

# Managing and Protecting Enterprise Big Data with FDR/UPSTREAM and Data De-duplication Techniques

Patrick Fitzsimmons  
INNOVATION Data Processing

August 6, 2012  
Session #11672



## Executive Overview

You hear and see two buzz words everywhere...  
“**BIG DATA**” and “**THE CLOUD**”

*Everything and everyone is facing challenges associated with... “Big Data” and “The Cloud”*

I’m going to give you two other words to remember...  
“**SPACE**” and “**TIME**”...the root to most of all the problems these buzz words conjure up

If you can help your organization solve their **SPACE** and **TIME** problems you’ll give them a **BIG JUMP** on everything else.



# Big Data and The Cloud

- **Cloud Computing is an effort to...  
make distributed computing more like mainframe computing**
- **Many customer sites want to or are considering putting their distributed Public/Private Clouds on the mainframe...**
- **Mainframe customers are moving to “The Cloud” with CICS, Websphere, Linux on System z and the new IBM z114/z196 zEnterprise**
  
- **“Big Data”...comes with moving to The Cloud...**
- **Too much data with no adequate way to protect it**
- **It takes too much *time* and too much *space* to protect it**



# Big Data...Enterprise Backup Pain Points

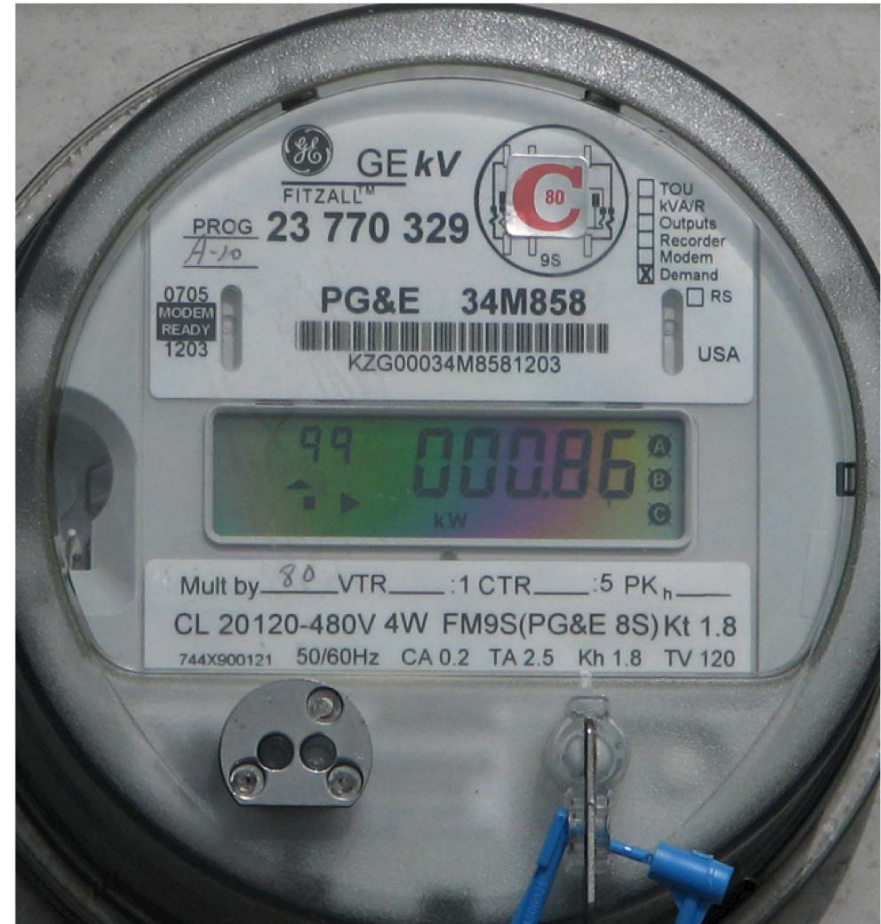
- Petabytes and Zettabytes here we come  
*A zettabyte is equal to 1 billion terabytes*
- RFID tags – 30 billion in use
- Smart Phones
- GPS
- Social Media Outposts
  - Facebook – 25 TB of logs data every day
  - Twitter – 12 TB per day





## Big Data...Here We Come

- Utility companies are now offering smart meters... that are doing meter readings once every hour or *less* compared to once a month...data collected then to data warehouse & then mined...how many TB? What is the retention period of the collections?



# Big Data...Enterprise Backup Pain Points

Big Data introduces new challenges:

- Airline – flight from France to JFK generates 640 TB of data  
Data Mining – Customer Loyalty Card
- Data Warehouse
  - Larger volumes are necessary to handle data growth
  - Backup taking longer
  - Enterprise backup congests communication networks



# Big Data...

## Affordable Care Act – Obamacare

- As some employers will stop offering healthcare benefits to employee, they will be forced to buy from exchanges.
- Big pharma will need to differentiate their products to payers and patients who may elect to pay for treatments themselves.
- Payers will aggressively advance the accountable care organization model to drive cost savings.
- 30 million previously uninsured Americans will have access to healthcare.
- This is going to be about data. Big data. And big supporting analytics.



# Big Data...Enterprise Backup Pain Points

Real Problems and new ideas:

- Many different backup solutions
  - Little islands with point solutions
- Backup taking longer
- Larger volume are necessary to handle data growth but require more time & space
- We need to reduce the amount of backup data
- Consolidating Linux puts more Linux data on the z/OS mainframe
- Enterprise backup congests communication networks





# Executive Summary

## Enterprise Backup Pain Points

- Do you have issues meeting your backup window?
- Meeting your Recovery Time Objective?
- Meeting your Recovery Point Objective?
- Need to get a better Total Cost of Ownership?
- Do your backups need to be crash consistent?
- Are you being asked to do more with less?

 #SHAREorg

# Save Time and Space *with* Storage Resource Optimization

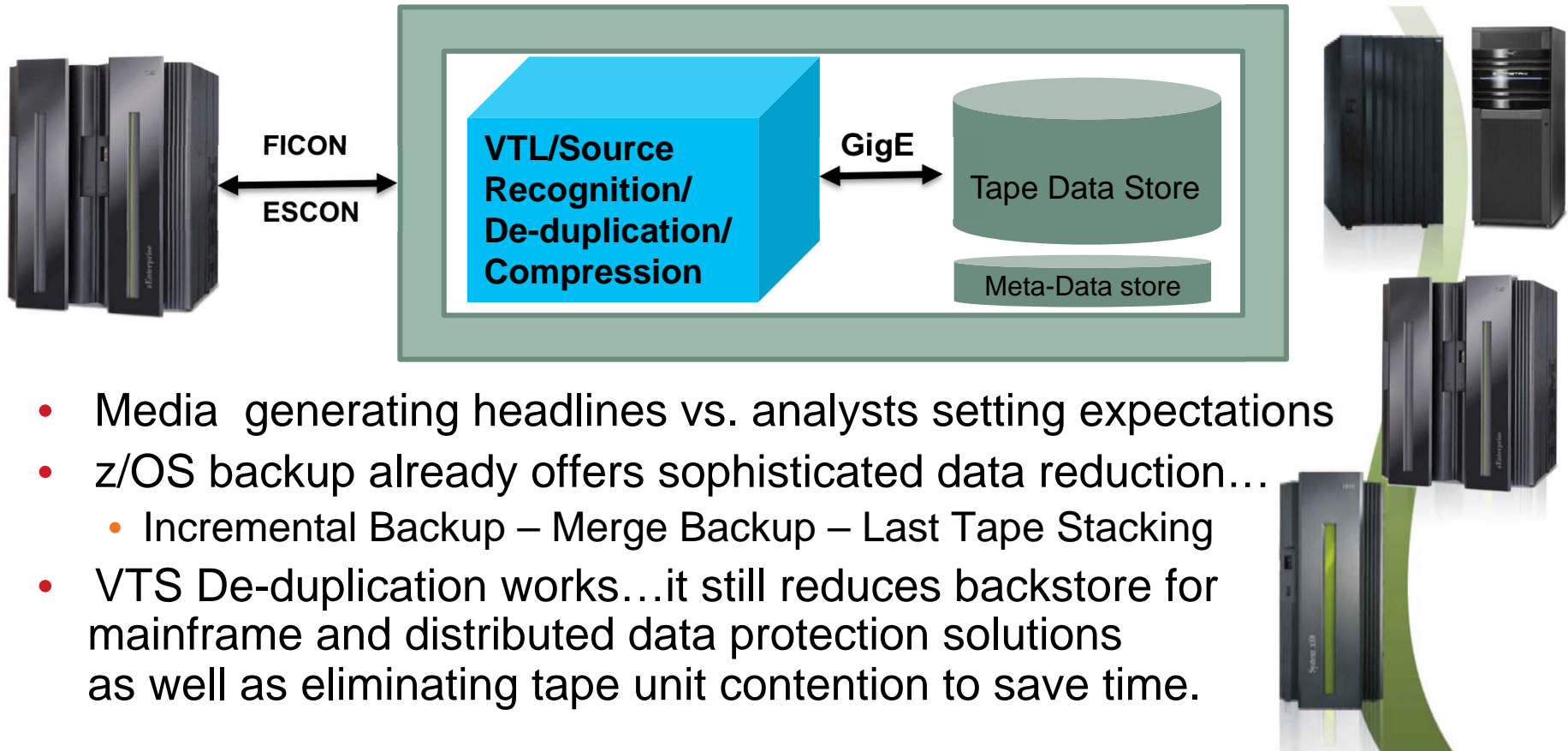


# Why Consider De-duplication?

- Media Attention – Why All The Noise?
  - Hype Versus Reality
- Reduce Storage Requirements
  - Reduced Footprint, Power, Cooling
  - Better Media Utilization
- Other Areas of Promise
  - Reduced Transmission Costs (Replication)
  - Shared Storage With Open Systems
  - Shared Storage With Other Mainframes
  - Reduced TCO
- What Factors Affect Data De-dupe Rates?
  - One Thing – Repetitive Data



