



Managed by Fermi Research Alliance, LLC for the U.S. Department of Energy Office of Science

Resources Analysis

Stu Fuess

Scientific Computing Portfolio Management Team (SC-PMT) Review

5 March 2015

Outline

- Processor requests
 - There was some information supplied about peak needs, but too difficult to analyze in a short time...
 - The changed needs resulting from large memory jobs is not included here; roughly, if an experiment needs 2x the memory then it needs 2x the CPU-hours... with $\frac{1}{2}$ the CPUs wasted
- Disk requests
- Tape requests
- There are a number of other requests that are not summarized here, but should go into a report
 - Systems for faster code builds
 - Better FTS servers

Processor needs analysis

- Partial year problem...
 - Experiments talk “slots”, which is a measure of needs at a point in time
 - Would like to talk in CPU-years, or the integral
 - For almost all cases, I’m assuming that slot needs are the average over the year, so just multiply $*24*365$
 - When showing FY15 Request numbers, assume that past = future
 - Exceptions are where experiments need to dramatically increase usage during remainder of FY15, e.g. Mu2e
 - I haven’t made this correction here

Processors: Past, Current, Requests

Experiment	14 Request	14 Actual	15 Quota	15 Request	16 Request	17 Request
MINERVA	21.99	9.29	14.02	22.00	28.60	34.32
MINOS	17.52	15.02	10.51	18.00	22.00	25.00
NOVA	7.88	10.18	11.39	16.00	16.00	0.00
MU2E	15.77	9.64	8.76	14.00	20.00	20.00
ELBNF	10.51	5.37	4.38	12.70	10.90	10.90
LARIAT	0.00	0.30	0.22	10.00	20.00	0.00
GM2	7.01	0.38	1.75	5.20	7.00	8.80
MARS	0.00	6.10	17.96	4.70	6.15	0.00
SEAQUEST	0.00	0.65	1.75	1.80	1.20	0.00
NUMIX	0.00	0.00	0.00	1.00	1.00	1.00
MICROBOONE	3.07	1.33	4.38	0.50	1.20	1.50
DES	9.64	0.35	8.58	0.49	0.62	0.00
CDMS	0.00	0.33	4.38	0.32	0.32	0.32
DS50	0.00	0.63	0.22	0.30	0.50	0.55
COUPP	0.00	0.18	1.75	0.30	0.40	0.00
Other	0.00	0.00	1.31	0.04	0.04	0.13
TOTAL	93.38	59.75	91.37	107.35	135.93	102.52

Units are Million CPU-hours

Processors: Past, Current, Requests

Same as previous, but just top 6 FY15 Request

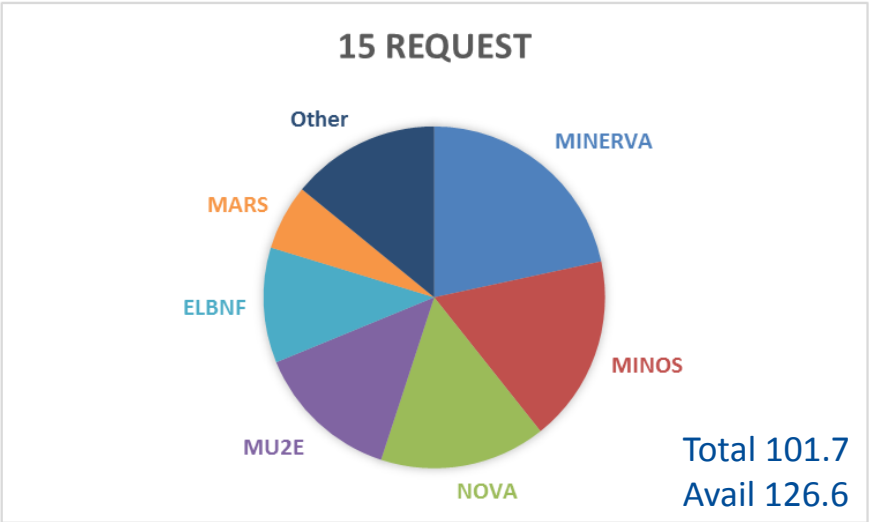
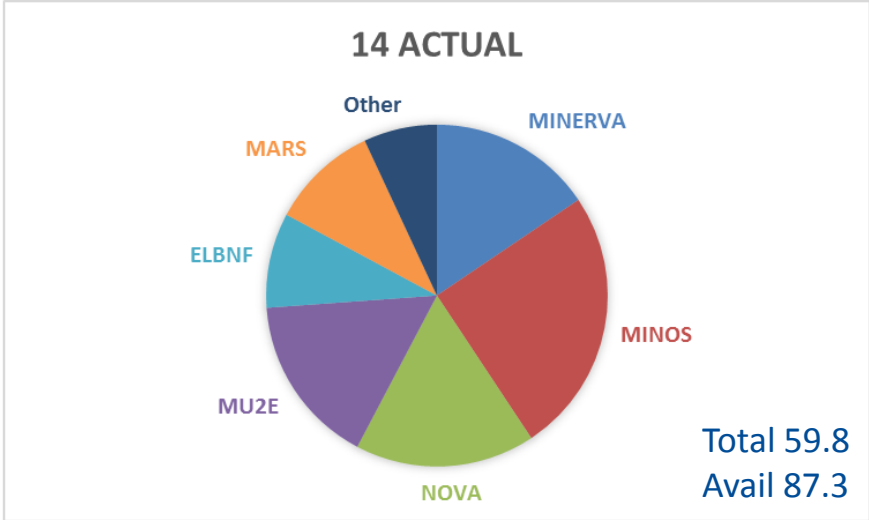
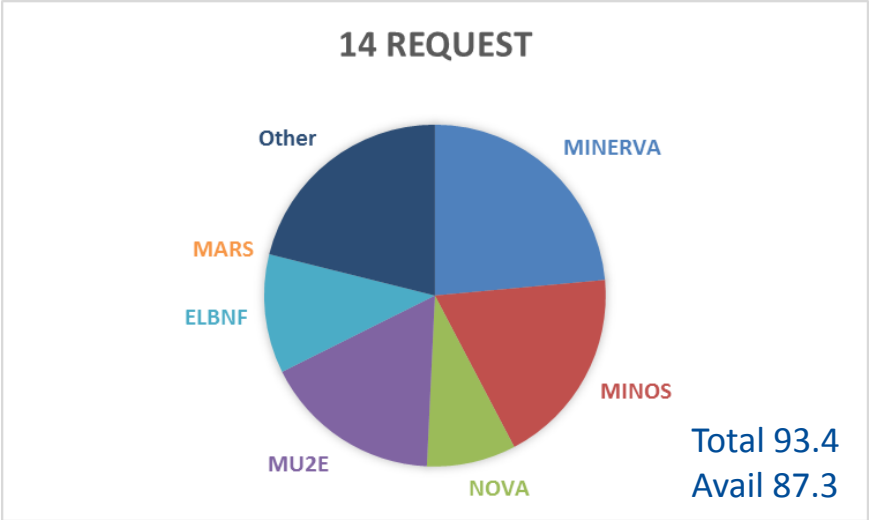
Experiment	14 Request	14 Actual	15 Quota	15 Request	16 Request	17 Request
MINERVA	22.0	9.3	14.0	22.0	28.6	34.3
MINOS	17.5	15.0	10.5	18.0	22.0	25.0
NOVA	7.9	10.2	11.4	16.0	16.0	16.0
MU2E	15.8	9.6	8.8	14.0	20.0	20.0
ELBNF	10.5	5.4	4.4	11.1	10.3	10.9
MARS	0.0	6.1	18.0	6.3	6.8	0.0
Other	19.7	4.2	24.4	14.3	21.1	12.4
TOTAL	93.4	59.8	91.4	101.7	124.7	118.6
Available	87.3	87.3	87.3	126.6	158.7	164.8

