



Large Synoptic Survey Telescope

www.lsst.org

LSST Data Management Infrastructure

D. Michael Freemon¹, J. Becla², Gregory P. Dubois-Felsmann², J. Kantor³, K. T. Lim², R. Plante¹

¹NCSA, ²SLAC, ³LSSTC

(on behalf of the LSST Data Management team)

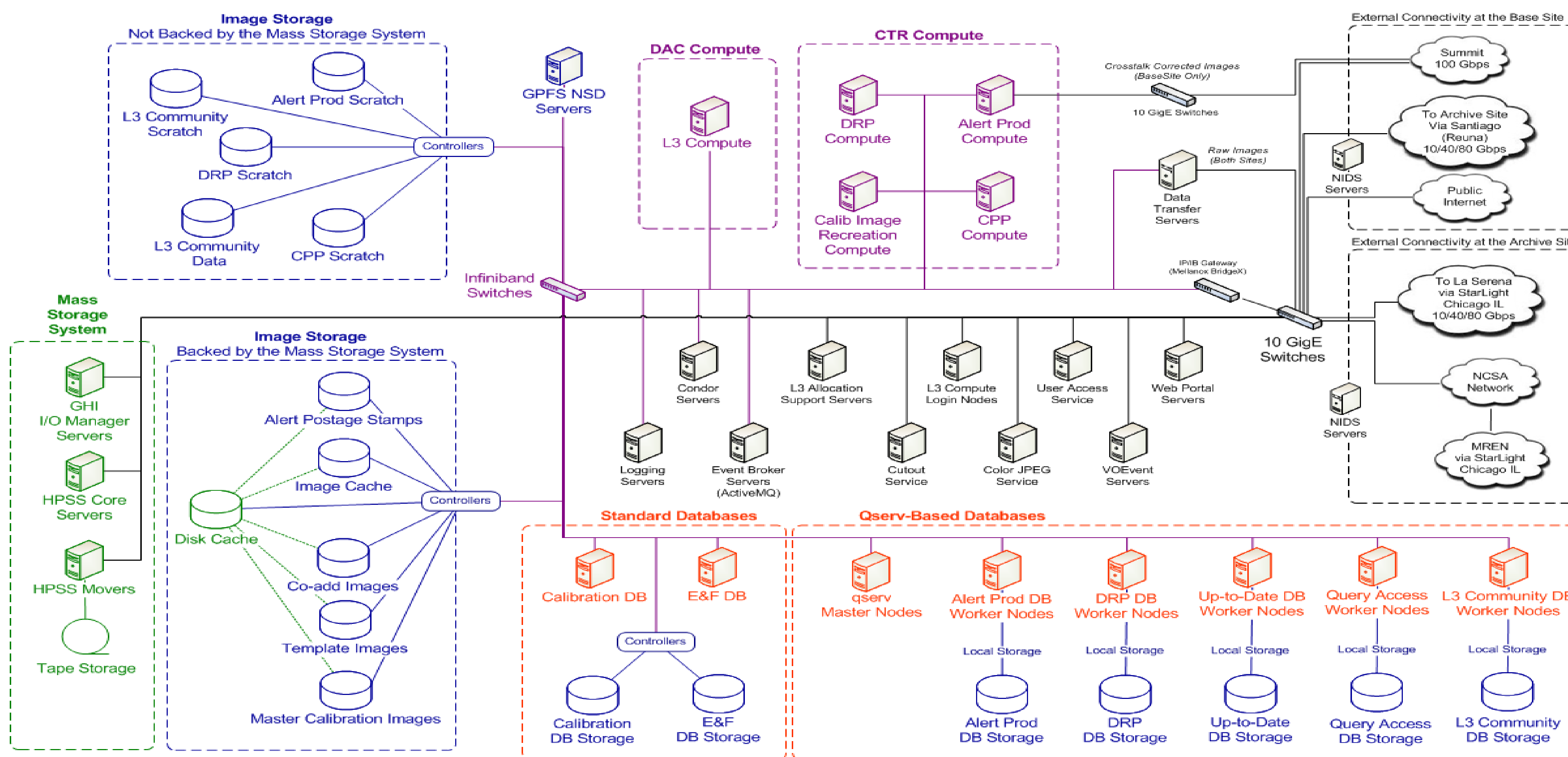
The Large Synoptic Survey Telescope (LSST) project is a proposed large-aperture, wide-field, ground-based telescope that will survey half the sky every few nights in six optical bands. The 8.4-meter telescope will be located in the Andes mountains near La Serena, Chile. The 3.2 Gpixel camera will generate 6.4 GB images at roughly a 17-second cadence, with alerts generated within one minute.

