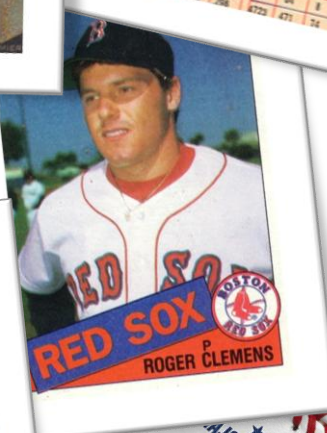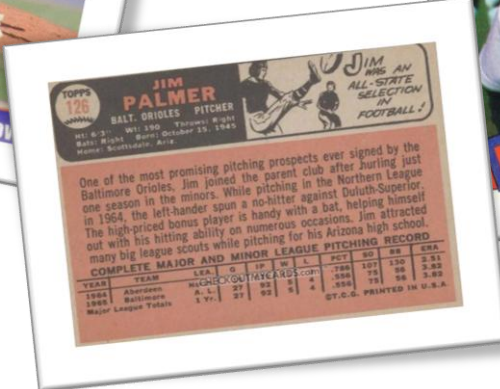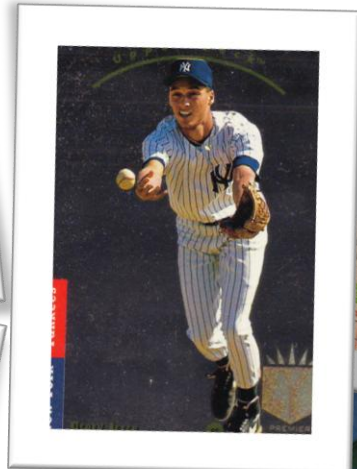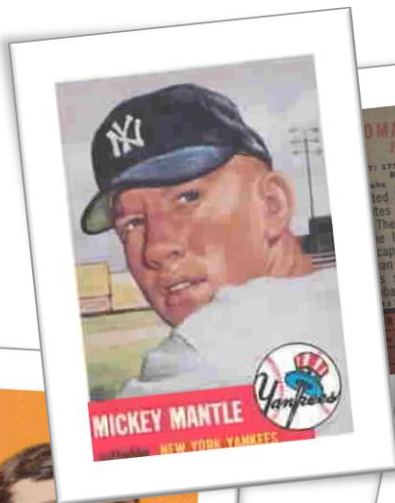# Fujifilm Presentation

Rob Sims
President & CEO
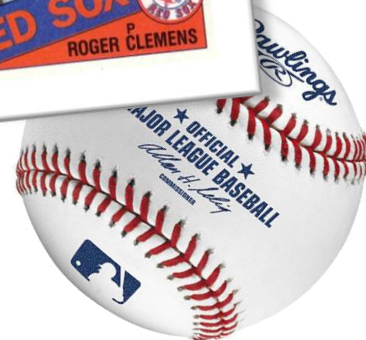
Homeruns

RBI'S

Batting average
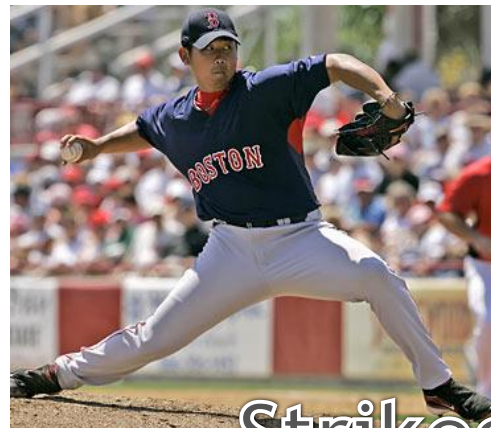
Stolen bases
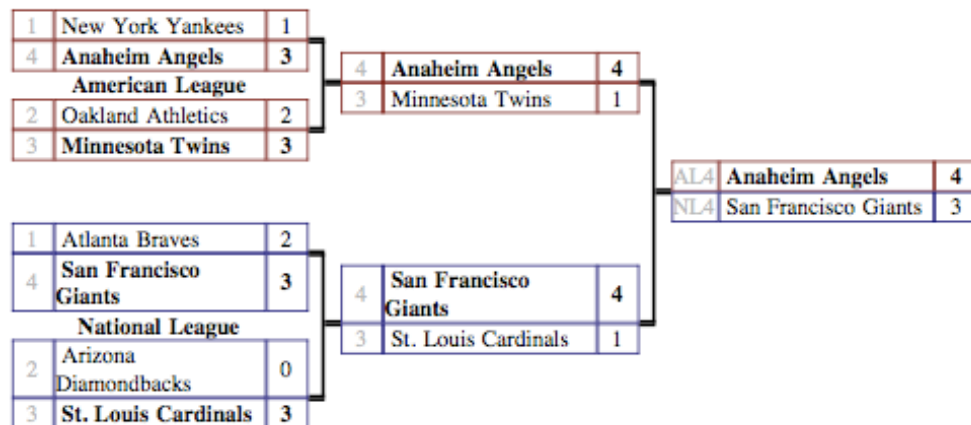
Strikeout

2002 MLB Playoffs

	American League		
1	New York Yankees	1	
4	**Anaheim Angels**	3	
2	Oakland Athletics	2	
3	**Minnesota Twins**	3	

4	**Anaheim Angels**	4
3	Minnesota Twins	1

	National League		
1	Atlanta Braves	2	
4	**San Francisco Giants**	3	
2	Arizona Diamondbacks	0	
3	**St. Louis Cardinals**	3	

4	**San Francisco Giants**	4
3	St. Louis Cardinals	1

AL4	**Anaheim Angels**	4
NL4	San Francisco Giants	3

Oakland Athletics Champion Run

2000	American League West Champion
2001	American League Wildcard
2002	American League West Champion
2003	American League West Champion
2006	American League West Champion

Team	Total payroll
New York Yankees	$125,928,583
Boston Red Sox	$108,366,060
Texas Rangers	$105,726,122
Arizona Diamondbacks	$102,819,999
Los Angeles Dodgers	$94,850,953
New York Mets	$94,633,593
Atlanta Braves	$93,470,367
Seattle Mariners	$80,282,668
Cleveland Indians	$78,909,449
San Francisco Giants	$78,299,835
Toronto Blue Jays	$76,864,333
Chicago Cubs	$75,690,833
St. Louis Cardinals	$74,660,875
Houston Astros	$63,448,417
Los Angeles Angels	$61,721,667
Baltimore Orioles	$60,493,487
Philadelphia Phillies	$57,954,999
Chicago White Sox	$57,052,833
Colorado Rockies	$56,851,043
Detroit Tigers	$55,048,000
Milwaukee Brewers	$50,287,833
Kansas City Royals	$47,257,000
Cincinnati Reds	$45,050,390
Pittsburgh Pirates	$42,323,599
Florida Marlins	$41,979,917
San Diego Padres	$41,425,000
Minnesota Twins	$40,225,000
Oakland Athletics	$40,004,167
Washington Nationals	$38,670,500
Tampa Bay Rays	$34,380,000