# De-duplication is a natural complement to z/OS mainframe and Multi-Platform z/OS Distributed Data protection...

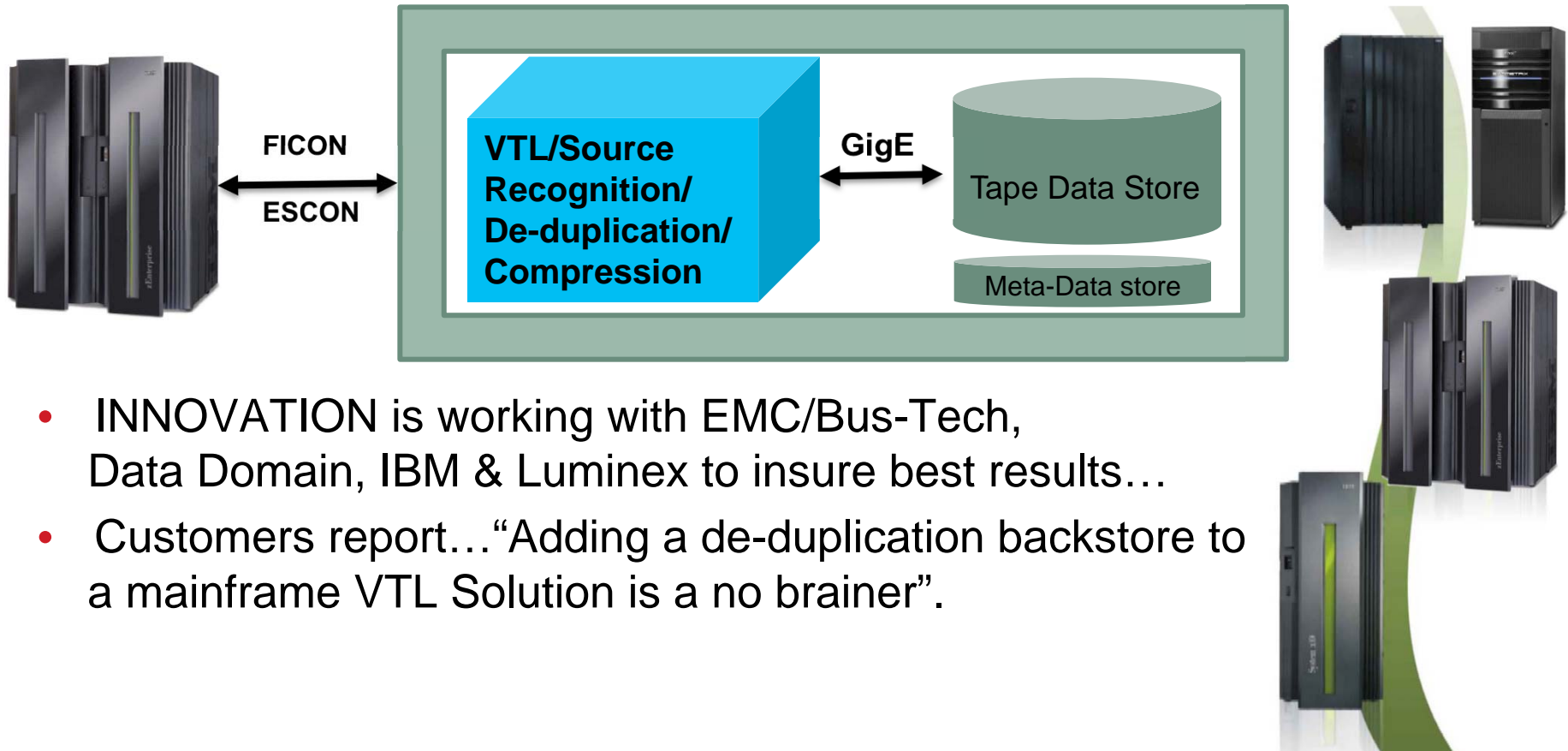


- Media generating headlines vs. analysts setting expectations
- z/OS backup already offers sophisticated data reduction...
  - Incremental Backup – Merge Backup – Last Tape Stacking
- VTS De-duplication works...it still reduces backstore for mainframe and distributed data protection solutions as well as eliminating tape unit contention to save time.





# FDR, FDRABR, FDRSOS, UPSTREAM and RESERVOIR... all benefit.



- INNOVATION is working with EMC/Bus-Tech, Data Domain, IBM & Luminex to insure best results...
- Customers report...“Adding a de-duplication backstore to a mainframe VTL Solution is a no brainer”.



# Case Study...

## z/OS and Distributed Data Backup to an EMC DLM

### ***FDR/ABR z/OS Backup***

- 50 z/OS volumes w/130 GB of data (a mix of 3390-1/3/9 & 27)
- z/OS IPL packs, System & ISV executables, JCL, Source, Listings, sysdump, FTP server, application work, DB2 systems & data bases, BookManager and product softcopy manuals.
- ABR daily incr (7% changing) & weekly fulls.

### ***FDR/UPSTREAM Distributed Backup***

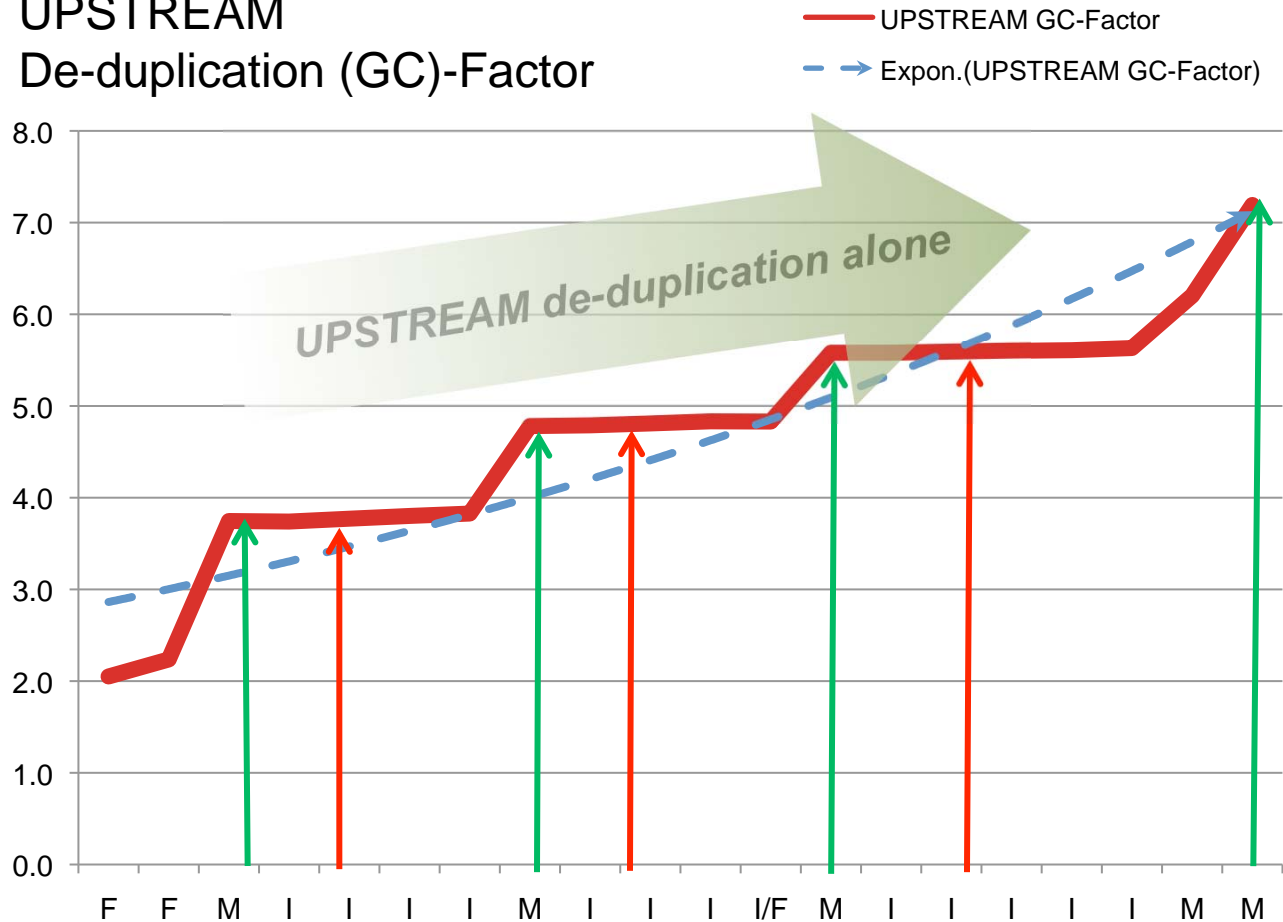
- 50 distributed servers with 2 TBs of data
- Windows Servers, Solaris on x-86, SuSE, & RedHat Linux on x-86 and Linux on System z.
- UPSTREAM daily incr (4% changing) & weekly full merge



