I/O Latency - Eliminating the Enemy of e-Transaction Performance

Mark W. Hall

Director of Product Management

Solid Data Systems, Inc.
2945 Oakmead Village Court
Santa Clara, CA 95051

Voice: 408.845.5747
Fax: 408.727.5496
mhall@soliddata.com

Overview

- Latency defined
- Explosive growth of e-transactions
- Anatomy of an e-transaction
- Understanding rotating disk latency
- Performance Effects of Solid-state Storage on e-transactions
- Emerging storage virtualization & QoSS
- Q&A

La-ten-cy [láyt'nsee] *noun*

- In general, the period of time that one component in a system is spinning its wheels waiting for another component. Latency, therefore, is wasted time.

Source: Webopedia (webopedia.internet.com)

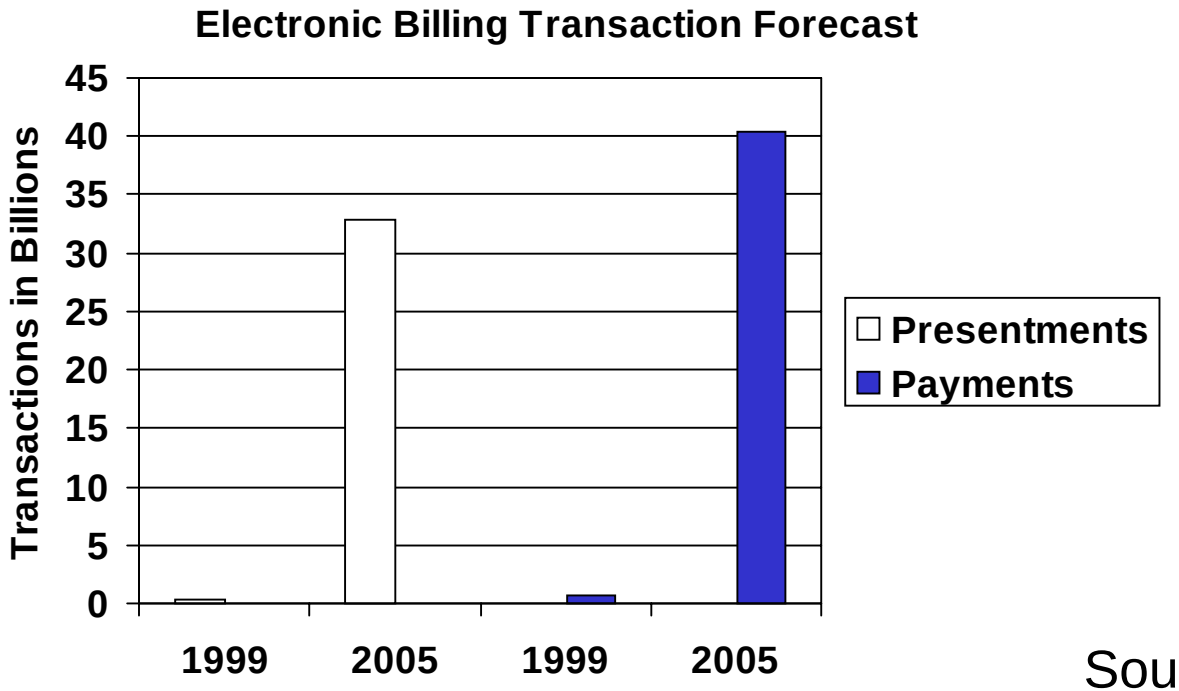
Explosive Growth of e-Transactions

- A study of large companies found that 73% are involved in online integration through the use of both commercial messaging technologies and in-house coding.

Source: Forrester Research

Explosive Growth of e-Transactions

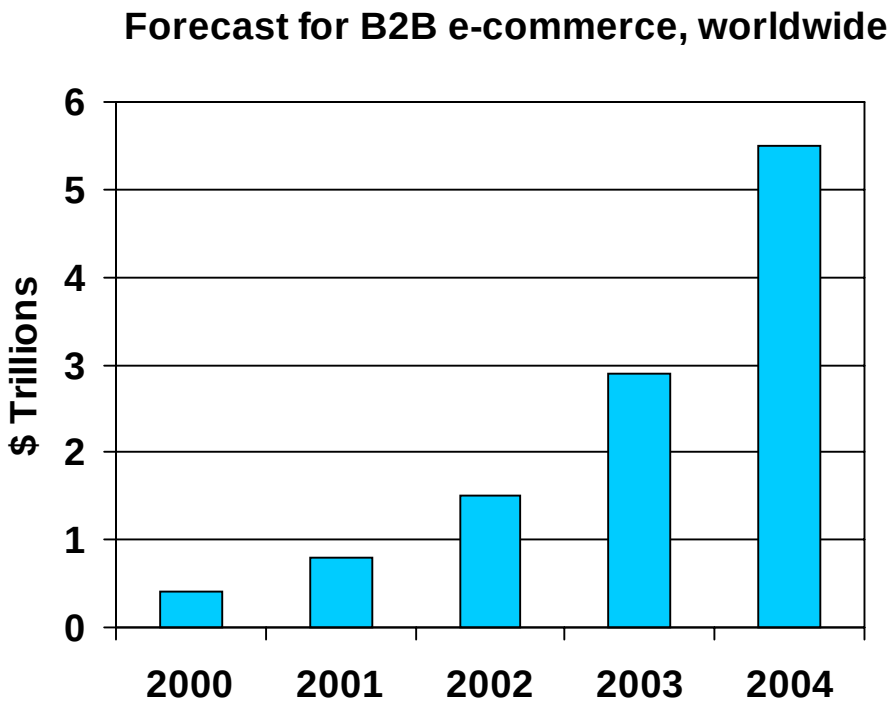
- The number of bills presented electronically will increase more than a hundredfold and the number of payments will reach 40.4 billion by 2005.



Source: Ovum Ltd.

