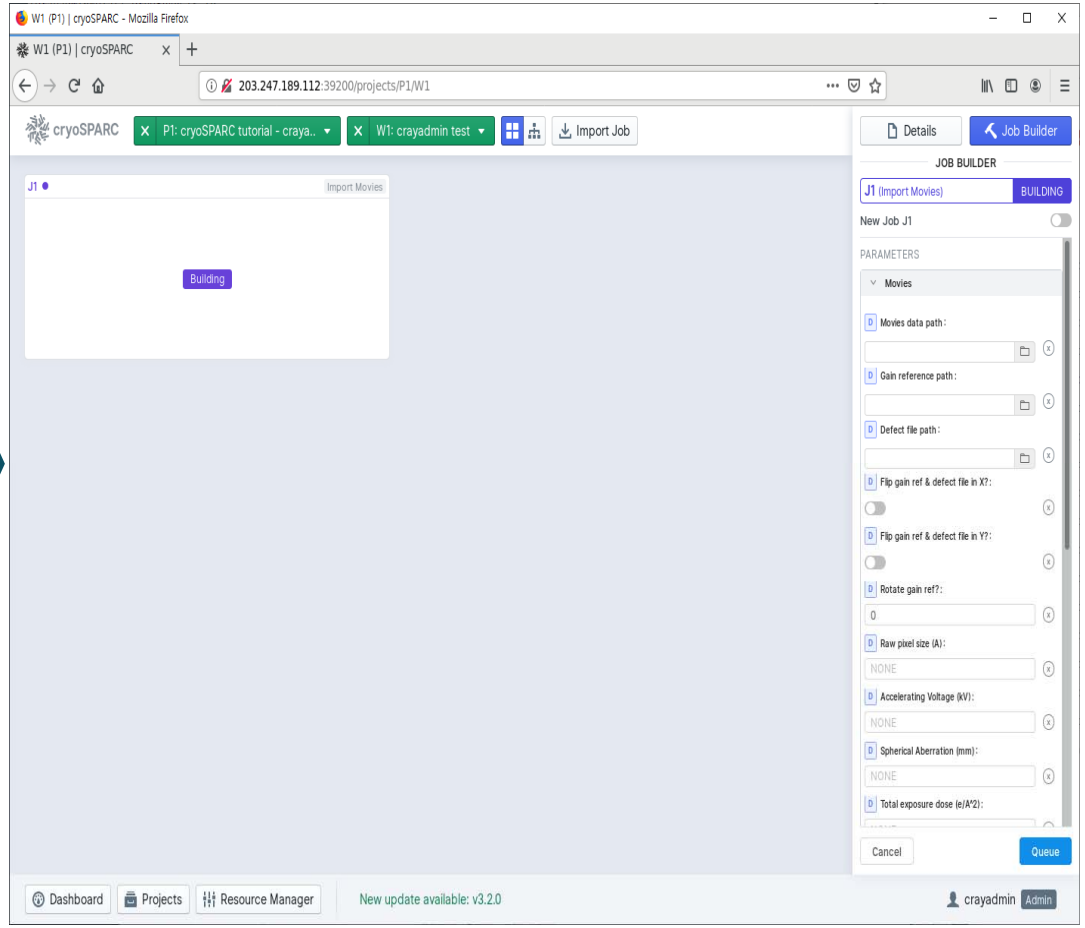




STEP 3

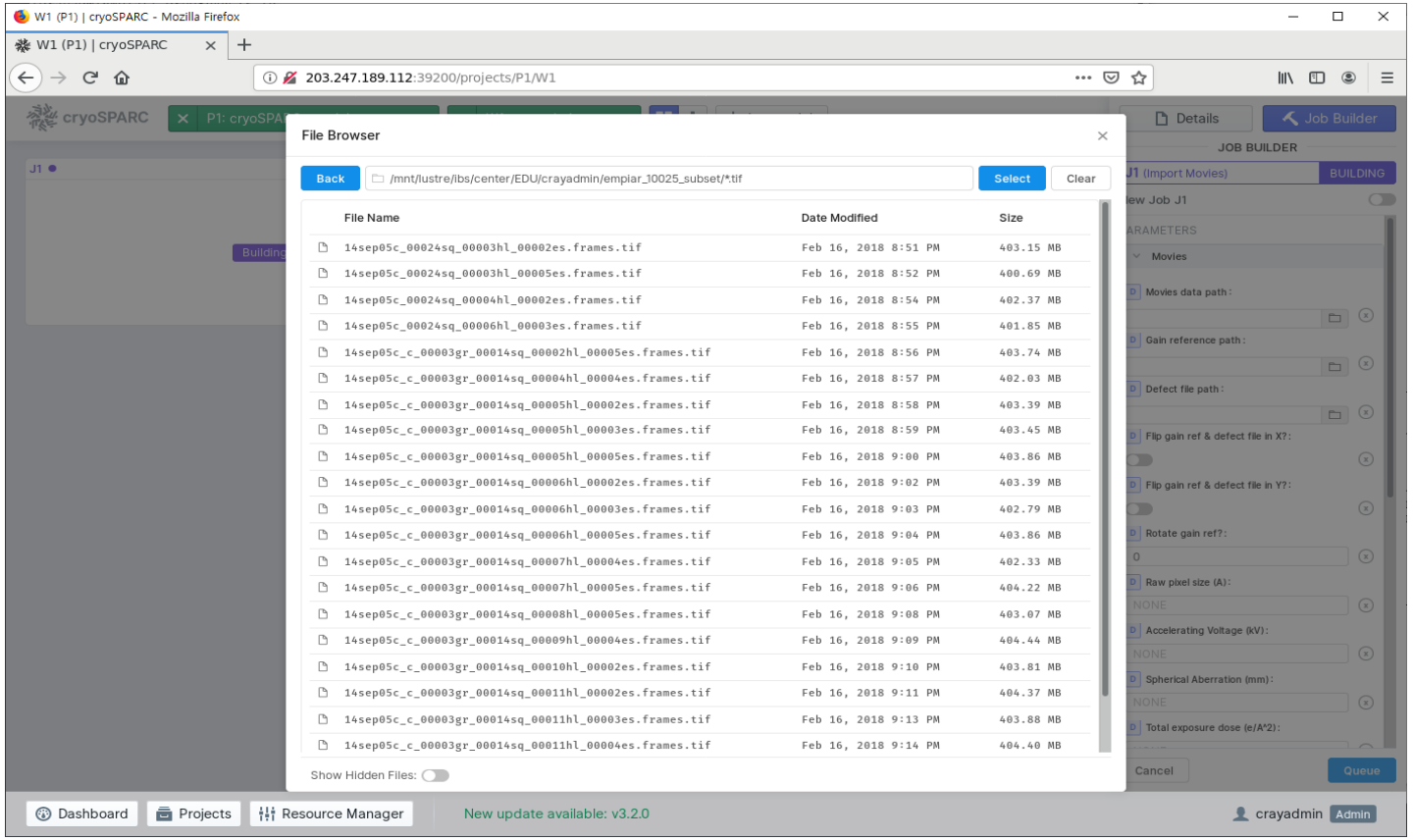


- Click Import Movies



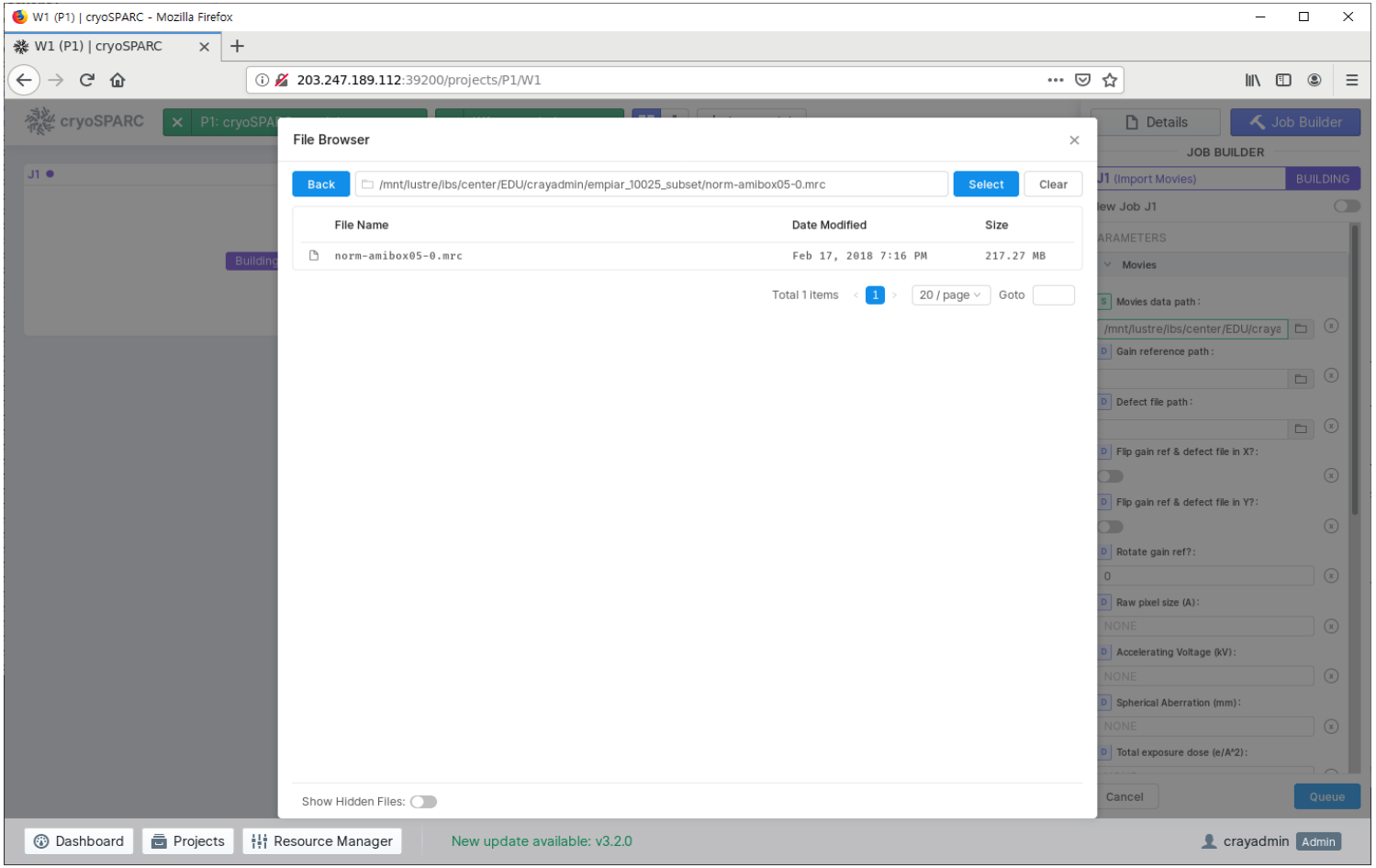
STEP 3

- Movies data path:
- /mnt/lustre/ibs/center/EDU/crayadmin/empiar_10025_subset/*.tif

