IDA based Virtual Appliance for Secondary Storage

Giridhar Lakkavalli / Subramani Nallusamy
MindTree Ltd

Growth of Digital Data

- An IDC reports states the following
  - In 2007, the amount of information created, replicated etc was 281 exabytes
  - In 2011, this is projected to grow to 1800 exabytes, a compound annual growth of almost 60%.
  - Out of the 1800 exabytes, about 800 exabytes is expected to be information that needs to be stored, the rest is transient data
  - Out of this the information emanating from an enterprise is seen at 35% of the entire data created, this would be around 300 exabytes in 2011

Enterprise Data - Types

- The data in an enterprise includes
  - Documents
  - Source Code
  - Mails
  - Etc…

Enterprise Data - Classification

- Primary Data
  - Data that is actively being worked on
  - Stored on the fastest, most expensive drives
- Secondary Data
  - Data that is older
  - Is referenced for e-discovery / recovery needs

Data Protection - Need

- An enterprise adopts various protection strategies to ensure that information is available even in the event of catastrophes like
  - Hardware failures
  - Site shutdown
- The enterprise also needs to protect the data for compliance requirements like
  - HIPAA
  - Sarbanes-Oxley

Data Protection – Technologies

- Some of the technologies that provide data protection are
  - Backup & Recovery
    - For recovering from disasters
  - Archival
    - To store important documents for e-discovery
  - Replication
    - A technology that works in conjunction with the above 2 technologies to create multiple copies

Data Protection - Technologies

- Data is also protected using the following hardware technologies
  - RAID (0, 1,3, 5 and 6)
  - Tape

Data Protection - Issues

- Some of the current issues we see with Data Protection are
  - Cost
  - Complexity
  - Recovery limitations

Data Protection - Cost

- Data Protection is costly
  - The software / hardware / appliances that provides data protection is not cheap
  - RAID technologies provide protection but reduce the amount of addressable storage
  - Tape is cheaper when compared to disks, but there are certain disadvantages

Data Protection - Complexity

- Data Protection setups are complicated and consist of
  - Backup Solution
  - Archival Solution
  - Replication

  It will help even if we eliminate one element of this complexity

Data Protection – Recovery Limitations

- RAID rebuild in the event of a failure
  - We currently have 1TB disks, if a RAID system with a 1TB disk fails, the rebuild time typically runs into tens of hours. The probability of another disk failure in the rebuild time is a real possibility

Data Protection – Reduction of Cost

- De duplication is seen as a technology that reduces the data protection cost
  - This technology basically identifies duplicate segments in the data and stores only one instance of these duplicates
  - This technology is seen as being able to reduce the data stored by at least 50%

Data Protection – Answers to current limitations?

- Are there are other solutions other than
  - De-dupe
  - Usage of open source data protection software

  To reduce the overall cost of Data Protection

Data Protection - A better solution

- There is a new technology that promises overall lower cost and better protection
  - This technology is called IDA

IDA – a brief introduction

- IDA or Information Dispersal Algorithm (en) codes and disperses the given chunk of data into slices. The chunk can be reconstructed from any subset of slices.
- IDA variants : Reed Solomon (RS), Cauchy RS, Rabin
- The advantages of a system built using IDA are
  - Withstand multiple (storage node/disk) failures
  - Secure, Reliable, Cost Effective
  - Based on dispersal and achieves the DR functionality as by-product