SABERMETRICS

Base Runs (BsR)
Draft round values (DRV)
Defense Independent Pitching Statistics (DIPS)
Equivalent average (EQA)
Fantasy Batter Value (FBV)
Late-inning pressure situations (LIPS)
On-base plus slugging (OPS)
PECOTA
Peripheral ERA (PERA)
Pythagorean expectation
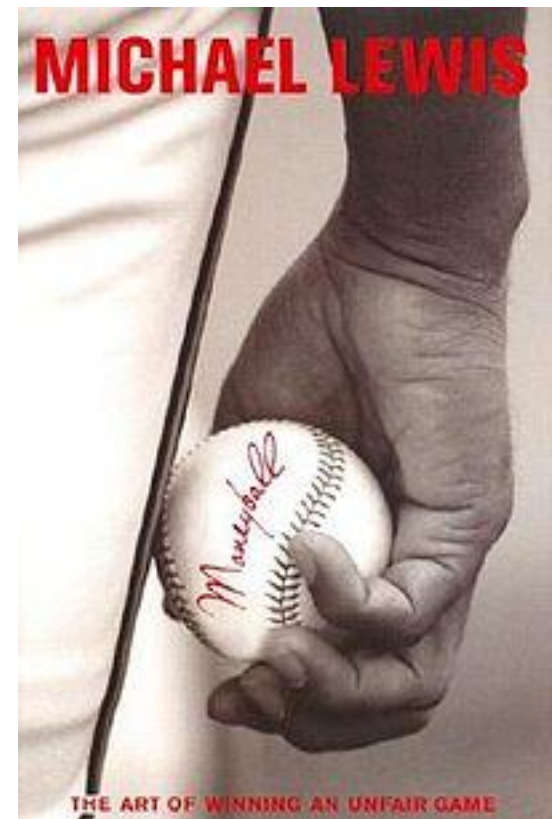Range Factor
Runs created
Secondary average
Similarity score
Speed Score
Super linear weights
Total player rating (aka PW/BFW)
Value over replacement player (VORP)

Moneyball: The Art of Winning an Unfair Game

Are you measuring . . .

the right things?

Are the things you're measuring . . .

the right things?

Crossroads - Proprietary

Current "Management" of Tape

Viscous cycle of Backup Window

Backup window to small

New Drive version Available ?

No → Buy more existing version drives

Yes → Buy new drives

Backup window doesn't improve much
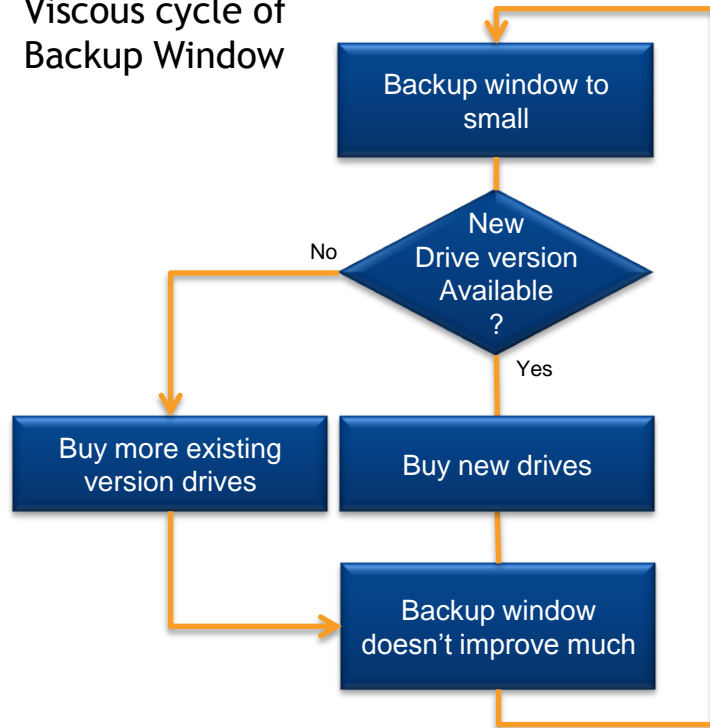
Viscous cycle of Backup Failure
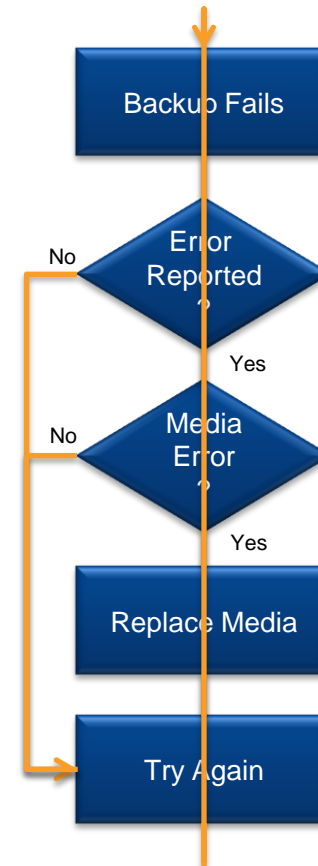
Backup Fails

Error Reported ?

No

Yes

Media Error ?