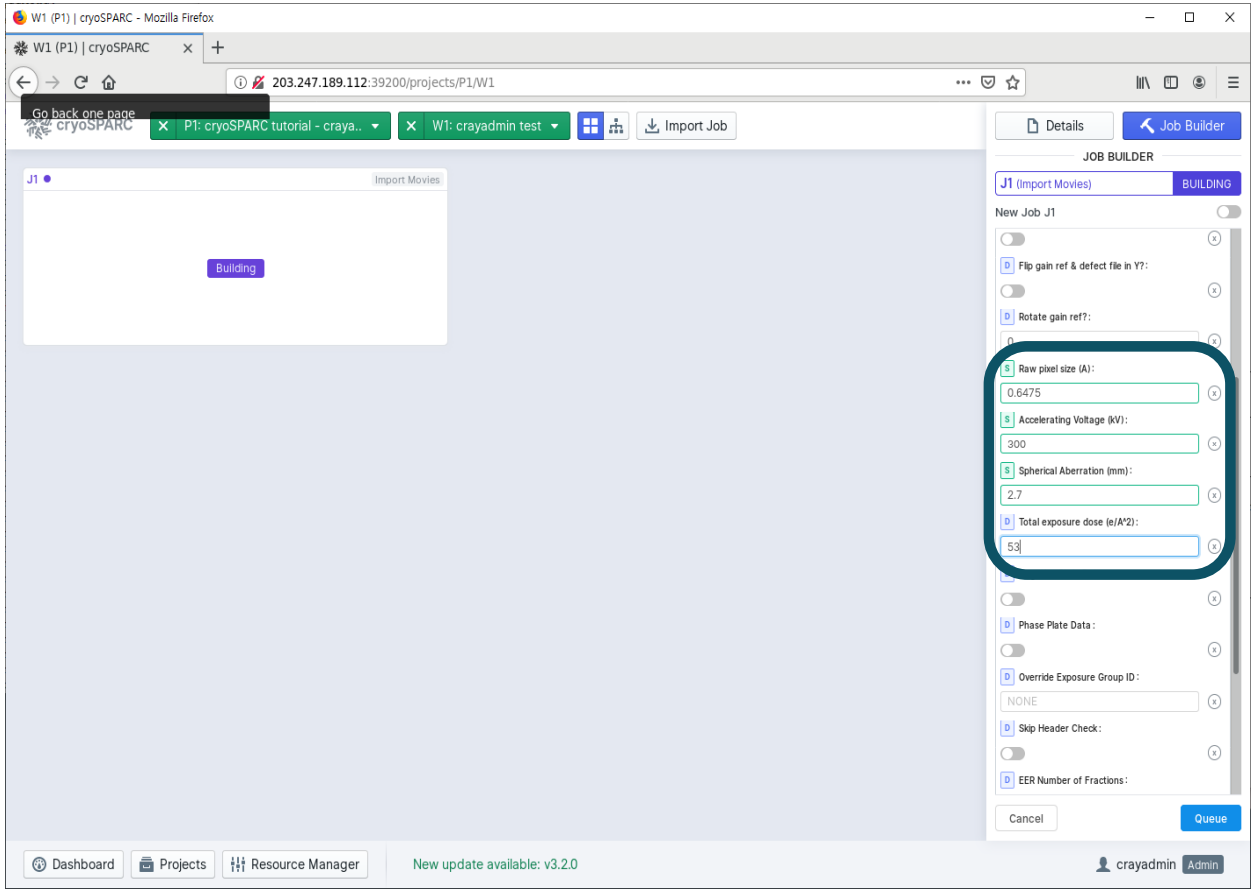


STEP 3

- Gain reference path:
- /mnt/lustre/ibs/center/EDU/crayadmin/empiar_10025_subset/norm-amibox05-0.mrc

