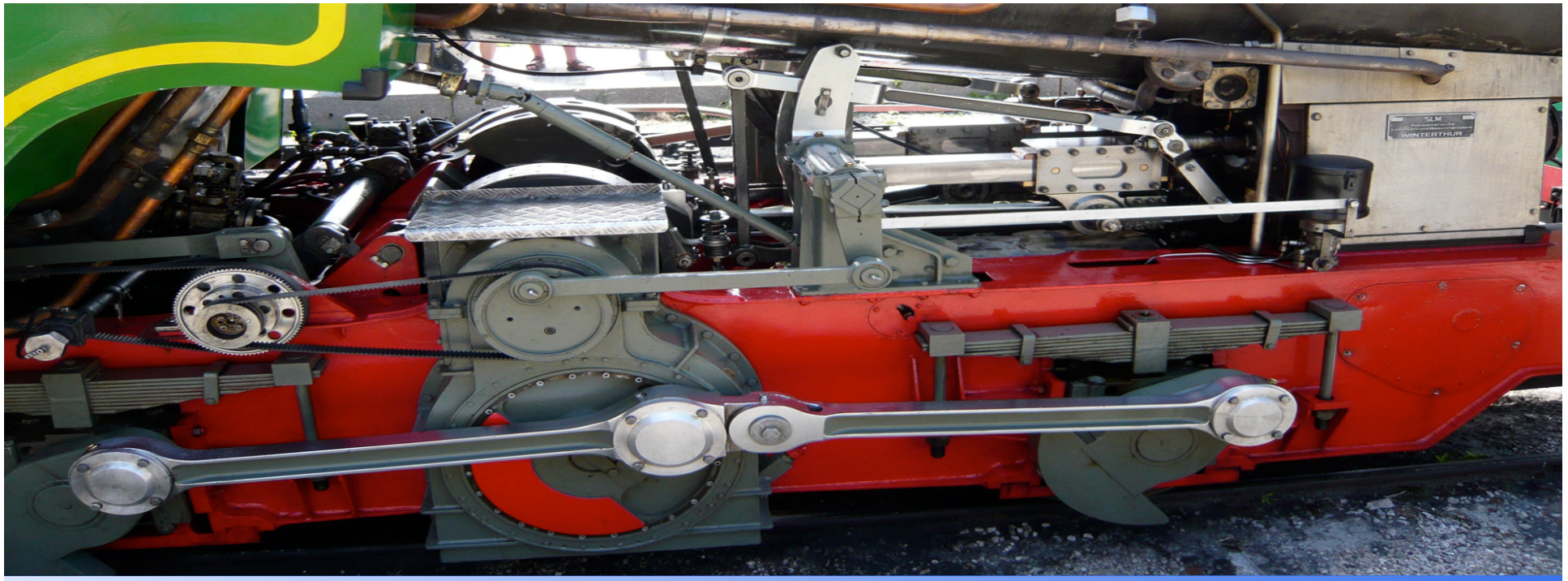


Create a backup site with opensource

Fabrizio Manfredi Furuholmen

Beolink.org

Agenda

- **Introduction (Part I)**
  - Disaster Recovery Plan
  - Business continuity Plan
- **Disaster Recovery Components (Part II)**
  - Starting point
  - Software list
  - Design Pattern
- **Case Studies (Part III)**

Disaster Statistics & Potential Implications

Beolink.org

- 90% of businesses that lose data from a disaster are forced to close down within 2 years since the disaster.
- 80% of businesses without a well structured recovery plan are forced to close down within 12 months since the flood or fire.
- 43% of companies experiencing disasters never recover.
- 50% of companies experiencing a computer outage will be forced to close down within 5 years
- Companies experiencing a computer outage lasting longer than 10 days will never recover its full financial capacity.
- Less than 50% of all organizations in the UK have a business continuity plan or disaster recovery plan.
- One out of 500 data centers experience a severe disaster every year.

What is a Disaster Recovery Plan ?

Beolink.org

Disaster recovery plan (DRP)

Disaster recovery plan consists of the precautions taken so that the effects of a disaster will be minimized and the organization will be able to either maintain or quickly resume mission critical functions

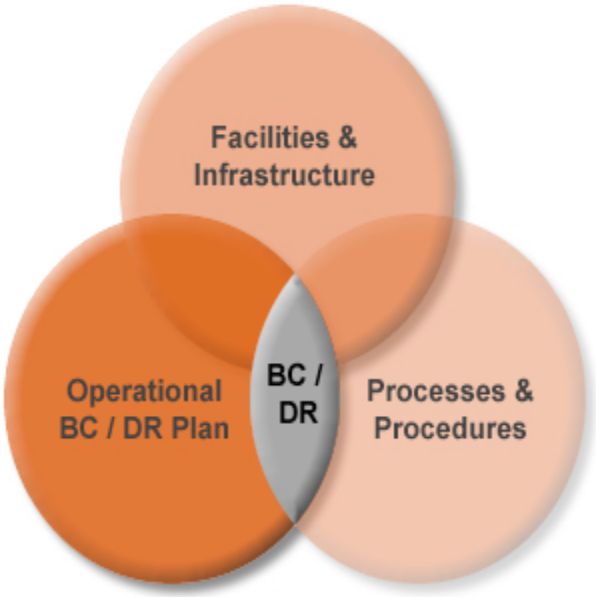
Classification of Risk

- **Natural disasters**
  - Such as tornadoes, floods, blizzards, earthquakes and fire
- **Accidents**
- **Sabotage**
- **Power and energy disruptions**
- **Service sector failure**
  - Such as communications, transportation …
- **Environmental disasters**
  - Such as pollution and hazardous materials spills
- **Cyber attacks and hacker activity**

BIA

- How many transactions can be lost without significantly impacting revenue or productivity?
- Does a majority of the businesses depend upon one or more mission critical applications?
- How much revenue per hour would be lost when these critical applications remain unavailable?
- Are there periods of time, for example the end of fiscal quarters or holiday seasons, when an outage would cause a greater disruption?
- How will productivity be affected if critical applications become unavailable?
- In the event of an unexpected outage, how will your partners, vendors and customers be affected ?
- Historically, what has been the total cost of lost productivity and revenue during downtime?

Disaster recovery planning (DRP)
is a subset of a larger process known as Business Continuity Plan.



Business Continuity Plan (BCP)
A business continuity plan enables critical services or products **to be continually** delivered to clients.

Technology problem ?

