Bacula: an Intro

K. M. Peterson

BBLISA 14 March 2012

Welcome

- Bacula is a client-server backup system that runs on a variety of platforms, mostly *NIX.

- Most of what I'll talk about is Bacula's functionality, along with a simple example implementation.

- But also: my issues, and some edge-cases.

- Abstract
- Functionality
  - Client-Server architecture
  - Third Server (Storage)
  - Client Support
  - Features
  - Implementation Details
  - Issues

- Example
  - Simple Server + 1 client
  - Configurations
  - Sample backup output
  - Sample recovery
- Edge-cases
- Summary
  - Questions?

What is it?

- Backups: Copy data from *here* to *there*.

  Complexity: Security

  Complexity: Managing resources

  space, bandwidth, MTTR

  Complexity: Recovery Scenarios

- What problem are we trying to solve?

  System Errors hardware, network, application, disk

  User Errors mistakes, errors

  Unknown errors System and User errors in time

- Protecting against unknown errors

  the more copies the better

overhead: space, lookup

- Backup vs. Archive

  Business issues

  Regulatory issues

  Key operational challenge: fresh metadata, optimized recovery.

Why Bacula?

- Clients and Servers
- Open
- That which is least ugly is most pretty.
- Flexibility

Functionality

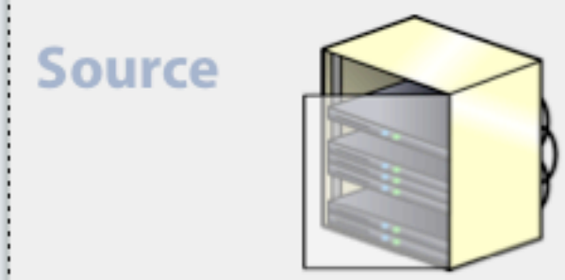
- Client-server
- The “Third Server” (SD)
- Client Support

Server(s)

Clients

Media

Source



Windows

Mac OS

*nix

Server(s)

SD

Storage Daemon
Media access

Dir

Director
Schedule, catalog

FD

File Daemon
Client

Server(s)