Explosive Growth of e-Transactions

- Efforts to reduce costs and streamline supply chains will fuel a 92% annual growth rate for e-commerce and push the worldwide B2B market to $5.5 trillion in 2004.



Source: InfoTech Trends

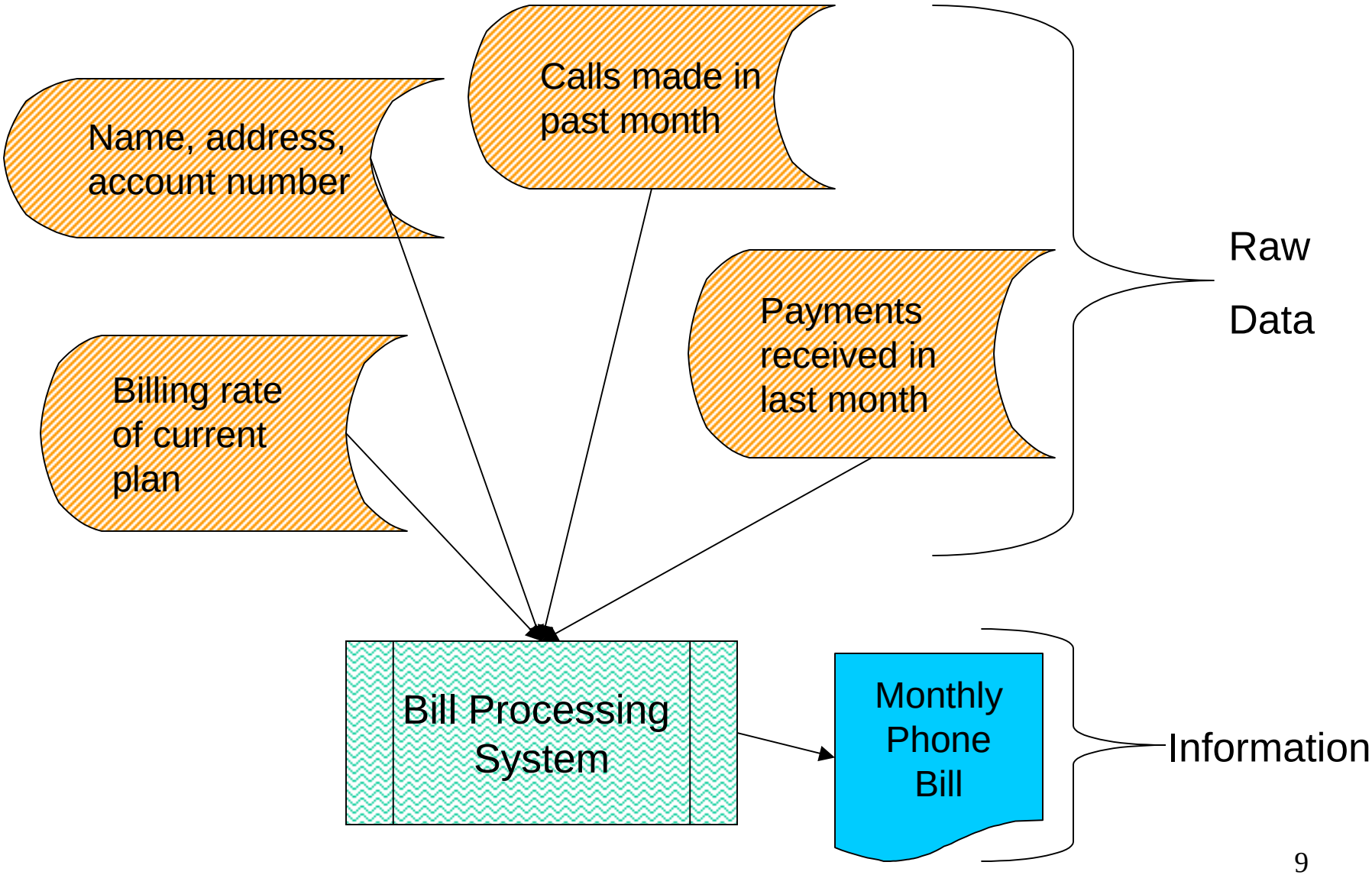
What is an e-Transaction?

- It is the process of transforming raw data into useful information, by applying business rules, through computerized methods.