Move CDF, D0 -> GP

Available = past / proposed GP Grid
with minimal retirements and ~ \$100K = 17M CPU-hr purchases
in 15 (available for use in 16) and 16 (for use in 17)

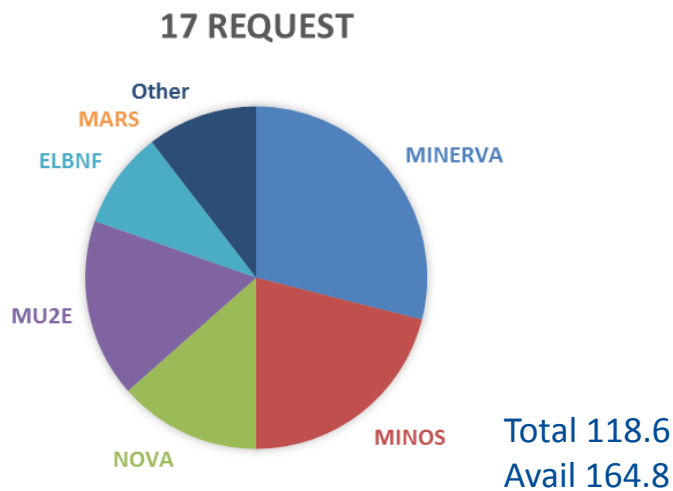
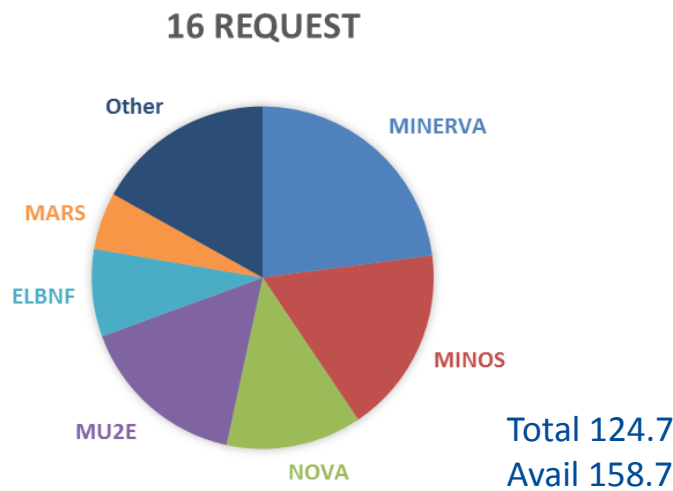
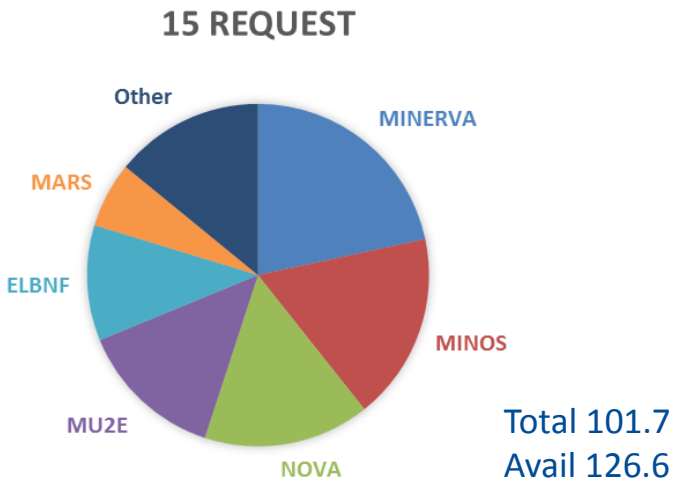
Units are Million CPU-hours