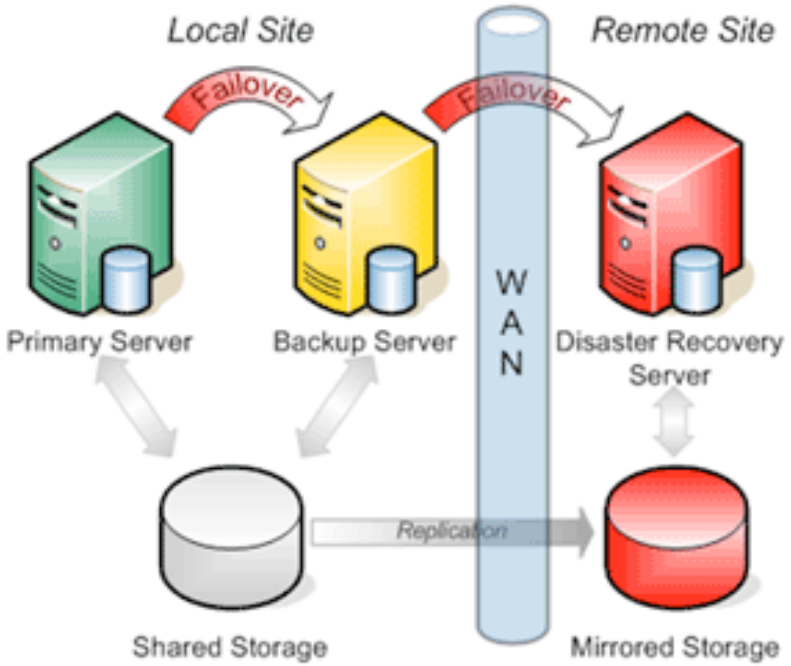
Beolink.org

Disaster recovery starting point

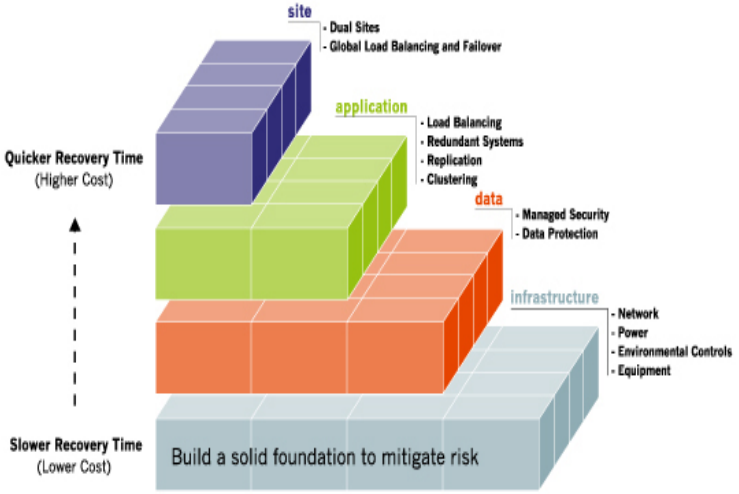
- **Recovery point objective (RPO)**
  Maximum data you can loose

- **Recovery time objective (RTO)**
  The amount of time you need to restart from the RPO

- System Duplication
  - Hardware
  - Network
  - Operating System
  - Application
- Data Replication
  - Filesystem
  - Database
  - Application Data
- Switch Procedure
  - Activation
- Rollback
  - How to switch back
  - How to merge data

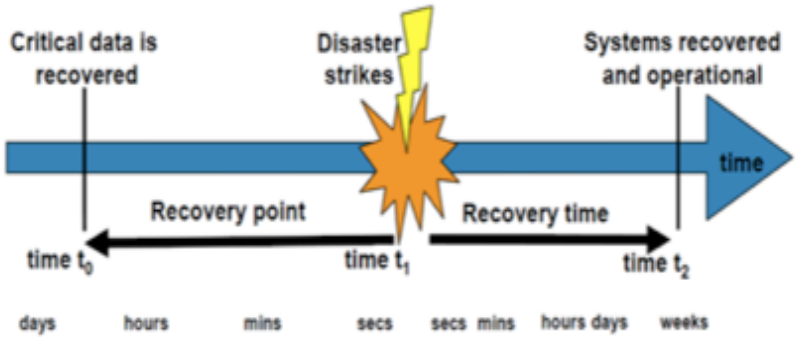


Disaster Recovery Components



Elements

- ❑ Routing
- ❑ Application
- ❑ Data
- ❑ Infrastructures



Working mode

- ❑ Off-line
- ❑ Near on-line
- ❑ Online

Type	OS	Database	Application
Off Line	OFF	OFF	OFF
Near Online	ON	OFF/ON	OFF
Online	ON	ON	ON