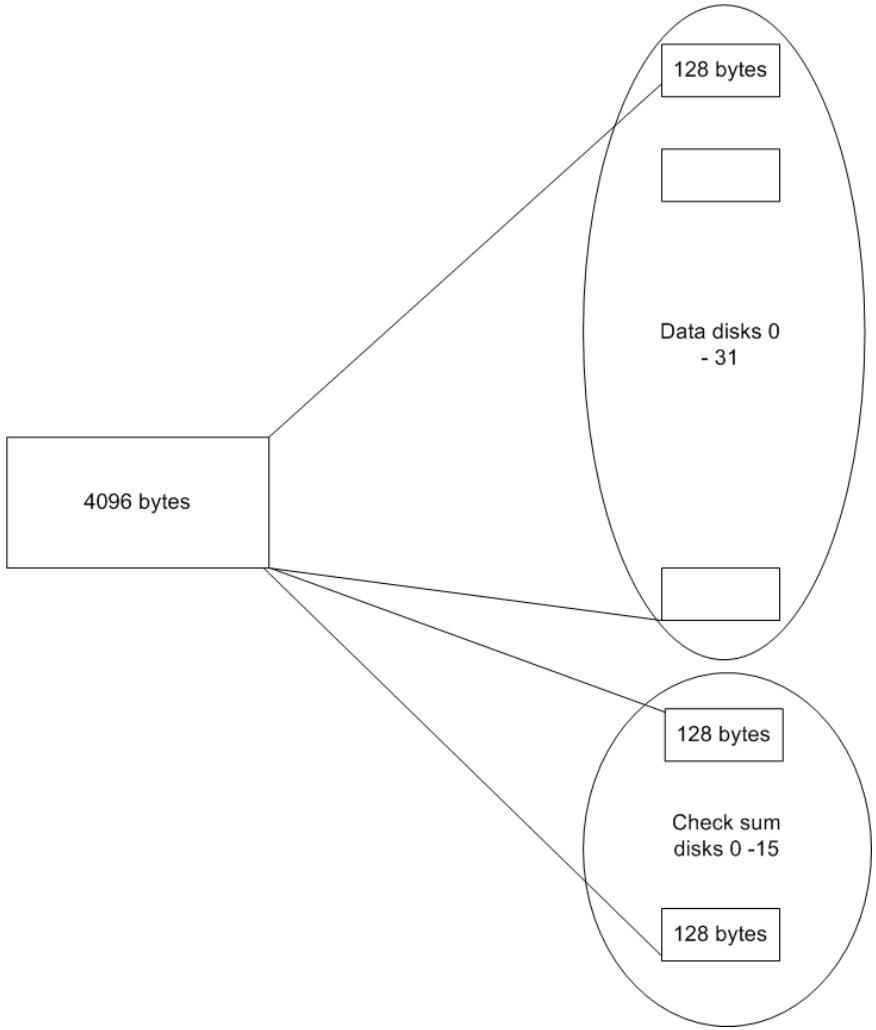
IDA – a brief introduction

- An IDA based architecture has the ability to with stand multiple hardware failures (erasures), this can be configured to say 16 failures
  - As an example, a system can be configured to have 32 Data Disks and 16 check sum disks giving a total of 48 disks
  - In this system, the original data block is split into 32 data disks.
  - The slices are processed to generated check sums that are stored on the 16 check sum disks
  - The original data can be reconstructed using any 32 disks of the 48 disks in the system

