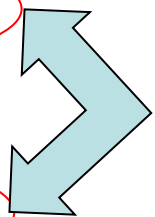# Enabling Big Data Applications with Flash Memory

Andy Walls

IBM Fellow, CTO and Chief Architect
Flash Systems

Big Data

- Not Just A Lot of Data
- More data than can be efficiently handled by traditional means
- The Three Vs
  - Volume
  - Variety
  - Velocity
  - Maybe a 4th – Veracity!
  - Maybe a 5th – Vulnerability!

Flash Most impacts Volume and Velocity

Flash Memory Summit 2014
Santa Clara, CA

3

Volume
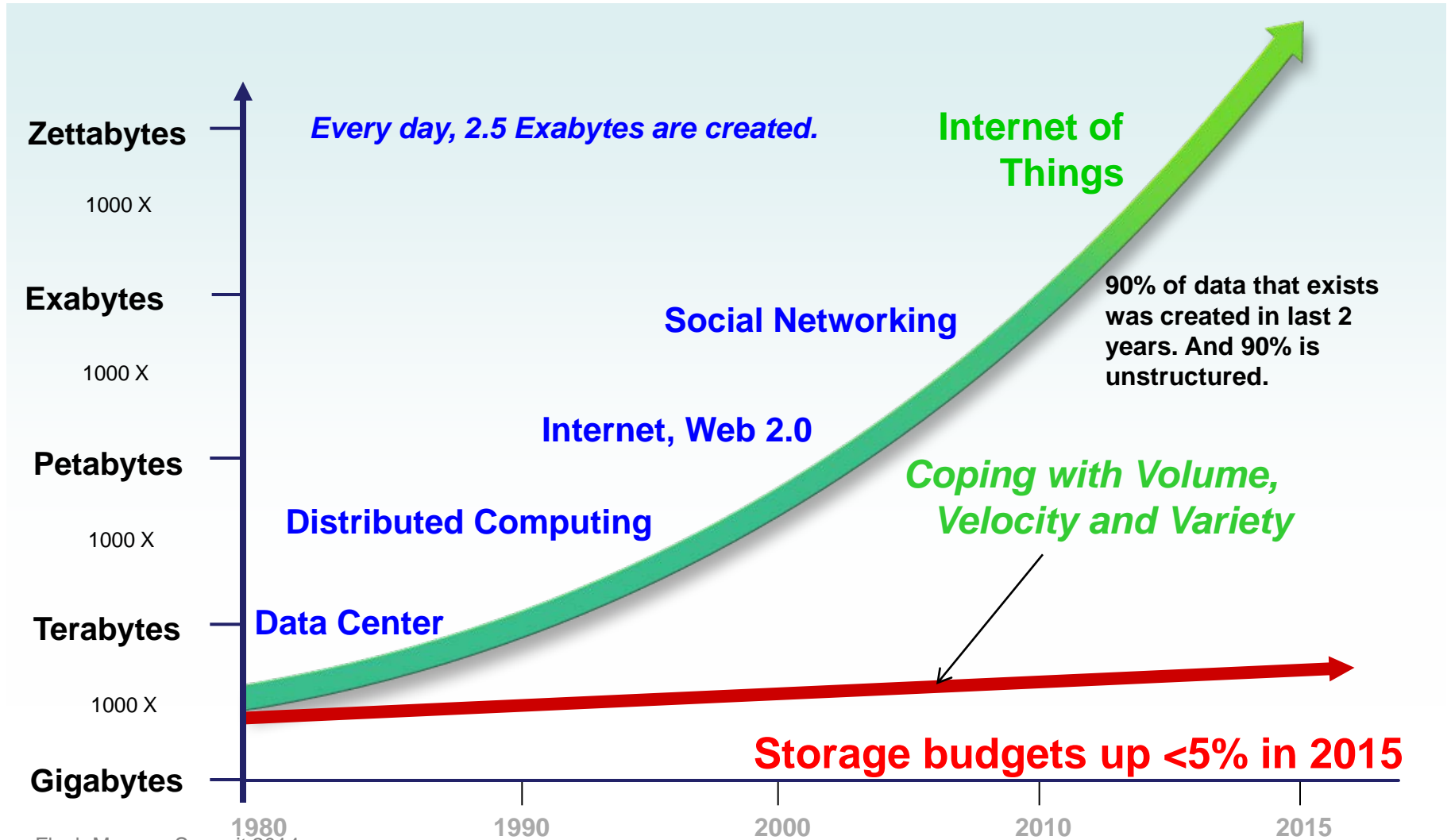
- 2.5 Exabytes a day are created
  - 62.5K Flashsystem 840s
  - Facebook, Linkedin, Twitter, Instagram, and others create enormous amounts of data
  - Sensor data can be overwhelming: Utility meter data, Geological data,
  - Call Data Records
  - Weather information
  - Retail and online auction transactions
  - Click records