Processors: Past, Current, Requests



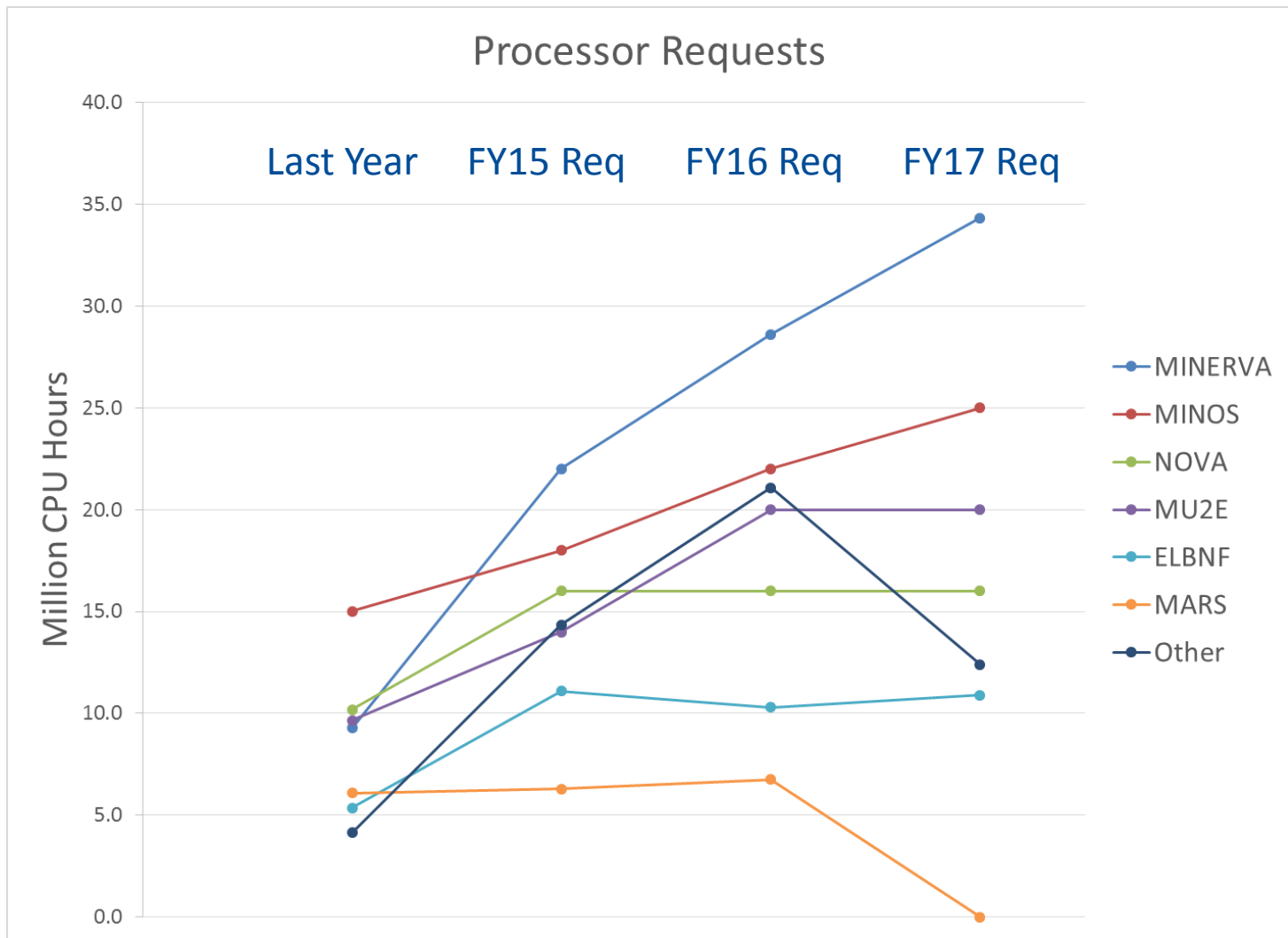
Units are Million CPU-hours

Processors: Past, Current, Requests



Units are Million CPU-hours

Processors: Evolution of needs



Disks: dCache Requests

Experiment	2014 Req	Actual	FY15 Req	FY16 Req	FY17 Req
MINERVA	950	251	1,330	2,380	3,380
NOVA	900	1,599	1,000	1,000	0
DS50	200	599	820	970	1,120
ELBNF	200	6	590	592	592
MINOS	450	63	570	920	1,020
MU2E	50	56	100	150	150
Other	520	511	179	254	240

Only the top 6 FY15 requestors are listed

Total requests (cumulative, i.e. amount requested to be available):

- FY15 4.6 PB
- FY16 6.3 PB
- FY17 6.5 PB

We currently have 3.3 PB in production, with another 2 PB on the floor to be deployed.

But this is cache... so what does any of this mean?