LSST will produce a data set suitable for answering a wide range of pressing questions in astrophysics, cosmology, and fundamental physics. What is dark energy? What is dark matter? How did the Milky Way form? What are the properties of small bodies in the solar system? Are there potentially hazardous asteroids that might impact the Earth and cause significant damage? Are there new kinds of undiscovered explosive phenomena?

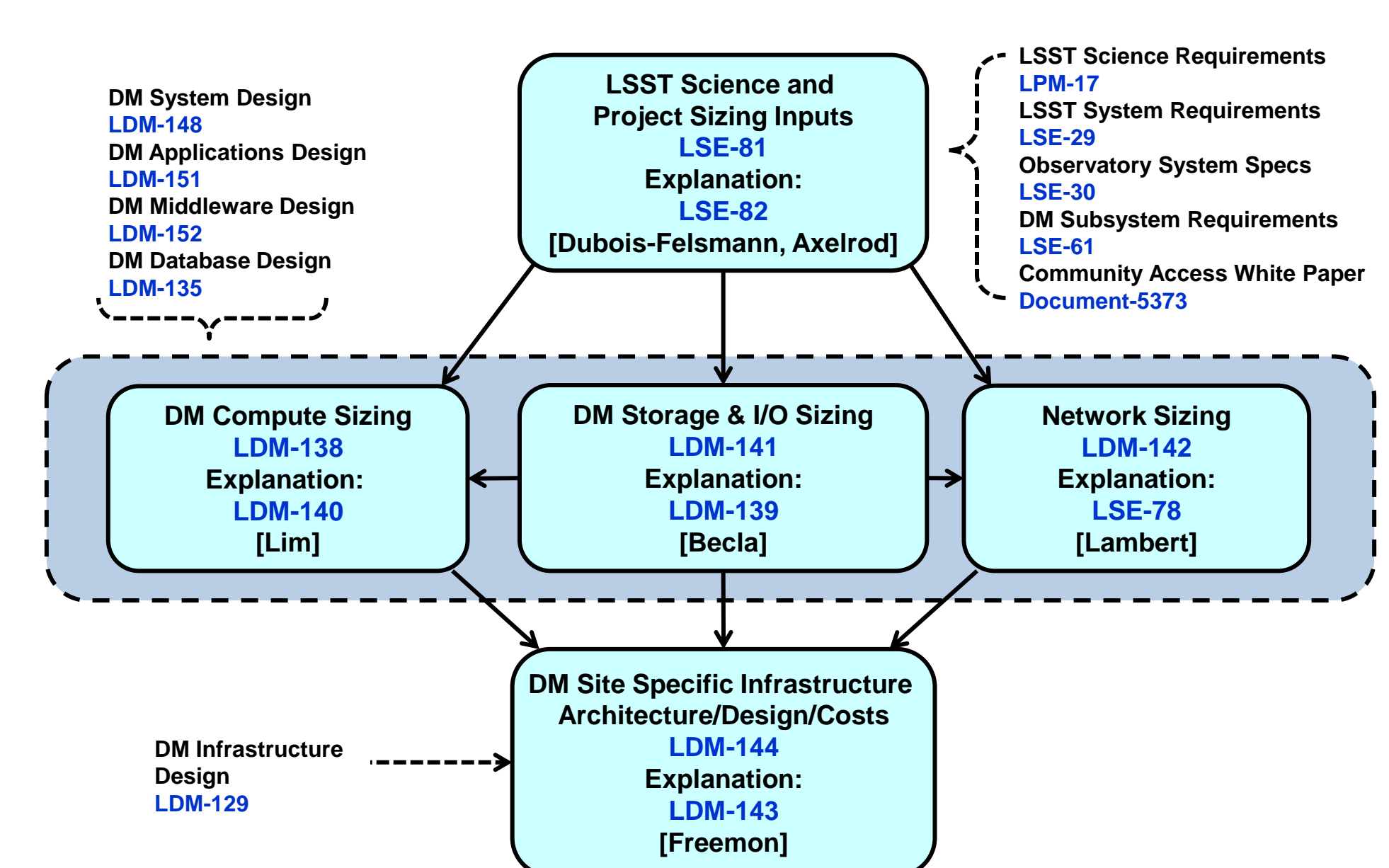
The Data Management system for the LSST will have to perform near-real-time calibration and analysis of acquired images, particularly for transient detection and alert generation; annual processing of the entire dataset for precision calibration, object detection and characterization, and catalog generation; and support of user data access and analysis.