Flash Impact on Volume

- Flash trends do help directly with volume
  - Enabling MLC has increased use in the enterprise
  - Data Reduction techniques increased the logical availability
  - <u>But – Flash can not hold all the data created</u>
- Flash Indirectly helps with volume:
  - Hybrid and caching can allow for use of cheaper media and still acceptable performance..
  - Flash can augment DRAM for In Memory databases.
- As the volume of data increases, passing this data through high velocity storage increases the rate of analysis

Velocity

- Data is being created and changed at a mind boggling rate
- Analysis of said data requires Enormous Performance attributes
  - 10, 40, 100Gb Fabrics
  - High IOPs
  - High throughput
  - Very Low Latency
- Elasticity
  - Even with processor speed improvements
  - Memory bandwidth is a gate
  - Network

What is Critical though is . . . .

- That in the midst of the Volume of data. . . . .
- Even though it is Various and is coming at high velocity . . . . .
- Despite the issues with Voracity and the vulnerabilities . . . .
- The point of Big Data is one more V

TO FIND VALUE . . . *In Real TIme*

Examples of Finding Value

- Recommendation Engine and pattern recognition
- Sentiment Analysis: Determines micro and macro trends with regard to brands, companies or products
- Risk Modeling
- Fraud Detection
- Social Graph Analysis
- Customer experience analysis
- Network monitoring

Analytics for Data at Rest

- Hadoop, Netezza
- Analyze previous months sales records, clicks, online auctions, groupings, across geographies and factor in weather, events, etc to then plan for current month.

Real Life Examples

- Retail companies finding trends from social media and combine with sales records to offer the right promotion or price geared to demographics or type of consumer.

- Analyzing Call Data Records of "group leaders" in social networks to affect tendency of these groups to leave network.

- Boost Marketing Campaign management solution. Increase campaign effectiveness, deliver a shorter time to market for new promotions, find new revenue streams, and enrich churn analysis.

- Fraud Detection for credit card usage

More Real Life Examples

- Internet Protocol Detail Records can be analyzed to predict issues and problems before they lead to a breakdown.
- Improved Law Enforcement: Detecting cyber attacks, intruder detection before they intrude.
- Analyze high volumes of trading and market data at ultra low latency across multiple markets and geographies for arbitrage trading and risk analysis.
- Look for trends and synergies in huge amounts of diagnostic data in critical care.

In Motion Analytics could make this scenario a reality. . . .

- A credit card company calls Mrs. Walls at home in San Jose. . . .
- Andy bought a cup of cappuccino in airport at San Jose at 5AM along with 2 glazed doughnuts
- At noon he had lunch in Phoenix and had a large Hamburger and Apple Pie
- At 7PM he had a 16 inch prime rib at Houston along with a Crème Brulee
- He is now trying to buy a quart of ice cream at the grocery store

Would you like us to deny his charge???

What it Means

- Even though example is fictitious it demonstrates what is needed for in motion analytics:
- Examine millions and maybe billions of pieces of information
- Group or categorize related items
- Perform analysis (Figured out calories of each item, added them up, also determined there were 3 deserts)
- Determine a course of action (Call wife)
- Act – send information to a real person to intervend

Examples of Velocity

- One example required the analysis of 8 Billion CDRs a day. 5PB of data generated a year.

- 12.7 Million option market trading messages

- Analyze 1,000 pieces of unique diagnostic information per second for babies in neonatal units.