No

Yes

Replace Media

Try Again

Conclusion…

- I get backup failures, restore failures – the system says media error, or some other helpful statement
  - I replace tapes
  - I tinker with things
  - I try again, things are ok for awhile and then it starts all over again
- My backup window is no longer being met
  - I buy the latest technology (2x faster)
  - Or I buy more of the existing drives (more is better right?)
  - I get marginal improvement or nothing at all
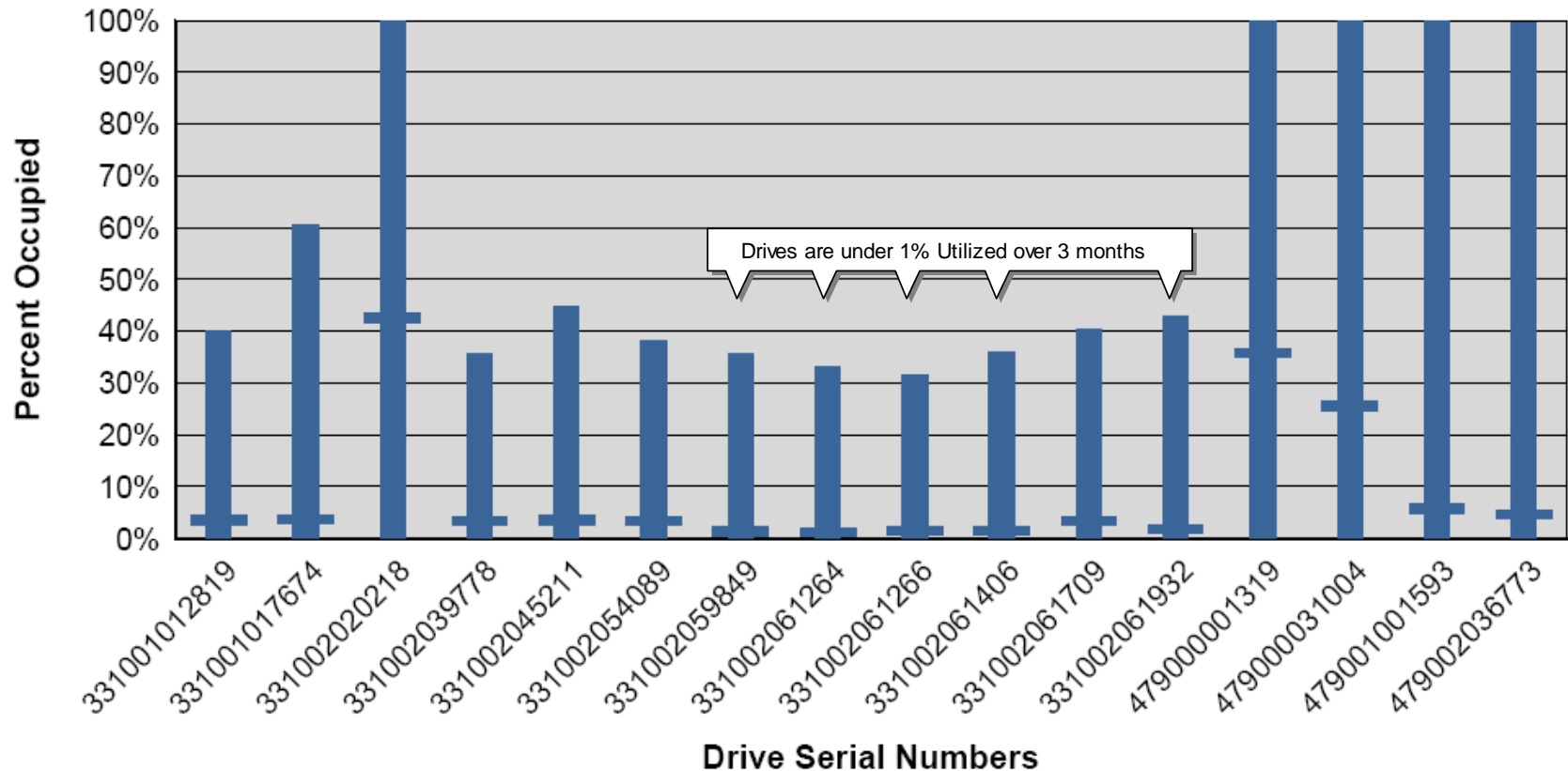
TAPE SUCKS!

Looking under the covers

Backup window no longer being met

▸ Before buying anything new or upgrading…
- How are the current drives being utilized?

Actual Drive Utilization

Max/Min/Average Percent Drive In Use Per Day — Tape Loaded into Drive

Hour by hour drive utilization

0% Utilization: ☐ 1 - 25% Utilization: ☐ 26 - 50% Utilization: ☐ 51 - 75% Utilization: ☐ 76 - 100% Utilization: ■

Drive	Avg. Util./Day
0110016763	16.26
0110103033	3.62
0110260164	0.00
0110262256	10.54
0110314533	20.62
0110321440	20.59
0110361813	14.01
0110360127A	16.55
0110026836	1.54
0110587632	21.81
0110069060	13.96
0110060716	9.26
0110179141	17.50
0110212545	14.04
0110214840	12.80
0110341846	2.31
0110360265	20.84
0110807404	1.31
0110807769	0.83
0110812064	16.87
0110812343	19.75
0110813310	12.23
0110813346	10.31
0110813461	11.93
0111801760	1.75
0111801861	26.34

Days and weeks go by with little or no use of certain drives

Looking under the covers

Backup window no longer being met

▶ Before buying anything new or upgrading…
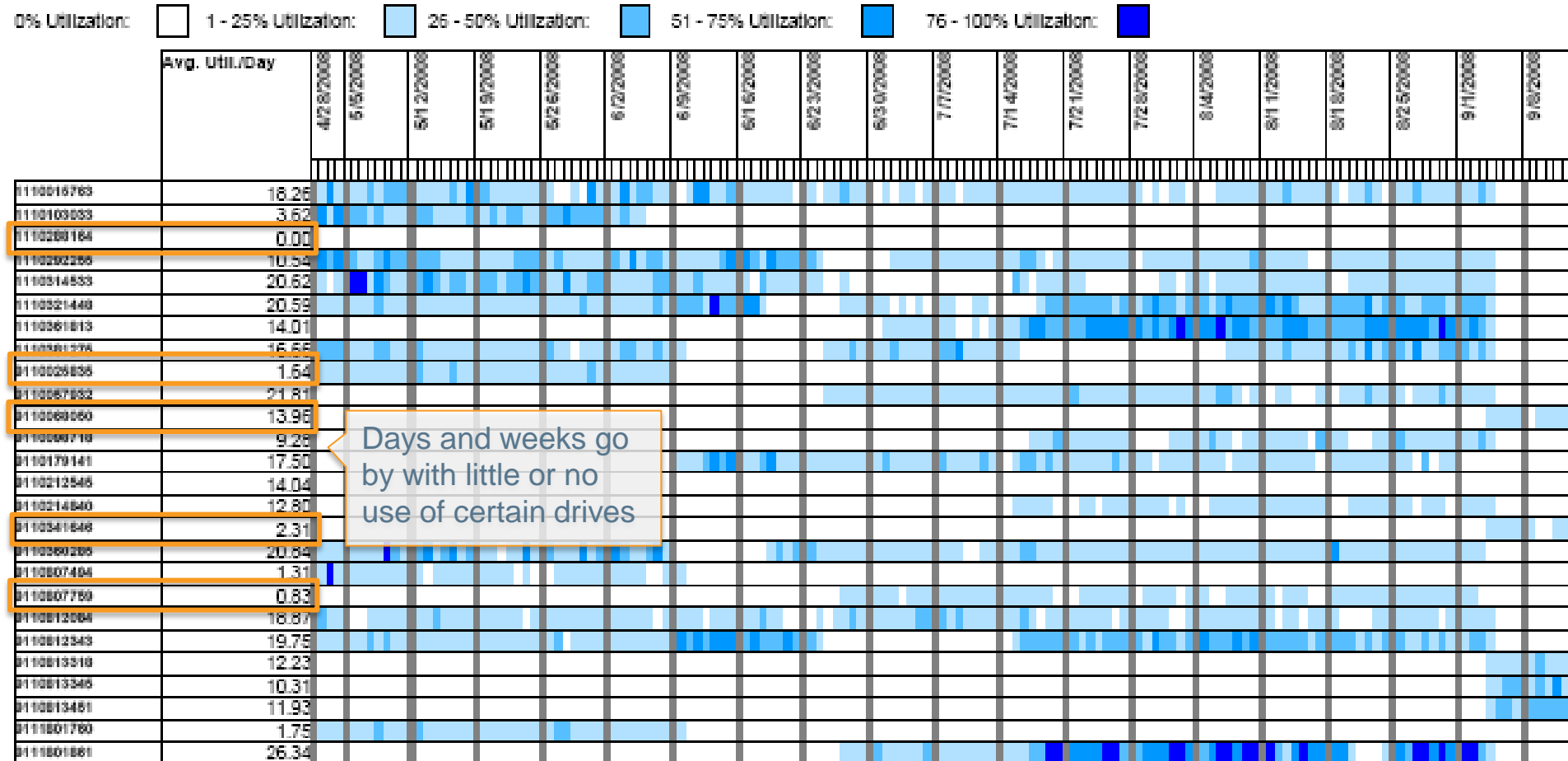
  − How are the current drives being utilized?