Processing such a large volume of data, converting the raw images into a faithful representation of the universe, automated data quality assessment, automated discovery of moving or transient sources, and archiving the results in useful form for a broad community of users is a major challenge, and requires a significant investment in computing infrastructure.

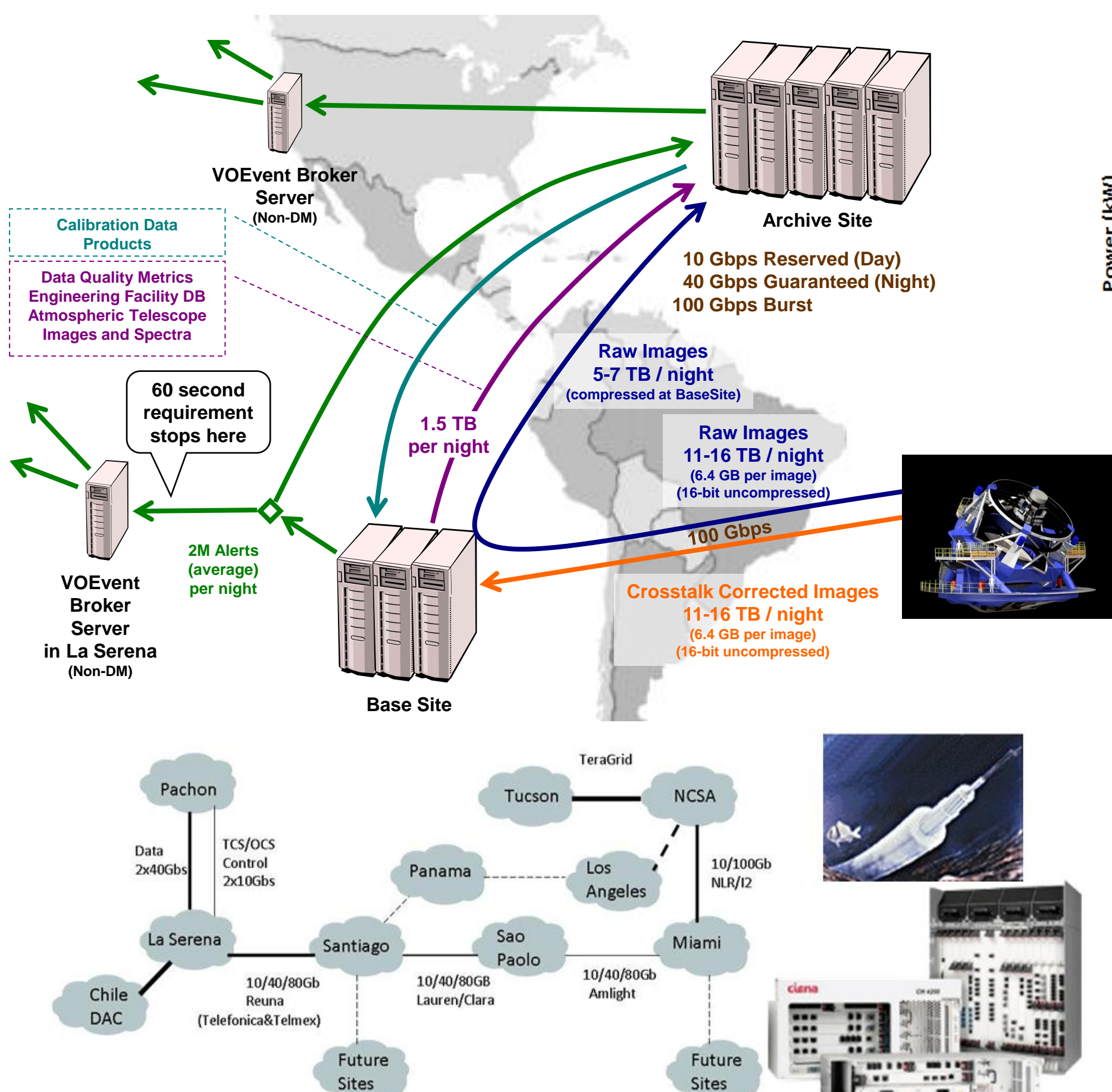
Computing Infrastructure at the Base and Archive Sites