- We are re-evaluating how we deploy in order to make requests/predictions useful

Disks: BlueArc Requests

- Showed the current BlueArc allocations in the introduction talk yesterday (included here in backup slides)
- Very few requests for additional space:

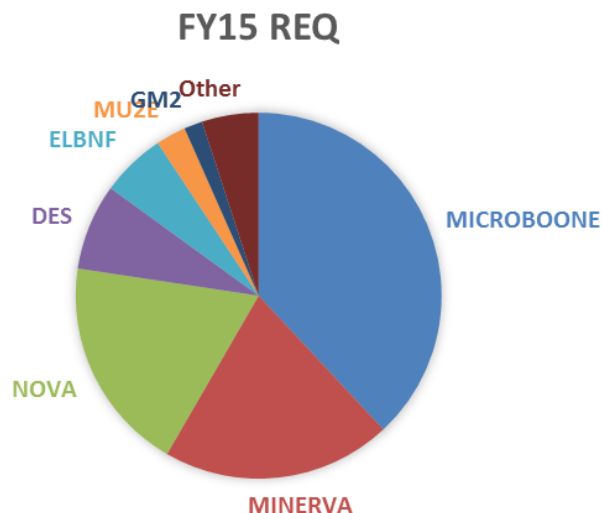
Experiment	Requested Addition (TB)
CHIPS	1
COUPP	10
G-2	50
LARIAT	2

- SCD needs to examine BlueArc replacements (for arrays out of warranty)
 - High replacement costs!
 - Annual maintenance costs (~10% of original cost) ~ cost of new dCache disks!

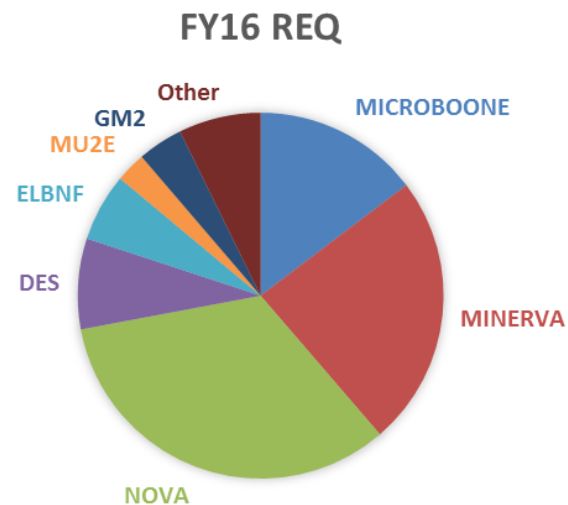
Tape storage requests

Experiment	FY15 Req	FY16 Req	FY17 Req
MICROBOONE	3000	1100	1600
MINERVA	1600	1800	2100
NOVA	1500	2500	0
DES	600	600	600
ELBNF	450	450	100
MU2E	210	200	300
GM2	128	300	500
Other	395	545	350

Units of TB



Total addition ~7.9 PB



Total addition ~7.5 PB

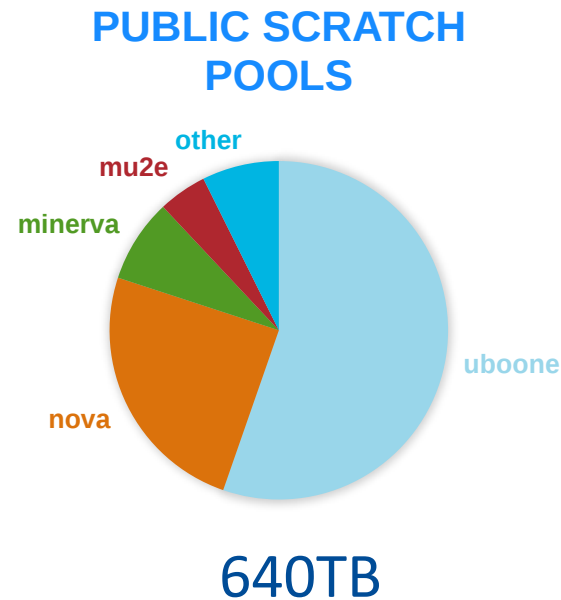
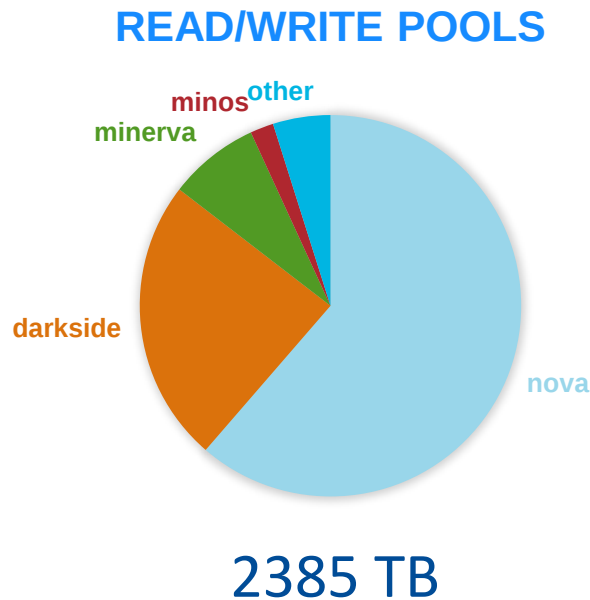
Backup slides

Disks: dCache Resources

Pool Type	Sum of Total TB	
readWritePools	2442.4	Tape backed
PublicScratchPools	656.6	Not tape backed
SfaPoolGroup	75.8	Internal use by tape facility
MinervaWritePools	65.5	Raw data from experiment DAQ
ArchivePools	59.9	Used by Active Archive Facility
DESPools	45.1	DES
NovaWritePools	39.9	Raw data from experiment DAQ
LQCDPools	39.9	LQCD to/from tape
RawDataWritePools	24.1	Raw data from experiment DAQ
EmptyPools	22.1	Currently not in use
ExpDbWritePools	22.1	Database backups
FermiGridVolPools	9.4	FermiGrid scratch area
Grand Total	3502.9	

3.5 Petabytes total
~ \$100/TB (direct cost)
~ 3x for surroundings
→ ~ \$1M investment

Disks: Major dCache Users



How do you interpret requests for cache disk space?

