[SRDP-303] 6.4.1.2 Socorro Operations Lustre estimates high

Created: 15/Mar/19 Updated: 22/Mar/19

Status:	Topic For Meeting
Project:	Science Ready Data Products
Component/s:	SRDP Operations Plan Review
Affects Version/s:	None

Type:	Review item Discrepancy	**Priority:**	Minor
Reporter:	James Robnett	**Assignee:**	James Robnett
Resolution:	Unresolved	**Votes:**	0
Labels:	None		

Page Number:	13
Suggested Solution:	Not sure, I guess at some level get with me and help me understand where that number comes from and, if possible, order more storage ASAP.

Description

This section states 255TB will be needed for the pilot and Wave 1 for EVLA calibrations. That's more than VLASS uses for calibration and imaging and roughly 4x more than the baseline pipeline operations usage so I have a hard time understanding how it was arrived at.

There is no free usable space on the NM lustre so whatever delta SRDP will contribute needs to be accounted for.

Comments

Comment by Mark Lacy [22/Mar/19]

This just assumes the raw data volume times the expansion factor for calibrations and that calibrated MSes will hang around for about 1 month (both from pipeline runs and user restores) - see the details of the resource model in the Appendix. Certainly, based on ALMA operations I think this is realistic, I would not like to try to squeeze it down as it would be bad for the user experience and make e.g. bursts of processings after moves impractical.

Comment by James Robnett [22/Mar/19]

We'll need to discuss this in the review. The steady state EVLA data rate is ~1TB/day, 30TB a month so with expansion roughly 100TB. This varies as a function of array configuration so it can be as high as 125TB and as low as 75.

The non-VLASS EVLA pipeline is not removing SDMs on completion (VLASS does) so there are 4 copies not 3 of the data which gives a 120TB expansion with a high and low of around 100 to 150TB. It's currently at 144TB.

I don't want to second guess or short change SRDP, merely expressing a counter experience, so if in fact it needs 255TB then we'll have to order 2 Lustre servers at ~30K each.

[SRDP-286] DA vs. user defined parameters

Created: 14/Mar/19 Updated: 25/Mar/19

Status:	Topic For Meeting
Project:	Science Ready Data Products
Component/s:	SRDP Operations Plan Review
Affects Version/s:	None

Type:	Review item Discrepancy	**Priority:**	Minor
Reporter:	Anthony Remijan	**Assignee:**	Anthony Remijan
Resolution:	Unresolved	**Votes:**	0
Labels:	None		

	Relates

Issue Links:	relates to	~~SRDP-284~~	Inconsistency in WF3 and WF6 regardin...	Done
Page Number:	16			
Suggested Solution:	This need to be better explained as is related to ~~SRDP-284~~.			

Description

It is described that the DAs will "suggest changes to the parameters if it is clear that doing so will improve the results." I find this in contradiction to WF3 and WF6 where the user has requested re-imaging or calibration. Wouldn't the user be the most informed on how the parameters should be changed to improve the results.

Comments

Comment by Mark Lacy [22/Mar/19]

I still would like to keep this statement in here - many of our users will not necessarily be very experienced, and input from the DAs might be very valuable.

Comment by Anthony Remijan [25/Mar/19]

I would like further discussion on this point as I feel there is a disconnect between normal DA operations and what SRDP would be asking the DAs to assess in the workflows.

[SRDP-297] Section 2.1 - Overview

Created: 14/Mar/19 Updated: 25/Mar/19

Status:	Topic For Meeting
Project:	Science Ready Data Products
Component/s:	SRDP Operations Plan Review
Affects Version/s:	None

Type:	Review item Discrepancy	**Priority:**	Minor
Reporter:	Anthony Remijan	**Assignee:**	Anthony Remijan
Resolution:	Unresolved	**Votes:**	0
Labels:	None		

Page Number:	8
Suggested Solution:	Make an addition to the last sentence that while the managers are responsible for allocating effort it is not at the expense of core telescope deliverables.

Description

It should be clearly pointed out that the SRDP operations must not interfere with core deliverables to ALMA or the VLA when the division heads are allocating resources.

Comments

Comment by Jeff Kern [18/Mar/19]

I actually disagree that this is the appropriate place to make that statement. The SRDP project doesn't set the overall priorities, that is done by the division head and the management chain above. I don't disagree with the concept, just that this is not the appropriate place to put it.