# De-duplication by itself reduces UPSTREAM backstore requirements...7 to 1



UPSTREAM  
De-duplication (GC)-Factor

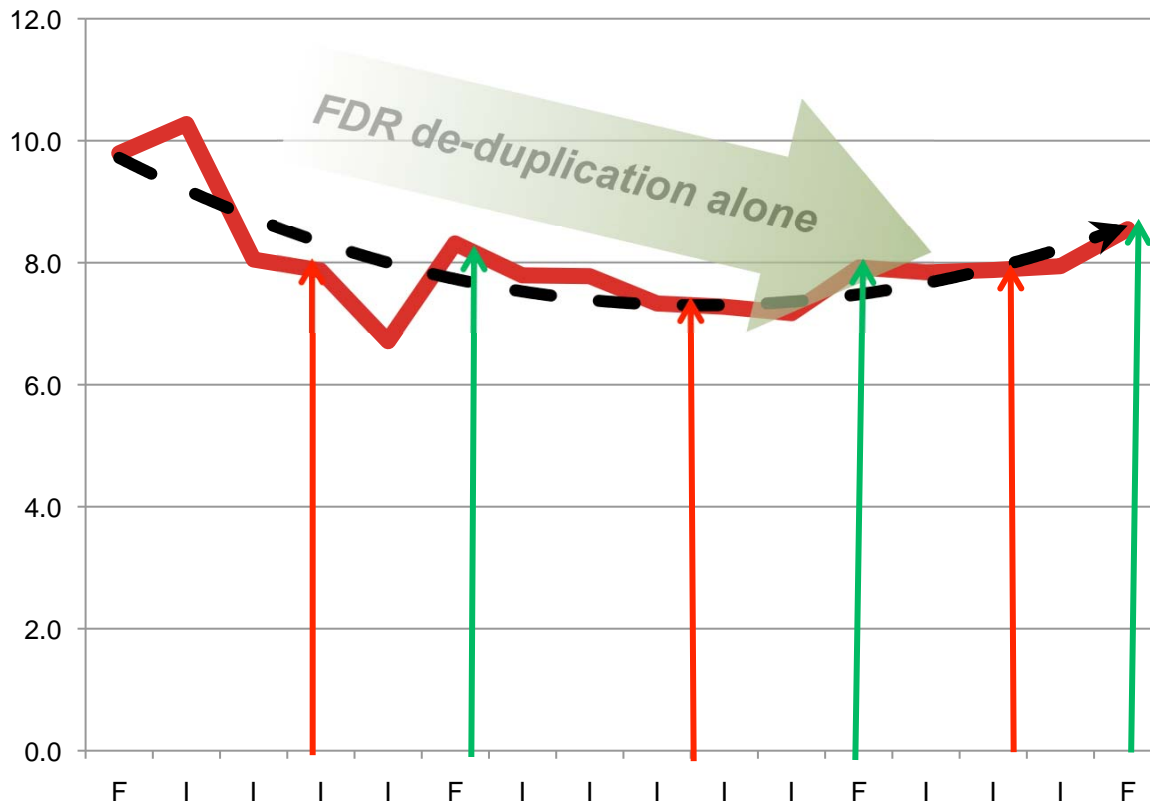


# De-duplication alone reduces FDR backstore requirements...8 to 1



FDR  
De-duplication (GC)-Factor

— FDR GC-Factor  
- - - Poly.(FDR GC-Factor)



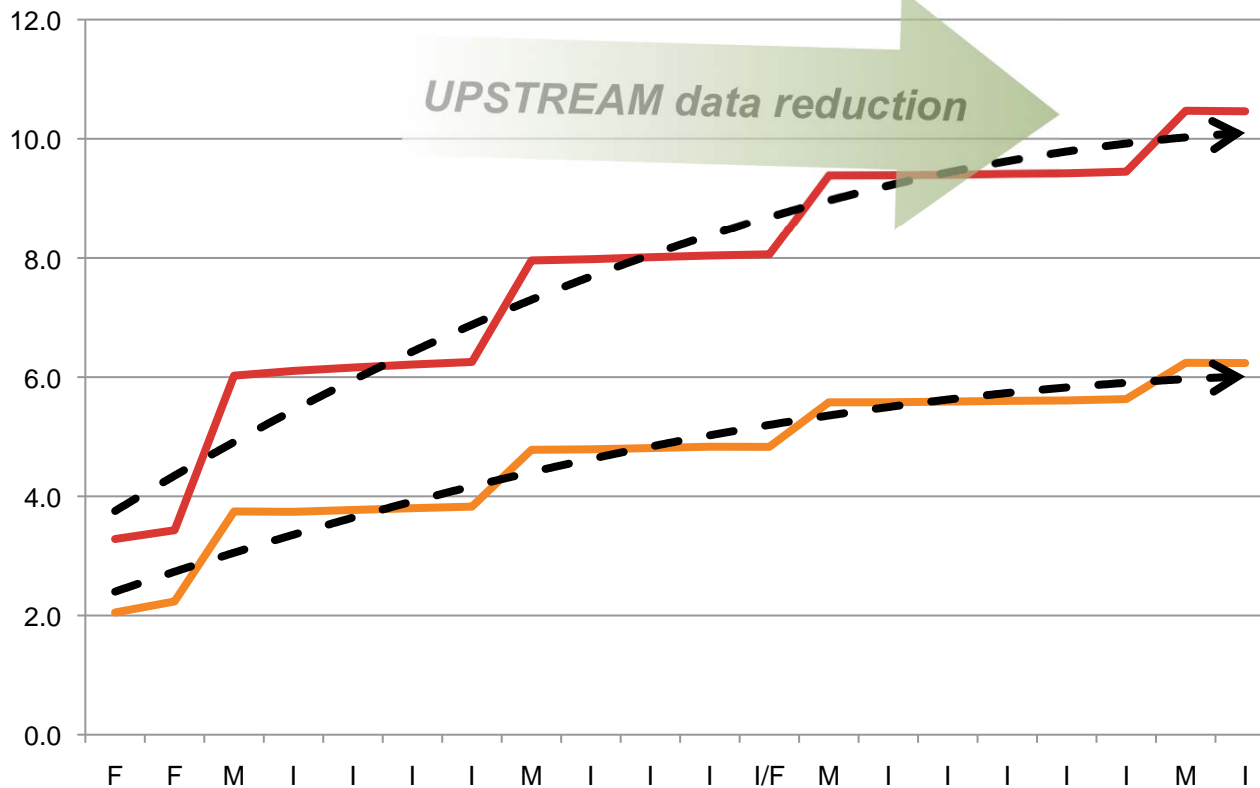


# The De-duplication appliance further reduces the UPSTREAM backstore to...10.5 to 1



UPSTREAM Overall Data Reduction Factor  
De-duplication (GC) and Compression (LC)

— UPSTREAM LC-Factor  
- - Poly.(UPSTREAM GC-Factor)  
- - Poly.(UPSTREAM LC-Factor)