- Transient space for writing raw or reconstructed data to tape
- Transient space for staging data for reconstruction / analysis passes
- Long-lived space for data that will be accessed multiple times

Note that “permanent” space not available in current model for dCache

- But working on design to provide this functionality

Tape

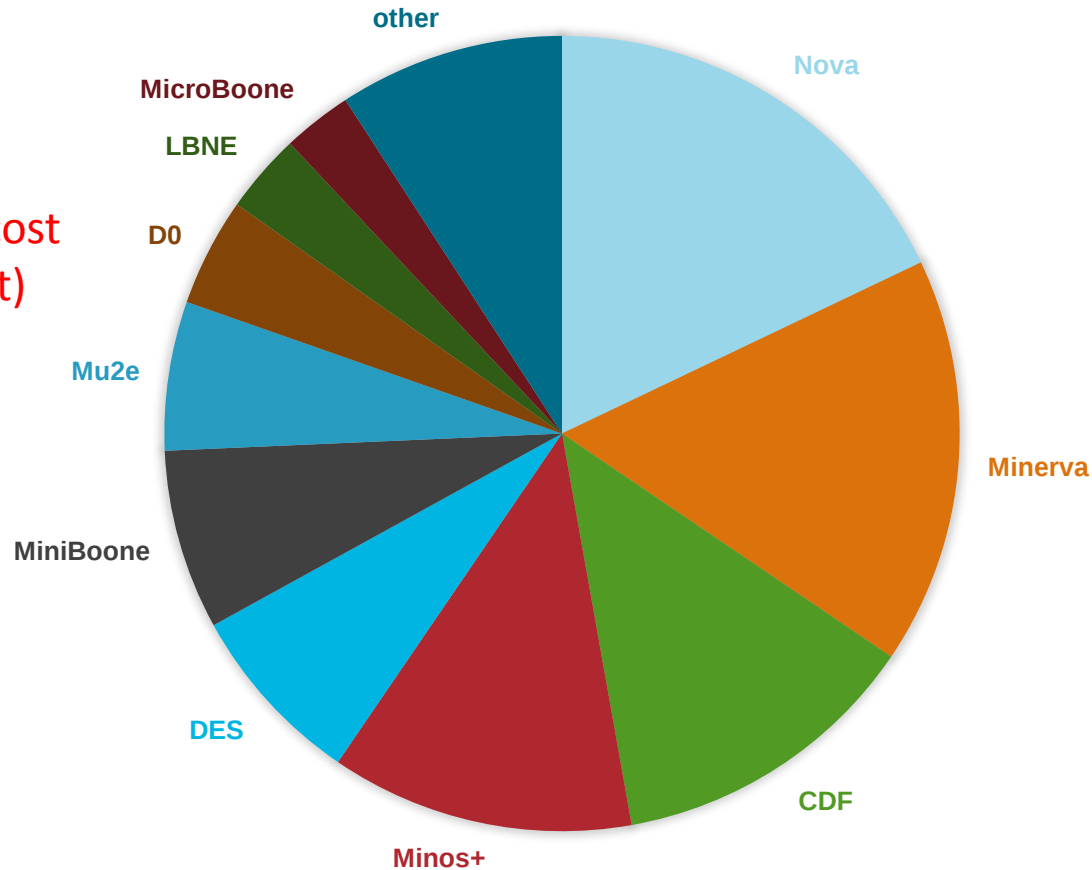
- Current complex has 7x 10,000 slot tape **libraries**
 - 3x are dedicated to CMS
 - Remaining 4x, with most recent tape technology (> 8 TB per) provides ~320 PB of capacity
 - So capacity is not an issue
- Assume that ~ \$50/TB for tape **media** and peripherals
 - So cost is only an issue when speaking PB size volumes
- Will track requests and provision appropriately, but otherwise need not pay much attention to tape storage requests