Comment by Anthony Remijan [25/Mar/19]

I think that this should be discussed at the meeting as to where to put this statement in. I understand that SRDP does not set the priorities but it does have its own deliverables (as do the sites) so what falls off the table if the necessary resources are not available?

[SRDP-310] 3.1 meeting frequency Created: 15/Mar/19 Updated: 26/Mar/19

Status:	Topic For Meeting
Project:	Science Ready Data Products
Component/s:	SRDP Operations Plan Review
Affects Version/s:	None

Type:	Review item Discrepancy	**Priority:**	Minor
Reporter:	Claire Chandler	**Assignee:**	Mark Lacy
Resolution:	Unresolved	**Votes:**	0
Labels:	None		

Page Number:	8
Suggested Solution:	Increase the frequency of SRDP meetings, especially for the Pilot and early operations.

Description

From experience with VLASS operations, it is necessary to meet much more often than once a week for 30mins. When we start an observing campaign, we meet twice a day, and I imagine that the DAs would require and appreciate much more interaction than Jira tickets and a weekly meeting.

Comments

Comment by Mark Lacy [19/Mar/19]

I discussed this with Drew when I was last in Socorro, and he thought a weekly meeting was fine (with the ability to Skype etc for further impromptu meetings as needed if things crop up).

Comment by Claire Chandler [19/Mar/19]

Right, but Drew also asked not to use Jira in favour of more direct interaction. Especially when you're starting up the pilot I recommend more frequent meetings, and add this to the risk register as something to track for full operations.

Comment by Mark Lacy [22/Mar/19]

I'm reluctant to add formal meetings, especially as there are already many for the DAs to attend, especially while VLASS is taking data. I agree we should track this as a risk though in case less formal methods of communication are not effective.

Comment by Anthony Remijan [25/Mar/19]

I am worried about the formality of reporting issues since there is no similar PR system at the VLA that I am aware of... although I can be completely wrong. Reporting on issues through meetings may be problematic as things may easily slip through the cracks.

Comment by Claire Chandler [26/Mar/19]

I think I'd like to see "a day in the life" of SRDP to understand how this will work. We can discuss it at the meeting.

[SRDP-319] remote processing recommendation Created: 16/Mar/19 Updated: 26/Mar/19

Status:	Topic For Meeting
Project:	Science Ready Data Products
Component/s:	SRDP Operations Plan Review
Affects Version/s:	None

Type:	Review item Discrepancy	**Priority:**	Minor
Reporter:	Claire Chandler	**Assignee:**	Claire Chandler

Resolution:	Unresolved	**Votes:**	0
Labels:	None		

Page Number:	14
Suggested Solution:	Re-write this section to remove consideration of QA at non-NRAO processing centers.

Description

In section 7.3, the idea is floated that non-NRAO resources may be trained in QA processes. Since none of these resources will be under NRAO control I highly recommend that this not be considered (at least, initially), otherwise we have no means of quality control.

Comments

Comment by Jeff Kern [19/Mar/19]

The intention is that this would be part of the Radial project and that before handing the QA process over to someone external they would certainly need to be trained (and observed for a period). Mark I suggest you re-write the section to explicitly state that there would need to be a training program before this could be executed.

Comment by Claire Chandler [19/Mar/19]

NRAO would still be taking responsibility for the scientific validity of products not validated by NRAO. I really think this needs to be re-considered.

Comment by Mark Lacy [22/Mar/19]

In a sense NRAO already relies on non-NRAO employees for QA, in that JAO employees regularly QA NA ALMA data products. I agree though that training would be required.

Comment by Claire Chandler [26/Mar/19]

The case for ALMA is different, because JAO are all part of ALMA. The use of outside resources for SRDP QA potentially puts NRAO's credibility at risk. We should discuss this at the meeting, and if it is decided that it will stay in the plan, it will need to be added to the risk register.

[SRDP-309] 2.2 VLA feedback

Created: 15/Mar/19 Updated: 27/Mar/19

Status:	Topic For Meeting
Project:	Science Ready Data Products
Component/s:	SRDP Operations Plan Review
Affects Version/s:	None

Type:	Review item Discrepancy	**Priority:**	Minor
Reporter:	Claire Chandler	**Assignee:**	Mark Lacy
Resolution:	Unresolved	**Votes:**	0
Labels:	None		

Issue Links:	**Relates**			
	relates to	SRDP-314	Clarify what will be covered in each ...	Topic For Meeting
Page Number:	8			
Suggested Solution:	Specify who will attend the meeting from SRDP. The Operations Manager?			