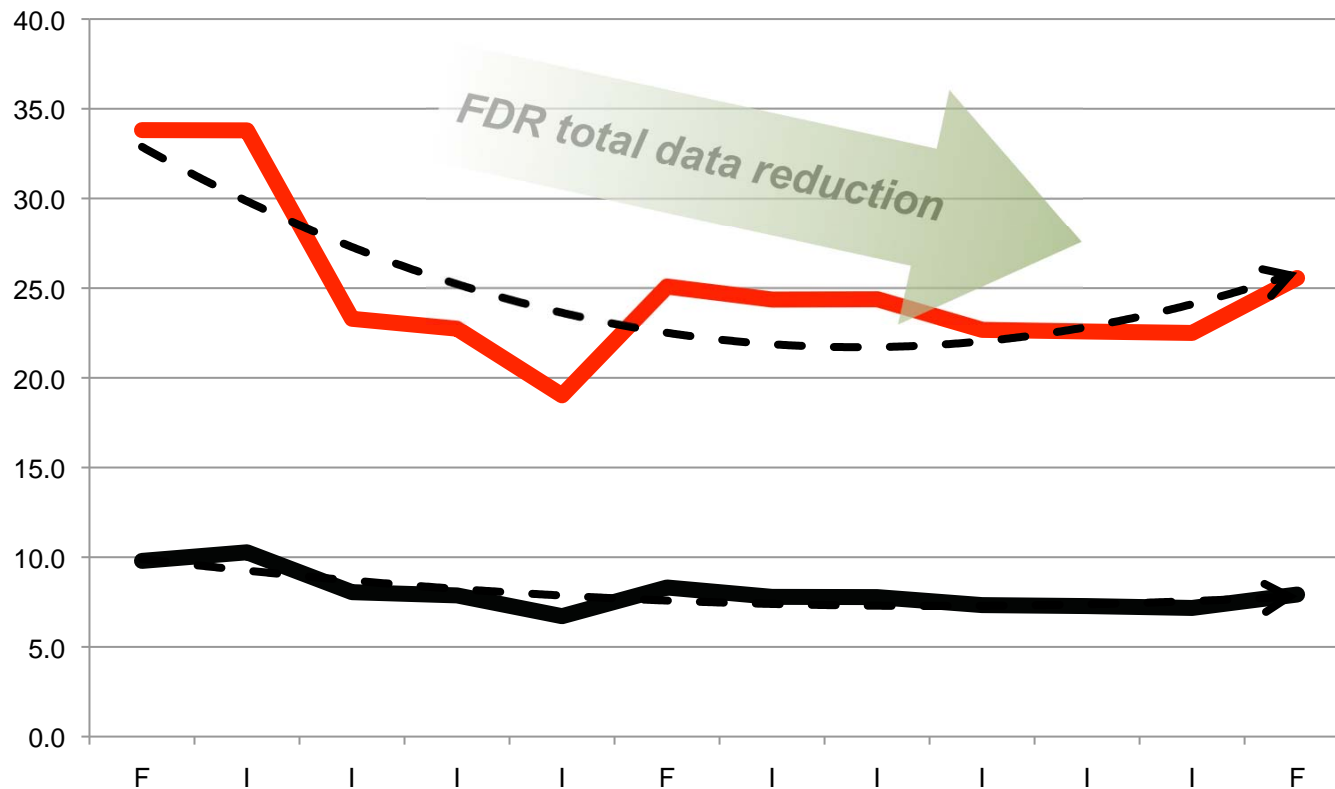


# Leveraging compression further reduces FDR backstore to...25 to 1



FDR Overall Data Reduction Factor  
De-Duplication (GC) and Compression (LC)

- FDR GC-Factor
- FDR LC-Factor
- - Poly.(FDR GC-Factor)
- - Poly.(FDR LC-Factor)



# Understanding Reduction Ratio and % Saved...

- 10 to 1 data reduction is a backstore savings of 90%
- Greater than 10 to 1 can be only a bit more...
- How much is a 20 to 1 reduction?



Customers Report Average

Ratio to 1	Saved	Stored
50	98%	2%
25	96%	4%
20	95%	5%
16	94%	6%
12	92%	8%
10	90%	10%



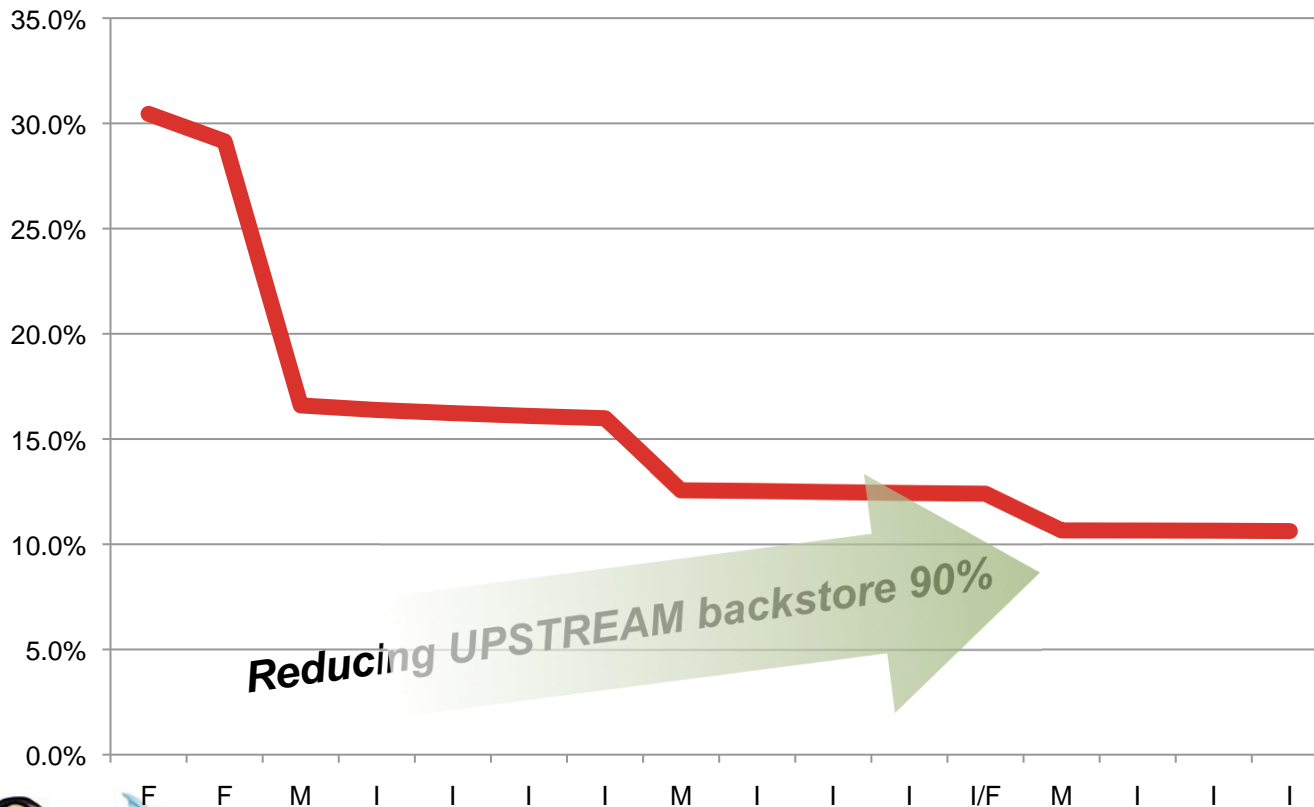
# How much can De-duplication appliances reduce overall UPSTREAM backstore?