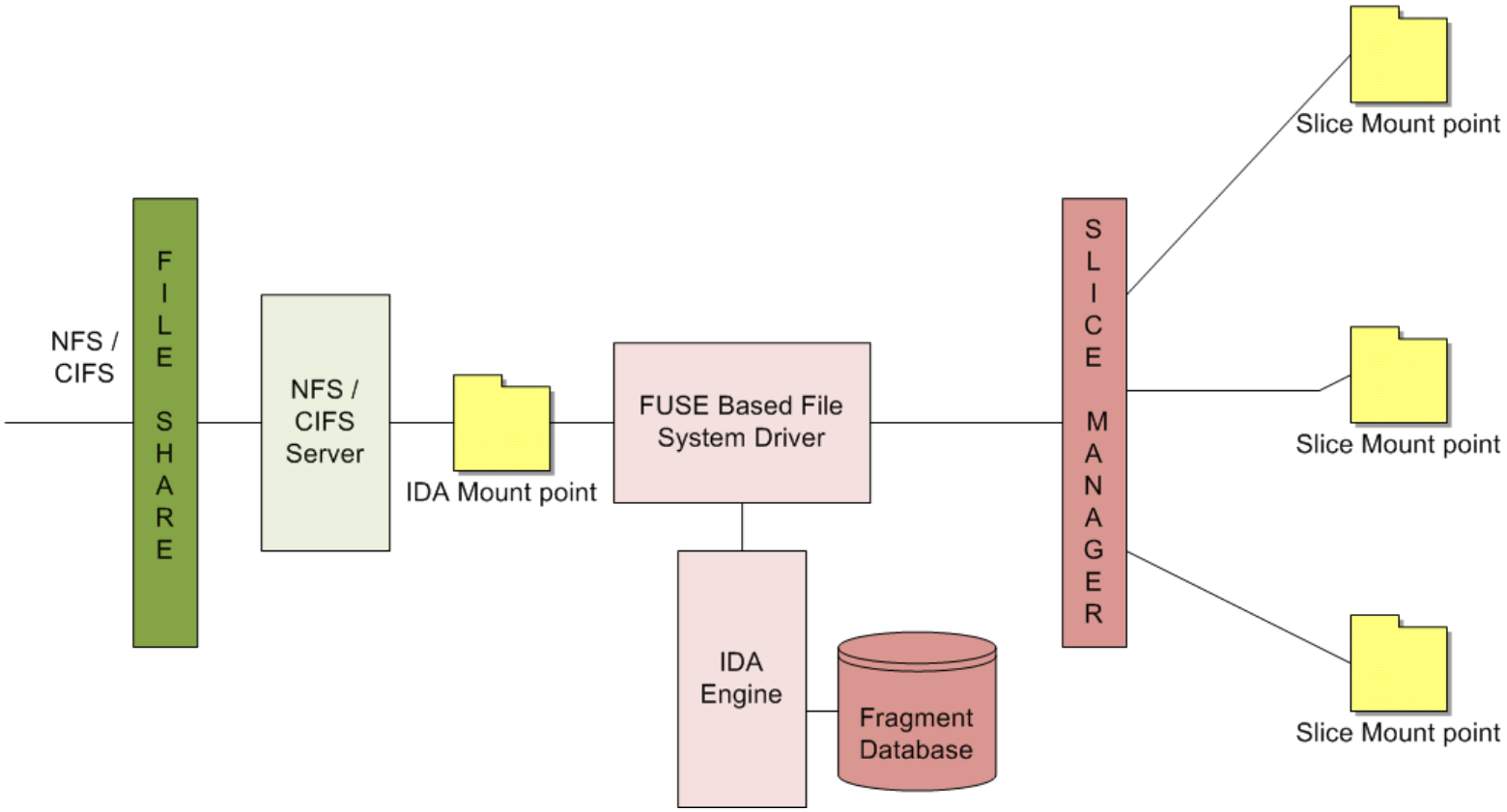
IDA – the algorithm

- If we consider devices D1, D2 …Dn as the data devices and devices C1, C2..Cm as the checksum devices,
  - C1 = F1(D1,D2….Dn)
  - C2 = F2(D1,D2…Dn)
  - ..
  - Cm = Fm(D1,D2…Dn)

IDA – the algorithm



IDA Appliance Architecture



IDA Appliance Architecture

- The components of the appliance includes
  - A Linux distribution with support for CIFS / NFS
  - FUSE
  - IDA Engine in Software
  - Custom File System Driver that integrates with the IDA Engine on the Read / Write paths
  - A database to store the relationships between the file chunks and the slice locations

IDA Appliance Architecture

- The Appliance does the following
  - Provides a File System Interface (CIFS / NFS) as an end point for Backup & Recovery or Archival Software
  - Slices the data it receives and spreads it across all the Slice Mount points
  - Reconstructs the data by accessing the quorum number of slices
  - Maintains the mapping of the data to the slices in a database

IDA Appliance Architecture

- Database
  - Need to store the relationship between the file chunks and the file slices
  - The file slice location is referenced as a 64 bit number
  - The database should handle addition / deletion of rows
  - The database should handle queries for files and mapping rows
  - If it is a multisite deployment, it should be easy to synchronize the databases between the sites