    • If drives are not used consistently (or at all)

      − Rebalance drive usage with backup application or system using drives

      − Look to change process such that more efficient transfers of data go to tape system

    • Buying more drives would probably reduce efficiency rather than improve it

Looking under the covers
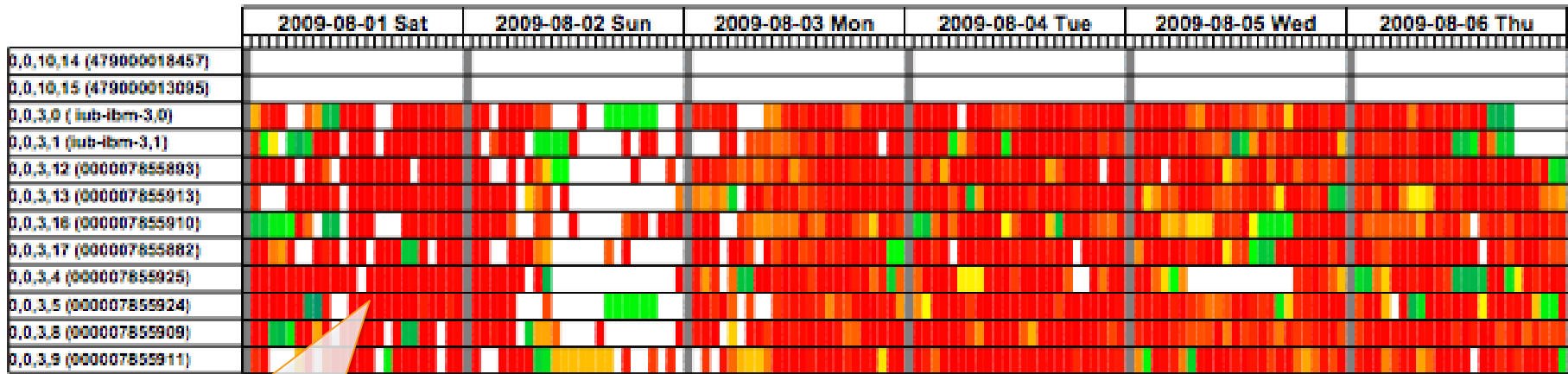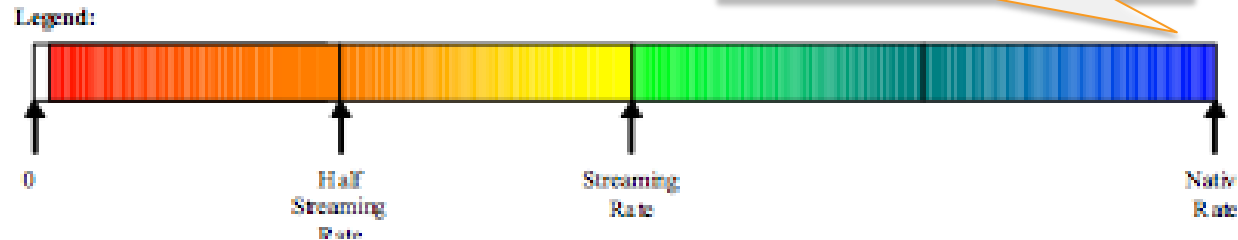
Backup window no longer being met

▶ Before buying anything new or upgrading…
  – How are the current drives being utilized?
  – How are the current drives performing?

Actual I/O performance of drives

Backup window no longer being met

▸ Before buying anything new or upgrading…

– How are the current drives being utilized?

– How are the current drives performing?

- Improving performance to streaming rate would give 400% improvement to window – native rates would provide 800% to 1000% improvement!

- Sometimes less is more – Systems writing data might not be able to stream data efficiently if too many drives are in use at the same time

- Buying newer technology is like buying a Ferrari when stuck on the highway – sure its faster, but your still going nowhere!

Relook at Conclusion…

▸ My backup window is no longer being met

- I buy the latest technology (2x faster)

  - If your system can't send data fast enough to existing technology than new drives won't give any improvement

  - Problem lies with systems and/or network configuration – not the tape system

- Or I buy more of the existing drives (more is better right?)

  - If current utilization is poor, buying more doesn't help much at all (in fact it might cause more problems)

  - Problem lies with systems and processes writing data not effectively using tape drives

- I get marginal improvement or nothing at all

  - Because focus was on the wrong statistics or no statistics at all

Looking under the covers

Endless cycle of backup and/or restore failures

▸ Before throwing out the tapes or just trying over and over again

  – Is the environment optimized for tape?

"The definition of insanity is doing the same thing over and over again and expecting different results"
- Albert Einstein

Performance Revisited

Daily Drive Performance
8/1/2009 to 8/18/2009

Below streaming rate causes drive to "shoe shine"

Most drives are constantly in "shoe-shining" state

What is "shoe-shining" and why is it bad?