UPSTREAM

% of Original Space Requirement Used

— % of Original Space taken



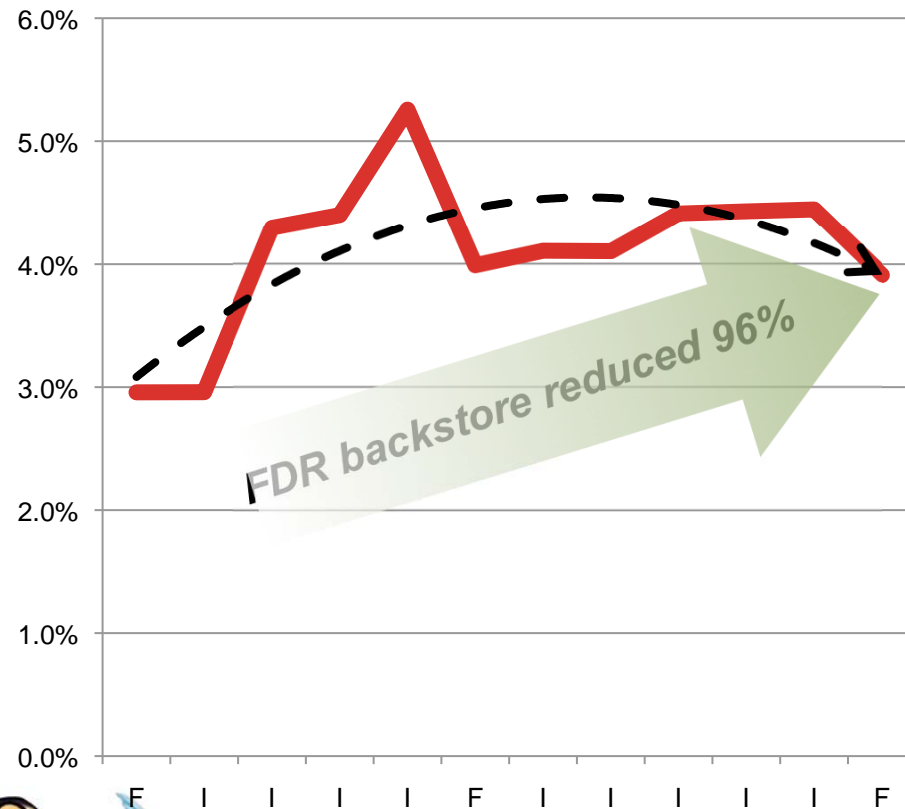
INNOVATION®  
DATA PROCESSING



# How much can De-duplication appliances reduce overall FDR requirements?



FDR  
% of Original Space Requirement Used

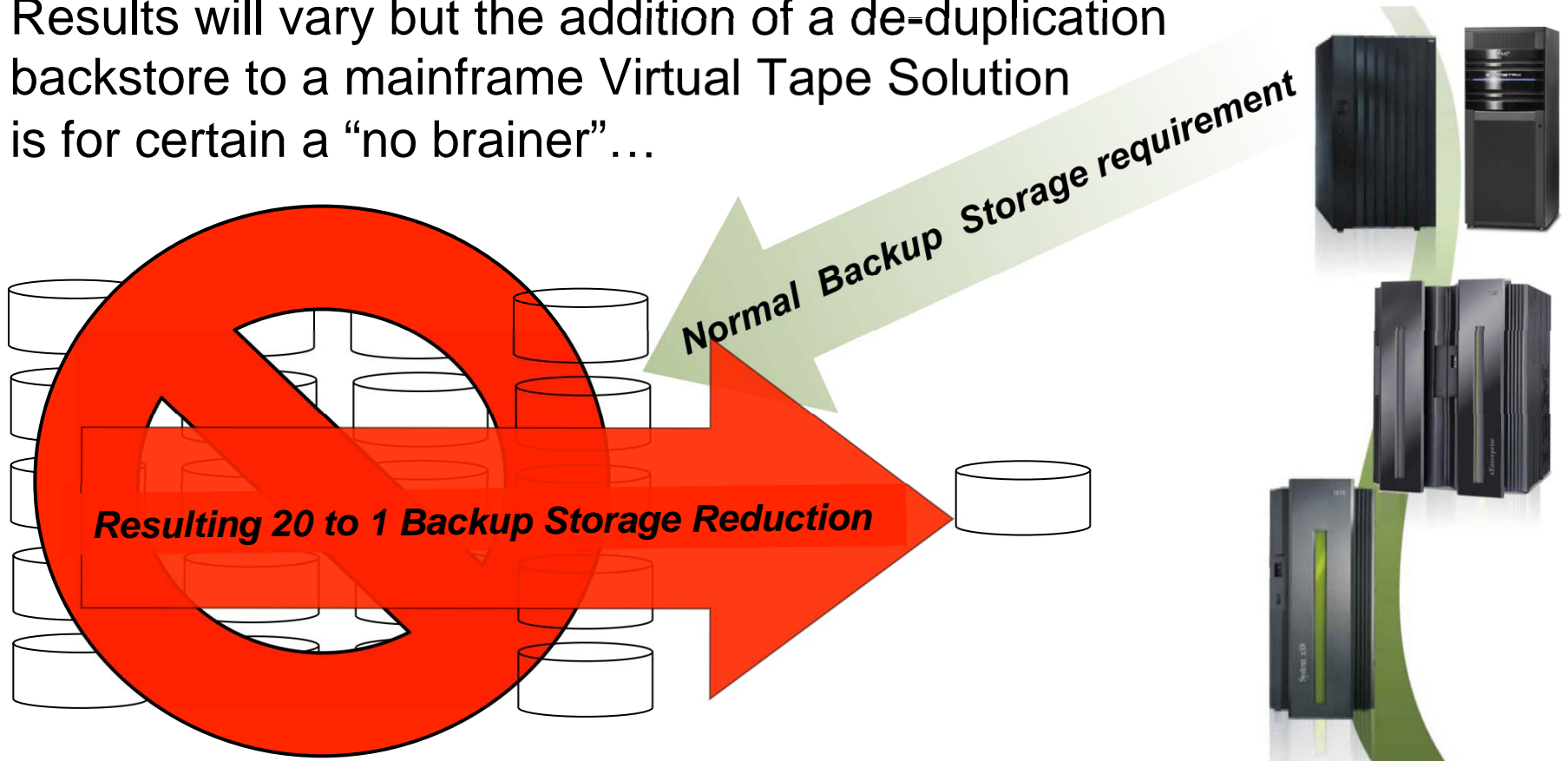


— % of Original Space taken  
- - -> Poly. (% of Original Space taken)



# What does 20 to 1 Mean in Practical Terms?

Results will vary but the addition of a de-duplication backstore to a mainframe Virtual Tape Solution is for certain a “no brainer” ...



# Data Reduction and De-duplication Technology Minimizes Data Transmission

- **Have too much data to backup?**
- **Advanced Data Reduction Technology Features employ De-Dupe Technology to reduce backup time:**
  - Data Compression
  - Synthetic Full-Merge Backup
    - Logical File Granularity
    - Incremental Backup Processing
  - Block Level Segmented Backup Support
  - Exclude/Include
  - Migration or Disk Grooming of inactive data
  - Support for EMC/Bus-Tech/Data Domain/IBM & Luminex



# INNOVATION's Enterprise Data Protection Solutions

- **FDR/UPSTREAM**  
a scalable, highly reliable, z/OS mainframe server-resident file level data protection solution featuring network and HiperSockets data transfer.
- **FDRSOS**  
a simple alternative direct cross-platform access solution for Open Systems disaster recovery backup. Bare metal recovery for virtualized environments.
- **UPSTREAM/SOS**  
combines the direct access technology of FDRSOS to extend the FDR/UPSTREAM mainframe resident solution with high speed robust off-network data transfer.

# Executive Summary

*FDRSOS leads way in Disaster Recovery for Distributed Data Protection.  
If planning on z114/z196 zEnterprise Hybrid Computing with zEnterprise  
BladeCenter® Extension (zBX) with blades.*

- IBM zDDB and EMC z/SOS with FDRSOS is the solution for...
  - High speed image backup across FICON channel instead of TCP/IP networks
  - No network traffic and no TCP/IP usage
  - Extend z/OS data protection to distributed servers
  - Consolidate mainframe and distributed data protection
  - Bare metal recovery for any array resident OS
- INNOVATION's FDRSOS...a mature, reliable player with a new look...
  - Extend the value of z/OS tape management & security systems
  - Improve distributed Data Disaster Recovery with reduced backup and recovery times
- INNOVATION SOSINSTANT is a new way to...
  - Advance distributed data disaster recovery backup with IBM FlashCopy and EMC TimeFinder.





# Centralized Enterprise Data Protection using z/OS Mainframe Infrastructure Strengths

- One data protection solution for all your enterprise platforms
- A single solution that provides fast efficient data protection for all your enterprise platforms
- **Benefits:**
  - Centralizes management, scheduling, tracking and auditing
  - z/OS controls backup from Open Systems server to z/OS tape
  - Data reduction helps solve backup window issues
  - Efficient tape drive and tape media utilization
  - Meet Disaster Recovery time objectives
  - Strategic Direction to offload UPSTREAM CPU processing to zIIP engine(s)



# Centralized Enterprise Data Protection using z/OS Mainframe for DR

- Leverages your existing z/OS procedures and methodology
- Need floor system, z/OS tape management system and system/user catalogs
- Recover UPSTREAM started task & control files
- Decide what to restore
- Start restores



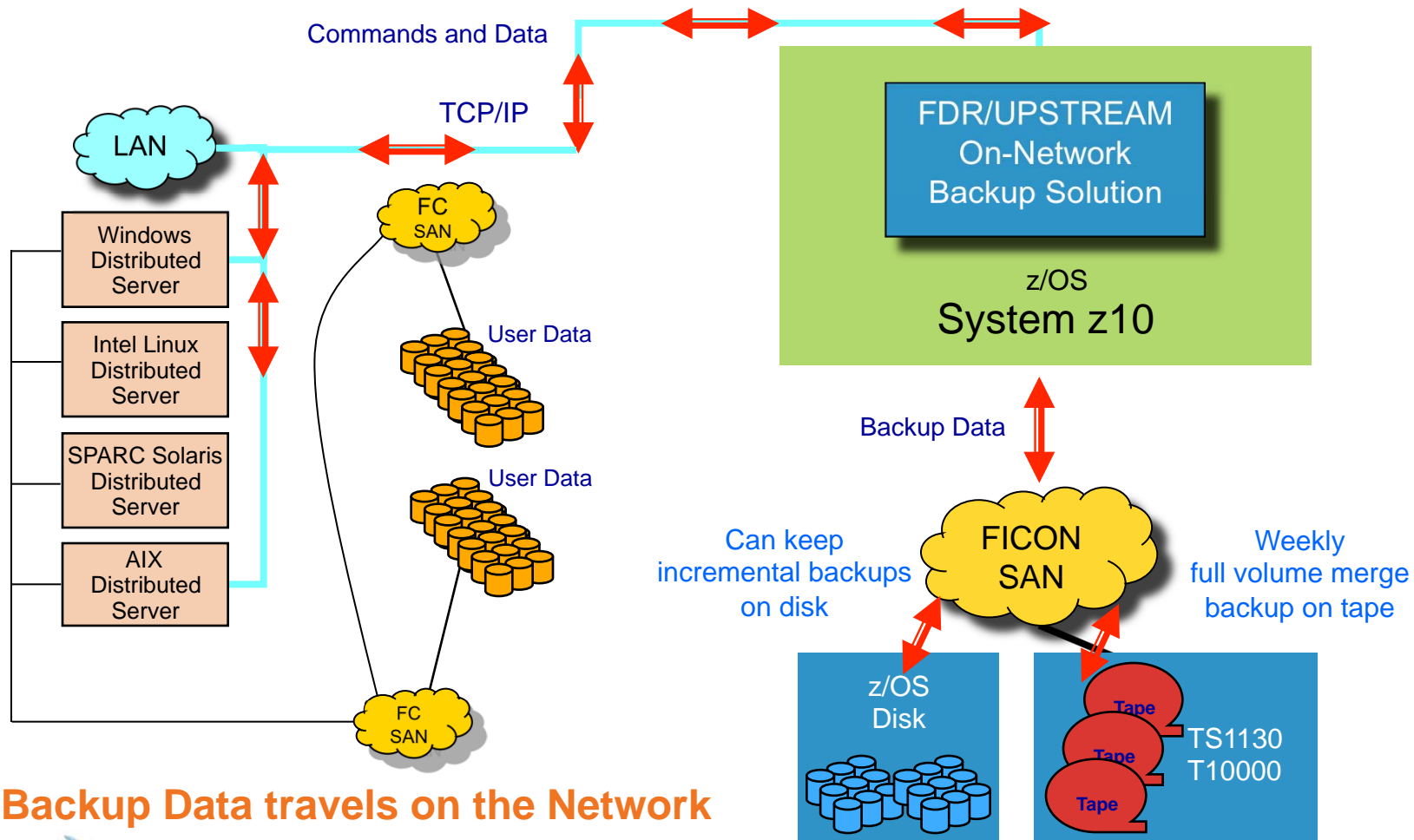


# FDR/UPSTREAM – On-Network Enterprise Data Protection

- Enterprise Data Protection for Linux on System z, AIX, Linux, UNIX and Windows storage
- Command Data, Metadata and Backup Data all Travel on the TCP/IP LAN or WAN Communications Network
- Centralized z/OS Tape Management, Security and Scheduling
- Online Database Support for DB2/UDB, ORACLE, LOTUS Notes, SQL Server, MS Exchange, etc.
- Provides Data De-duplication Methods, Synthetic Merge Backup, Block Level Segment Support