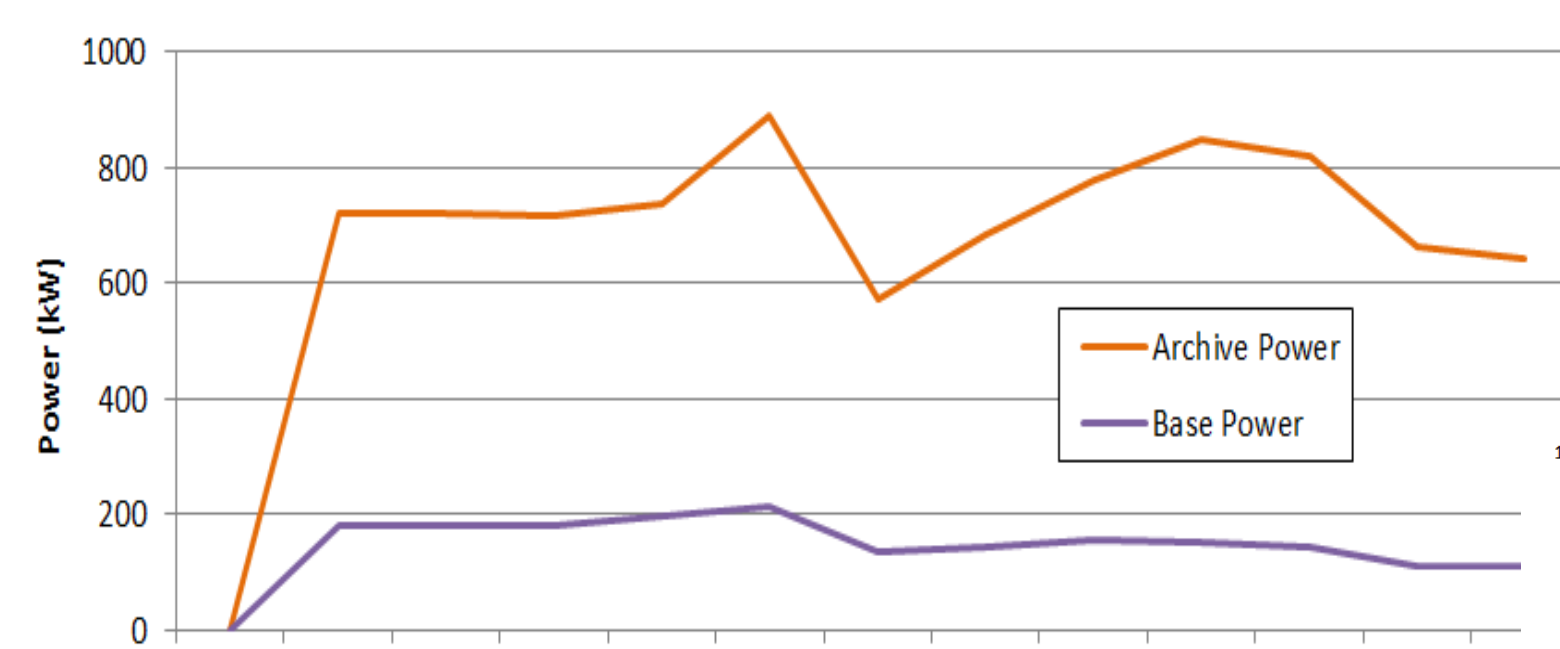
Infrastructure Sizing Model