Constant stopping and starting of tape (back hitch)

▶ Excessive wear of media
- Tape media ideally moves forward and rewinds out of data path
- LTO heads touch media so back hitch scrapes a rock over your data

▶ Previous written data isn't checked
- Read after write only checks the current data
- Back hitch goes over previous data, so any damage will go undetected

Looking under the covers

Endless cycle of backup and/or restore failures

▶ Before throwing out the tapes or just trying over and over again
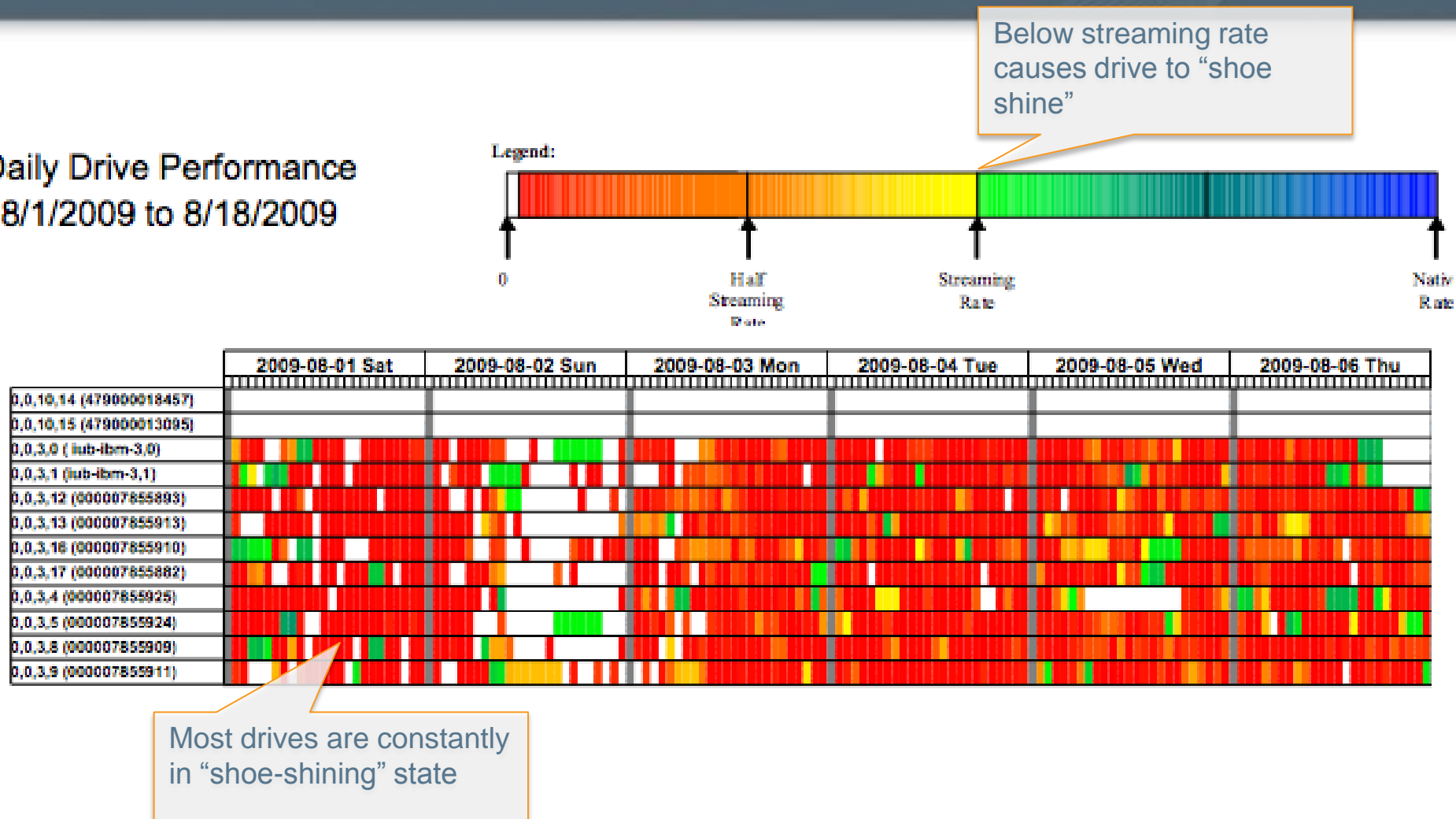
   – Is the environment optimized for tape?

   – Is the problem really the tape media or could it be the tape drive?
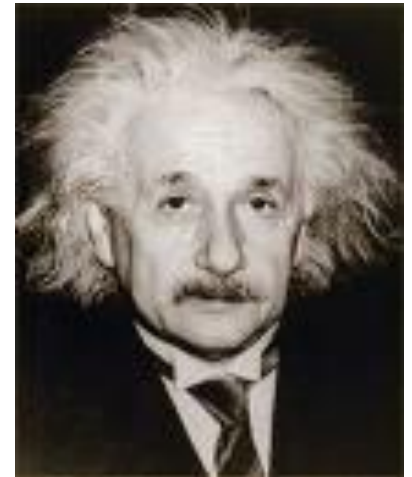
"The definition of insanity is doing the same thing over and over again and expecting different results"
- Albert Einstein

Tape Errors per Mbyte plot

Telecom Tape Err/MByte

460 Suspect
Tapes out of 1498

Normally you would throw these tapes away – the statistics show they have high error rates