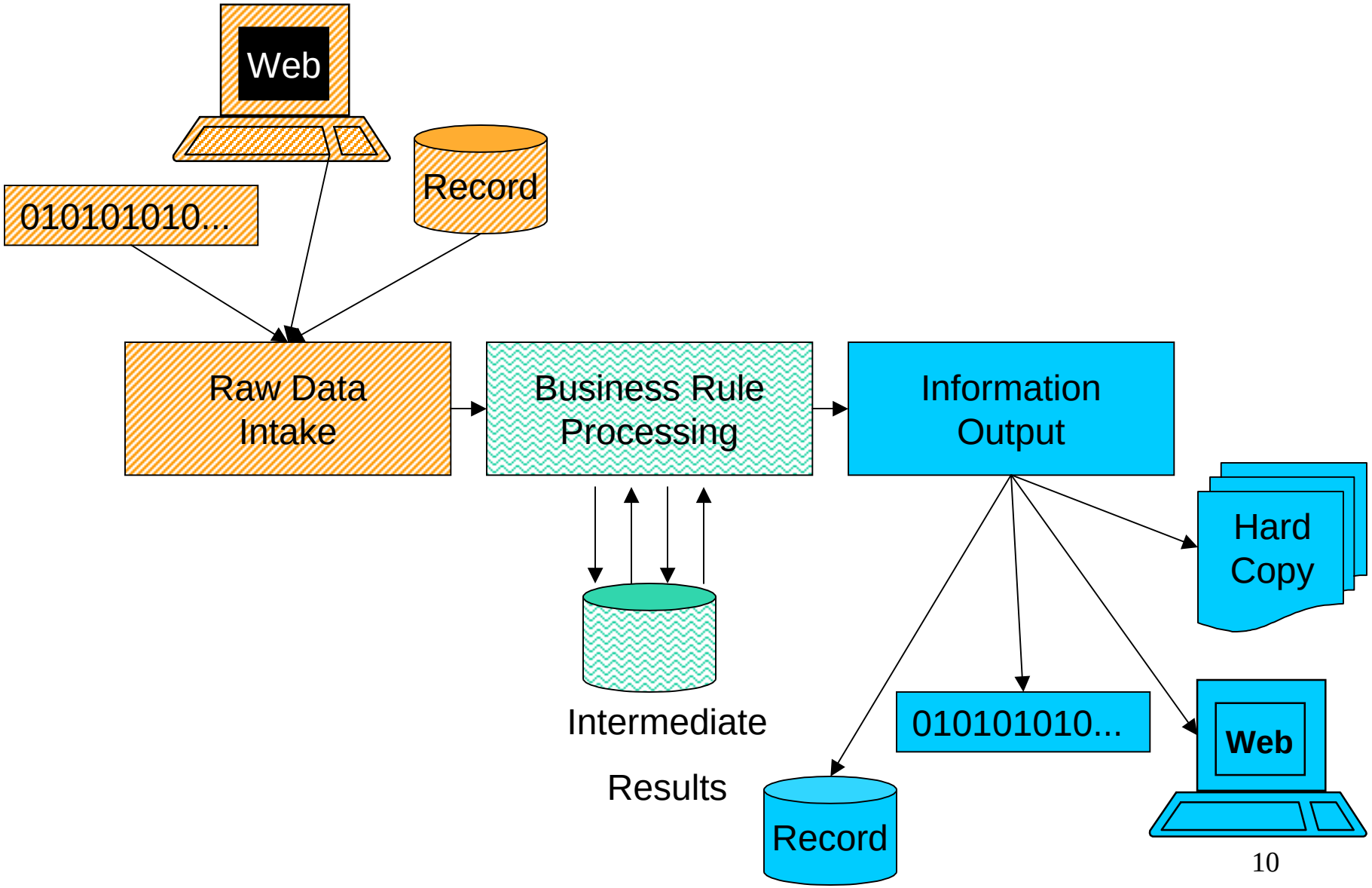
Examples of e-Transactions

- online trading / offline trading
- e-mail
- wireless messaging / prepaid cellular
- credit card processing / fraud detection e-commerce Web sites - B2C
- m-commerce (wireless devices)
- industry specific trading exchanges - B2B
- CRM / ERP
- batch processing (payroll, billing, etc.)

Anatomy of an e-Transaction



Anatomy of an e-Transaction

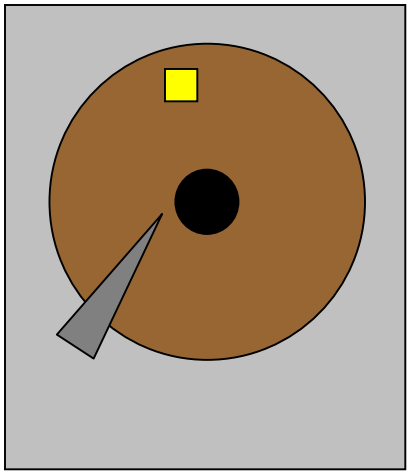


Why is e-Transaction Processing Important?

- It is an evolving business requirement aimed at producing tangible business returns, maximizing competitive advantage, and providing the highest levels of customer satisfaction.

Source: IBM White Paper - e-transaction processing

Understanding Rotating Disk Latency



Seek Time

the time it takes to move the disk arm to the required cylinder.