Disks: Major BlueArc Users

BLUEARC ALLOCATIONS

Total ~1.9 PB

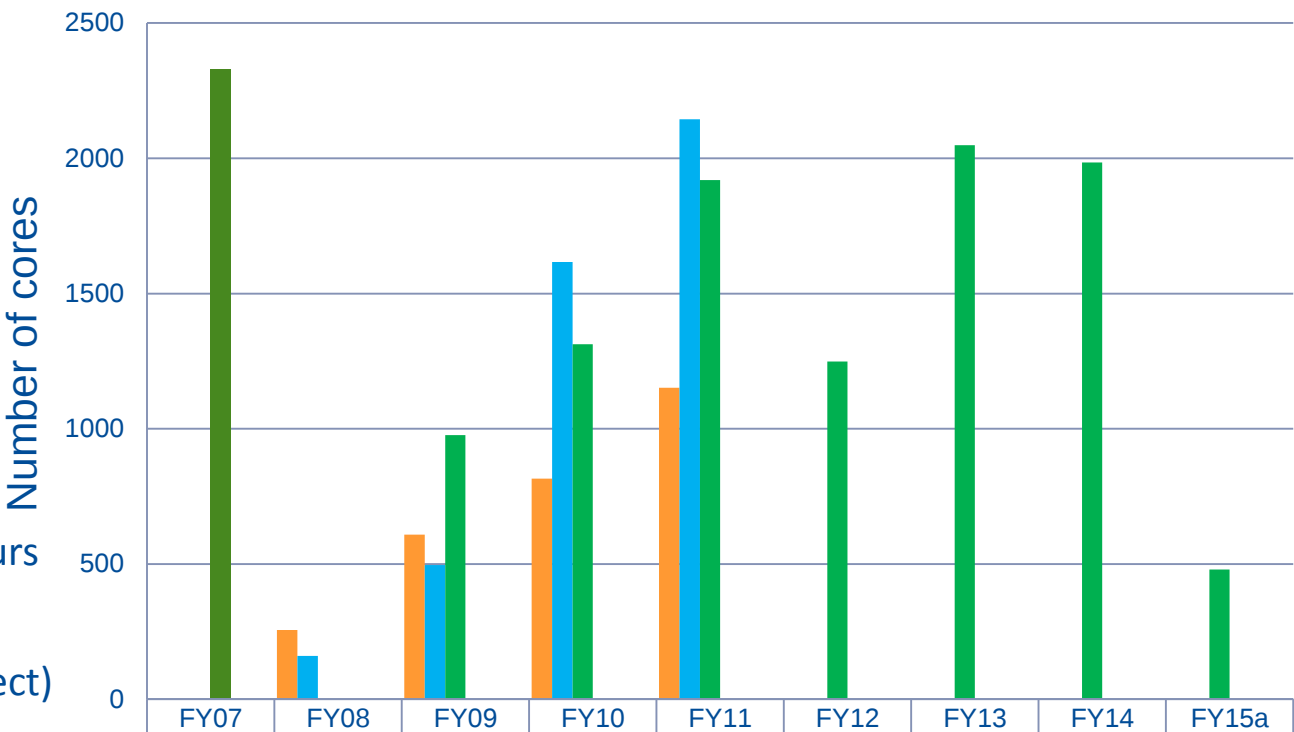
Replacement cost
~ \$2M (direct)