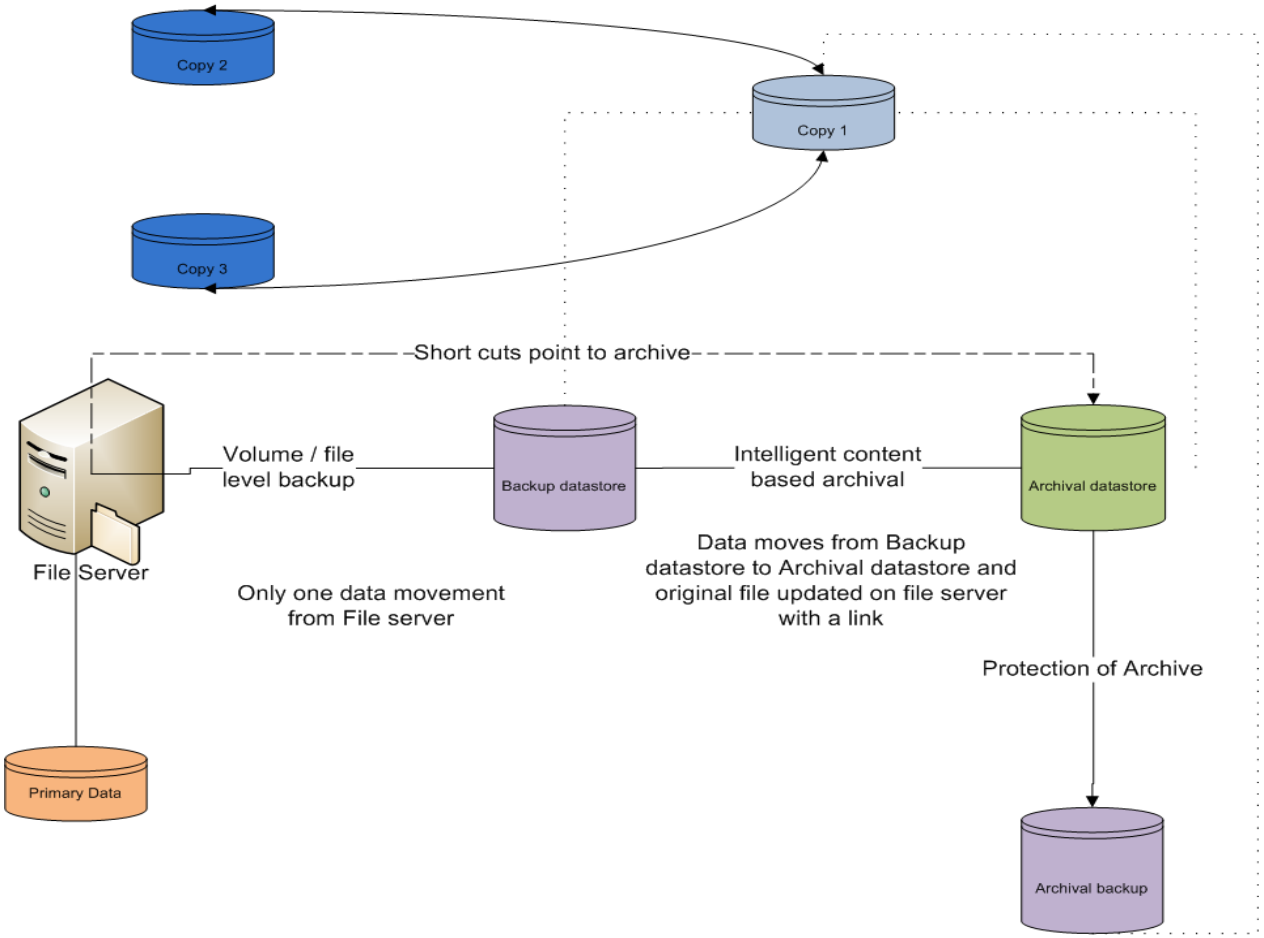
IDA Appliance Architecture

- Database (contd..)
  - Also, in the multi site deployment where the inserts can happen at any location, the next row identifiers should be unique across all locations
- The overhead per slice is 16 bytes.
  - 8 bytes for the slice location
  - 8 bytes for the file identifier

IDA – Why as an appliances?