Err/MB Limit
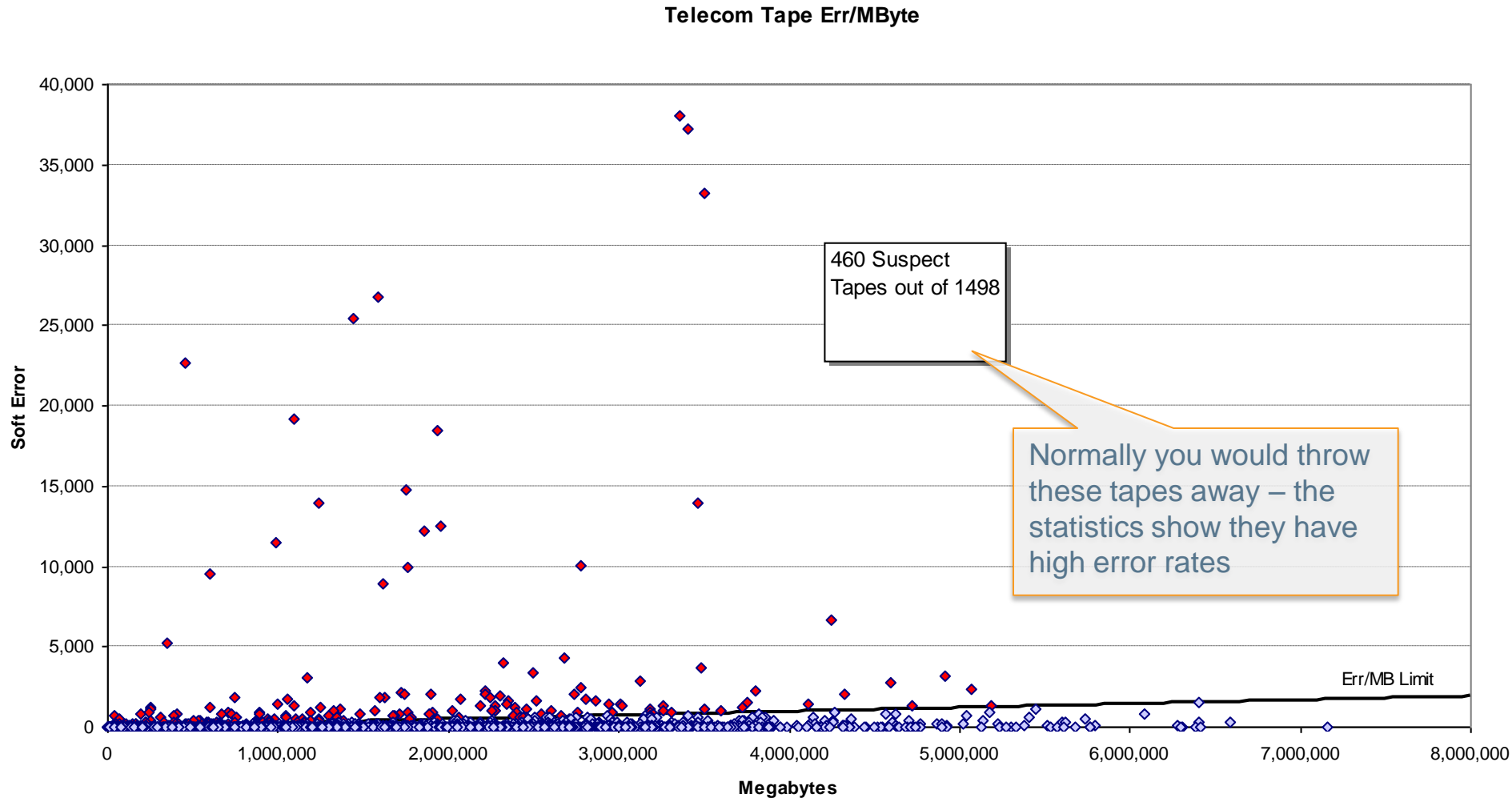
Tape Barcode	9110XXXXXX Err/MiB	Affect	9110XXXXXX Err/MiB	Affect	9110XXXXXX Err/MiB	Affect	1110XXXXXX Err/MiB	Affect	9110XXXXXX Err/MiB	Affect	9111XXXXXX Err/MiB	Affect	9110XXXXXX Err/MiB	Affect	1110XXXXXX Err/MiB	Affect	1110XXXXXX Err/MiB	Affect	9110XXXXXX Err/MiB	Affect
IP3480L2	99.33%	●	0.00%	■											0.01%	■			0.00%	■
IP4415L2	95.65%	●																		
IP4446L2	87.30%	●			0.00%	■	0.00%	■							0.00%	■			0.00%	■
IP5487L2	63.09%	●															0.00%	■	0.01%	■
IP5110L2	61.39%	●																		
IP3380L2	56.59%	●																	0.00%	■
IP4545L2	27.96%	●	0.00%	■													0.00%	■		
IP6735L2	23.95%	●																		
IP4384L2	15.30%	●	0.00%	■											0.00%	■				
IP5266L2	11.13%	●																		
IP4628L2	10.68%	●	0.00%	■	0.01%	■	0.00%	■							0.00%	■	0.00%	■	0.00%	■
IP5109L2	9.08%	▼	0.00%	■																
IP3638L2	8.27%	●					0.00%	■												
IP4268L2	7.66%	●	0.00%				0.00%	■												
IP4257L2	6.44%	●					0.00%	■												
IP3619L2	6.19%	●					0.00%	■												
IP4624L2	6.15%	●	0.00%	■																
IP5064L2	6.07%	●																		
IP5397L2	5.88%	●									0.00%	■			0.14%	■			0.02%	■
IP5403L2	5.46%	●	0.00%												0.05%	■			0.00%	■
IP4019L2	3.41%	●	0.00%										0.04%	■	0.00%	■	0.00%	■	0.00%	■
IP5237L2	3.40%	●																		
IP5352L2	3.27%	●											0.00%	■						
IP5174L2	2.56%	●																		
IP4567L2	2.20%	●	0.00%												0.01%	■				
IP5150L2	1.55%	●	0.00%																	
IP3696L2	1.10%	●													0.00%	■			0.00%	■
IP4080L2	1.09%	●																		
IP5259L2	0.97%	●			0.00%	■									0.00%	■	0.05%	■	0.00%	■
IP5172L2	0.93%	●					0.00%	■			0.00%	■								
IP5095L2	0.85%	●			0.00%	■														
IP4156L2	0.75%	●																		
IP5349L2	0.74%	●					0.00%	■												
IP5353L2	0.72%	●	0.23%	■			0.00%	■	1.13%	■										
IP4393L2	0.60%	●	0.00%	■	0.01%	■	0.00%	■	0.37%	■			0.01%	■						
IP5090L2	0.58%	●	0.00%	■			0.00%	■			0.00%	■								
IP4230L2	0.57%	●											0.00%	■						
IP4511L2	0.56%	●					0.00%	■			0.00%	■								
IP4332L2	0.56%	●			0.00%	■														
IP3543L2	0.55%	●																		
IP3363L2	0.53%	●	0.00%	■																
IP4319L2	0.51%	●	0.00%	■	0.00%	■	0.00%	■												
IP5161L2	0.49%	●	0.00%	■	0.04%	■	0.00%	■					0.11%	■						
IP4395L2	0.47%	●					0.00%	■												
IP4000L2	0.47%	●	0.00%	■			0.00%	■							0.11%	■				
IP5408L2	0.45%	●															0.00%	■	0.00%	■
IP5484L2	0.44%	●	0.00%	■											0.00%	■				
IP5121L2	0.44%	●									0.00%	■								

5 Drives account for over 90% of the suspect tapes

The problem is systemic is that these drives keep creating "defective" media which gets replaced, the problem goes away for awhile and then comes right back

Legend:
- **?** Tape was only in this drive
- **●** Tape Error rate is below limit if drive is removed from population
- **●** Tape Error rate is below limit if drive is removed from population (errors only occurred on read)
- **▼** Tape Error rate is reduced by more than 50% if drive is removed from population

Relook at Conclusion…

▸ I get backup failures, restore failures – the system says media error, or some other helpful statement

▸ My backup window is no longer being met

> Root cause of problem in many cases is never addressed and therefore the problem always returns

TAPE SUCKS!