Several minutes

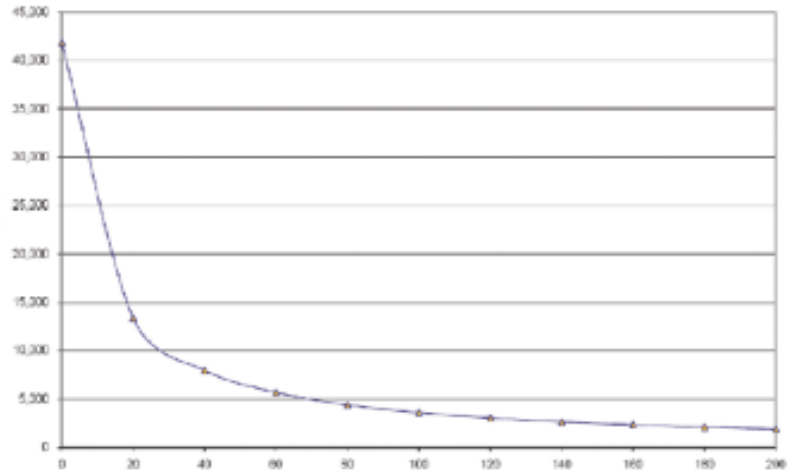
Low Cost

High Cost

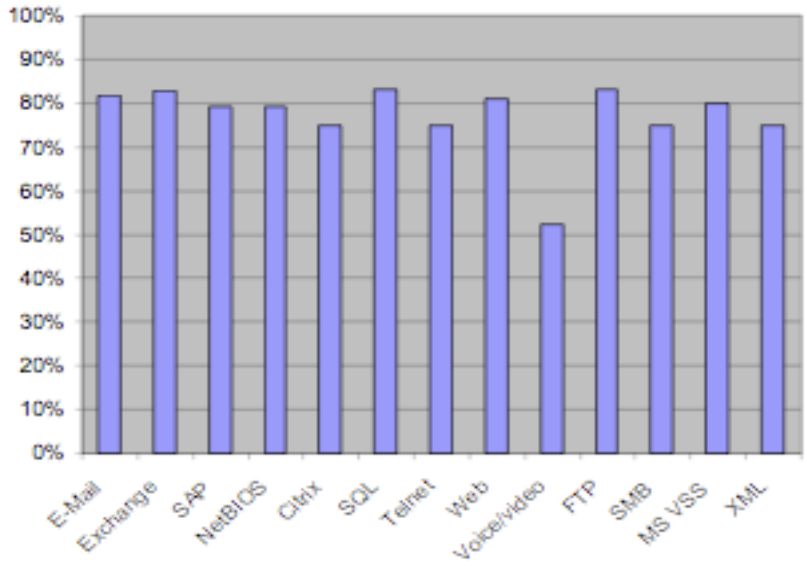
Zero

Keep your DR site unreachable until switch

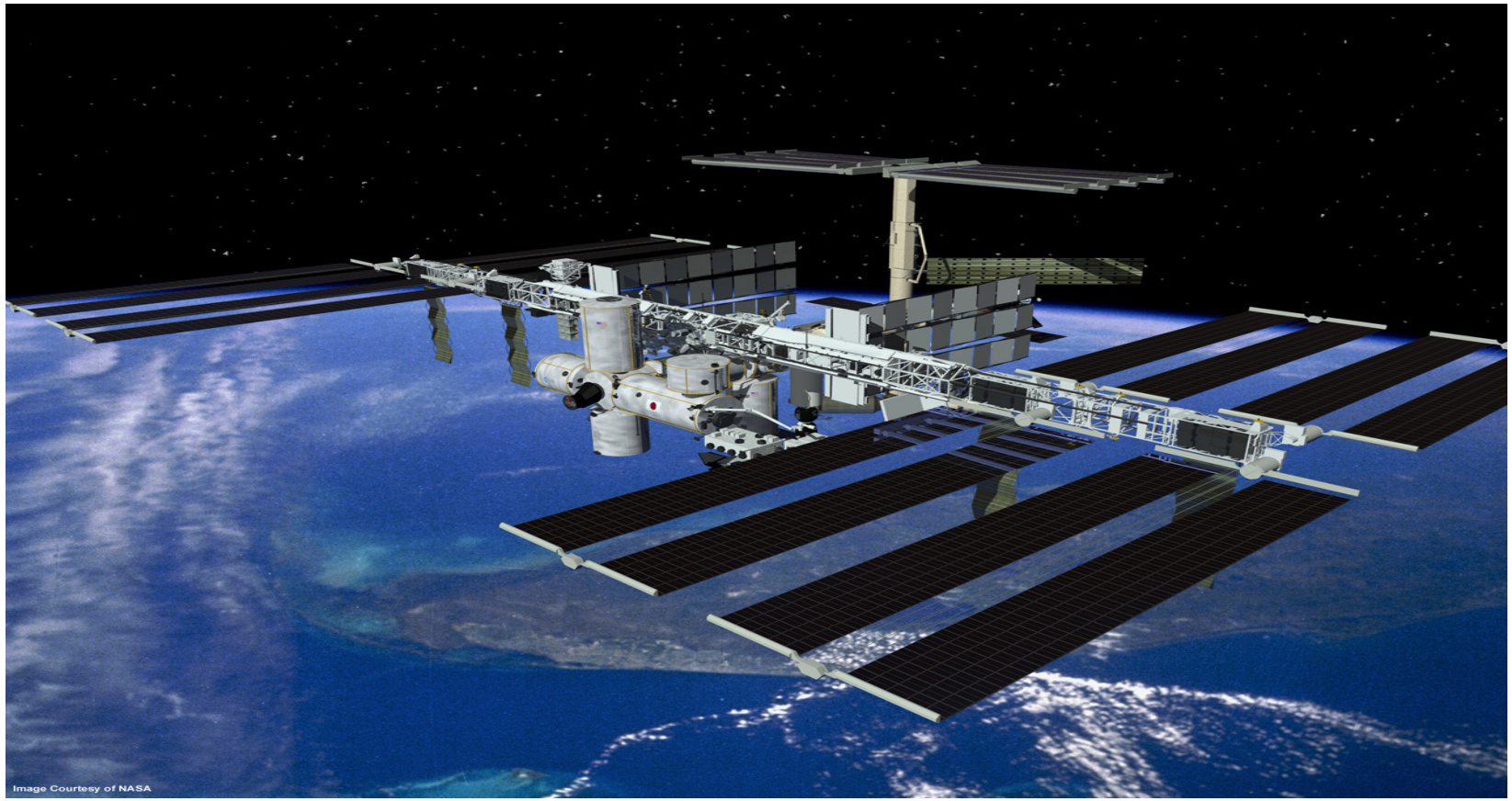
- Latency (speed of light)
  - Asynchronous operation
  - MultiStream/Parallel



- Bandwidth
  - Caching
  - Compression (AVG 60%)
  - Diffs transmission



What do you need ?



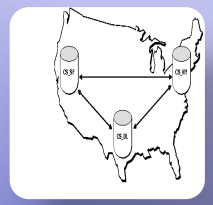
Infrastructure Software

Componets	Software
Virtualization	Vmware, XEN, KVM, Cloud computing, jeos
Firewall	Pfsense, iptable
System Configuration	Cfengine, puppet
Snapshoot	ZFS,LVM
Montoring	Nagios, Zenoss, Zabbix

Data Layer

Off-Line

- Amanda, Bacula, tar : incremenetal, os dump



Near on-line

- rsync: compression, incremental

On Line

- GlousterFS, openAFS, DB Replication

Tape Backup — **Days**

Periodic Replication — **Hours**

Asynchronous Replication — **Mins**

Synchronous Replication — **Secs**

Data Layer: Amanda

Compression
- Client Fast/Best
- Server Fast/Best

Multi Stream:
- Number of parrells dump

Diskbase
- Holding disk
- changer file base

Bandwith
- Global
- Per interface

Encryption
- SSH
- KRB5

Amanda.conf:

```
inparallel 4          # maximum dumpers that will run in parallel (max 63)
                      # this maximum can be increased at compile-time,
                      # modifying MAX_DUMPERS in server-src/driverio.h

netusage  600 Kbps    # maximum net bandwidth for Amanda, in KB per
      se
…

define dumptype comp-high {
   global
   comment "very important partitions on fast machines"
   compress client best
   kencrypt no
   priority high
}

define interface lo0 {
   comment "a local disk"
   use 1000 kbps
}
```

DiskList:

```
Dbserver  /dumps comp-user-tar
…
```

Data Layer: rsync

Diffs

Only actual changed pieces of files are transferred, rather than the whole file.

Compression

The tiny pieces of diffs are then compressed on the fly, further saving you file transfer time and reducing the load on the network.

Encryption

The stream from rsync is passed through the ssh protocol to encrypt your session.