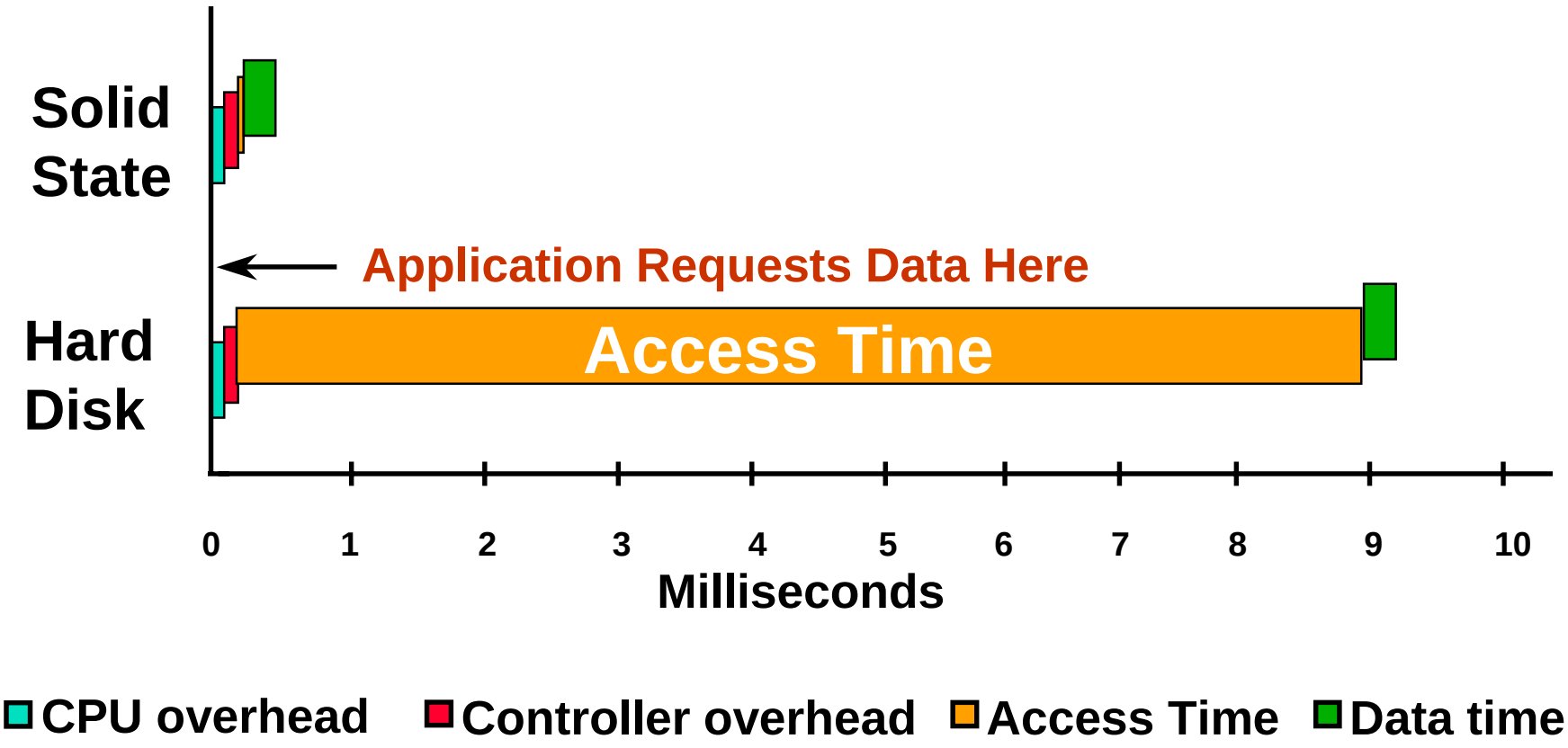
Rotational Latency

the time it takes for the disk to rotate so the required sector is under the head.

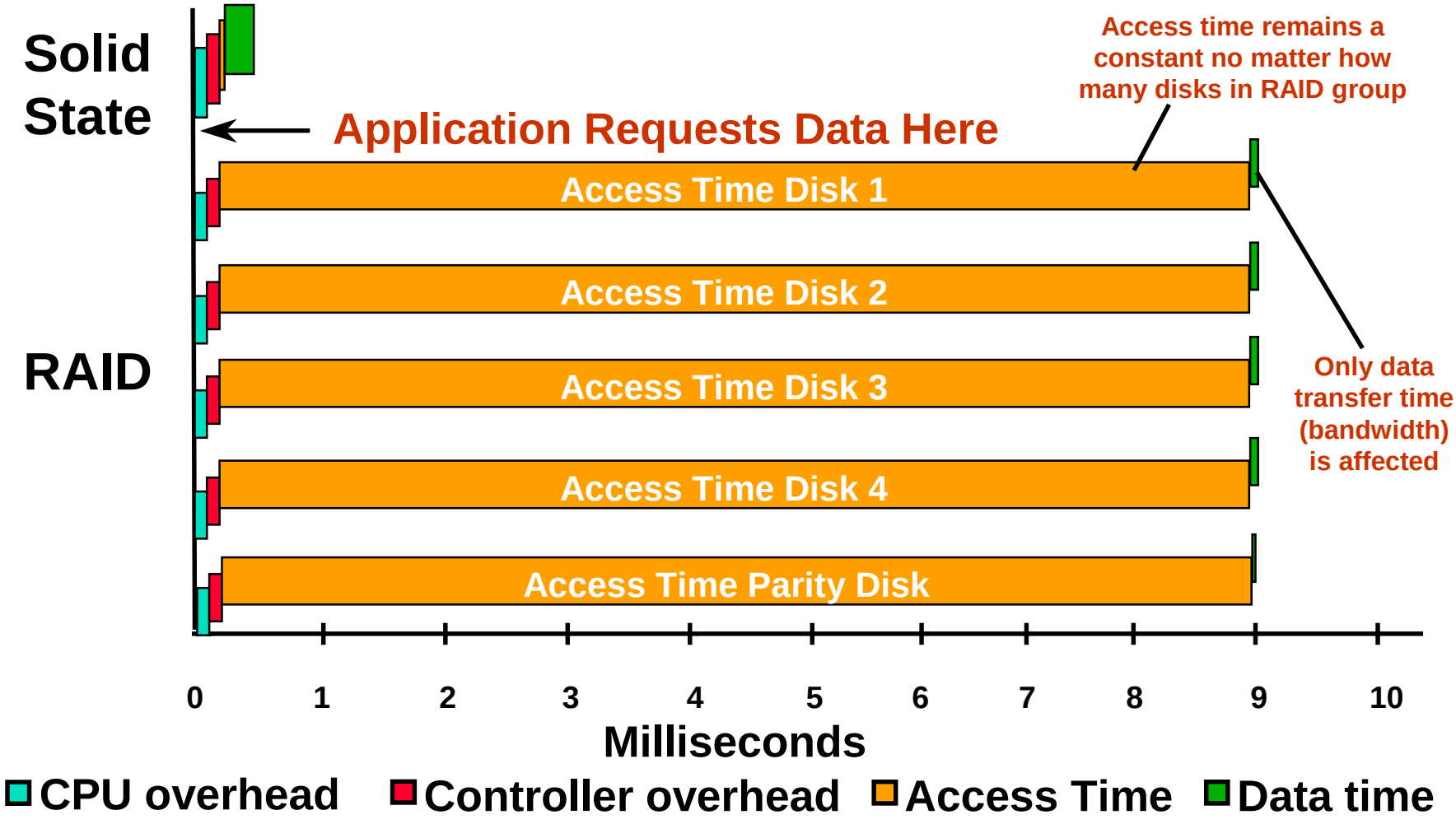
Access Time

the time it takes to get in position to read or write (seek time + rotational latency).