Tape Monitoring of HPSS at LBNL/NERSC

Fujifilm Tape Summit

Jason Hick

Storage Systems Group Lead

jhick@lbl.gov

2/24/2010

NERSC is the Production Facility for DOE SC

◆ **NERSC serves a large population**

Approximately 3000 users,

400 projects, 500 code instances
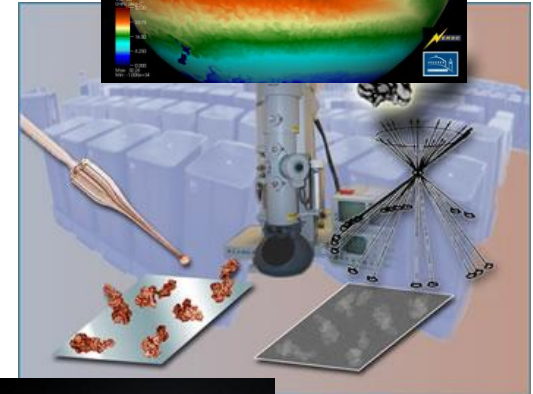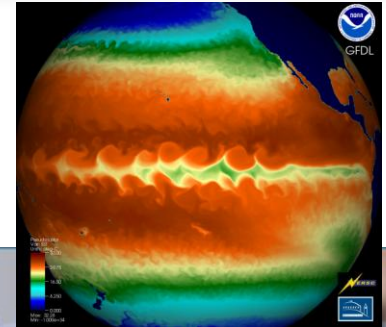
◆ **Focus on "unique" resources**

▸ **High end computing systems**

▸ **High end storage systems**

– **Large shared file system**

– **Tape archive**

▸ **Interface to high speed networking**

– **ESNEt soon to be 100 Gb/s**

◆ **In 2003, NERSC changed from being a data source to being a data sink**

2009 Allocations


ASCR 7%
BER 19%
BES 29%
FES 18%
HEP 17%
NP 10%
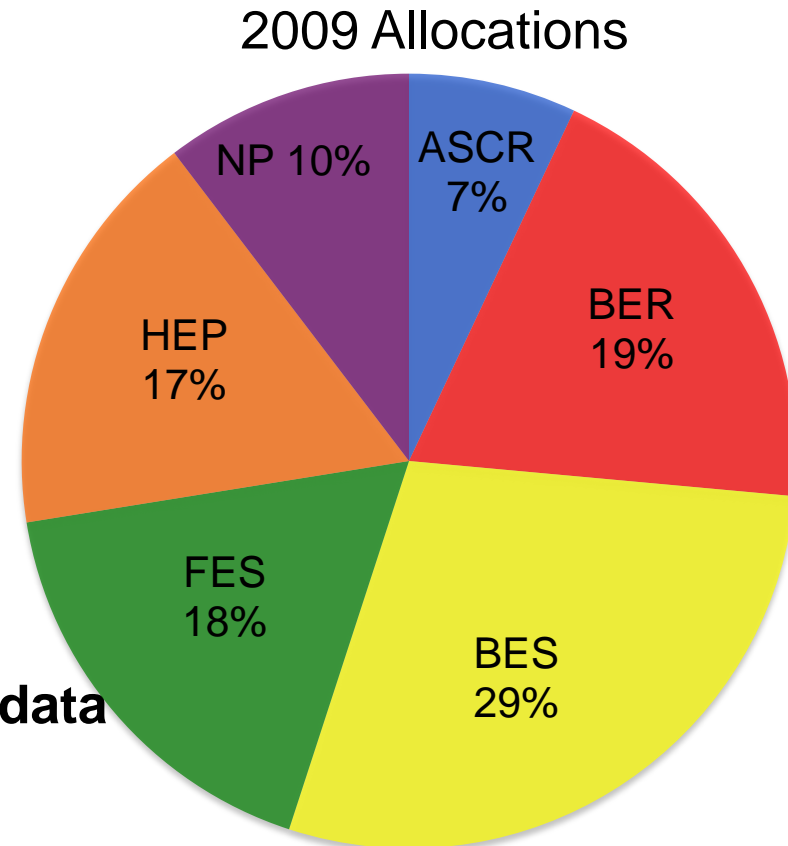
Data Needs Continue to Grow

◆ **Scientific data sets are growing exponentially**
- **Simulation systems and some experimental and observational devices grow in capability with Moore's Law**

◆ **Petabyte (PB) data sets will soon be common:**
▸ *Climate modeling:* **estimates of the next IPCC data is in 10s of petabytes**
▸ *Genome:* **JGI alone will have .5 petabyte of data this year and double each year**
▸ *Particle physics***: LHC are projected to produce 16 petabytes of data per year**
▸ *Astrophysics***: JDEM alone will produce .7 petabytes/year**

◆ **We will soon have more data than we can effectively store and analyze**

Tape Archives: Green Storage

Cumulative Storage by Month and System

Scientific data at NERSC increases by 1.7X per year
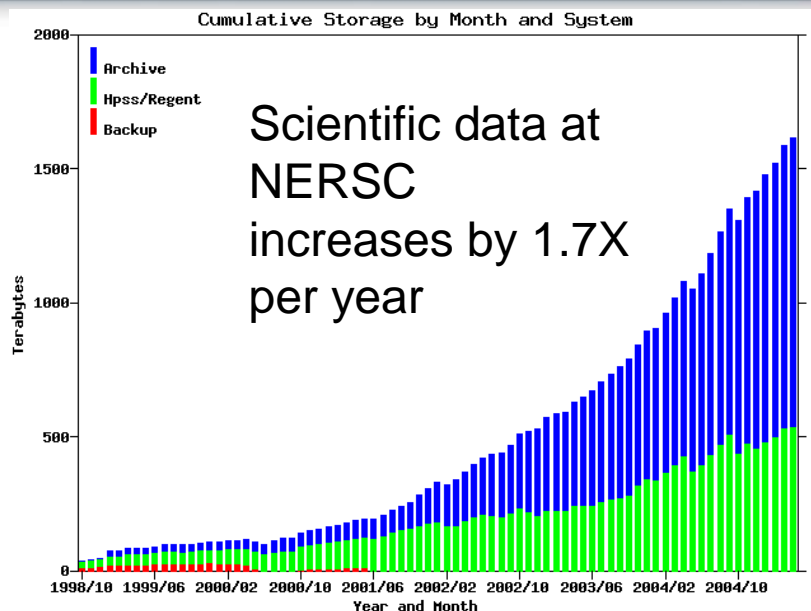
- ◆ **Tape archives are important to efficient science**
  - ▸ **2-3 orders of magnitude less power than disk**
  - ▸ **Requires specialized staff and major capital investment**
  - ▸ **NERSC participates in development (HPSS consortium)**

Tape Hardware & Software

- ◆ **6 x 9310 Powderhorns (read only)**
  - ▸ **34 x 9840A**
  - ▸ **32 x 9940B**
- ◆ **4 x SL8500 (new data)**
  - ▸ **84 x T10KB**
  - ▸ **28 x 9840D**
- ◆ **Some Statistics**
  - ▸ **20-40 TB I/O per day**
  - ▸ **1.7 PB growth in 2009 (archive)**
  - ▸ **0.5 PB growth in 2009 (backups)**
- ◆ **Tape related software**
  - ▸ **HPSS 6.2**
  - ▸ **ACSLS 7.3**
  - ▸ **Crossroads RVA/AV for tape subsystem monitoring**
  - ▸ **Software Delivery Platform (SDP) by Sun/STK for tape subsystem monitoring and remote resolution**
  - ▸ **Locally developed tape monitoring**