- Estimated that Twitterers Tweet 350K tweets every 60 seconds. 700K Facebook posts and 100M emails, 200K photos uploaded
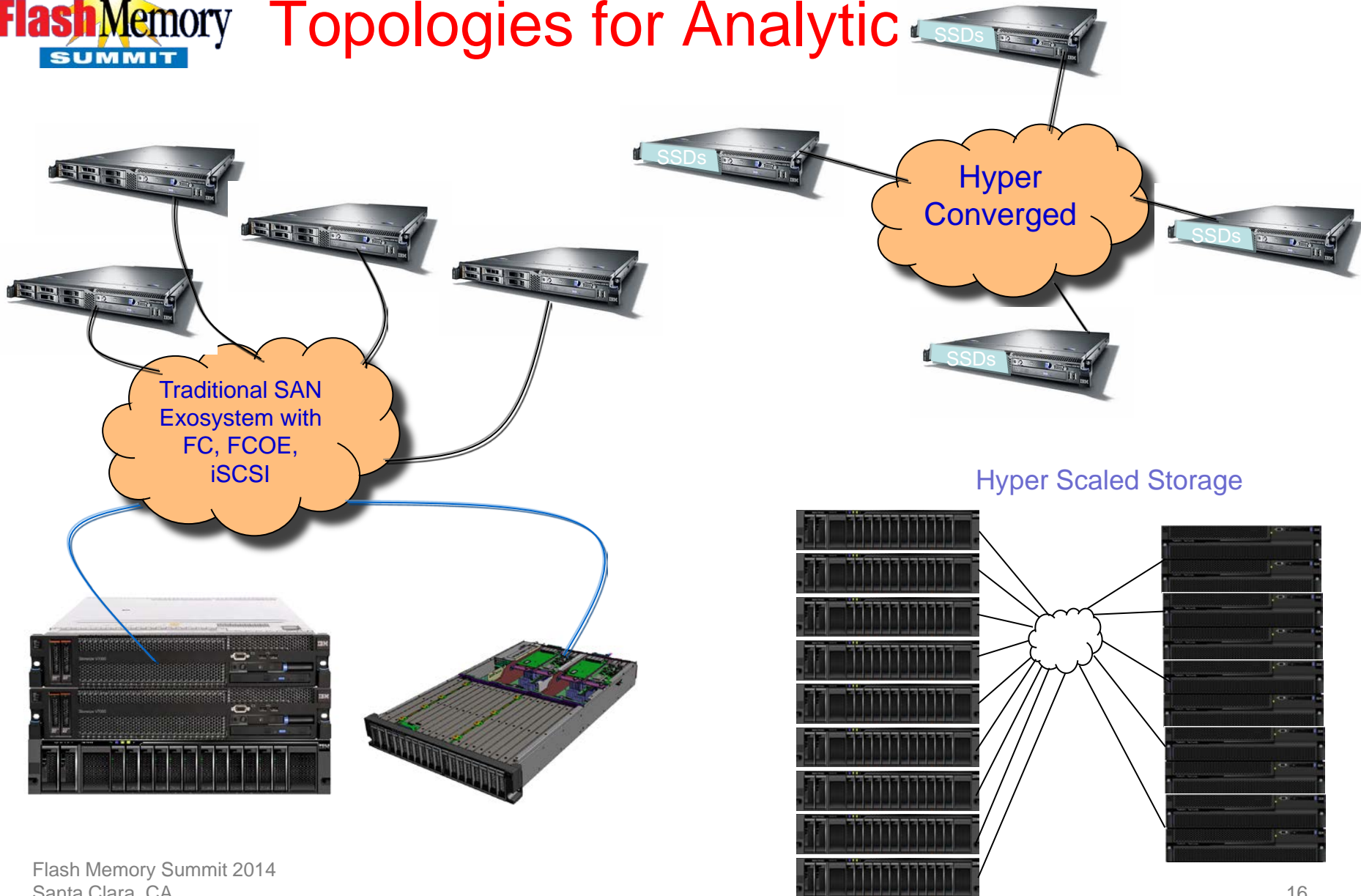
How Flash affects Analytics for data at rest

- Very high read throughput
- Very high read IOPs
- Very low latency

How Flash affects Analytics in motion

- Mixed IOPs but still read heavy
- Very low latency

Topologies for Analytic

SSDs

SSDs

Hyper Converged

SSDs

SSDs

Traditional SAN Exosystem with FC, FCOE, iSCSI

Hyper Scaled Storage

Key Considerations

- Applications are typically not write intensive.

- Consistent and low response time are vital

- Power is the somewhat overlooked benefit that significantly adds to cost advantage

- Automation and ease of management are vitally important.

- Offload is critical
  - RAID Offload
  - Data Reduction offload
  - Analytics offloads