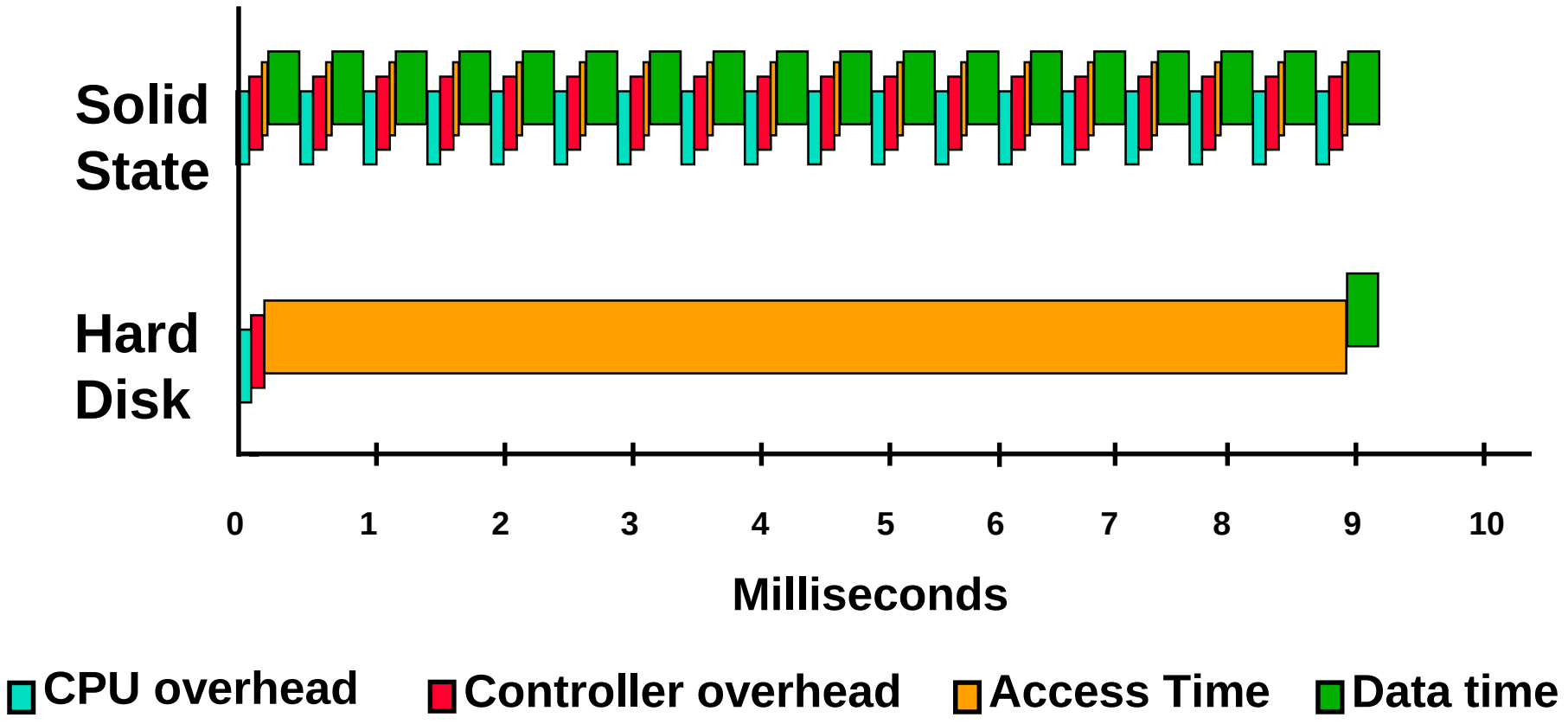
Latency During Data Access



RAID Impacts Bandwidth, Not Latency



Solid-state Storage Performance Advantage is Orders of Magnitude



Solid State Storage vs Rotating Disk

	Speed	Seek Time (ms)	Rotational Latency (ms)	Access Time (ms)	Max IOPS
Hard Drive	7,200 RPM	7.1	4.17	11.27	88
	15,000 RPM	2.0	3.6	5.6	177
RAID	Slow*	N/A	N/A	N/A	1,680
	Fast*	2.99	5.2	8.19	5,200
Solid State	Slow	0	0	0.35	6,247
	Fast	0	0	0.14	12,639

*Requires multiple host adapter cards & storage device channels

All numbers represent average random read/write results

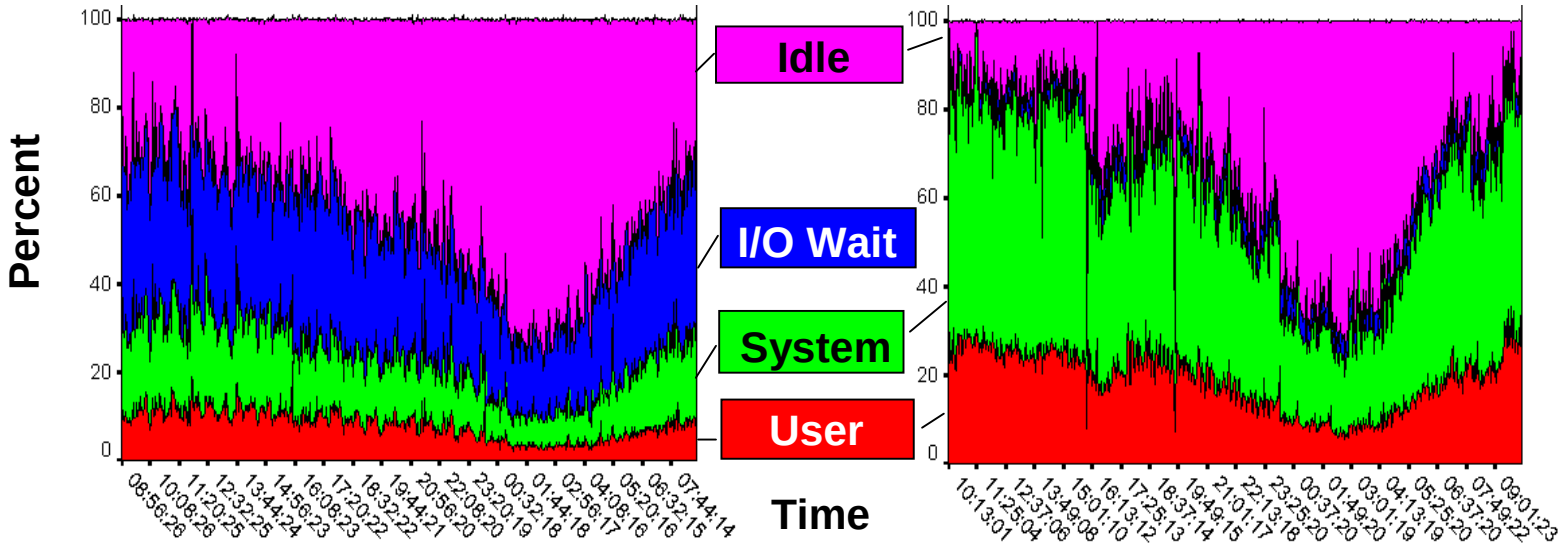
Sources: Solid Data Systems, Maxtor, Seagate, Sun, Dell