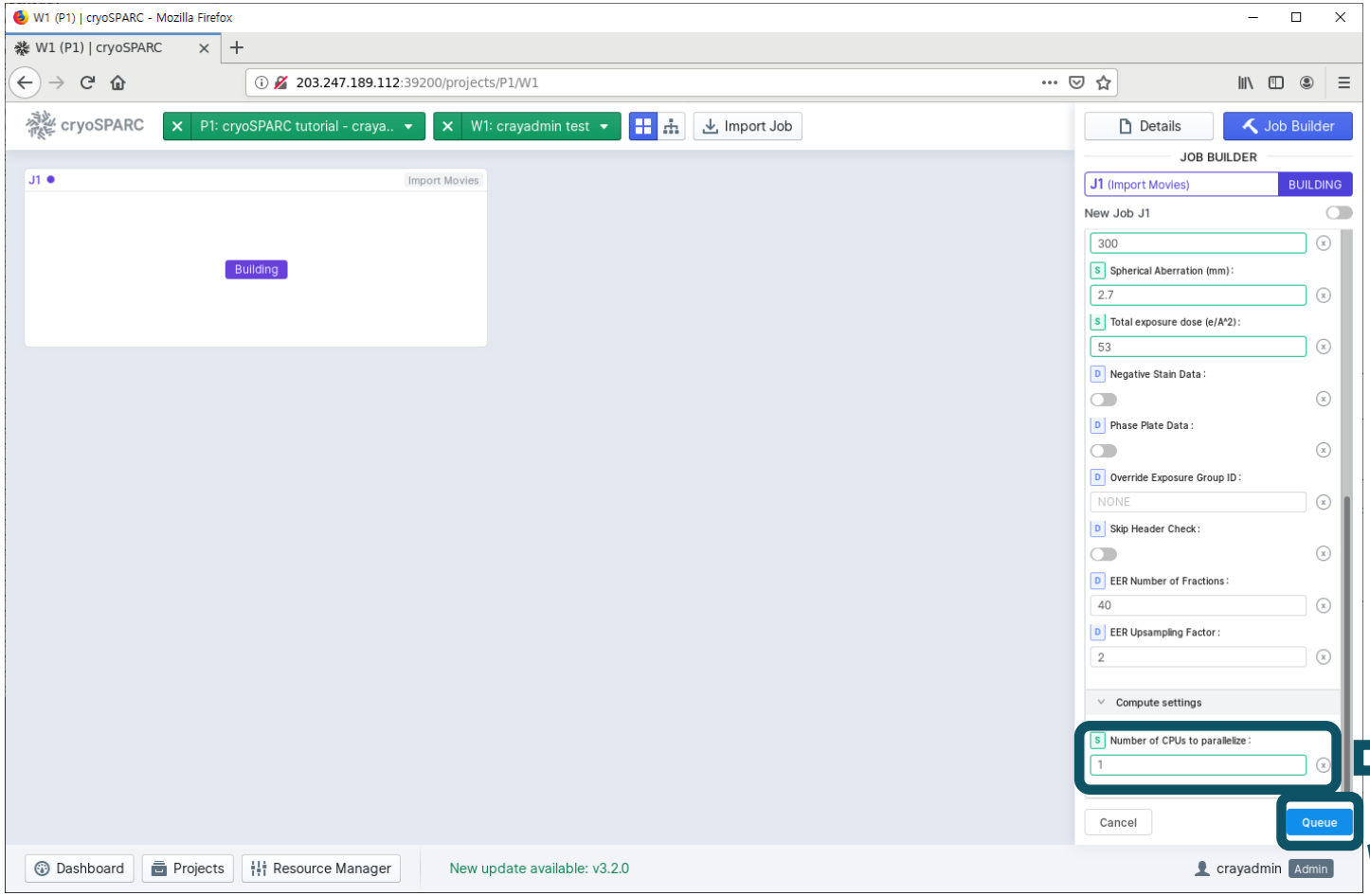


STEP 3



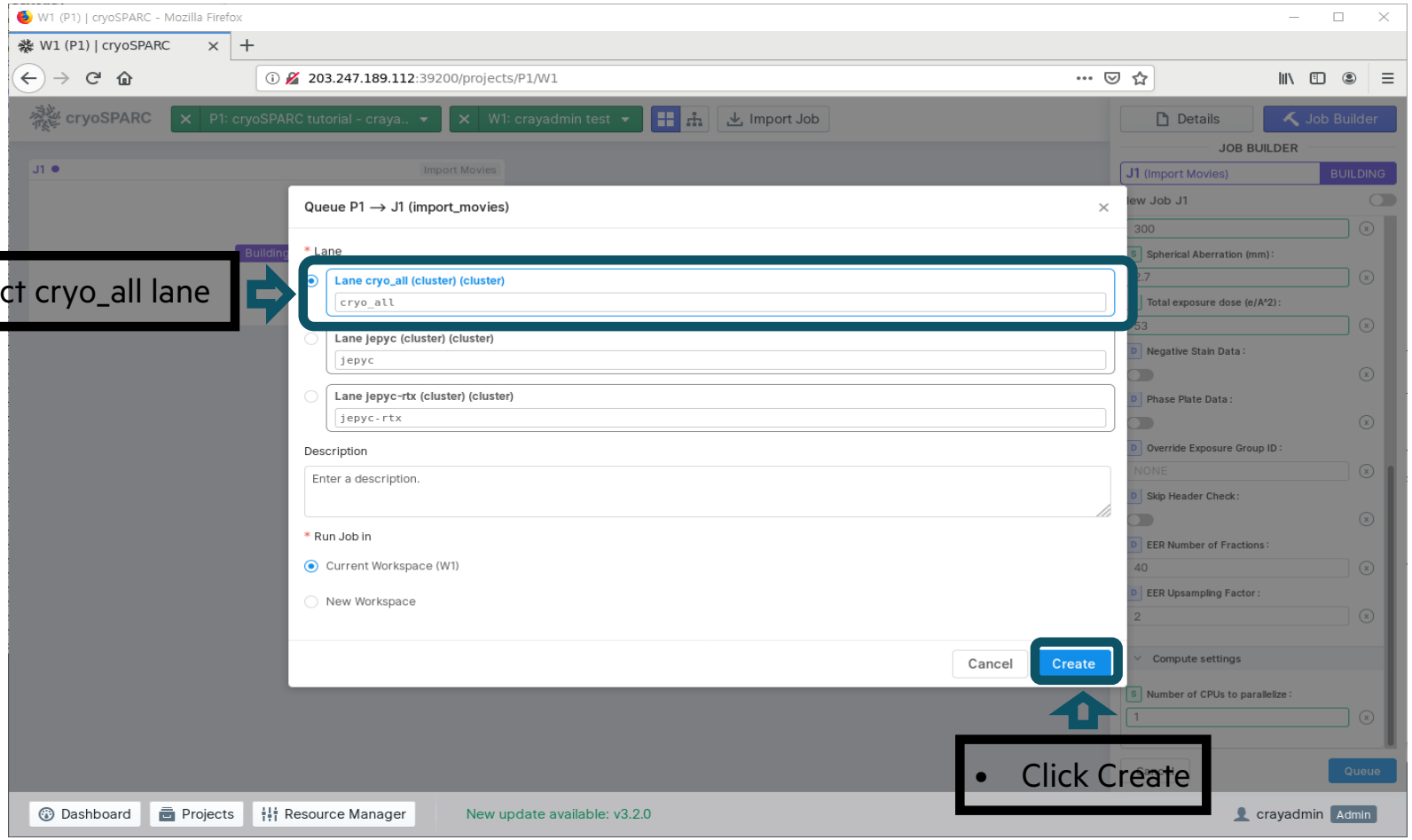
- Raw pixel size (A): 0.6475
- Accelerating Voltage (kW): 300
- Spherical Aberration (mm): 2.7
- Total exposure dose (e/A^2): 53

STEP 3



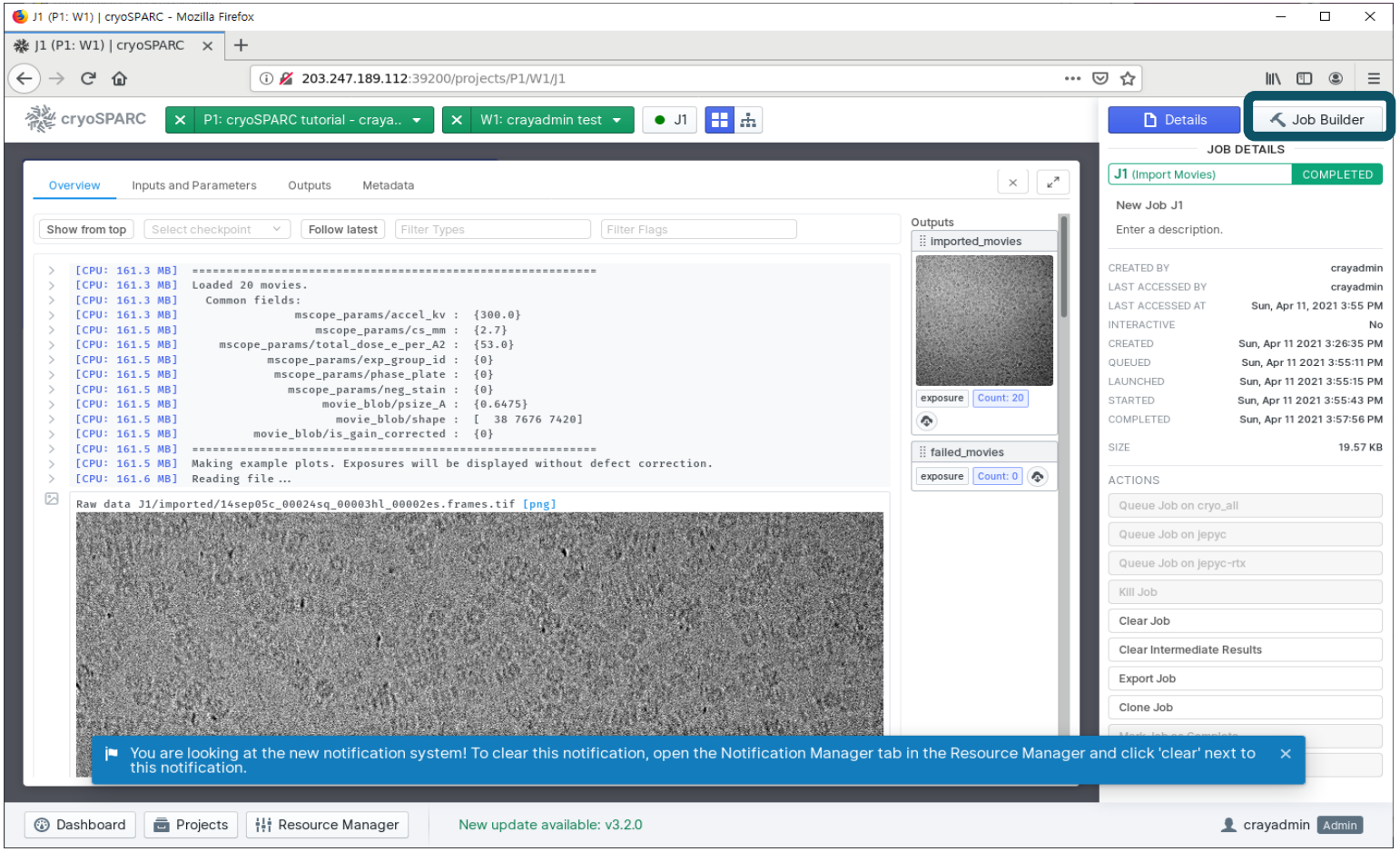
- Number of CPUs to parallelize: 1
- Click Queue

STEP 3


W1 (P1) | cryoSPARC - Mozilla Firefox
W1 (P1) | cryoSPARC
203.247.189.112:39200/projects/P1/W1
cryoSPARC
P1: cryoSPARC tutorial - craya..
W1: crayadmin test
Import Job
Details
Job Builder
JOB BUILDER
J1 (Import Movies)
BUILDING
Queue P1 → J1 (import_movies)
* Lane
Lane cryo_all (cluster) (cluster)
cryo_all
Lane jepyc (cluster) (cluster)
jepyc
Lane jepyc-rtx (cluster) (cluster)
jepyc-rtx
Description
Enter a description.
* Run Job in
Current Workspace (W1)
New Workspace
Cancel
Create
Queue
Dashboard
Projects
Resource Manager
New update available: v3.2.0
crayadmin
Admin
• Select cryo_all lane
• Click Create