- Packaging the entire solution as an appliance helps as
  - To control the content
    - OS
    - Database
    - Utilities
  - Deployment is easier
  - Patching / maintenance is easier
  - Making it as a virtual appliance, means that there is no new hardware to be purchased as well

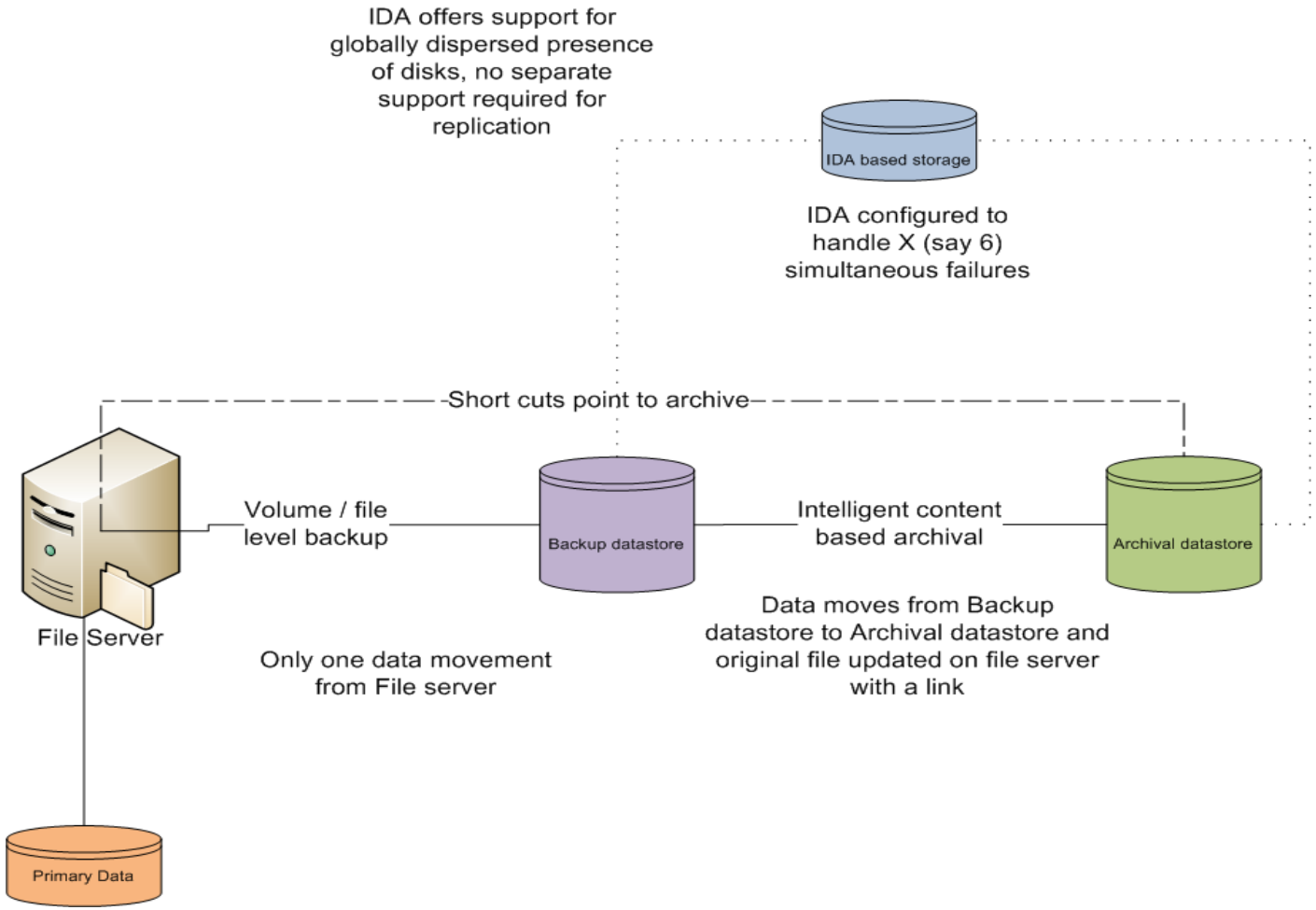
Current Backup / Archival / Replication Architecture