Tape Archive Requirements

- ◆ **Reliability**
  - ▶ **Maintain user data beyond changes to computing environment, until the user deletes it**
    - – **Identify and protect against tape failures**
    - – **How often is hardware swapped out, and when?**
  - ▶ **Identify root cause of read/write errors to improve/eliminate**
    - – **Is it the tape cartridge or the drive… or the combination due to variant drives?**
- ◆ **Performance & Capacity**
  - ▶ **All files to disk then migrated to tape**
    - – **Match speed between disk and tape**
    - – **Numbers of tape drives by type needed for peak ingest**
    - – **Are the drives in the right location to optimize tape mount time**
- ◆ **Availability**
  - ▶ **Strive for 99 or 99.9's availability**
    - – **Root cause analysis of outages (software, hardware, device, …)**
  - ▶ **Minimize system downtimes**
    - – **Shield users from tape subsystem**

Our Quest in Running a Production Tape Archive

- ◆ **Identify and protect against tape failures**
  - ▸ **Sun SDP was supposed to help with identifying problem**
  - ▸ **Some local solutions have helped (fault symptom code analysis, database of error reports)**
- ◆ **How often is hardware swapped out, and when? Do these affect error rate (i.e. if we swap out an error causing drive)?**
  - ▸ **Manual record keeping, helped on a few occasions, but required months to enter into a database and analyze for trends**
- ◆ **Is it the tape cartridge or the drive… or the combination due to variant drives?**
  - ▸ **A local solution (fault symptom code analysis) was most useful, but still fell short**
- ◆ **Match speed between disk and tape. Are we optimally configuring tape and disk resources?**
  - ▸ **Tape drive bandwidth determined periodically through analysis of logs and statistics**
- ◆ **How many tape drives by type are needed for peak ingest? (concurrent user reads/stages, migration from disk, data movement to new technology)**
  - ▸ **Analyze tape library manager mount logs**
- ◆ **Are the drives in the right location to optimize tape mount time?**
  - ▸ **Difficult to determine, but could analyze tape library manager mount logs**
- ◆ **Root cause analysis of outages (software, hardware, device, …)?**
  - ▸ **Manual process that took 9 months, results were mixed**

Lessons Learned

◆ **After two years of several FTEs worth of work, modest results**

◆ **Custom scripts and programs drawing on data from multiple sources and locations to maintain**

◆ **Analysis led us to make several changes in system configuration, improving user experience**

◆ **But there were many things we didn't have time for or a way to determine**

▸ **Why is migration from our disk to tape so slow?**

▸ **Where are the problematic drives (tape works in one drive but not another)?**

▸ **Moving data from bad tape to good sometimes takes three or more tries before succeeding, is it the tape or the drive?**

Tape Environment Analysis

- **Provided broad set of service offerings along with system**
  - save on precious staff time and effort
- **Archive verify service to validate readability of the entire archive**
  - **analyzing approximately 40,000 tapes**
  - **five different generations of drives**
  - **media up to ten years old**
- **Quarterly reports to provide detailed analysis of operational performance**
  - **drives being swapped out (actual service life)**
  - **statistical determination of whether the tape or drive is problematic**
  - **tape drive bandwidth per transfer**
  - **numbers of tape drives needed for peak ingest/load**
  - **passthrough and long mount activity identified for drive relocation**
  - **preemptive media failure analysis to prioritize data movement to new media**
- **Archive requirements and usage of tape is now gaining interest in industry**
  - systems and services are being tailored to work well for archive systems
- **Applying the results will improve user experience with tape, improve interaction with vendor service and support, and reduce tape problems**

Quarterly Report

Repair/Replace

- T10000B: 1,2,1,0 (572000400375)
- T10000B: 1,3,1,4 (572000400508)
- T10000B: Currently Removed (572004000429)

Watch List

- T9840D: 1,8,1,1 (5700GU004603)
- T9840D: 1,4,1,3 (5700GU003030)
- T9840D: 1,4,1,6 (5700GU003020)
- T10000B: 1,3,1,6 (572004000693)
- T10000B: 1,2,1,1 (572004000535)
- T10000B: 1,6,1,5 (572004000507)

Error Rate

Percentage of soft errors caused by the drives on the watch and repair lists:

90%

- **Identified error producing drives**
  - **3 T10KB drives that need replacement**
  - **Addresses the most severe and important problem to us, and something**
  - **We have months of effort devoted to figuring out the same problem**
  - **Replacing should reduce soft/hard errors in next report**

Quarterly Report

- **Identified that 9840Ds weren't being used as well as T10KBs**
    - **We identified this just prior to the report with tape type import/slot statistics that we analyze**
    - **We adjusted the size of data going to 9840D and now strike a better balance. The next report should confirm.**

Chart 24: T9840D Simultaneous Drives In Use
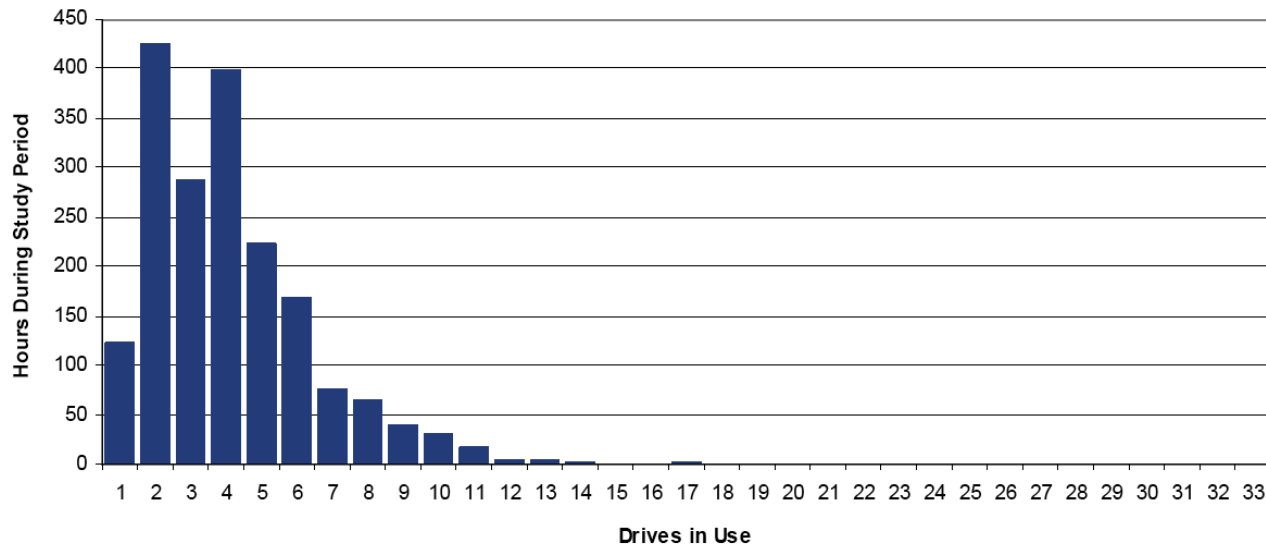
Questions?

Jason Hick

jhick@lbl.gov

2/24/2010