STEP 3

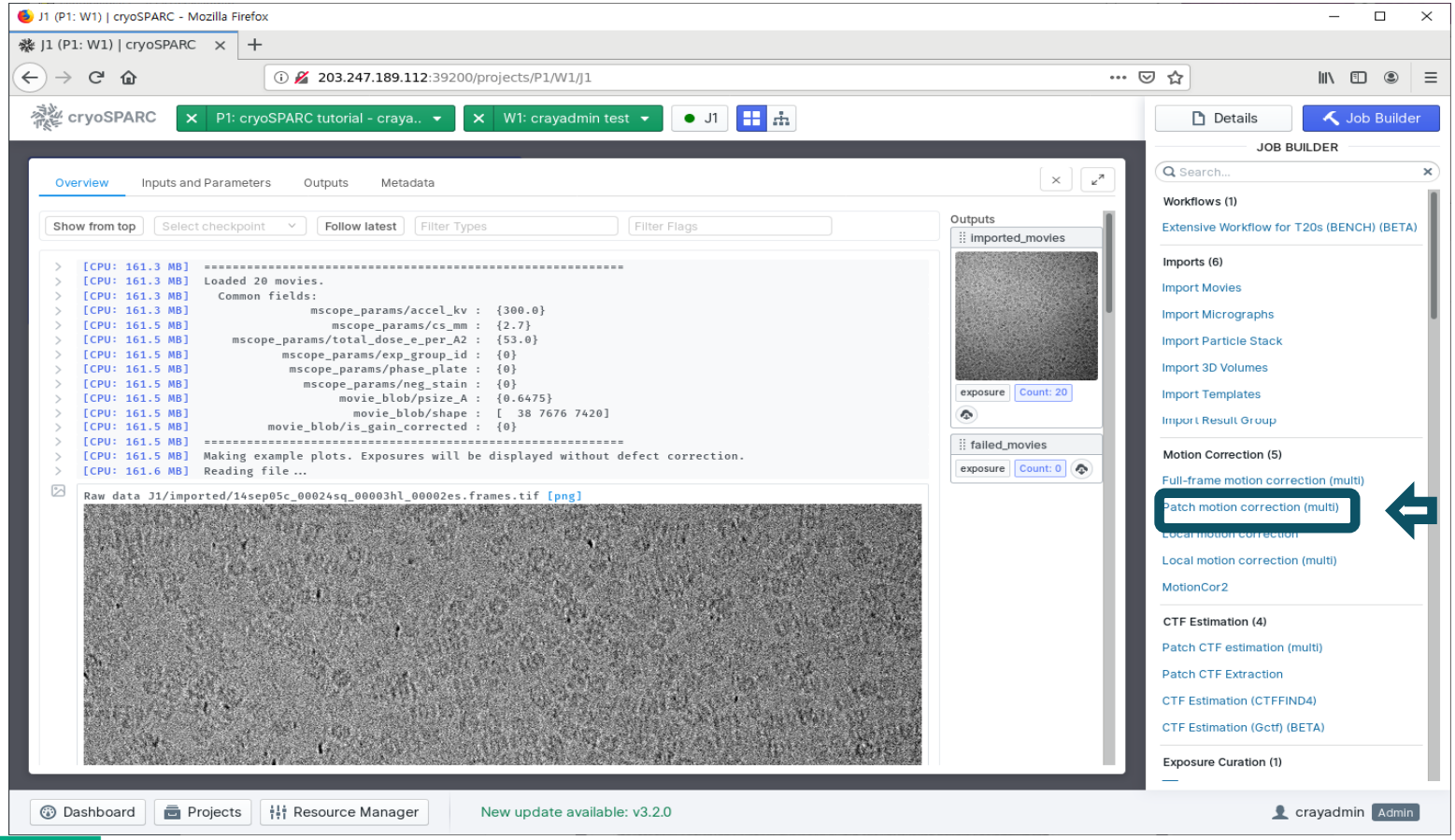
- Import Movies - Done



- Click Job Builder to go to next step

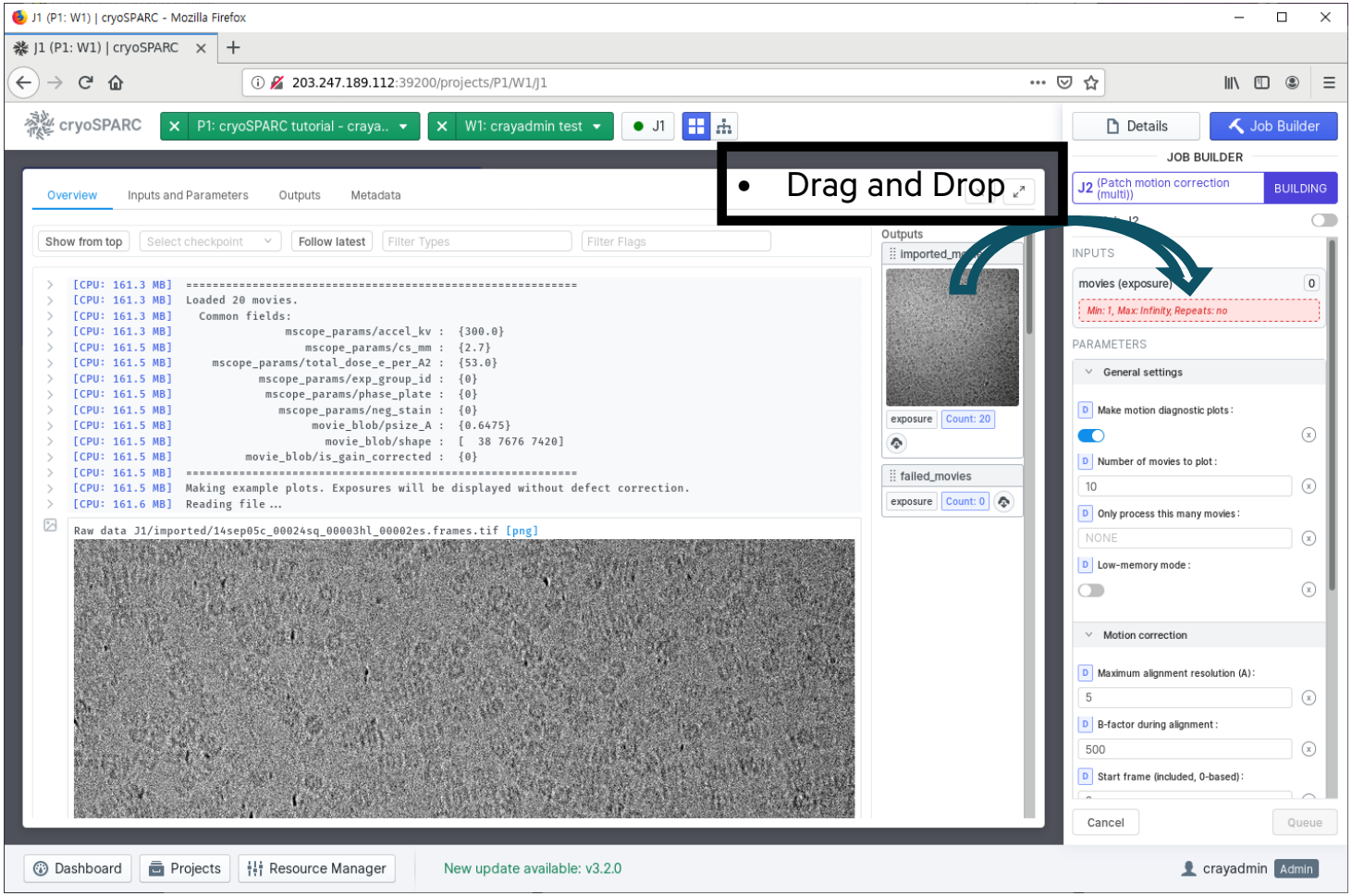
STEP 4

- Motion Correction

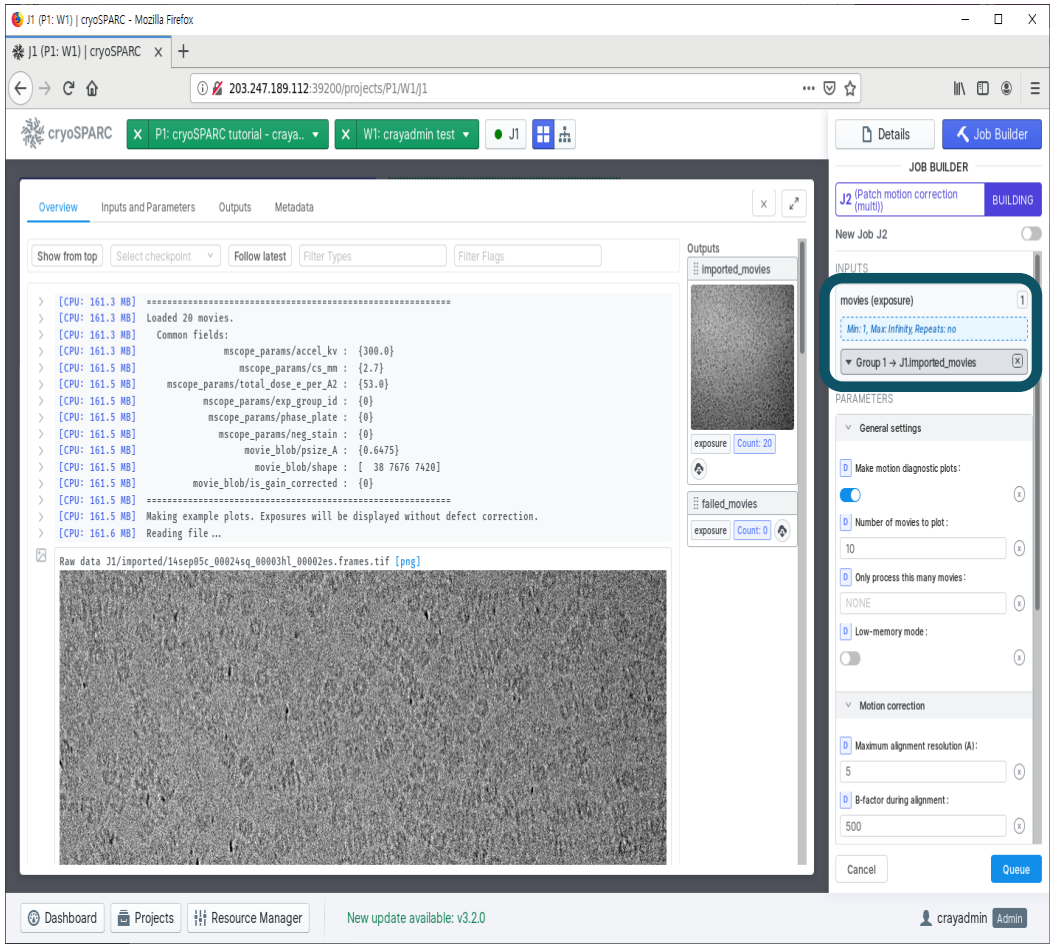


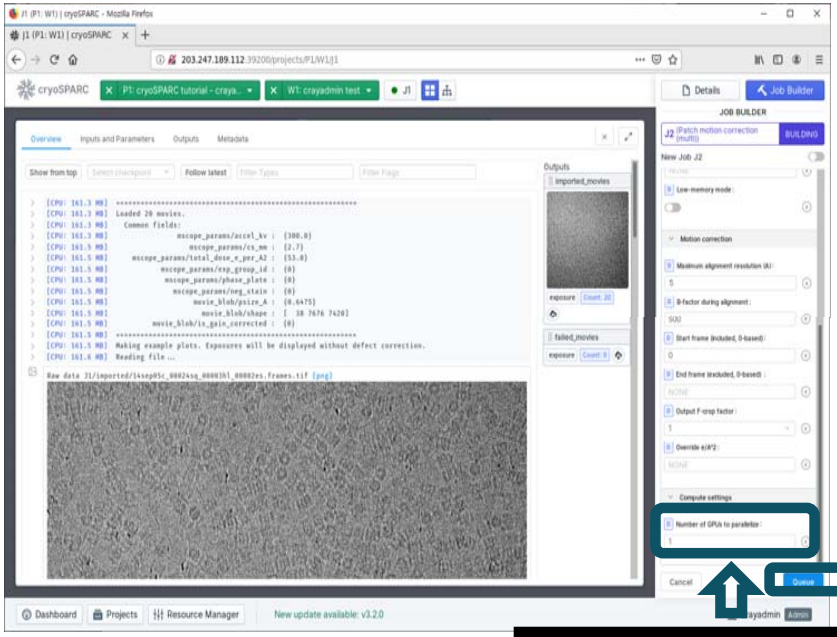
- Select Patch motion correction(multi)

STEP 4


J1 (P1: W1) | cryoSPARC - Mozilla Firefox
203.247.189.112:39200/projects/P1/W1/J1
cryoSPARC
P1: cryoSPARC tutorial - craya..
W1: crayadmin test
J1
Details
Job Builder
JOB BUILDER
J2 (Patch motion correction (multi))
BUILDING
Drag and Drop
Overview
Inputs and Parameters
Outputs
Metadata
Show from top
Select checkpoint
Follow latest
Filter Types
Filter Flags
[CPU: 161.3 MB] Loaded 20 movies.
[CPU: 161.3 MB] Common fields:
[CPU: 161.3 MB] mscope_params/accel_kv : {300.0}
[CPU: 161.3 MB] mscope_params/cs_mm : {2.7}
[CPU: 161.5 MB] mscope_params/total_dose_e_per_A2 : {53.0}
[CPU: 161.5 MB] mscope_params/exp_group_id : {0}
[CPU: 161.5 MB] mscope_params/phase_plate : {0}
[CPU: 161.5 MB] mscope_params/neg_stain : {0}
[CPU: 161.5 MB] movie_blob/psize_A : {0.6475}
[CPU: 161.5 MB] movie_blob/shape : [38 7676 7420]
[CPU: 161.5 MB] movie_blob/is_gain_corrected : {0}
[CPU: 161.5 MB] Making example plots. Exposures will be displayed without defect correction.
[CPU: 161.6 MB] Reading file ...
Raw data J1/imported/14sep05c_00024sq_00003hl_00002es.frames.tif [png]
Outputs
exposure
Count: 20
failed_movies
exposure
Count: 0
INPUTS
movies (exposure)
0
Min: 1, Max: Infinity, Repeats: no
PARAMETERS
General settings
Make motion diagnostic plots :
Number of movies to plot :
10
Only process this many movies :
NONE
Low-memory mode :
Motion correction
Maximum alignment resolution (A) :
5
B-factor during alignment :
500
Start frame (included, 0-based) :
Cancel
Queue
Dashboard
Projects
Resource Manager
New update available: v3.2.0
crayadmin
Admin