Processors: Current GRID Configurations

3/2015 status: Cores vs Year Purchased

Grid	Cores
CDF	2832
D0	4416
GP	9968
CW	2328

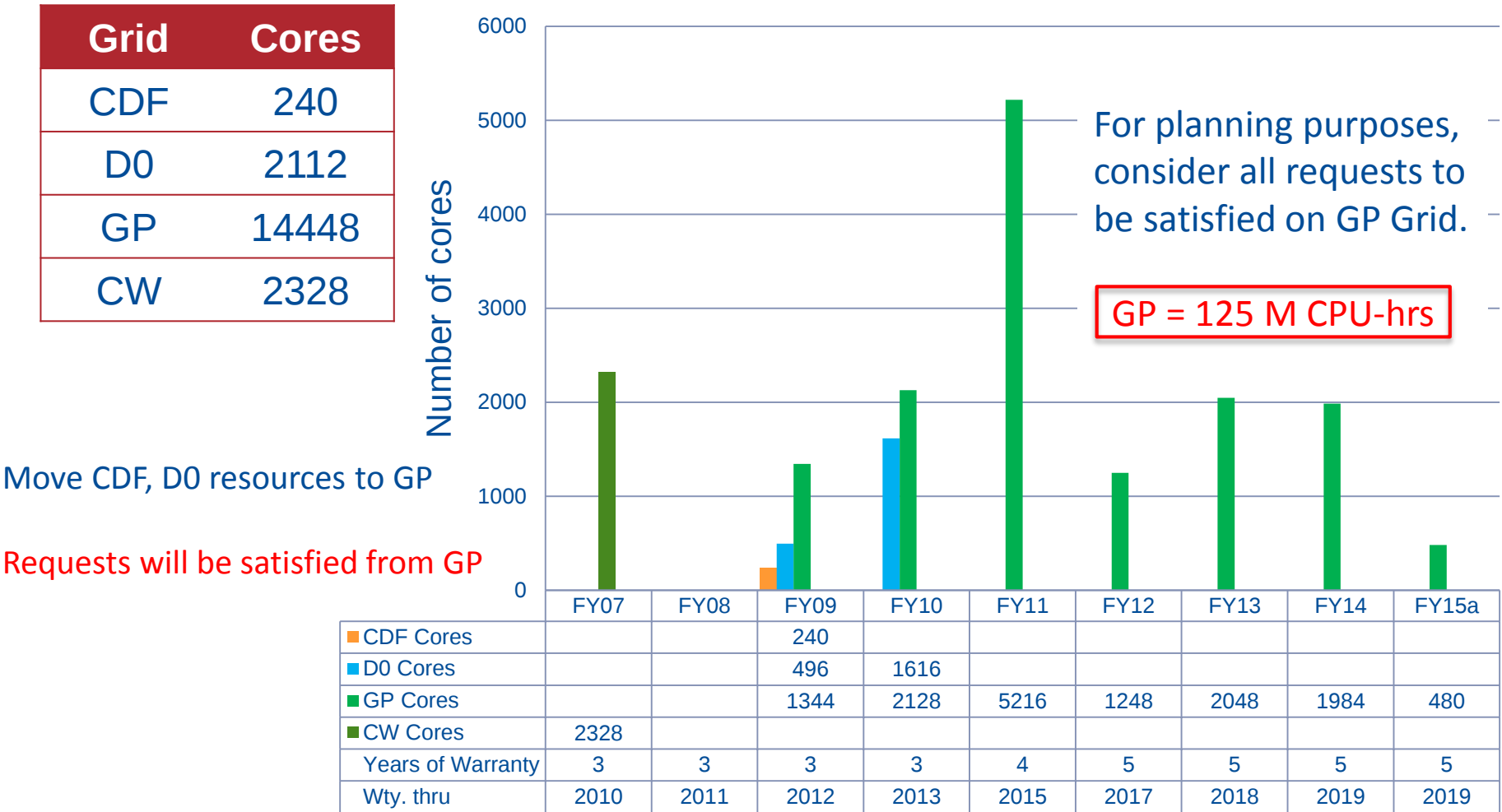


Equivalent to ~145M CPU-hours in one year

Replacement cost ~ \$2M (direct)

CDF Cores		256	608	816	1152				
D0 Cores		160	496	1616	2144				
GP Cores			976	1312	1920	1248	2048	1984	480
CW Cores	2328								
Years of Warranty	3	3	3	3	4	5	5	5	5
Wty. thru	2010	2011	2012	2013	2015	2017	2018	2019	2019

