PO.DAAC Cloud Scenarios for SWOT

Myche McAuley

ESIP Summer 2016

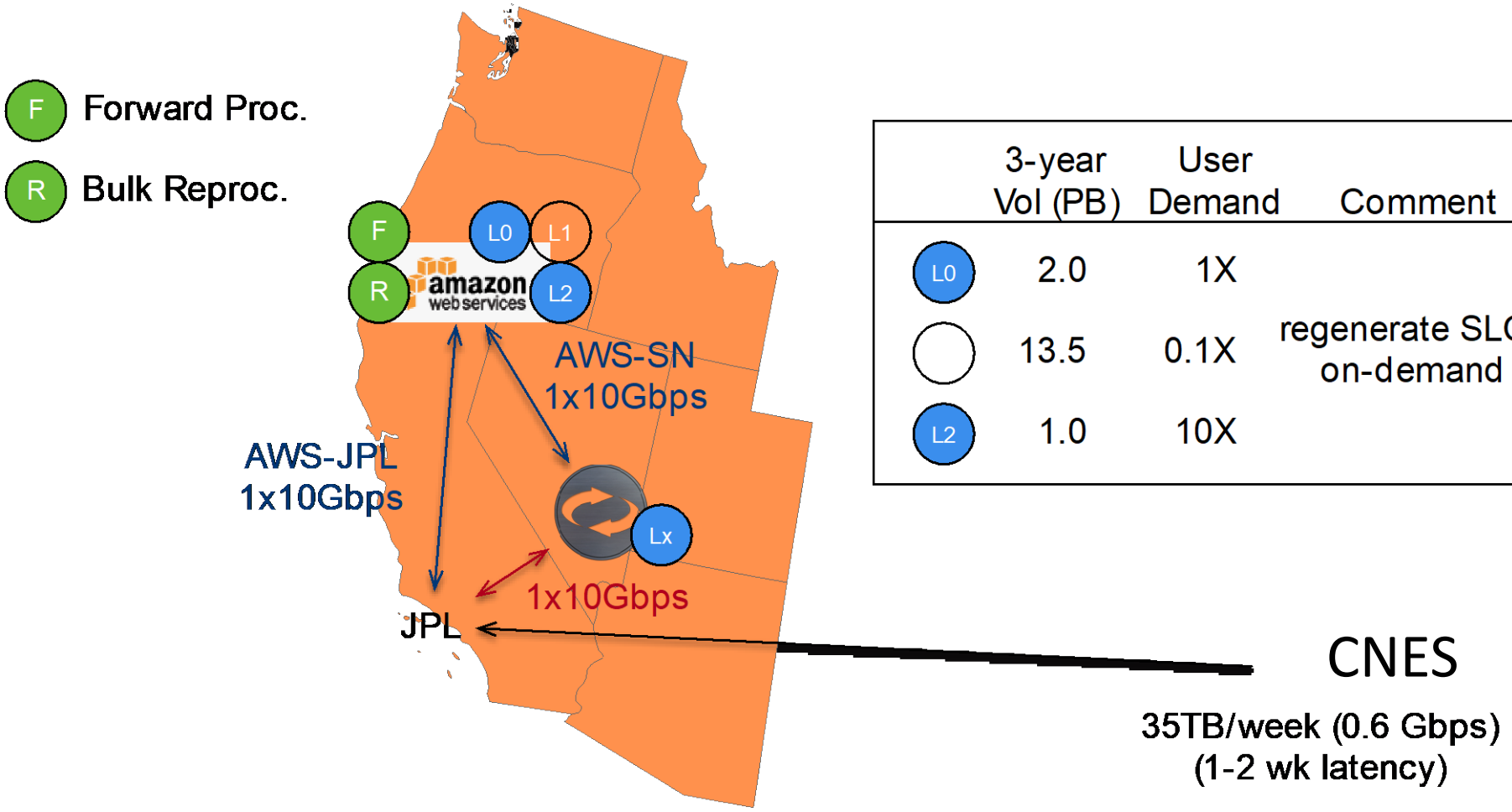
2016-07-21

Jet Propulsion Laboratory, California Institute of Technology.

© 2016 California Institute of Technology. Government sponsorship acknowledged.

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsements by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

Candidate Deployment Scenario



	3-year Vol (PB)	User Demand	Comment
L0	2.0	1X	
	13.5	0.1X	regenerate SLCs on-demand
L2	1.0	10X	

Cost Considerations:

- Egress costs:
  - Users bring their algorithms to the data (behavioral change)
  - Replicate most downloaded data to external object store (e.g. SuperNAP)
    - Conceivable to simply duplicate all L2+ data to external store (helps with multiple copy requirement)
- Regenerate large, rarely-accessed products on-demand
  - Huge cost savings with some added complexity (all previously released product versions)
- Store only latest versions after reprocessing
  - Large cost savings – but at the expense of archive completeness
    - If we can regenerate previous versions, then everybody wins
- Use highly aggressive storage lifecycle management
  - Hottest X% of data in S3; the rest in S3-IA or Glacier (X to be tuned over time)
  - Monitor hazard feeds and pre-promote data in selected areas

PO.DAAC Solution May Change Over Time:

- The solution PO.DAAC implements <u>may</u> <u>change</u> over time, because:
  - The egress costs are effectively unbounded, and
  - Archive storage doesn’t scale like compute
- The “best” solution changes depending on:
  - when you ask the question; and to whom
  - what you value most, therefore:
- PO.DAAC needs a malleable solution, one that can migrate over time
- PO.DAAC is looking to “buy time”™