The Performance Impact of Latency

CPU utilization analysis -- E-mail server peak workload

a) Before file cache

b) After hot files moved



I/O Wait = 45% of CPU time at peak load

13 messages/second

I/O Wait = virtually eliminated

55 messages/second

>4x Improvement

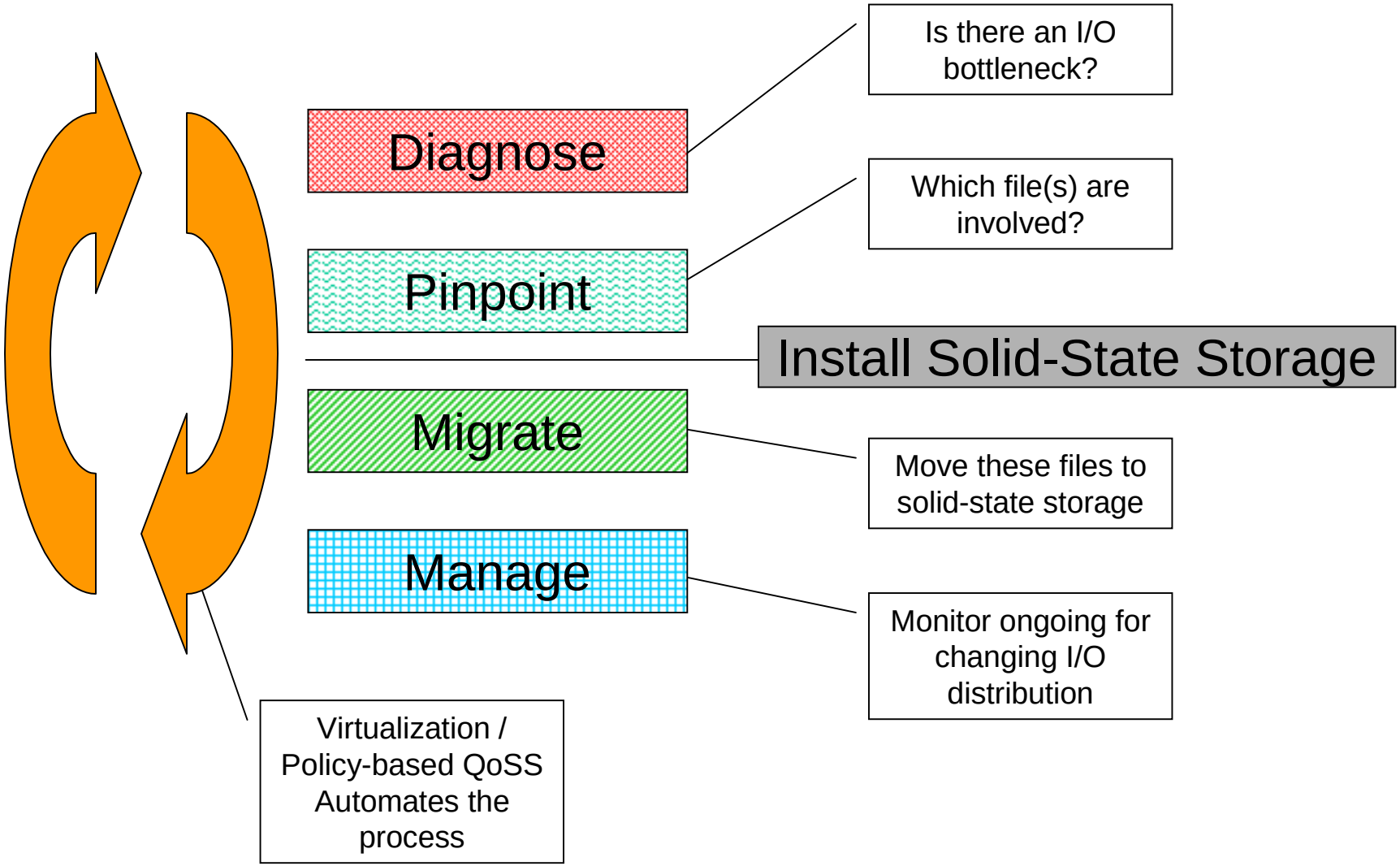
Benefits of Solid-state Storage in e-Transaction Environments

- Enable I/O bound servers to do the work of 4 servers - at the cost of 1
- Increase e-transaction load capacity - meet growth & peak demands
- Reduce time/costs spent on tuning efforts
- Improve customer experience and meet QoS requirements
- Enable higher e-transaction revenue generation / lower cost of e-transactions