# How FDR/UPSTREAM® Works... On-Network Enterprise Backup



**Backup Data travels on the Network**



## FDR/UPSTREAM Clients

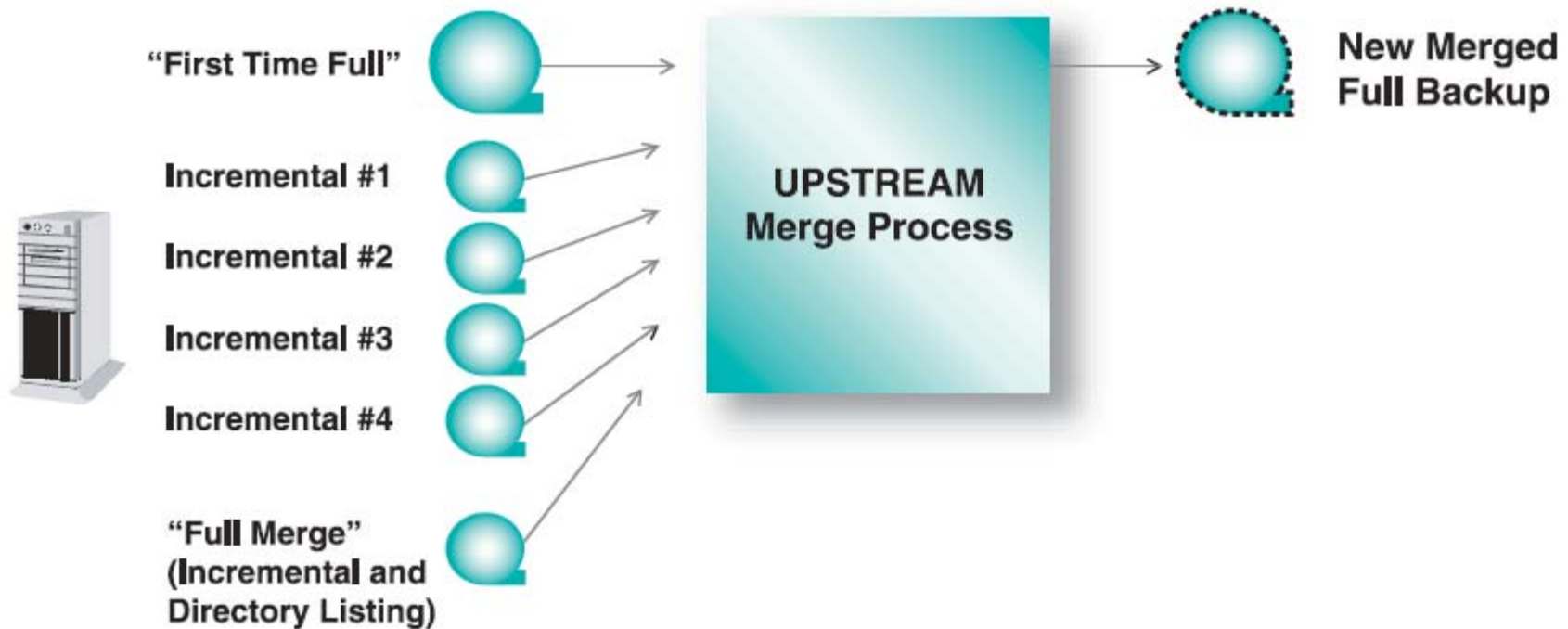
- Linux on System z
- z/OS UNIX
- Windows
- Novell NetWare
- Novell OES2 Linux
- IBM AIX
- VMware
- HP-UX
- Oracle Solaris
- Sun x86
- x86 Linux
- Linux for Power

***FDR/UPSTREAM can work in many virtualized environments including: • z/VM • VMware • Windows HiperV • XEN***





# Full Merge Synthetic Technology Minimizes Data Transmission



# Backup Types and Methods

- Sequential DASD and Tape – physical or Virtual Tape
  - First Time Full
  - Non-Merge (looks like an incremental)
  - Merge
    - Incremental
    - Full
      - *Deferred*
- Backups are keyed by Profile Name and version date
- Every backup consists of File Sets
- Backup parameters are specified in the UPSTREAM batch job input stream and communicated to the client.



# Backup Types and Methods

*Continued...*

- Combinations
  - All tape – Incremental and full backups to tape
  - Fulls on tape, incrementals on DASD
  - Deferred backups to DASD, merged to tape
- DASD Backup dataset migration considerations
  - Available space
  - Tape unit availability

## BATCH Interface – USTBATCH

- Initiate Backups, Restores, File Transfers, Run Functions, and *UPSTREAM commands*
- Can wait for process completion
- Can initiate multiple, concurrent operations
- Used as the interface to host based JOB scheduling systems (CA-7, OPCA, etc.)
- Allows return code checking and automation support
- Allows for the full specification or override of all workstation parameters
- Central point for all logging

## USTBATCH Utility

- Batch job requester to the UPSTREAM Started Task – does not communicate directly with the “target” Client
- Uses SNA/APPC LU6.2 conversation to the UPSTREAM started task
- Can wait to report on process completion via the “CONV=WAIT” parameter
- Supports “Restarts” and “Retries”
- Retrieves and logs the UPSTREAM started task log messages relevant to this request



# USTBATCH Utility

- Can initiate one or several processes from one batch job
- Use the ISPF panels to generate the USTBATCH job input parameters
- Can target a client by
  - IP address
  - UPSTREAM “Registered Name”
  - DNS name

# z/OS BATCH JOB for Backup

```
//USTBATCH      EXEC      PGM=USTBATCH
//STEPLIB       DD        DISP=SHR,DSN=USTSALES.UPSTREAM.LOAD
//SYSUDUMP      DD        SYSOUT=*
//USTLOG        DD        SYSOUT=*
//USTPARM       DD        *
  TARGNAME=LINUX1
  ACTION 1          *  BACKUP
  BACKUPPROFILE  ISYS
  MERGE 1
  STORAGETYPE 3    *  SEQ. TAPE
  COMPRESSLEVEL 0 *  NO COMPRESSION
  RESTARTTYPE 2   *  RESTART IF NOT COMPLETED
*
  SPEC /*
  SPECTYPE 0      *  INCLUDE
  SUBDIRECTORIES Y
  SPEC /opt/user/*
  SPECTYPE 1      *  EXCLUDE
  ENDPARM
/*
```



# FDRSOS – Off-Network Cross Platform Data Protection

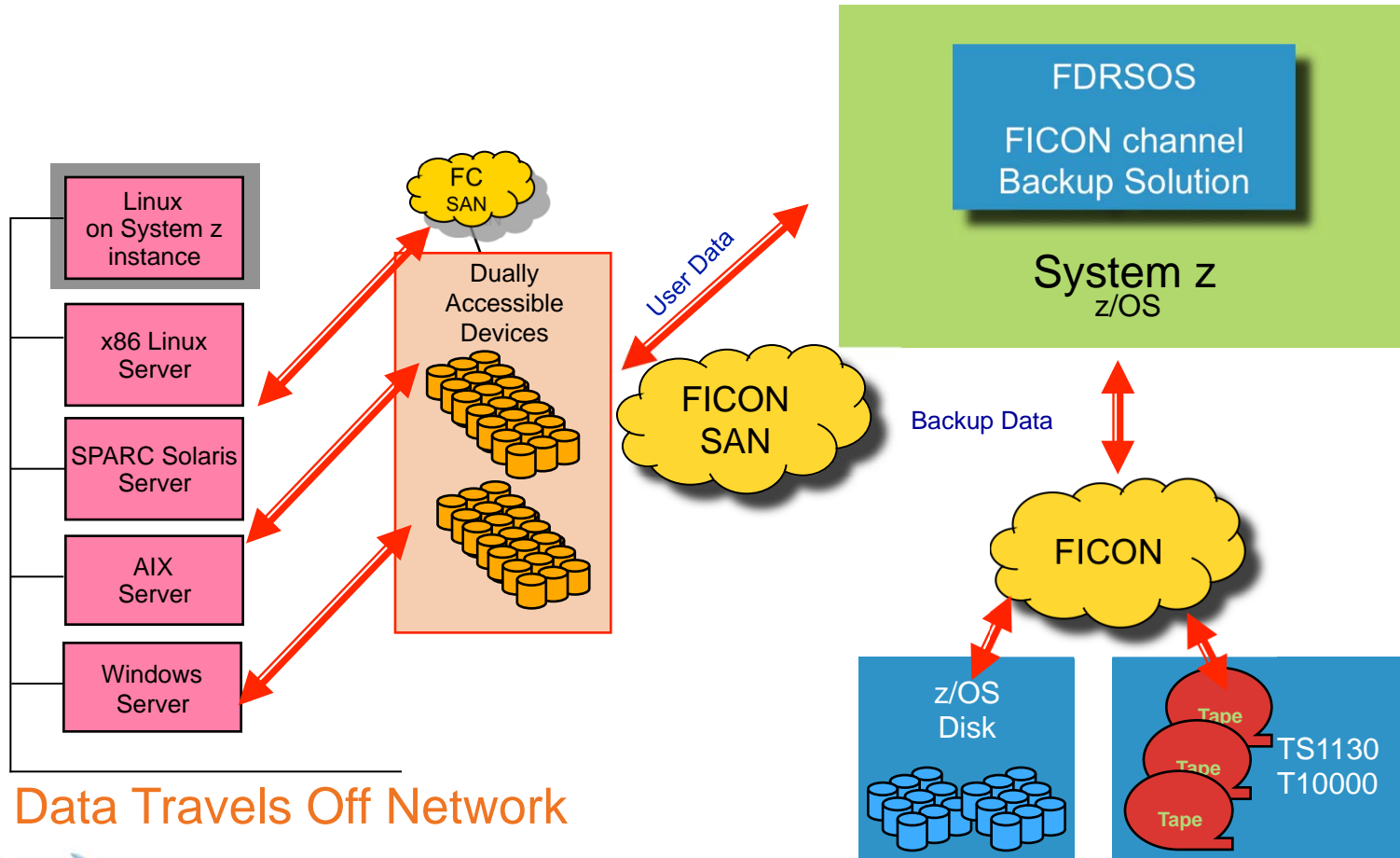
- High Speed FICON Image level backup/restore
- Takes the z/OS shared DASD concept and extends it to Open Systems platforms
- Supports IBM and EMC arrays
- Agent-less technology
- No network traffic...i.e. no TCP/IP usage
- Straight forward bare metal recovery of the Open Systems operating system and data files
- Data travels to z/OS storage devices across FICON channels
- Mainframe data protection for Linux, UNIX, Windows, etc.