Current Backup / Archival / Replication Architecture

- As seen in the diagram
  - Primary data is protected by a Backup technology.
  - The backup image is typically also replicated
  - Some of backup images would be deleted after a point in time (say 90 days etc..)
  - A portion of the data in the backup images may be archived for longer (compliance requirements)
  - The archive itself will again need to protected by a backup technology and also replicated

IDA based Backup / Archival / Replication Architecture



IDA Deployment Scenarios

- Single Site
  - All the backup / archival data is at a single site in the Enterprise
- Multi Site
  - The Backup / archival data is spread across at least 3 sites

IDA Deployment Scenarios – Single Site

- In single site deployments, IDA can give much better protection against data loss when compared to RAID 5 or RAID 6 at a lesser cost
- The ability to configure the number of check sum disks, allows the user to choose 3, 4 or any number of check sum disks
- An example could be 12 data disks and 4 check sum disks

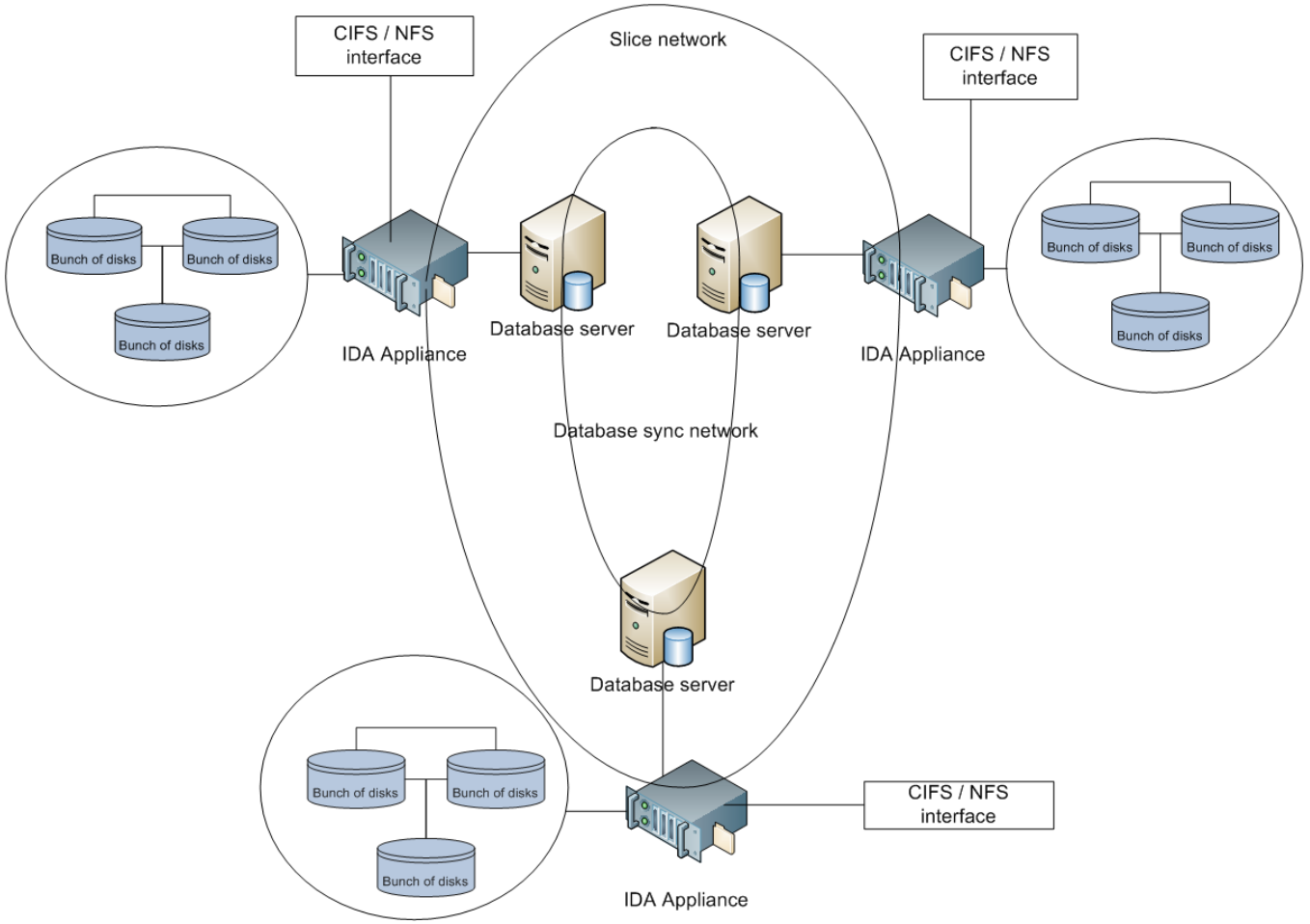
IDA Deployment Scenarios – Multi Site

- In Multi site deployments of IDA, the unique factors are
  - The ability to recover from an entire site going down
  - The ability to recover data with out hitting all the sites where the data is stored
  - An example could be 32 data disks and 16 check sum disks. The data can be recovered if any of the 32 disks are available
  - If we choose 3 sites, then each site contains 16 disks, the recovery is possible from 2 sites only

IDA Deployment Scenarios – Multi Site

- IDA based architecture
  - Lends itself well to distributed storage and handles replication as well, hence so separate cost for replication
  - Can be configured such that data can be retrieved even a site goes down

IDA deployment architecture



IDA Deployment Architecture

- The deployment architecture shows
  - An IDA deployment where there the data is sliced up and stored across disks spread across 3 locations
  - The data can be written into the IDA system or retrieved from any of the 3 locations
  - There are 2 internal networks
    - Slice network – For storage of the data slices
    - Data base sync network – For syncing the mapping of data to slices

IDA Deployment Architecture

- The slice network will be built on NFS / CIFS protocols
- For the database, we require something that can work in a distributed manner and also not be very complex to manager and Cassandra is our current choice. The database network consists of a network of Cassandra servers