Description

It is not clear whose responsibility it is to attend the weekly Thursday meeting to provide feedback from the pipeline to the VLA observatory staff.

Comments

Comment by Mark Lacy [22/Mar/19]

Currently, Emmanuel is attending these meetings, sentence added to the end of Section 2.2:

The head of Science User Support in Socorro collates feedback from the data analysts and presents any issues at these meetings.

Comment by Claire Chandler [26/Mar/19]

If the head of SUS is to represent SRDP at the weekly meetings this needs to be included in the required scistaff effort. Holding for general discussion of scistaff effort requirements at the meeting.

[SRDP-321] definition of science quality

Created: 16/Mar/19 Updated: 27/Mar/19

Status:	Topic For Meeting
Project:	Science Ready Data Products
Component/s:	SRDP Operations Plan Review
Affects Version/s:	None

Type:	Review item Discrepancy	**Priority:**	Minor
Reporter:	Claire Chandler	**Assignee:**	Mark Lacy
Resolution:	Unresolved	**Votes:**	0
Labels:	None		

Issue Links:	**Relates**			
	relates to	SRDP-285	Unrealistic/non-existent estimation o...	Topic For Meeting
	relates to	SRDP-314	Clarify what will be covered in each ...	Topic For Meeting
Page Number:	16			
Suggested Solution:	Add description of who decides that a result is of scientific quality or is scientifically-valid.			

Description

In section 8.1 the concept of a calibration or image being of "science quality" or a "scientifically-valid result" is introduced, but there is no mention of who or how the decision that something is of science quality is made. Nominally, I would think that this requires the input of a scientist, but as mentioned elsewhere, such resources have not been evaluated.

Comments

Comment by Mark Lacy [22/Mar/19]

Scientist resources have now been added (Section 6.3).

Comment by Anthony Remijan [25/Mar/19]

I don't think this completely addresses Claire's concern about an image being "science quality". Agree that the scientist effort has been added but the initial question is not addressed. Going to reopen this issue and leave it to Claire to determine if this needs to be discussed at the f2f - I believe it should be.

Comment by Anthony Remijan [25/Mar/19]

See added comment...

Comment by Claire Chandler [26/Mar/19]

I think that this question relates closely to SRDP-283. Who decides that something is "science ready"? Is it the user (the scientist who is going to use the data) or an NRAO scientist? For PI science, it probably should not be a DA, because it can require a detailed understanding of the science in order to understand whether a product can be used for science. We can hold this discussion for the meeting.

[SRDP-314] Clarify what will be covered in each wave

Created: 15/Mar/19 Updated: 27/Mar/19

Status:	Topic For Meeting
Project:	Science Ready Data Products
Component/s:	SRDP Operations Plan Review
Affects Version/s:	None

Type:	Review item Discrepancy	**Priority:**	Minor
Reporter:	Claire Chandler	**Assignee:**	Mark Lacy
Resolution:	Unresolved	**Votes:**	0
Labels:	None		

Issue Links:	**Relates**			
	relates to	SRDP-285	Unrealistic/non-existent estimation o...	Topic For Meeting
	relates to	SRDP-309	2.2 VLA feedback	Topic For Meeting
	relates to	SRDP-321	definition of science quality	Topic For Meeting
Page Number:	11			
Suggested Solution:	Add further detail as to what is covered in each wave, pp 11-14.			

Description

It was not at all clear what is to be covered in the different waves described. For example, where to spectroscopy and polarimetry for the VLA appear? Do these more complex observations require additional resources for the QA? The plan seems to assume that wave 2 = steady state, but I foresee considerable more effort needed unless a lot of automated QA is developed, which is not described.

Comments

Comment by Mark Lacy [22/Mar/19]

SRDP operations will execute the pipeline capabilities in each Wave as they become available. The capabilities roadmap for the pipeline is yet to be developed, and will be the responsibility of the SRDP Project Scientist, with advice from the SRDP Stakeholders Committee.

Comment by Claire Chandler [26/Mar/19]

It is not clear how you can estimate the resource requirements without knowing what will be covered. A complex spectroscopy project could take much more time than a simple continuum observation, similarly for large surveys vs. single fields.