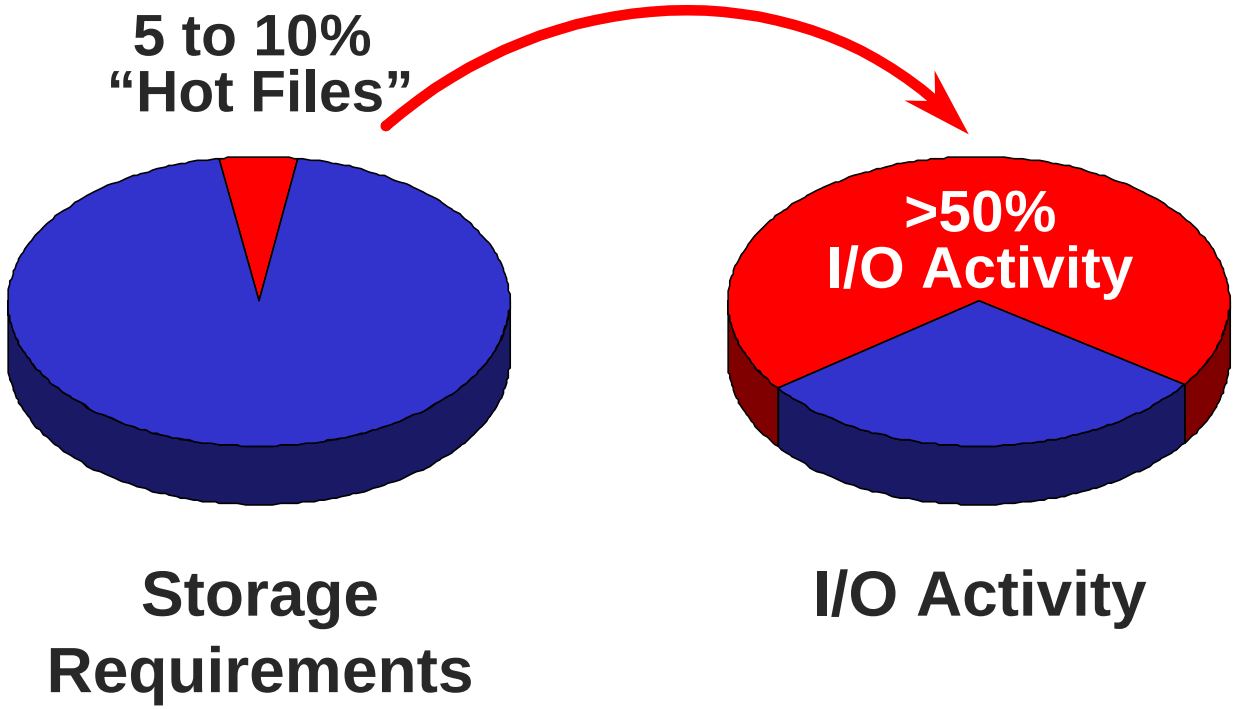
Real-world e-Transaction Performance Gains with Solid-state Storage

- Email server message capacity increased from 5/sec to over 40/sec
- Satellite-based stock data recording backlog reduced from 4 hours to 10 milliseconds
- Billing application reduced from 5 days to 2 days
- Batch job reduced from 8 hours to 2 hours
- Batch job reduced from 72 hours to 8 hours
- System response time after data entry reduced from 15 seconds to 3 seconds