STEP 4

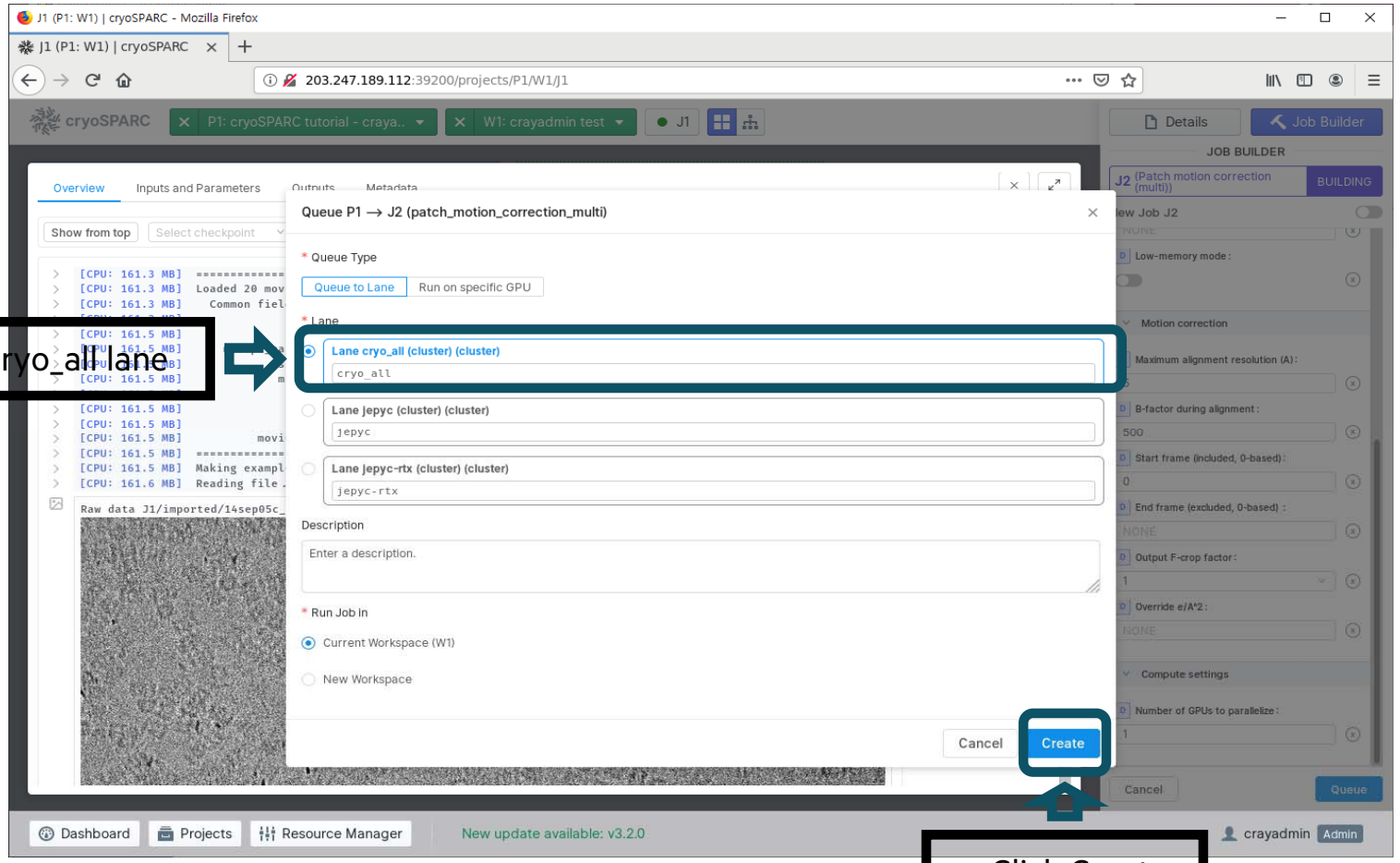




Click Queue

- Number of GPUs to parallelize: 1

STEP 4

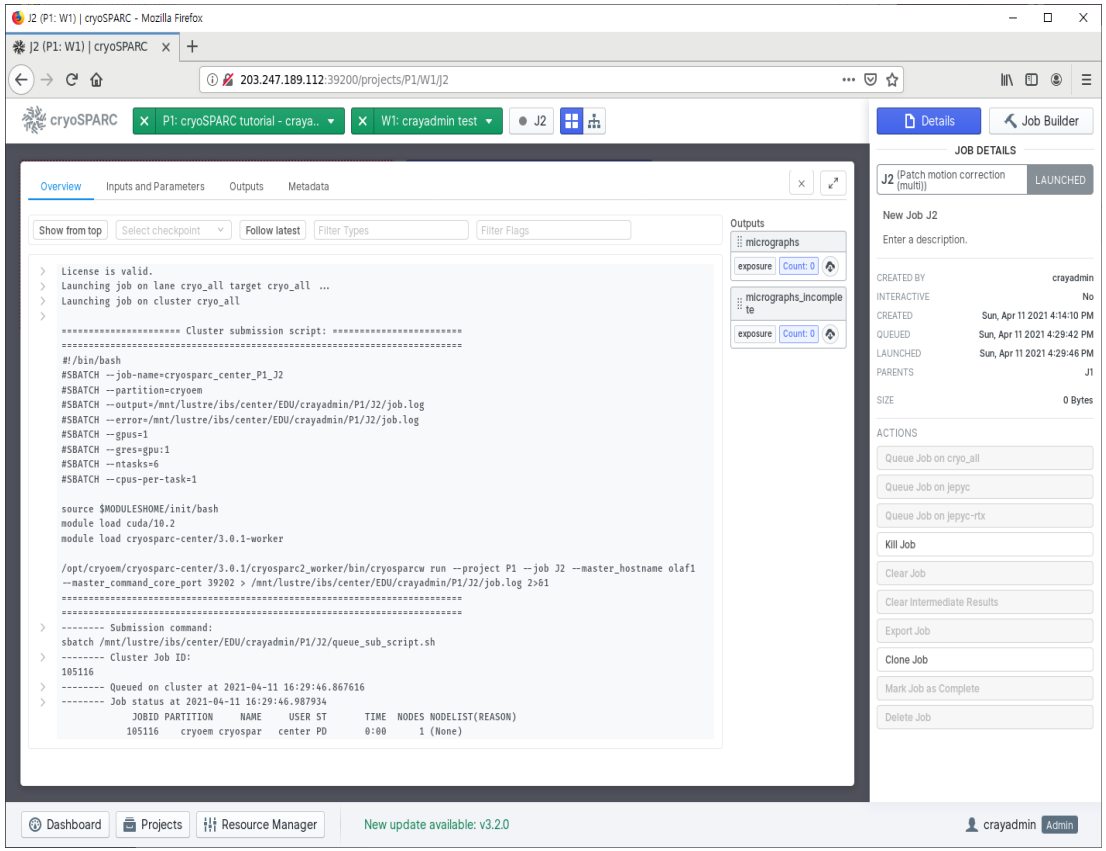


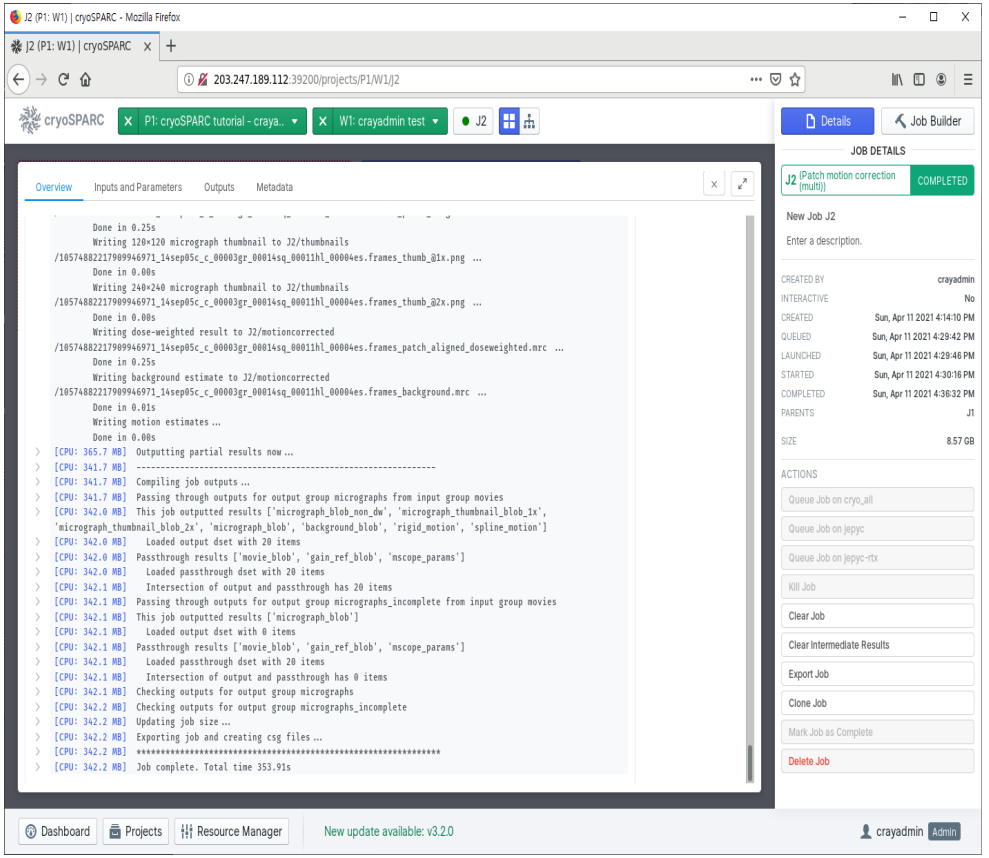
- Select cryo_all lane

- Click Create

STEP 4

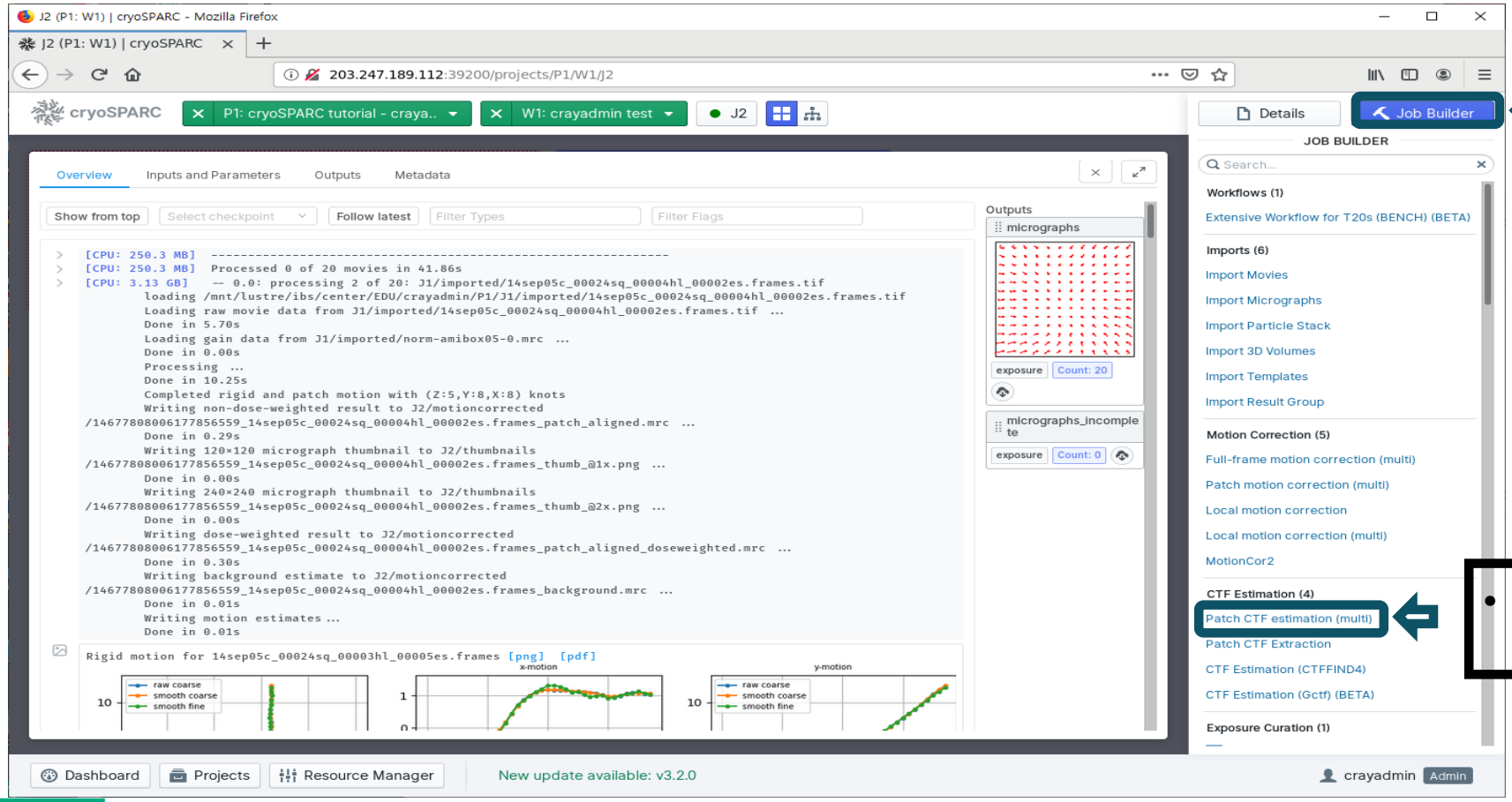
- Motion correction - Done





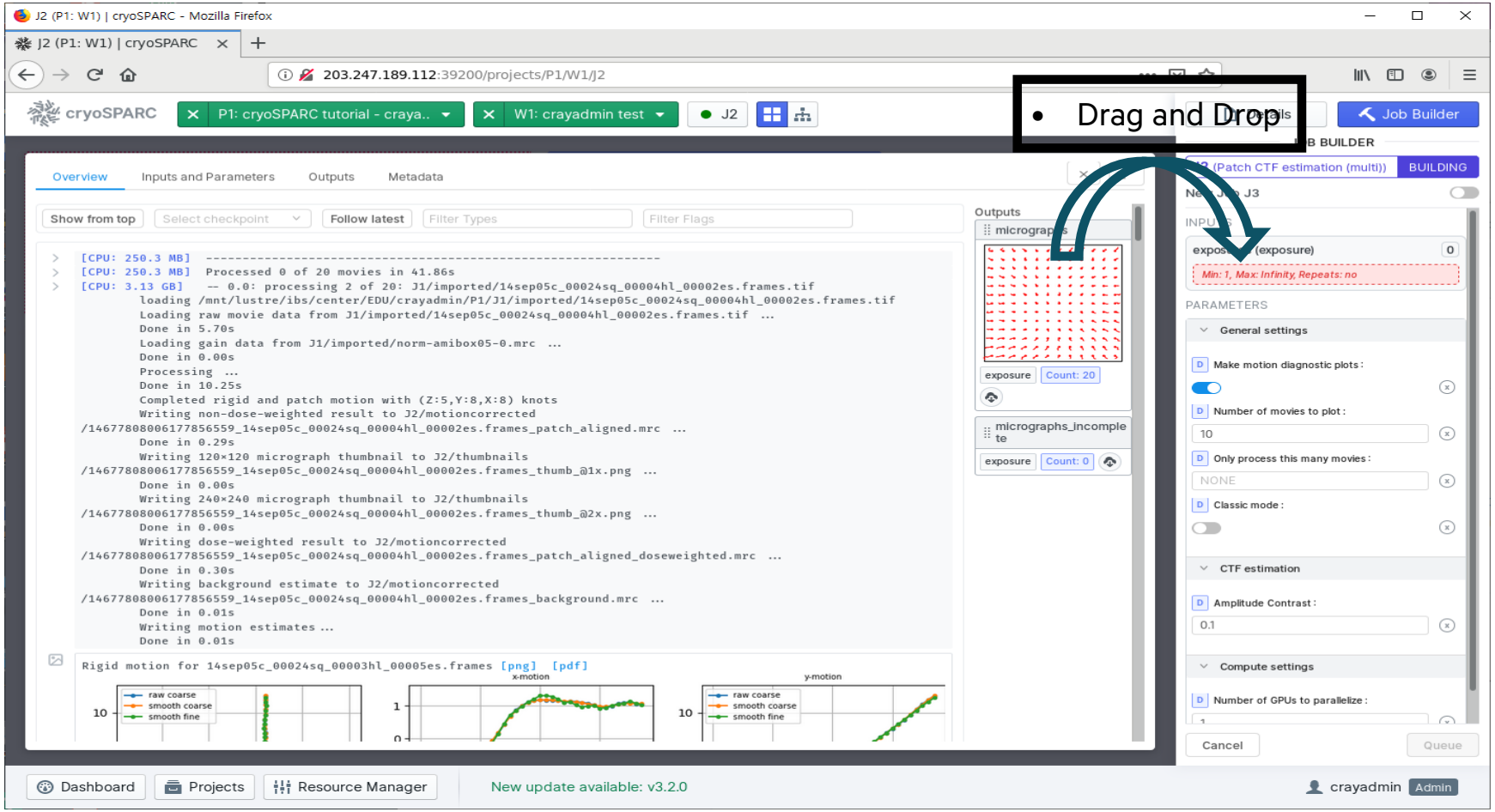
STEP 5

- CTF estimation

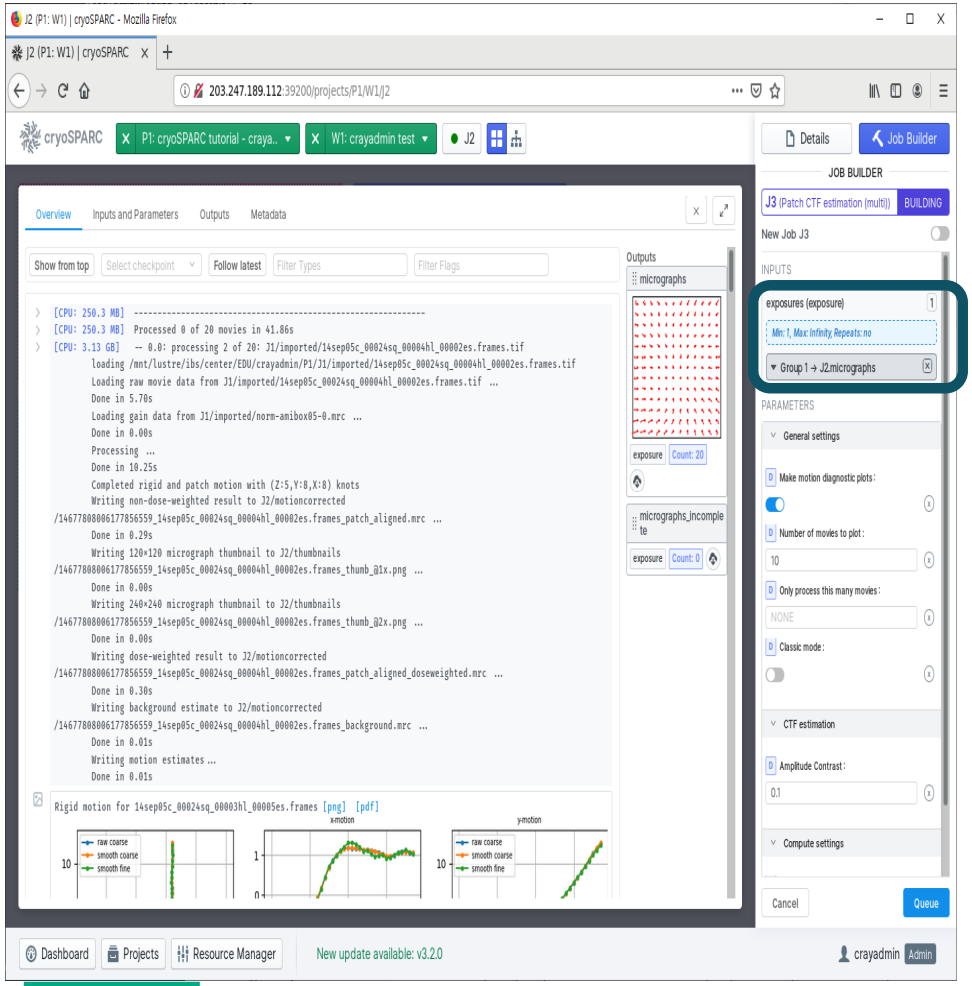


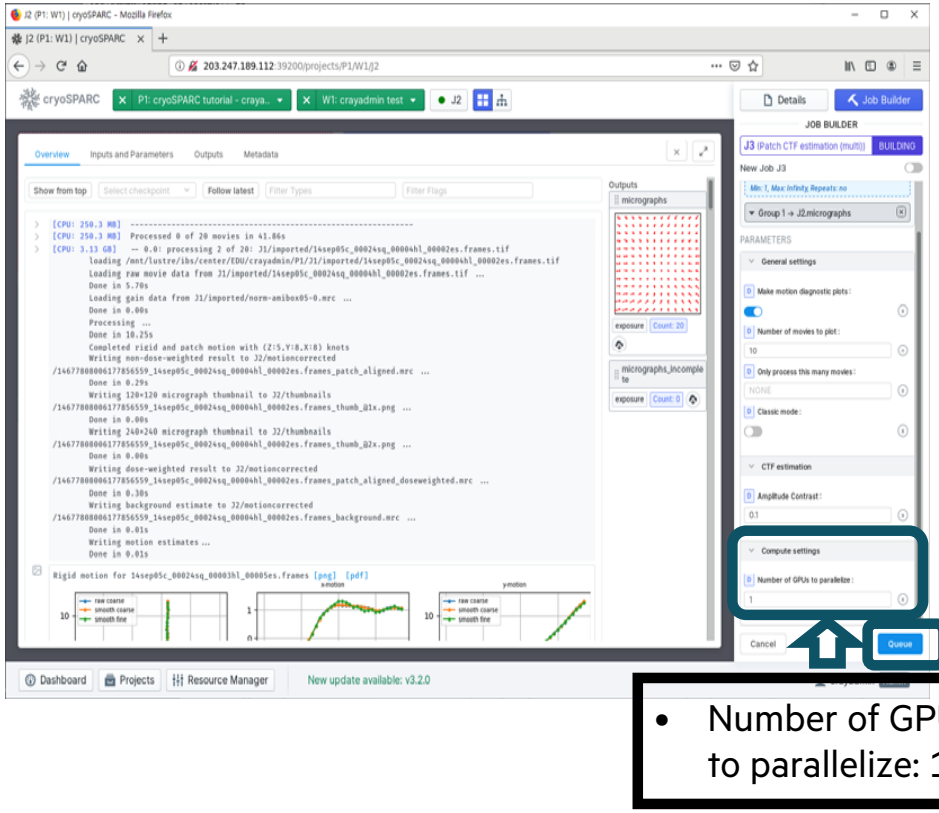
- Select Job Builder
- Select Patch CTF estimation (multi)