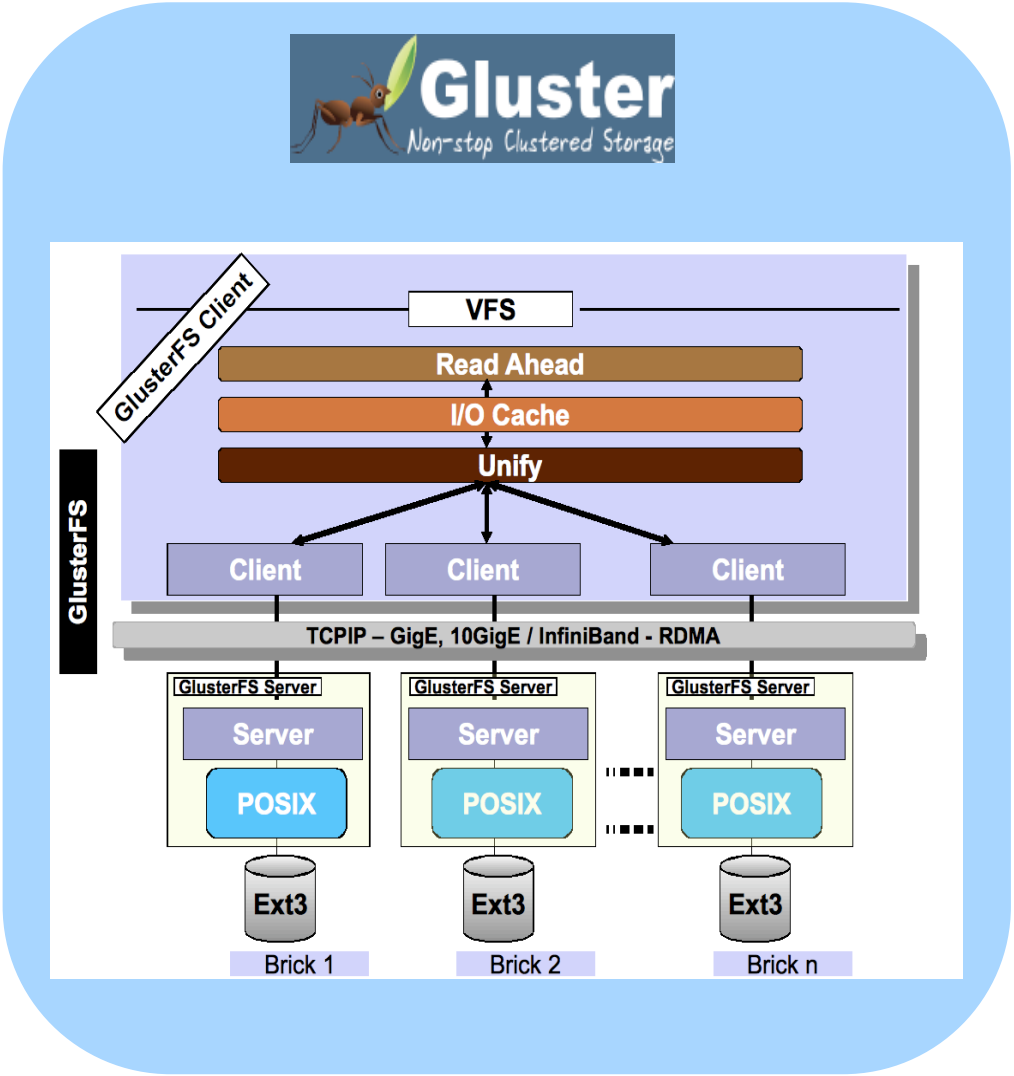
Latancy

Pipelining of file transfers to minimize latency costs

Simple Usage


```
rsync      --verbose  --progress --stats
           --compress
           --rsh=/usr/local/bin/ssh
           --recursive
           --times
           --perms
           --links
           --delete
           --exclude "*bak"
           --exclude "*~”
            /www/* webserver:path_name
```

Data Layer: GlusterFS



Features

- Automatic Replication
- Aggregation
- Scalable Striping
- Distributed Locking
- Performance Modules
- Pluggable I/O Scheduler
- Pluggable Transport
- Pluggable Auth

Administration

- NFS-like Backend
- Self-healing
- Staking Vol/modules

And much more ..

Data Layer: GlusterFS

Replication
- Client side
- Server side

Raid
- Raid 1
- Stripe
- Raid 10

Performance
- I/O modules
- Cache modules

```
volume posix
        type storage/posix
        option directory /data/export
end-volume
```

```
volume locks
        type features/locks
        subvolumes posix
end-volume
```

```
Volume brick
        type performance/io-threads
        option thread-count 8
        subvolumes locks
end-volume
```

```
volume server
        type protocol/server
        option transport-type tcp
        option auth.addr.brick.allow *
        subvolumes brick
end-volume
```

```
volume remote1
                type protocol/client
                option transport-type tcp
                option remote-host
                storage1.example.com
                option remote-subvolume brick
end-volume

volume remote2
 …
end-volume
```

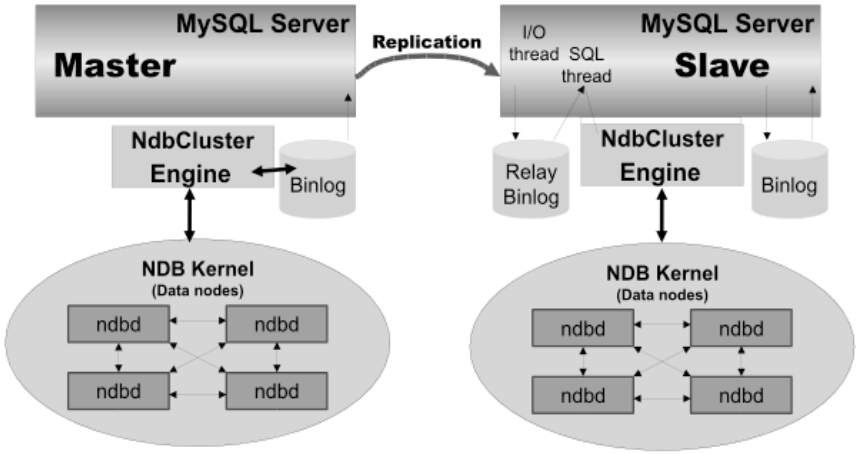
```
volume replicate
                type cluster/replicate
                 subvolumes remote1 remote2

end-volume
```

```
volume writebehind
                type performance/write-behind
                option window-size 1MB
                subvolumes replicate
end-volume
```

```
volume cache
                type performance/io-cache
                option cache-size 512MB
                subvolumes writebehind
end-volume
```