Comment by Mark Lacy [26/Mar/19]

The mitigation for hardware resources in cases like these is to adjust the maximum image size constraint for SRDP (MaxImgSize in the resource model). For some complicated projects though the extra resource need will be on the human side rather than the hardware side. For this, I agree we will need to be cautious that introducing new capabilities comes with improvements to the pipeline and weblog that ensure that QA does not become significantly more burdensome.

Comment by Claire Chandler [26/Mar/19]

Sorry, I should have specified that I was referring to human (DA and scistaff) resources, not compute resources.

Comment by Jeff Kern [27/Mar/19]

Putting this on the agenda for discussion at the meeting, I think it is related to SRDP-285, SRDP-309, SRDP-321

[SRDP-285] Unrealistic/non-existent estimation of the role of the scientific staff in SRDP operations

Created: 14/Mar/19 Updated: 27/Mar/19

Status:	Topic For Meeting
Project:	Science Ready Data Products
Component/s:	SRDP Operations Plan Review
Affects Version/s:	None

Type:	Review item Discrepancy	**Priority:**	Major
Reporter:	Anthony Remijan	**Assignee:**	Anthony Remijan
Resolution:	Unresolved	**Votes:**	0
Labels:	None		

Issue Links:	**Relates**			
	relates to	SRDP-321	definition of science quality	Topic For Meeting
	relates to	SRDP-314	Clarify what will be covered in each ...	Topic For Meeting
Page Number:	16			
Suggested Solution:	Think critically about the role the scientific staff plays in SRDP operations especially in the Pilot an early waves. The very limited use of the NRAO staff as described in the document is underestimated.			

Description

The only place where scientist involvement is described in the operations plan is in WF2 in section 8.1. I think this grossly underestimates the role of the scientific staff in SRDP operations. Especially during the Pilot and Wave 1, support will be needed from the scientific staff in order to help in operations.

Comments

Comment by Claire Chandler [15/Mar/19]

I concur that the level of scientific staff support needed for SRDP operations is distinctly lacking. Several sections need to recognize that scientific staff will be needed (2.1, 2.3,8.1/WF3) and others mention scientific staff involvement but not how much (3.1, 3.3, 8, 8.1). I suggest adding a subsection to section 6 "Resource requirements" describing the scientific staff support needs, as is currently in place for the Data Analysts.

Comment by Mark Lacy [21/Mar/19]

Scientists have been added to the Operations Model, and a new Section 6.3 added:

Scientist effort to help with SRDP operations (for the data analysts to consult if they have questions about QA, and also for responding to user queries that lie beyond a DA's expertise) is estimated at 10% of the DA effort (based on NA ALMA data processing operations). For the pilot and Wave 1, scientist effort in Socorro has been identified, and the Operations Manager will provide science support in Charlottesville. Any further resource needs will be assessed at the review of Wave 1 operations.

Comment by Anthony Remijan [25/Mar/19]

Just to be clear, the estimate of 10% of the DA efforts gives:

CVL: Scientist effort for the Pilot will be 9.4 hours and for Wave 1, 9.4 hours/quarter

Socorro: 26.7 hours for the Pilot and 119.5 hours/quarter for Wave 1

Do I understand this correctly?

Comment by Mark Lacy [25/Mar/19]

In the latest Tables 2-5 we have slightly less for Cville, Socorro is as you inferred though:

CVille: 7 hrs for pilot, 7hr/quarter for Wave 1

Socorro: 27hr for pilot, 120hr/quarter for Wave 1

Comment by Anthony Remijan [25/Mar/19]

Thanks for the clarification Mark and it aligns with what I thought you suggested. I want to hear from James, Lewis and Claire if they believe this assessment is adequate before closing as complete. It may need some discussion at the f2f before closing out this issue.

Comment by Claire Chandler [26/Mar/19]

I suspect the estimates are still too low, especially for the pilot as things are ramping up. I can easily imagine a scientist needing to spend an hour a day during the pilot for the VLA, which amounts to 65 hours.

The sections I noted previously that should include scientific staff resources (2.1, 2.3, 8.1/WF3) seem not to have been updated yet.

Comment by Anthony Remijan [26/Mar/19]

Let's discuss f2f... Thanks Claire! That was my assessment as well.