STEP 5


J2 (P1: W1) | cryoSPARC - Mozilla Firefox
203.247.189.112:39200/projects/P1/W1/J2
cryoSPARC
P1: cryoSPARC tutorial - craya..
W1: crayadmin test
J2
Job Builder
Drag and Drop
Overview
Inputs and Parameters
Outputs
Metadata
Show from top
Select checkpoint
Follow latest
Filter Types
Filter Flags
[CPU: 250.3 MB] Processed 0 of 20 movies in 41.86s
[CPU: 3.13 GB] -- 0.0: processing 2 of 20: J1/imported/14sep05c_00024sq_00004hl_00002es.frames.tif
Rigid motion for 14sep05c_00024sq_00003hl_00005es.frames [png] [pdf]
raw coarse
smooth coarse
smooth fine
x-motion
y-motion
Outputs
micrographs
exposure
Count: 20
micrographs_incomplete
Count: 0
(Patch CTF estimation (multi))
BUILDING
INPUTS
(exposure)
Min: 1, Max: Infinity, Repeats: no
PARAMETERS
General settings
Make motion diagnostic plots :
Number of movies to plot :
10
Only process this many movies :
NONE
Classic mode :
CTF estimation
Amplitude Contrast :
0.1
Compute settings
Number of GPUs to parallelize :
Cancel
Queue
Dashboard
Projects
Resource Manager
New update available: v3.2.0
crayadmin
Admin