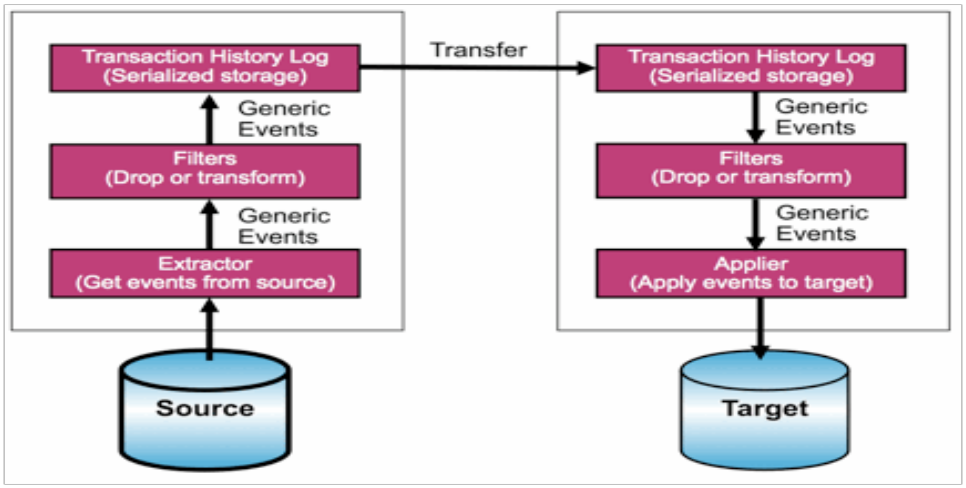
Builtin

- Mysql Master – Slave
- Mysql Multi-Master

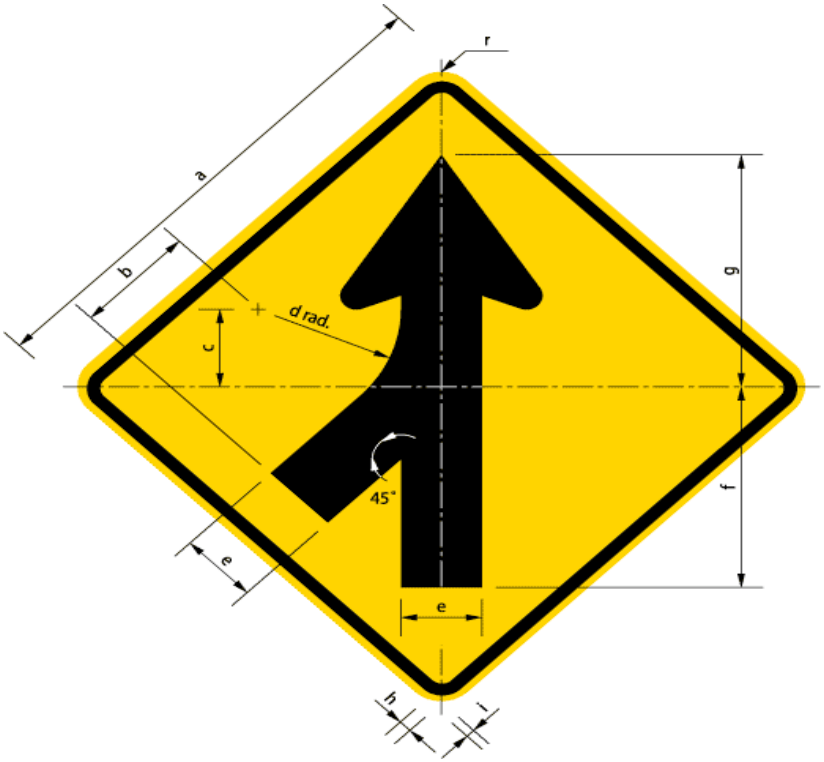


External

- Tungsten Replicator
- Slony-I (postgresql)

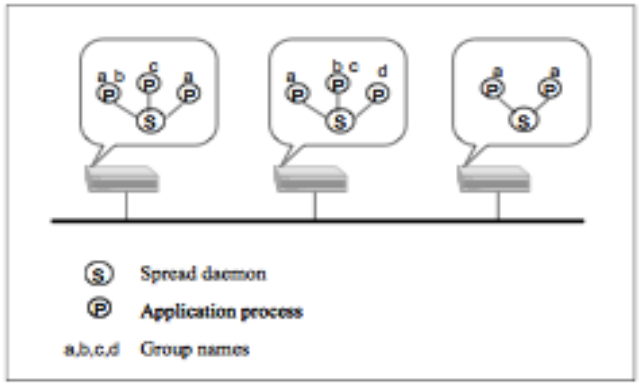


- **Data replication**
  - Cluster JDBC
  - Group messaging (spread, jgroup, ..)
  - Custom protocol
- **Distributed Application Design**
  - Data must be Unique
  - Primary key unique for all sites (hash, mixing key ..)

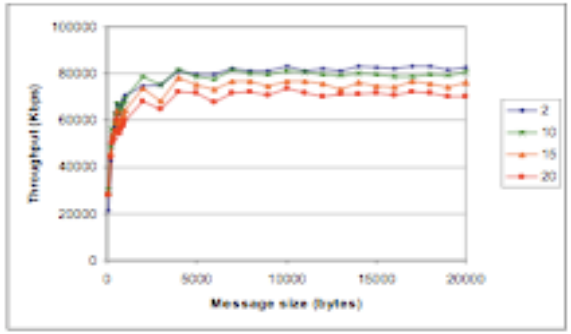


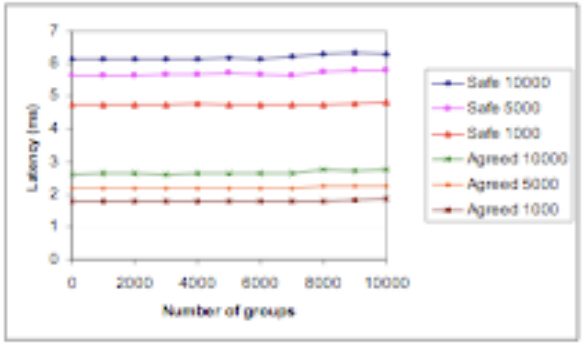
Spread

functions as a unified message bus for distributed applications, and provides highly tuned application-level multicast, group communication, and point to point support.



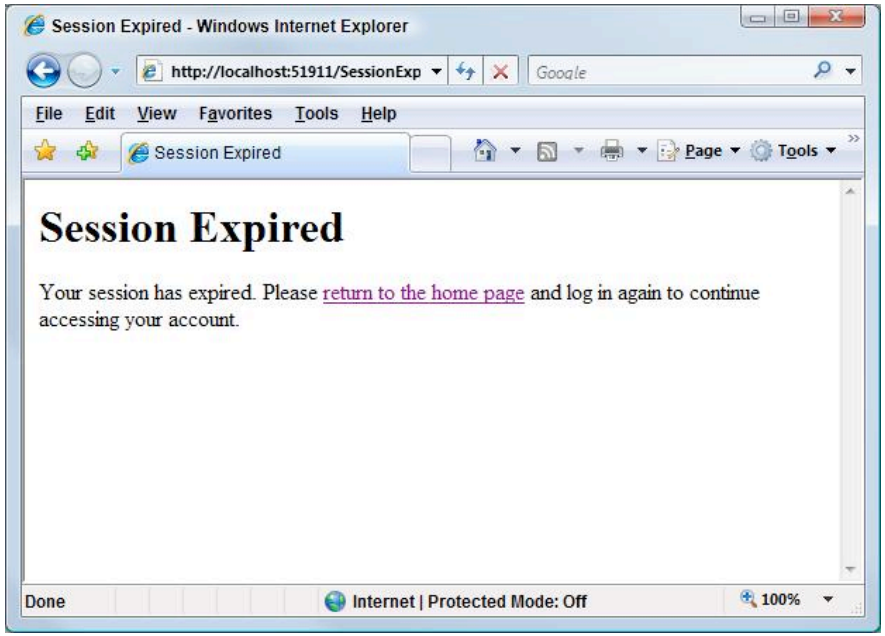
- Reliable and scalable messaging and group communication
- Easy to use, deploy and maintain.
- Highly scalable from one local area network to complex wide area networks.
- Supports thousands of groups with different sets of members.
- Enables message reliability in the presence of machine failures, process crashes and recoveries, and network partitions and merges.
- Provides a range of reliability, ordering and stability guarantees for messages.
- .Completely distributed algorithms with no central point of failure.