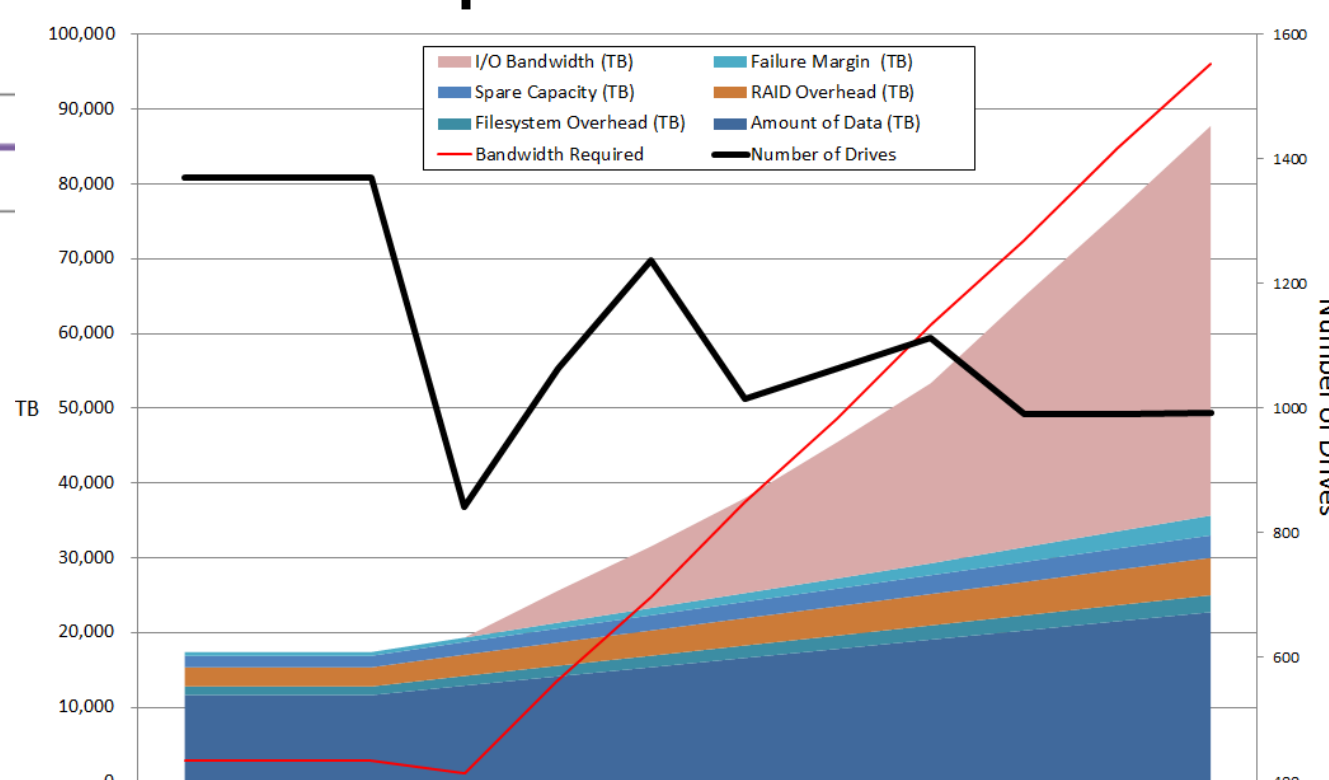
The Long Haul Network & International Data Flows