STEP 5

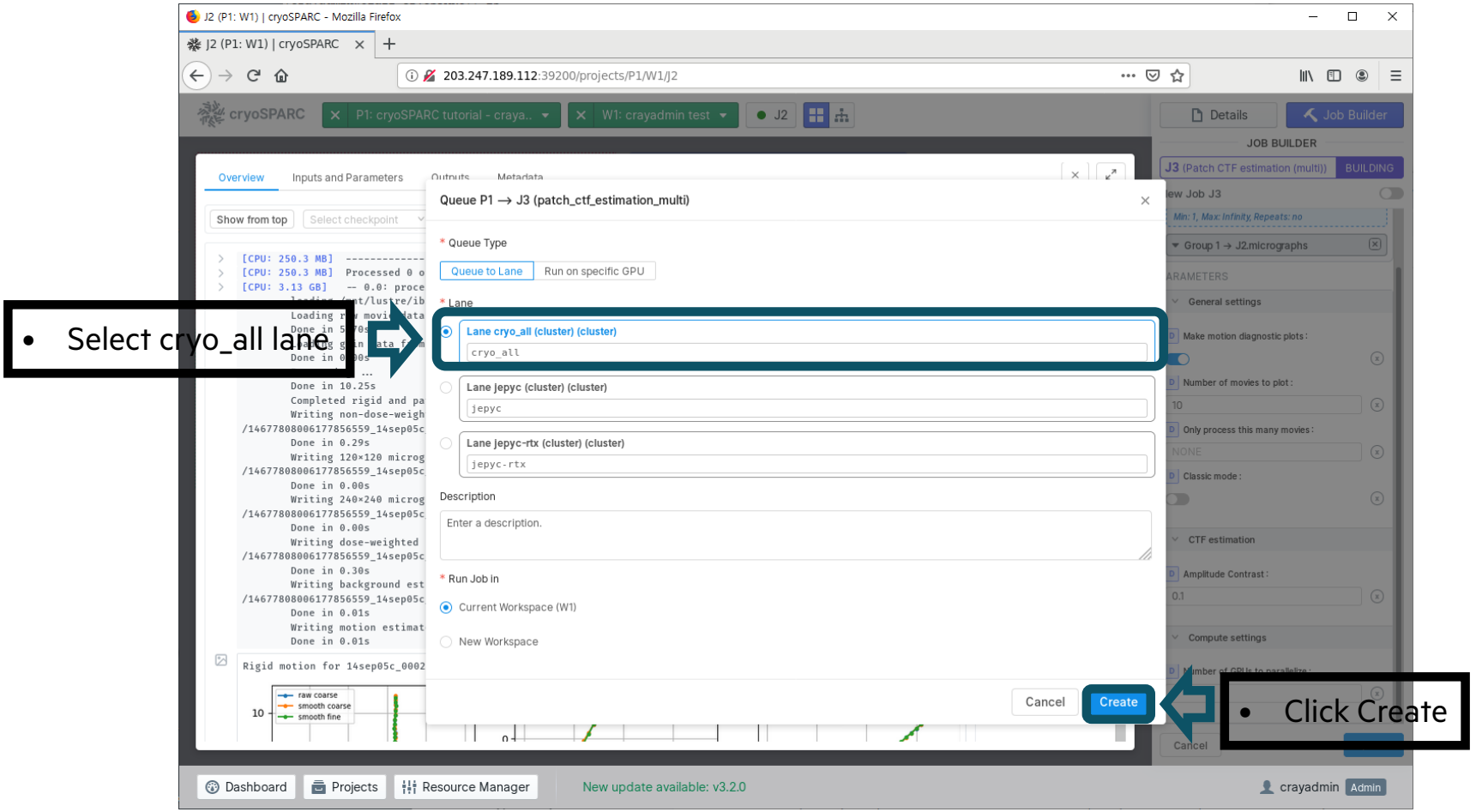




- Number of GPUs to parallelize: 1

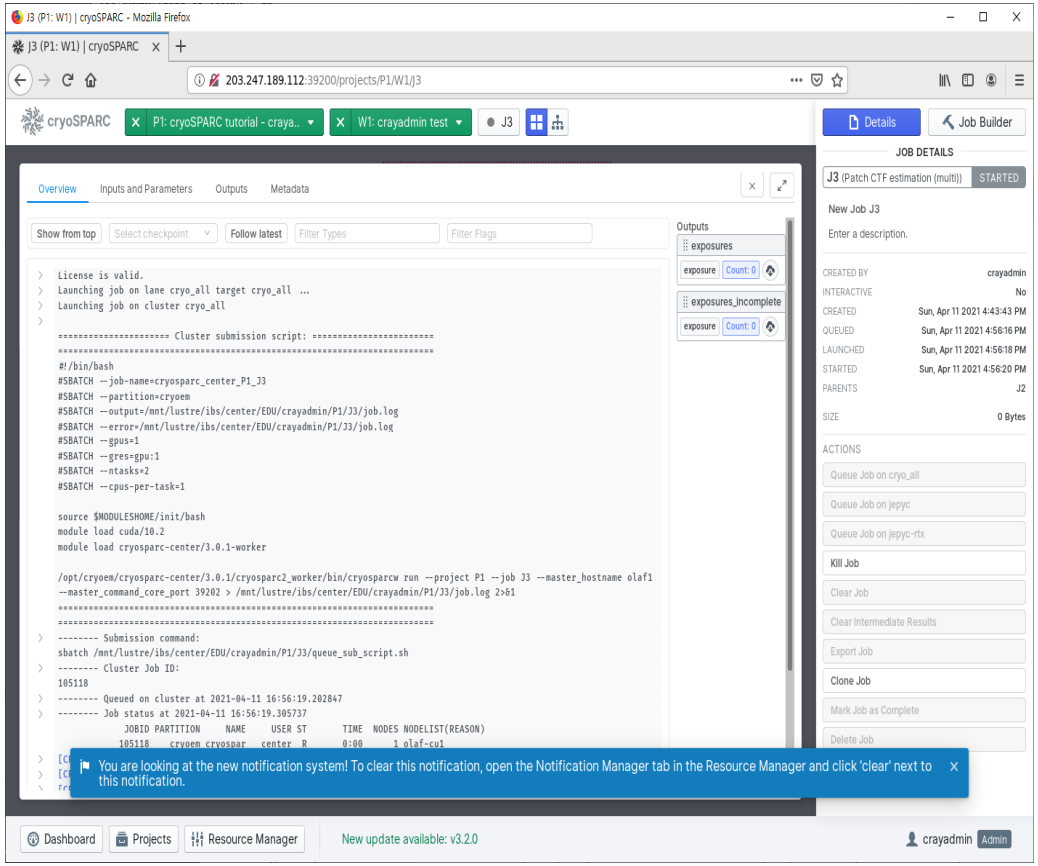
Click Queue

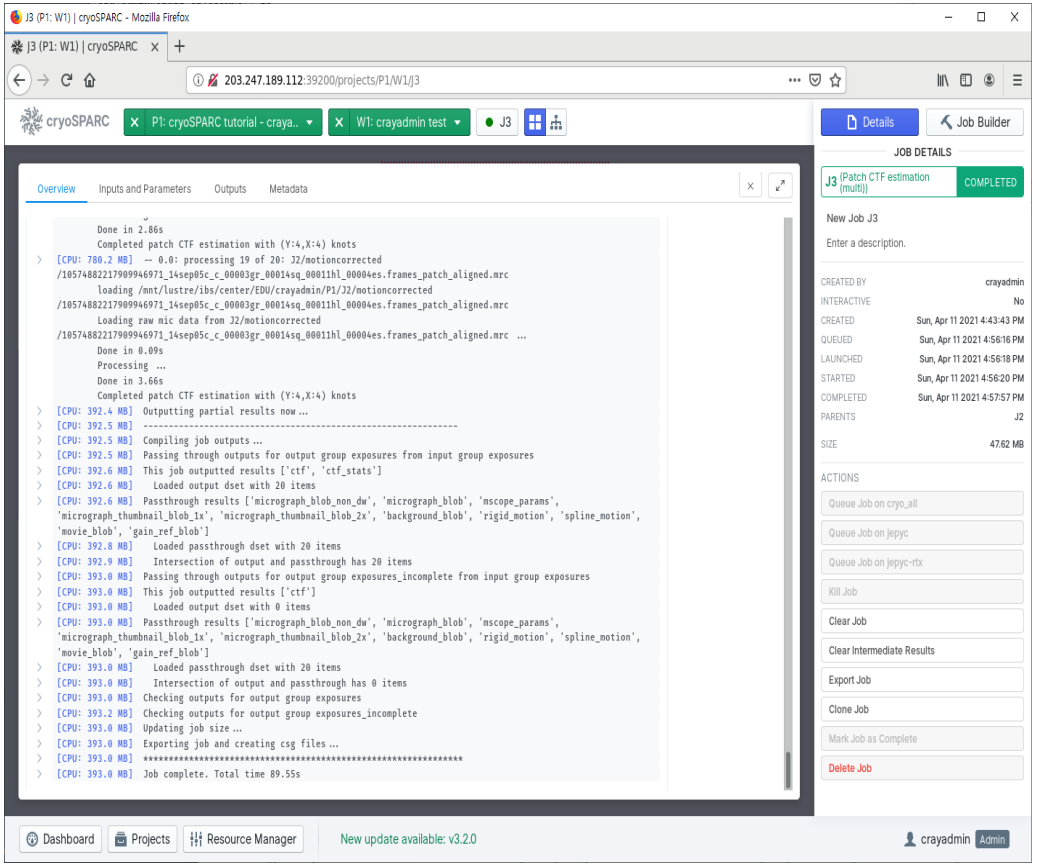
STEP 5


Select cryo_all lane
Click Create
Queue P1 → J3 (patch_ctf_estimation_multi)
Queue Type
Queue to Lane
Run on specific GPU
Lane
Lane cryo_all (cluster) (cluster)
cryo_all
Lane jepyc (cluster) (cluster)
jepyc
Lane jepyc-rtx (cluster) (cluster)
jepyc-rtx
Description
Enter a description.
Run Job in
Current Workspace (W1)
New Workspace
Cancel
Create
Dashboard
Projects
Resource Manager
New update available: v3.2.0
crayadmin
Admin

STEP 5

- CTF estimation - Done





THANK YOU