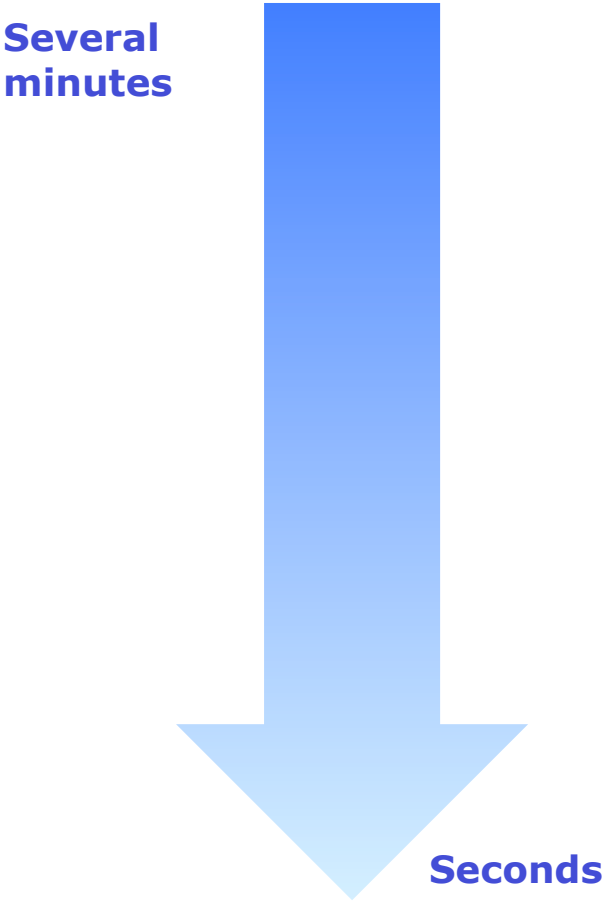


Session

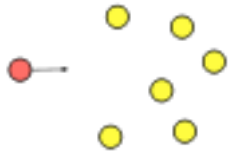
- Distributed file system
- Replicated Database
- AppServer Cluster

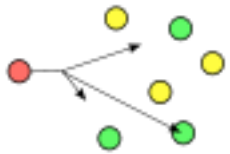


- **DNS: Bind**
  - keep low value of TTL
  - dynamic update
- **MPLS**
  - Change MPLS configuration, keep the same ip configuration
- **Routing Table : quagga**
  - OSPF/BGP announce new routing for your side



- Quagga
  - CISCO IOS syntax
  - All routing protocol
  - Device base interface





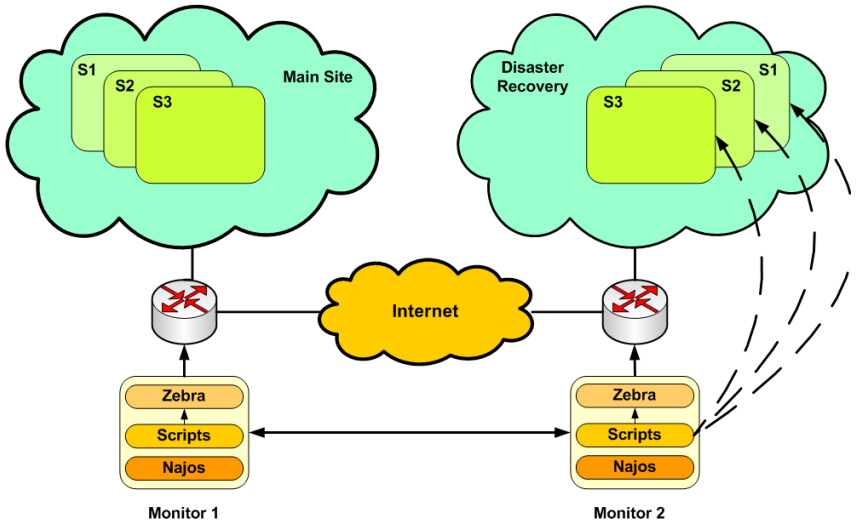
! Work only for !
  - Autonomous System
  - Large network

Quagga.conf:

```
interface eth0
        ip address XXX.XXX.XXX.XXX
!
…
Ospf.conf


router ospf
        ospf router-id XXX.XXX.XXX.XX
        log-adjacency-changes
        compatible rfc1583
        auto-cost reference-bandwidth 10000
        network XXX.XXX.XXX.XXX/XX area …
```

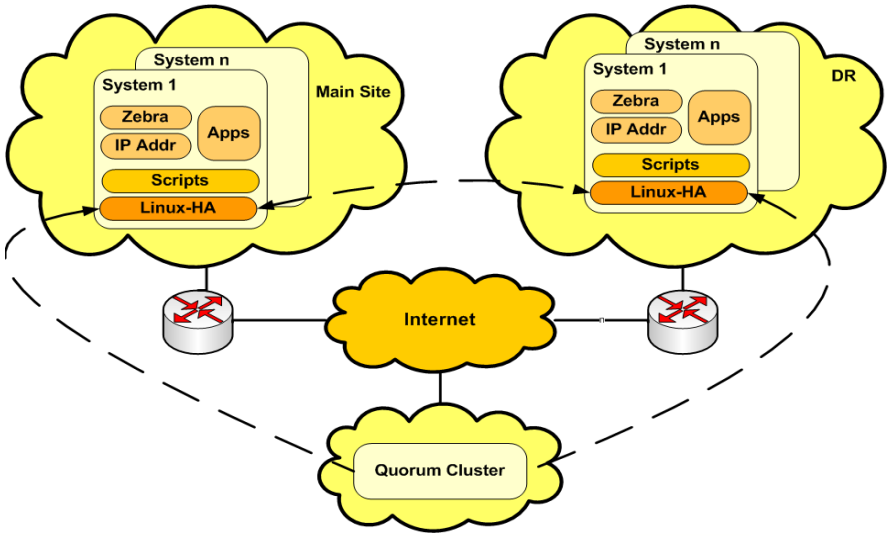
Monitoring System (outside)