Virtualization / Automated QoSS

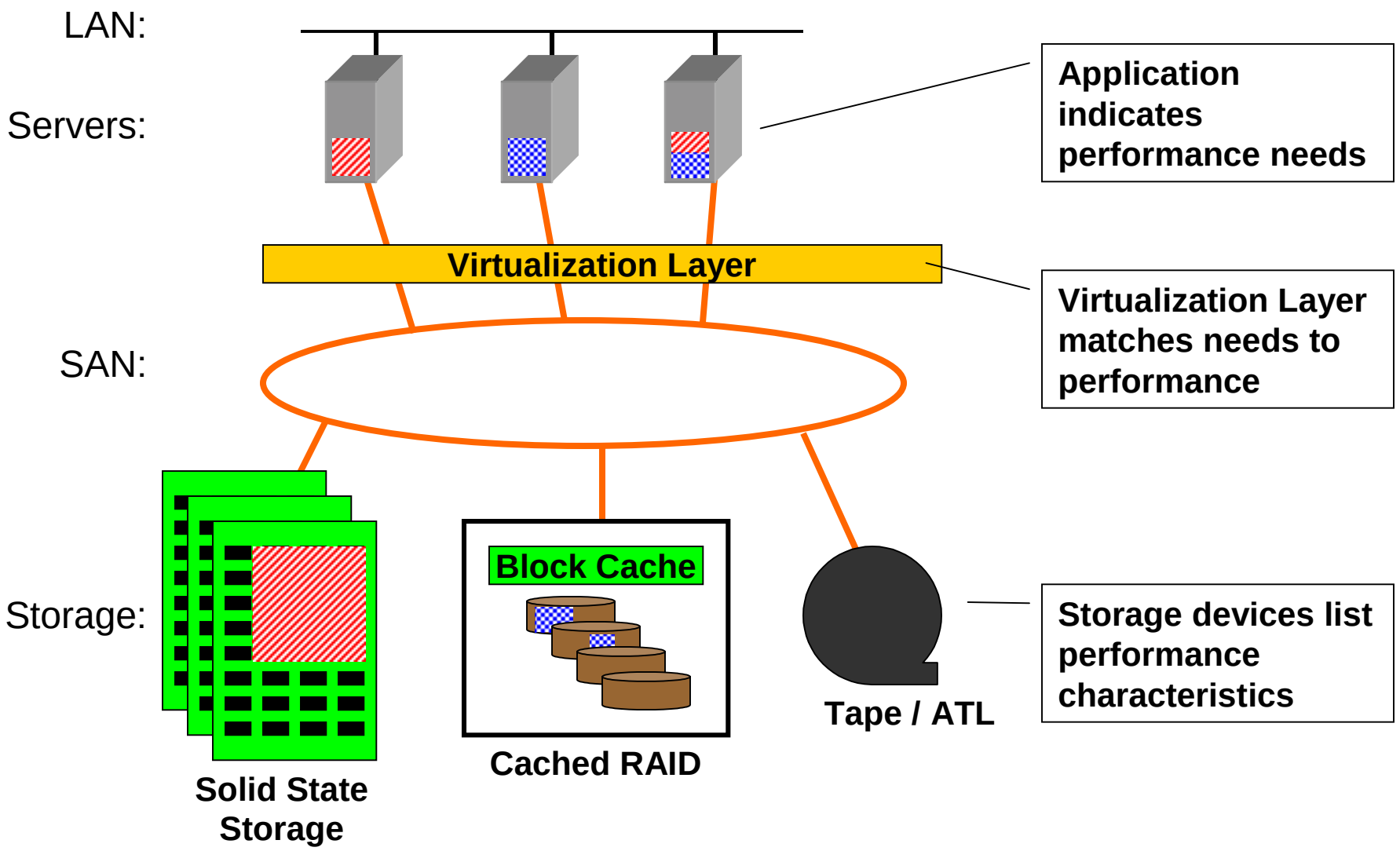


Solid-state Storage Criteria



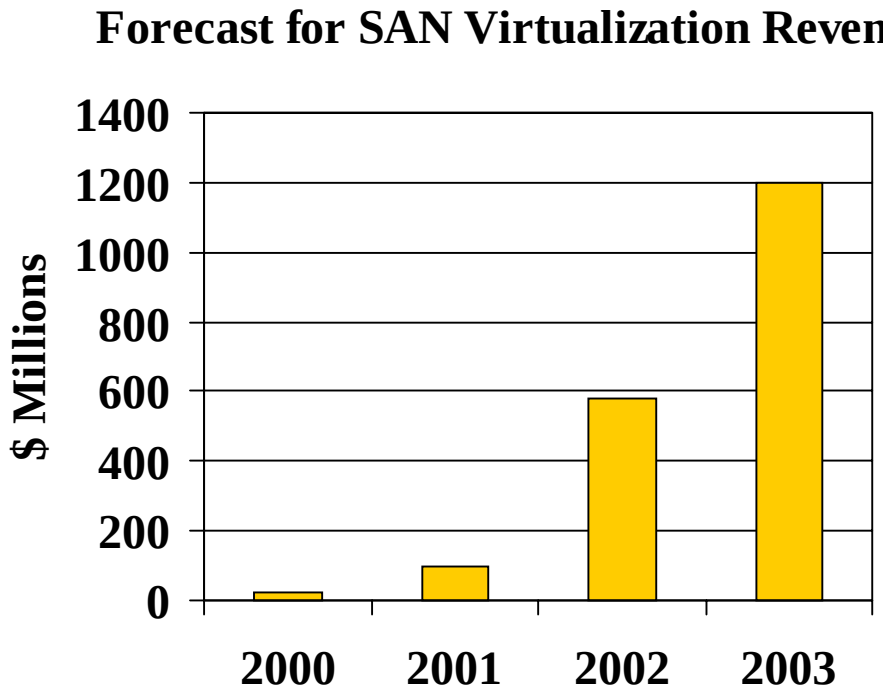
When a small number of files represent a large percentage of total I/O activity, solid-state storage is highly likely to multiply performance by 200%-800%+

Storage Virtualization



SAN Virtualization Forecast

- The need to more easily manage the growing enterprise storage pool will cause SAN Virtualization revenue to surge to $1.2 billion in 2003.



Source: IDC

Virtualization Vendors

Vendor	Product
• Adic	- SAN Gateway
• Compaq	- VersaStor
• DataDirect Networks	- SAN Data Director
• DataCore	- SANsymphony
• Dell	- PowerVault 530F
• IBM	- StorageTank
• StoreAge Networks	- SAN Volume Manager
• StorageApps	- SAN Link
• Veritas	- SANpoint
• Vicom	- SAN Virtualization Engine
• XIOtech	- Magnitude

- And the list is growing continuously...

Summary

- The Internet has fostered an explosion of e-transactions of many types
- Rotating disks have mechanical limitations that cause latency
- Solid-state storage is a cost efficient way to increase e-transaction capacity by eliminating mechanical latency
- Storage virtualization and automated QoSS will serve to accelerate the adoption of solid-state storage for e-transaction intensive applications

Q&A

Thank You!

For a copy of this presentation or answers to other questions please send e-mail request to me:

mhall@soliddata.com

Appendix

Features to look for in Solid-state Storage

- Appropriate form factor - *3.5", desktop, rack-mount*
- Low latency - *industry range is 14-50 microseconds*
- Non-volatile architecture
- Connectivity - *Fibre or SCSI*
- LUN mapping / masking
- Redundancies to match availability requirements
- Upgradable capacities to meet future growth
- Ease in serviceability
- Field proven reliability - high MTBF
- Out-of-band monitoring / SNMP / Web support
- Available 4 hour onsite & 24x7 phone support

Solid State Storage vs Cached RAID

- Entire file in RAM
- Latency as low as 0.014 ms
- Consistent performance
- No tuning required
- Best for small block random I/O
- No moving parts - very high MTBF
- Independently scalable

- Selected blocks in RAM - other blocks on disk
- Latency of 5.000 ms or greater
- Performance depends on cache algorithms
- Extensive tuning sometimes required
- Best for large block streaming I/O
- Mechanical complexity reduces MTBF
- Limited cache to RAID ratio

Symptoms of Potential I/O Bottlenecks in e-Transactions

- Performance drops radically as users are added to a system
- System no longer able to “keep up”
- Batch jobs don’t complete within the available processing time window
- Month-end close takes days instead of hours
- Extensive tuning has not solved performance issues
- SAR (Unix) or PerfMon (NT) indicate >50% I/O Wait