## FDRSOS Supports...

- Backup and recover large amounts of Open Systems data in less time with no negative impact on corporate networks
- Production less constrained by a shrinking backup window
- Protect your current mainframe hardware and software investments by using **existing** mainframe resources to protect Linux on System z and distributed enterprise storage
- **Extend** mainframe security and automated operations to non-disruptive Open Systems data protection
- Well-suited for large databases
- Makes duplicate offsite copy simultaneously
- Supports TimeFinder & FlashCopy volumes



# How FDRSOS® Works... Off-Network Agent-less Backup Solution



Data Travels Off Network







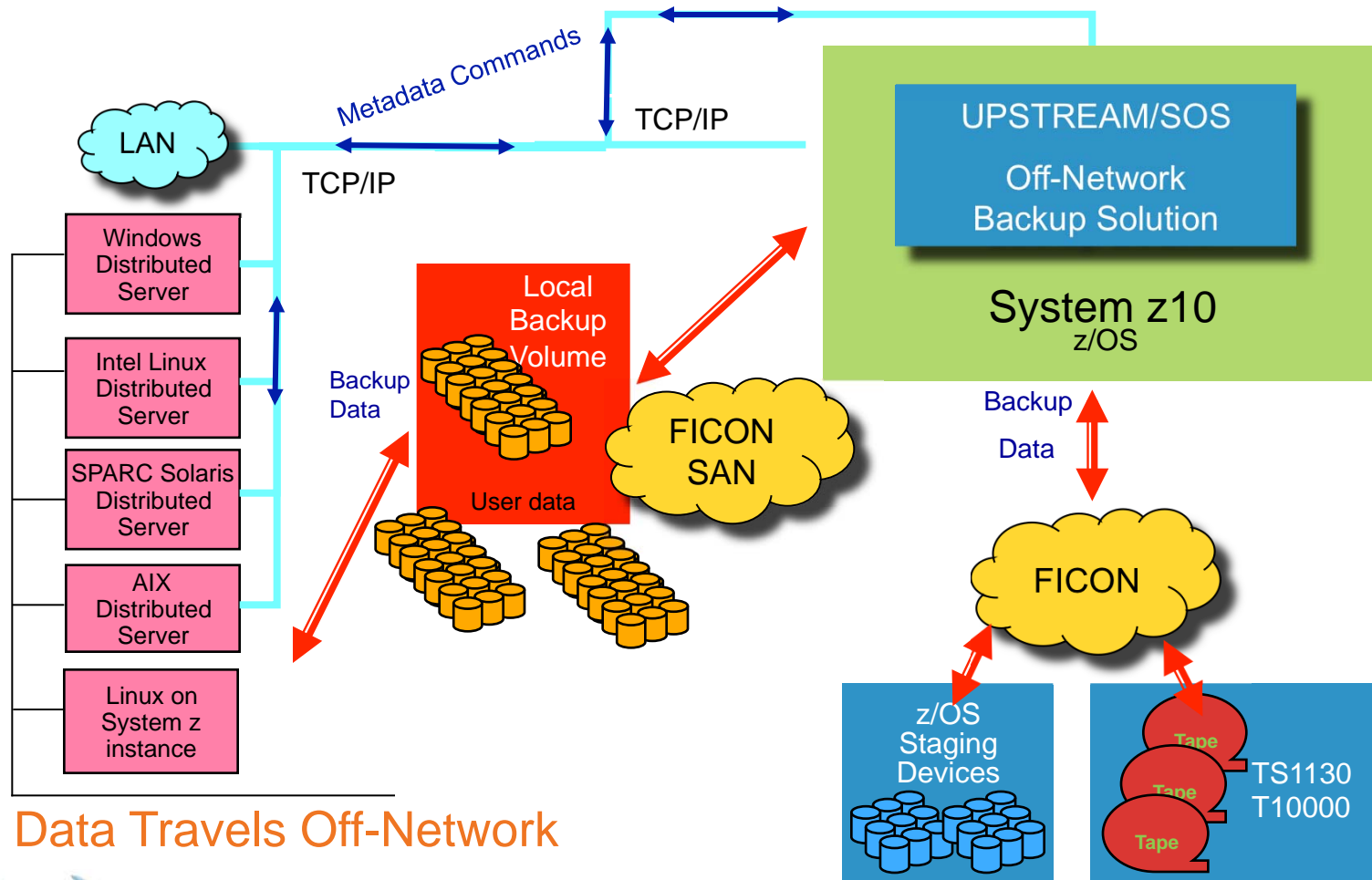
## UPSTREAM/SOS® – Off-Network, File Level Data Protection

- Provides high speed FICON data connection between Open Systems data and z/OS storage devices
- Only UPSTREAM Metadata travels on the TCP/IP network
- **All data transfer travels off-network**
- Enterprise Data Protection at the file level for AIX, Linux, UNIX, Windows and Linux on System z
- Supported by all UPSTREAM clients (except z/OS UNIX)



# UPSTREAM/SOS

## Storage Controller Data Transfer



# GREEN...

## Performance (MB/s) and Efficiency (MIP/MB)

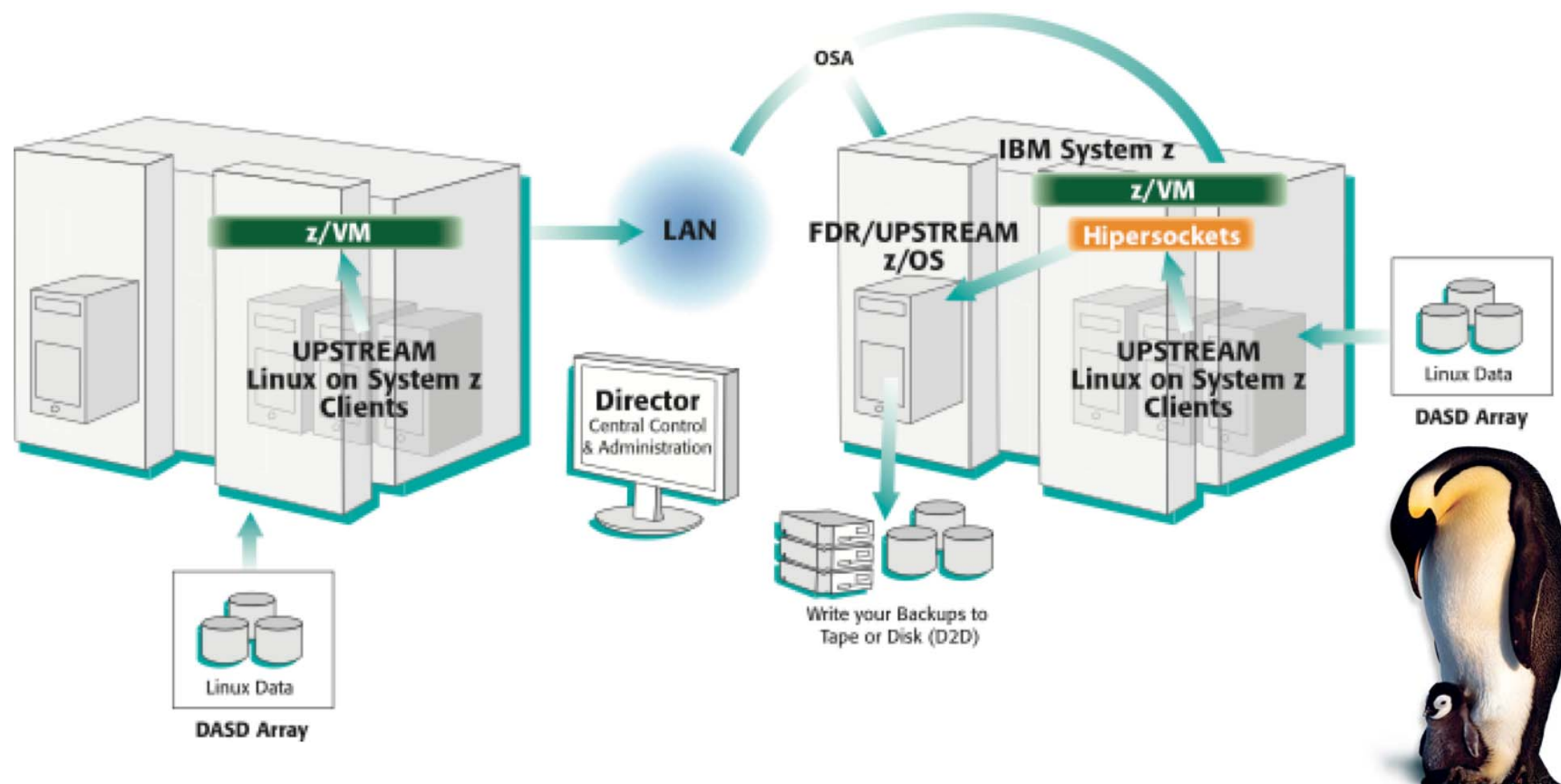
UPSTREAM TCP/IP Network Performance limited by 100 mb connection	MB Calc	MB/Sec Calc	CPU MIP/MB (in millions) per MB Task Only	Mainframe CPU Seconds	Elapsed Time
TCP/IP Avg 16GB 16 Files	16,384	8	1.23	157.5	34.4
TCP/IP Avg 16GB 65,536 Files	16,384	7	1.24	158.6	36.0