- Probe/check external
- Remote Invocation scripts
- Manual/Automatic

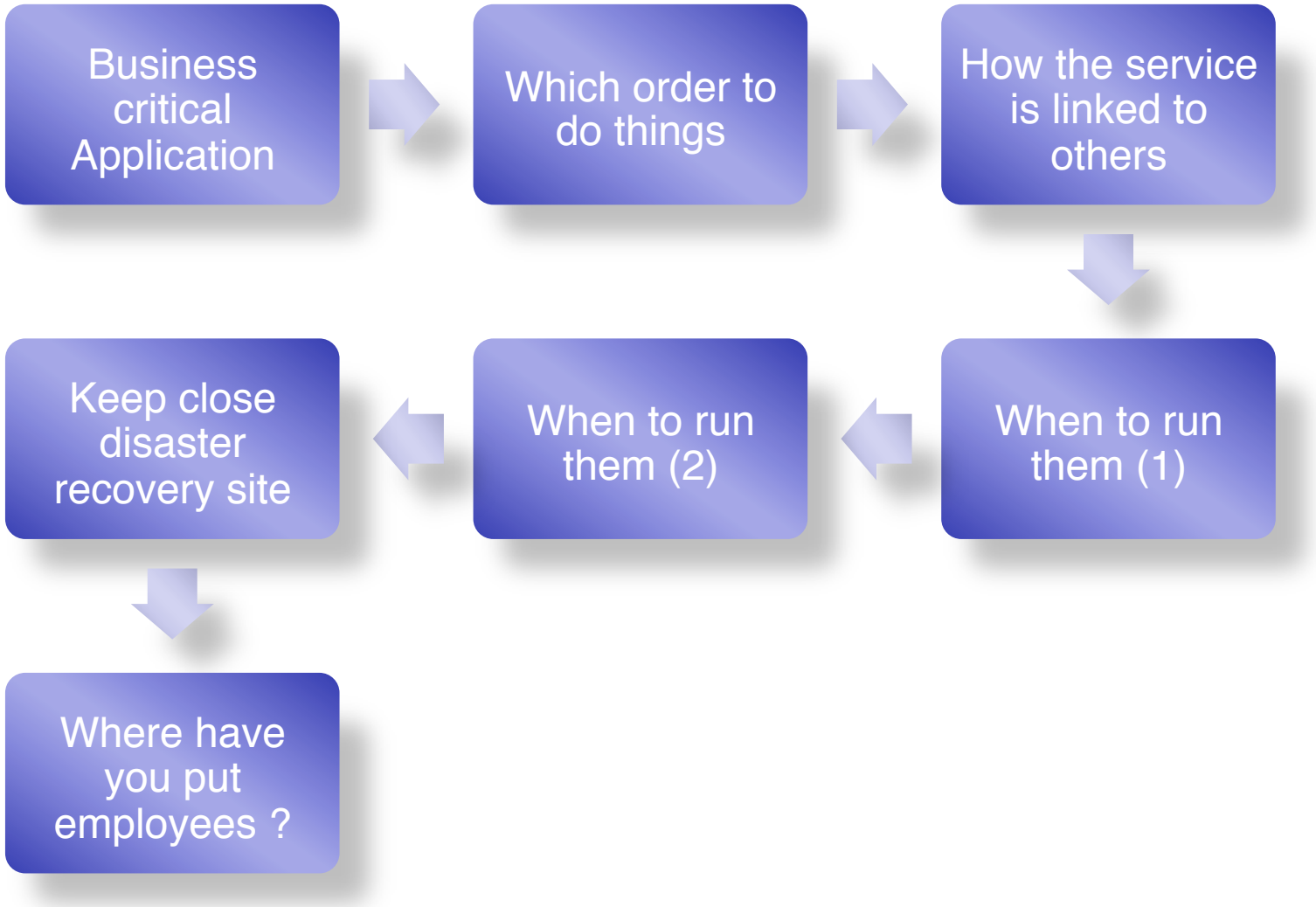


Linux HA (inside)

- Cluster definition with Quorum Server
- Local Scripts / Fencing script
- Manual/Automatic



Important Things



Case Studies

Company type: Heavy industry
Location: Single production site
Web Farm/IT: Inside the production site
APPS: Order management, production control
RPO/RTO: 2h
DR site: distance 400 km
Cost: a Lot

Disaster Event: Flooding of the production site

GOOD Solution ?

What happened ?

Perfect plan, everything OK

But…

Production restart after 6 months …

WRONG SOLUTION

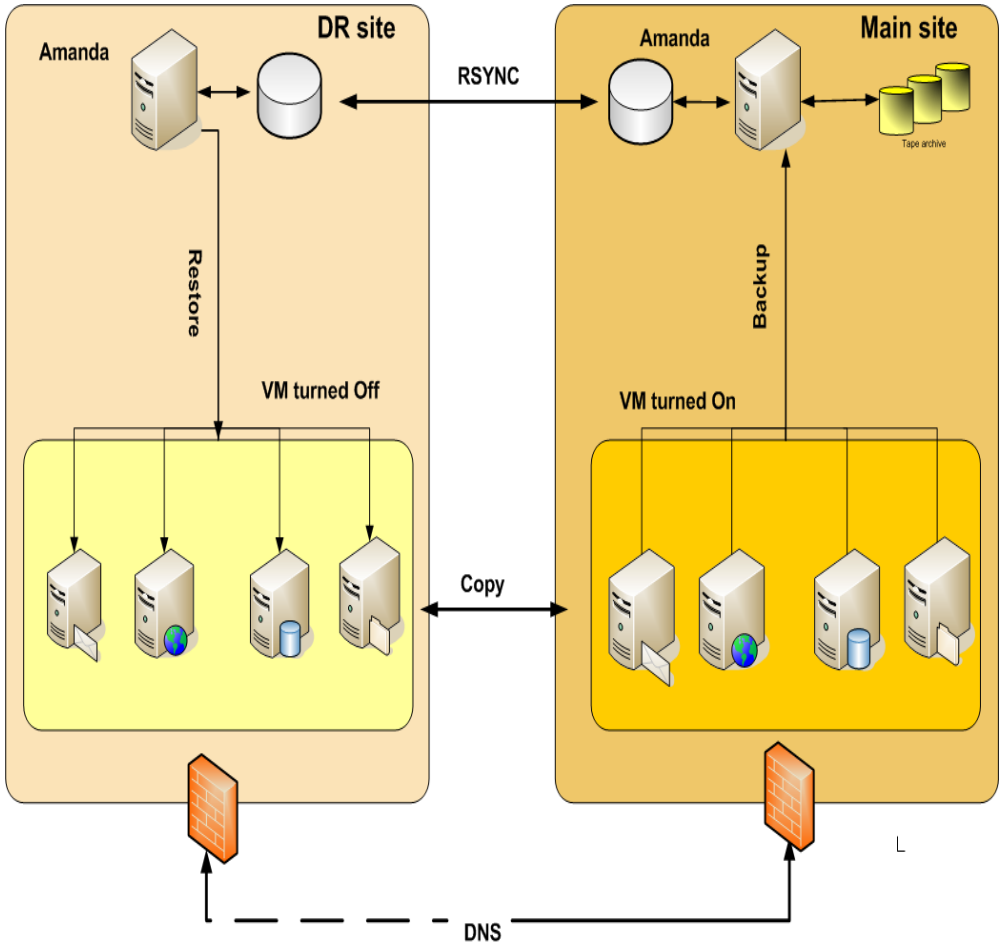
- Disaster not related to Business
- Misunderstanding between BCP and DRP