Advantages of IDA

- Reduction of Complexity
  - With IDA, replication is inbuilt so there is one less component in the Data Protection Architecture
  - IDA basically combines Replication and RAID into one single solution

Advantages of IDA

- Security
  - Since the original data is split into slices, even if a set of disks or a site get compromised, the original data cannot be reconstructed without the quorum slices
  - You can add encryption on top of the data slices and make it even more difficult to decipher the data

Advantages of IDA

- Configure level of protection
  - The IDA system allows the admin to choose the number of disk failures that the system should with stand
  - The ability to withstand multiple failures (say 6) reduces the probability of permanent data loss during rebuild time

Advantages of IDA

- Cost
  - Compared to technologies like RAID, the cost / TB is cheaper with IDA for a higher level of protection

Parameter	Cost in USD
1 TB Raw disks	400
1 TB for RAID 5 (3 Data + 1 parity)	533
1 TB for RAID 5 + 1 copy	1066
1 TB for RAID 5 + 2 copies	1566
1 TB for IDA – 14 Data disks + 6 Check sum disks	640
1 TB for IDA – 32 Data disks + 16 Check sum disks	900

Disadvantages of IDA

- We are looking at using off the shelf JBOD's, so a management solution that can work across sites needs to built
- Though it comes across as a better alternative to RAID and makes secondary data management easier, there is no backing from any of the big storage companies, this will make adaptation more difficult

Issues to consider

- Uninterrupted Connectivity at least between the sites
- Since the sites are expected to be geographically separated, the read time would be higher.
  - Since the solution being proposed here is for backup / archives which are typical secondary storage solutions, read time can be higher, we can address it to a certain extent using read ahead
- In order to handle the issues of all the sites not getting updated on a write, we will need to maintain a local slice store

Issues to consider

- If an IDA system has been setup to handle a certain number of failures, it is difficult to change that configuration
- The Database should also be configured to handle site level failures

References

- http://www.cs.utk.edu/~plank/plank/papers/FAST-2005.pdf
- http://apache.cassandra.org
- http://fuse.sourceforge.net
- http://www.cleversafe.org/dispersed-storage/idas
- http://www.enterprisestorageforum.com/technology/features/article.php/3839636?comment=16804-0

Q&A