UPSTREAM/SOS Off network transfer disk	MB Calc	MB/Sec Calc	CPU MIP/MB (in millions) per MB Task Only	Mainframe CPU Seconds	Elapsed Time
UPSTREAM/SOS Avg 16GB 16 Files	16,384	34	0.35	44.3	7.9
UPSTREAM/SOS Avg 16GB 65,536 Files	16,384	21	0.45	59.3	13.3

FDRSOS Cross Platform Access	MB Calc	MB/Sec Calc	CPU MIP/MB (in millions) per MB Task Only	Mainframe CPU Seconds	Elapsed Time
FDRSOS Avg 16GB 16 Files	18,432	93	0.12	17.0	3.3
FDSOS Avg 1616GB 65,536 Files/65	18,432	91	0.12	17.0	3.3

...MILEAGE MAY VARY

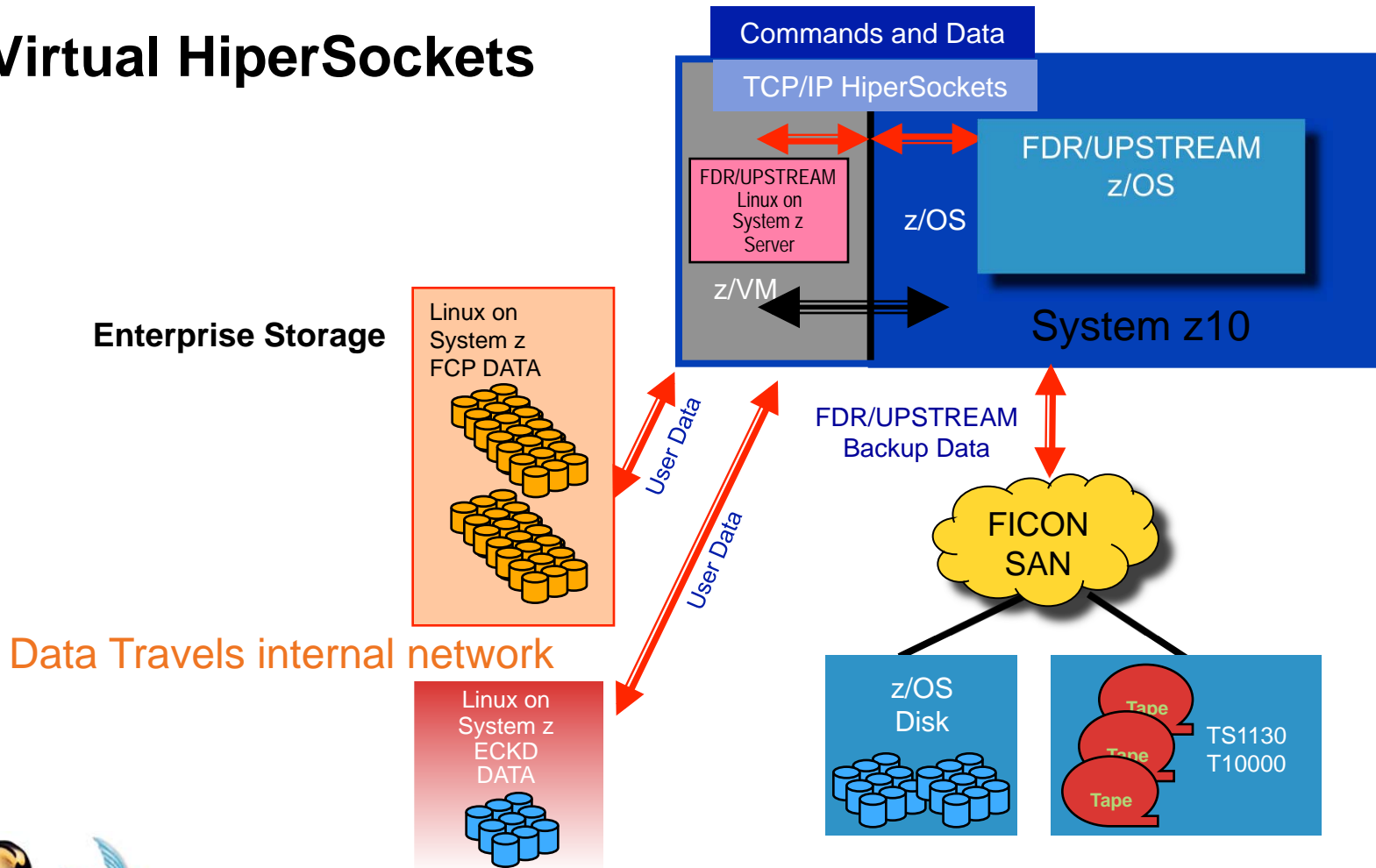
# UPSTREAM Linux on System z



# FDR/UPSTREAM for Linux on System z Internal Network Backup



## Virtual HiperSockets



## UPSTREAM Linux on System z

- SuSE and Red Hat Linux on System z in z/VM or LPAR
- A z/OS server-based backup solution
- Employs Centralized control using existing z/OS tape management, scheduling, automation, and security systems
  - Extensive database and file system support
  - Admin, user and z/OS Batch interfaces
  - De-duplication/data transfer reduction
  - Familiar z/OS operational procedures
  - File level Client support
  - “Rescuer” File level Bare-metal recovery





## UPSTREAM Linux on System z

- Exploits HiperSockets connectivity
- On-Line Database Agents for Oracle, DB2, and Domino
- Extensive file system features:
  - Hard links, symbolic links, owners,
  - Reiser, ext2, ext3, GFS, XFS, NFS, MAPFS, GPFS, single file system support, etc.
  - Support for Extended attributes and Security Enhanced Linux (SELinux)
  - Support for user specified file systems





# System z Cross Platform Access Benefits Summary

- Totally integrated Business Solution for Enterprise Data Protection built on z/OS proven technology.
- Non-disruptive backup solutions supporting continuous business operation for Linux on Systems z and Open Systems data platforms.
- Extends investment in existing mainframe resources; CPU, ATL tape silos VTL, tape drives and staff.
- Protects the customers software investment by utilizing existing tape library management software.
- Extends mainframe security to non-disruptive Open Systems data protection.
- Extends mainframe automated operations to non-disruptive Open Systems data protection.

Trademarks and statements

FDR, FDRSOS, SOSINSTANT, FDR/UPSTREAM and UPSTREAM/SOS are service marks, trademarks or registered trademarks of Innovation Data Processing Corporation. EMC, DLm, SYMMETRIX VMAX, and DLm and TimeFinder are trademarks or registered trademarks of the EMC Corporation. IBM, z/OS, ProtecTIER, zDDB, FlashCopy, System z and FICON are trademarks or registered trademarks of International Business Machines Corporation. All other service marks, trademarks or registered trademarks are the property of their respective owners.



**CORPORATE HEADQUARTERS:** 275 Paterson Ave., Little Falls, NJ 07424 • (973) 890-7300 • Fax: (973) 890-7147  
**E-mail:** [support@fdrinnovation.com](mailto:support@fdrinnovation.com) • [sales@fdrinnovation.com](mailto:sales@fdrinnovation.com) • <http://www.innovationdp.fdr.com>

<b>EUROPEAN OFFICES:</b>	FRANCE 01-49-69-94-02	GERMANY 089-489-0210	NETHERLANDS 036-534-1660	UNITED KINGDOM 0208-905-1266	NORDIC COUNTRIES +31-36-534-1660
--------------------------	--------------------------	-------------------------	-----------------------------	---------------------------------	-------------------------------------