IT Manager was Fired

Base Solution: small-middle companies

Beolink.org

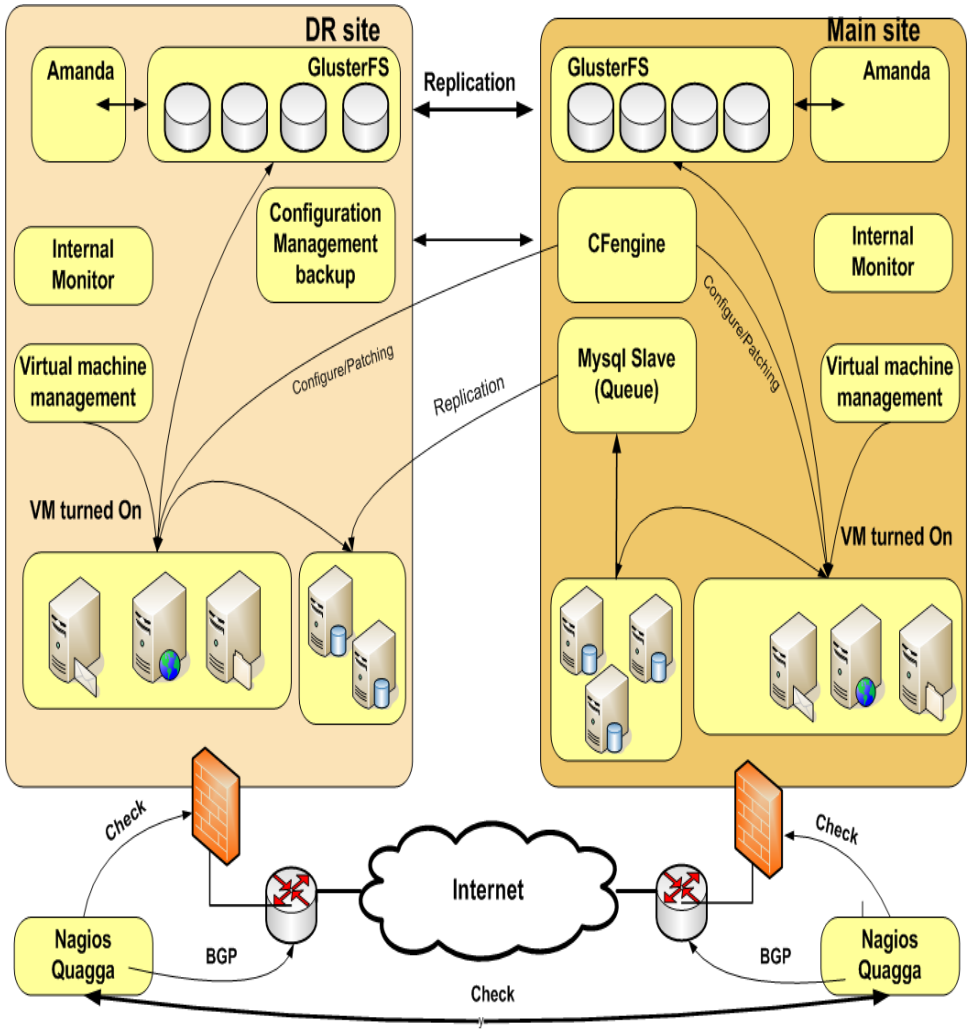
- Switch by hand
  - Configuration
  - Restore
  - Starting point
- RTO/RPO Hours
  - Linear to storage dimension
- Synchronization
  - Backup
- No more licenses needed



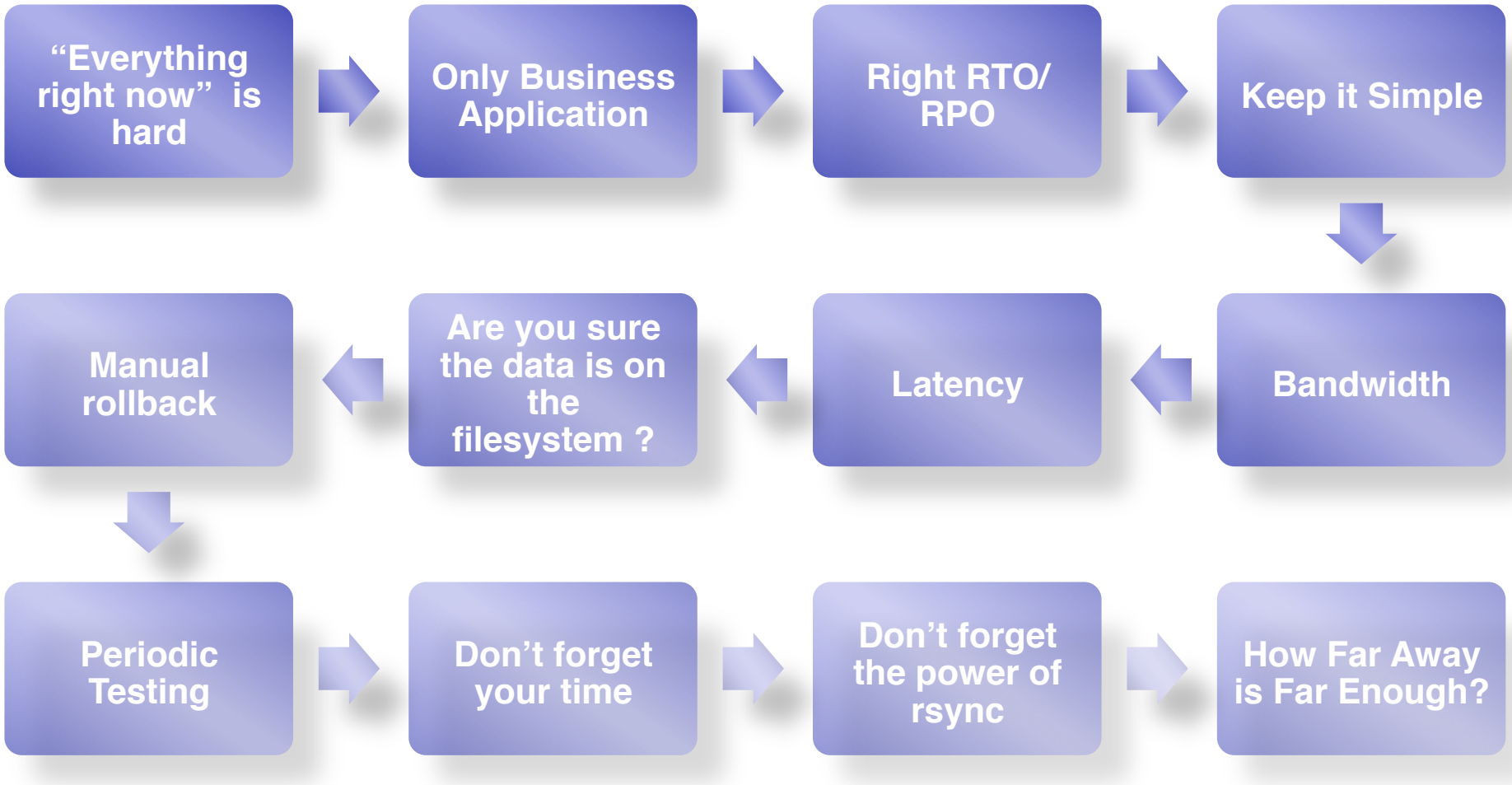
Medium/High Solution: Enterprise

Beolink.org

- Automatic Switch
  - IP migration (one IP per service)
  - Change Firewall role
- RTO/RPO seconds
- Synchronization
  - Multi-master FS
  - Database Replication
  - Configuration Management
- Double licenses needed



Conclusion


"Everything right now" is hard
Only Business Application
Right RTO/ RPO
Keep it Simple
Bandwidth
Latency
Are you sure the data is on the filesystem ?
Manual rollback
Periodic Testing
Don't forget your time
Don't forget the power of rsync
How Far Away is Far Enough?

XVI European AFS meeting 2009
Rome: September 28-30

Who should attend:

- Everyone interested in deploying a globally accessible file system
- Everyone interested in learning more about real world usage of Kerberos authentication in single realm and federated single sign-on environments
- Everyone who wants to share their knowledge and experience with other members of the AFS and Kerberos communities
- Everyone who wants to find out the latest developments affecting AFS and Kerberos

More Info:

www.openafs.it

Thank you

manfred@freemails.ch
www.beolink.org

Beolink.org