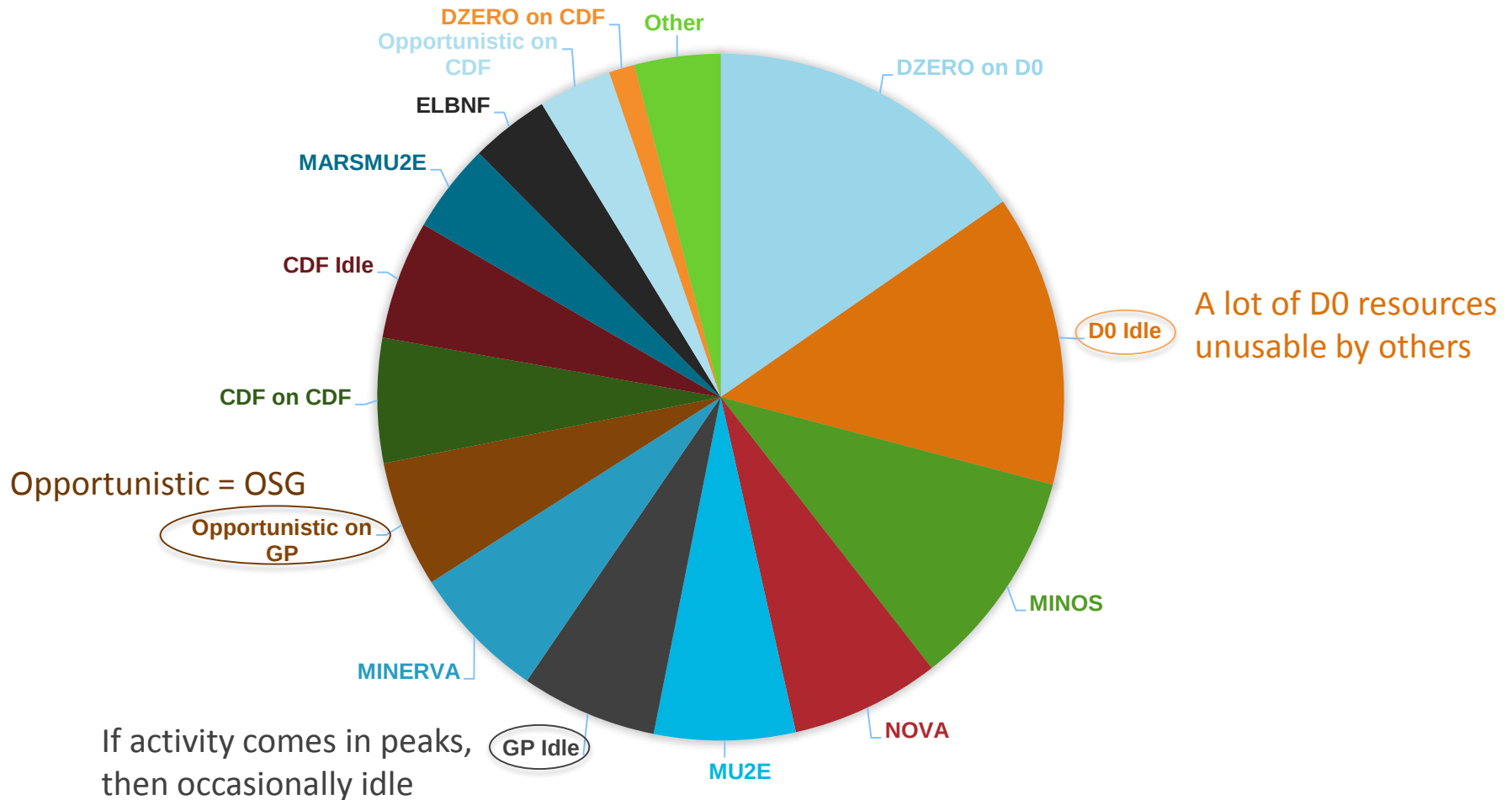
Processors: Plan for remainder FY15



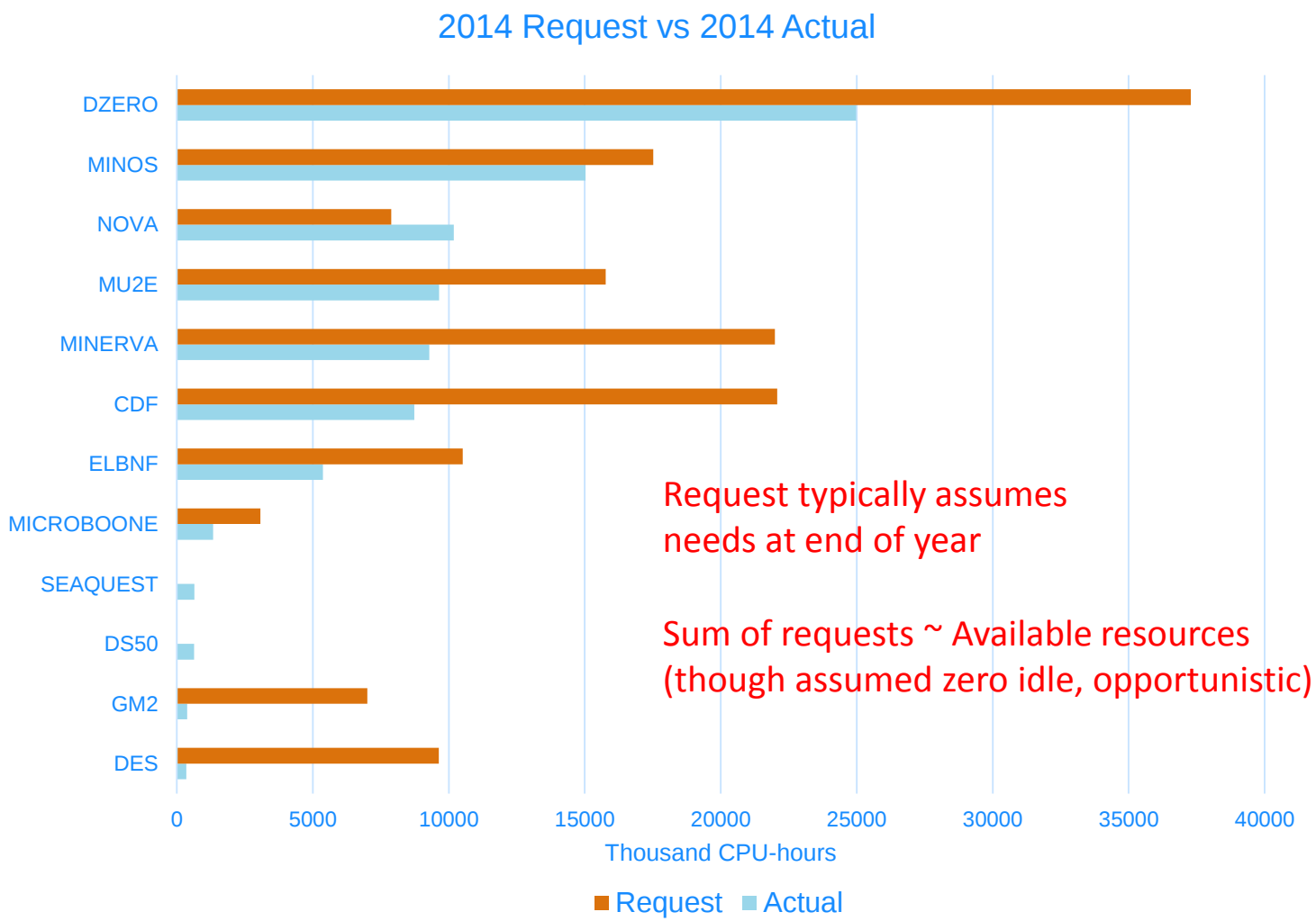
Processors: Past Usage (Mar 2014 – Feb 2015)

Total ~145M CPU-hours

2014 FERMIGRID USAGE



Processors: Last Year's Request vs Usage



Tape Storage

Values in TB

Warning: interpretation
of starting point inconsistent

Experiment	2014 SC-PMT Request thru end FY15	Current	FY15 Request (Increment)	FY15 Request (Cumulative)	FY16 Request (Increment)	FY16 Request (Cumulative)	FY17 Request (Increment)	FY17 Request (Cumulative)
CDMS		216	40	256	25	281	25	306
CHIPS		0	0	0	0	0	0	0
COUPP		11	0	11	30	41	0	41
DS50	400	639	50	689	150	839	150	989
DES	1,200	519	600	1,119	600	1,719	600	2,319
ELBNF	200	1	450	451	450	901	100	1,001
GM2		0	128	128	300	428	500	928
HOLOMETER		68	50	118	50	168	0	168
LARIAT		0	90	90	180	270	0	270
LQCD		2,553	0	2,553	0	2,553	0	2,553
LSST		0	0	0	0	0	0	0
MARSACCEL		0	0	0	0	0	0	0
MARSGM2		0	0	0	0	0	0	0
MARSLBNE		0	0	0	0	0	0	0
MARSMU2E		0	0	0	0	0	0	0
MICROBOONE	500	25	3,000	3,025	1,100	4,125	1,600	5,725
MINERVA	2,760	689	1,600	2,289	1,800	4,089	2,100	6,189
MINOS	590	470	90	560	90	650	50	700
MU2E	120	34	210	244	200	444	300	744
NOVA	2,500	2,265	1,500	3,765	2,500	6,265	0	6,265
NUMIX		0	20	20	40	60	60	120
PATRIOT		6	0	6	0	6	0	6
SEAQUEST		38	55	93	10	103	65	168
TOTAL	8,270	7,534	7,883	15,417	7,525	22,942	5,550	28,492