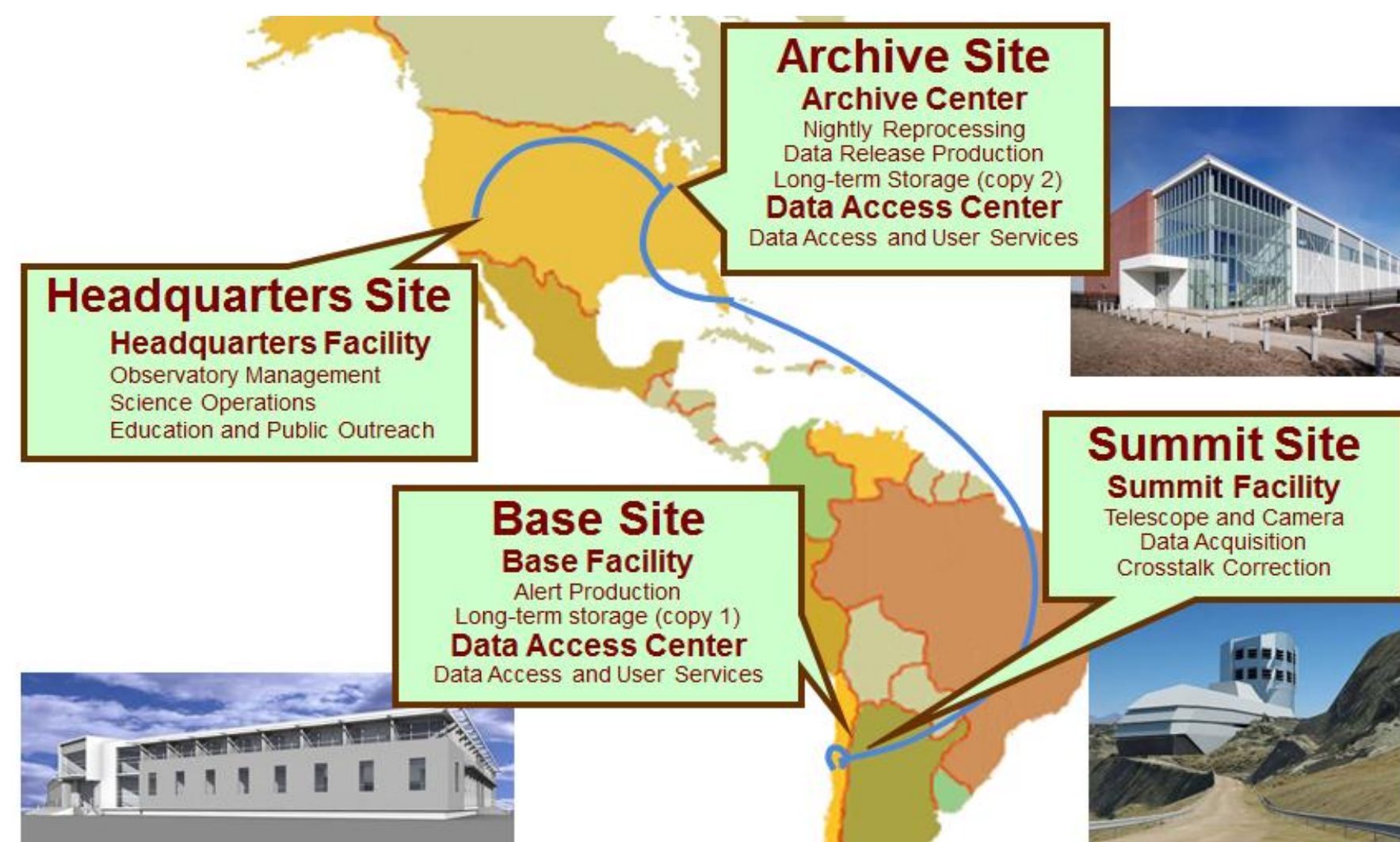
Power Consumption at the Sites



Spinning Disk Capacities & Drives



Sites, Facilities, and Centers



Estimates from Sizing Model

		Archive Site	Base Site
Compute	Teraflops (sustained)	175 → 1000 (required) 270 → 1600 (on-the-floor)	30 → 45 (required) 50 → 80 (on-the-floor)
	Nodes	1700 → 1400 (on-the-floor)	300 → 60 (on-the-floor)
	Cores	70K → 280K (on-the-floor)	12K → 13K (on-the-floor)
	Memory Bandwidth	25 → 130 TB/s (required)	4 → 6 TB/s (required)
Storage	File I/O Bandwidth	120 → 425 GB/s (required)	25 → 30 GB/s (required)
	Capacity	13 → 25 PB (required) 19 → 100 PB (on-the-floor)	8 → 8 PB (required) 12 → 23 PB (on-the-floor)
	Drives	1500 → 1100 (on-the-floor)	950 → 275 (on-the-floor)
	Disk Bandwidth	120 → 425 GB/s (required)	27 → 31 GB/s (required)
Mass Storage	Capacity	7 → 75 PB (required) 8 → 83 PB (on-the-floor)	7 → 75 PB (required) 8 → 83 PB (on-the-floor)
	Tapes	1000 → 3800 (near line) 1000 → 3800 (offsite)	1000 → 3800 (near line)
	Tape Bandwidth	6 → 24 GB/s (required)	6 → 24 GB/s (required)
	Teraflops (sustained)	16 → 193 PB (required)	12 → 126 PB (required)
Database	Storage Capacity	16 → 199 PB (on-the-floor)	12 → 133 PB (on-the-floor)
	Disk Drives	4 → 50 PB (required)	3 → 40 PB (required)
	Disk Bandwidth (sequential)	10 → 128 PB (on-the-floor)	7 → 95 PB (on-the-floor)
	Database Nodes	1400 → 2600 (on-the-floor)	1000 → 2000 (on-the-floor)
Facilities	Floorspace	125 → 625 GB/s (required)	95 → 425 GB/s (required)
	Power	100 → 190 (on-the-floor)	80 → 130 (on-the-floor)
	Cooling	950 → 850 sq ft (1100 high water mark) 720 → 640 kW (880 high water mark) 2.4 → 2.1 mmbtu (3.0 high water mark)	460 → 350 sq ft (470 high water mark) 180 → 110 kW (210 high water mark) 0.6 → 0.4 mmbtu (0.7 high water mark)