SD

Dir

FD

Clients

Windows
FD

Mac OS
FD

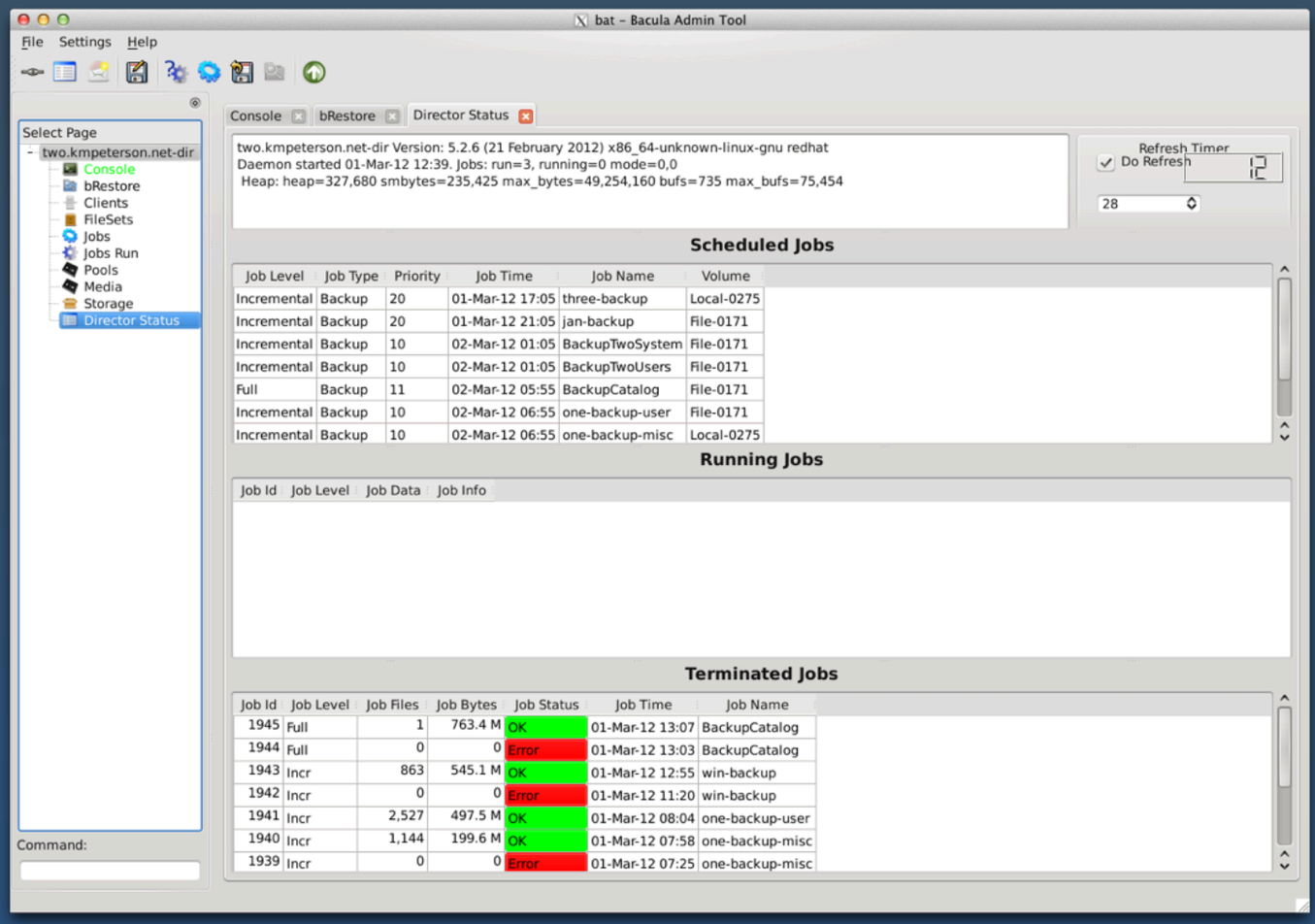
*nix
FD

KMP

Features

- Encryption
- Authentication - Shared Key between daemons
- Encryption at FD (encrypt at source)
- Supports TLS between daemons

Features UI


bat – Bacula Admin Tool
File Settings Help
Select Page
two.kmpeterson.net-dir
Console
bRestore
Clients
FileSets
Jobs
Jobs Run
Pools
Media
Storage
Director Status
Command:
Console
bRestore
Director Status
two.kmpeterson.net-dir Version: 5.2.6 (21 February 2012) x86_64-unknown-linux-gnu redhat
Daemon started 01-Mar-12 12:39. Jobs: run=3, running=0 mode=0,0
Heap: heap=327,680 smbytes=235,425 max_bytes=49,254,160 bufs=735 max_bufs=75,454
Refresh Timer
Do Refresh
28
Scheduled Jobs
Job Level | Job Type | Priority | Job Time | Job Name | Volume
Incremental | Backup | 20 | 01-Mar-12 17:05 | three-backup | Local-0275
Incremental | Backup | 20 | 01-Mar-12 21:05 | jan-backup | File-0171
Incremental | Backup | 10 | 02-Mar-12 01:05 | BackupTwoSystem | File-0171
Incremental | Backup | 10 | 02-Mar-12 01:05 | BackupTwoUsers | File-0171
Full | Backup | 11 | 02-Mar-12 05:55 | BackupCatalog | File-0171
Incremental | Backup | 10 | 02-Mar-12 06:55 | one-backup-user | File-0171
Incremental | Backup | 10 | 02-Mar-12 06:55 | one-backup-misc | Local-0275
Running Jobs
Job Id | Job Level | Job Data | Job Info
Terminated Jobs
Job Id | Job Level | Job Files | Job Bytes | Job Status | Job Time | Job Name
1945 | Full | 1 | 763.4 M | OK | 01-Mar-12 13:07 | BackupCatalog
1944 | Full | 0 | 0 | Error | 01-Mar-12 13:03 | BackupCatalog
1943 | Incr | 863 | 545.1 M | OK | 01-Mar-12 12:55 | win-backup
1942 | Incr | 0 | 0 | Error | 01-Mar-12 11:20 | win-backup
1941 | Incr | 2,527 | 497.5 M | OK | 01-Mar-12 08:04 | one-backup-user
1940 | Incr | 1,144 | 199.6 M | OK | 01-Mar-12 07:58 | one-backup-misc
1939 | Incr | 0 | 0 | Error | 01-Mar-12 07:25 | one-backup-misc

bat – Bacula Admin Tool

File Settings Help

Select Page

- two.kmpeterson.net-dir
  - Console
  - bRestore
  - Clients
  - FileSets
  - Jobs
  - Jobs Run
  - Pools
  - Media
  - Storage
  - Director Status

Command:

Console | bRestore | Director Status

two.kmpeterson.net-dir Version: 5.2.6 (21 February 2012) x86_64-unknown-linux-gnu redhat
Daemon started 01-Mar-12 12:39. Jobs: run=3, running=0 mode=0,0
Heap: heap=327,680 smbytes=235,425 max_bytes=49,254,160 bufs=735 max_bufs=75,454

Refresh Timer
Do Refresh 12
28

Scheduled Jobs

Job Level	Job Type	Priority	Job Time	Job Name	Volume
Incremental	Backup	20	01-Mar-12 17:05	three-backup	Local-0275
Incremental	Backup	20	01-Mar-12 21:05	jan-backup	File-0171
Incremental	Backup	10	02-Mar-12 01:05	BackupTwoSystem	File-0171
Incremental	Backup	10	02-Mar-12 01:05	BackupTwoUsers	File-0171
Full	Backup	11	02-Mar-12 05:55	BackupCatalog	File-0171
Incremental	Backup	10	02-Mar-12 06:55	one-backup-user	File-0171
Incremental	Backup	10	02-Mar-12 06:55	one-backup-misc	Local-0275

Running Jobs

Job Id	Job Level	Job Data	Job Info

Terminated Jobs

Job Id	Job Level	Job Files	Job Bytes	Job Status	Job Time	Job Name
1945	Full	1	763.4 M	OK	01-Mar-12 13:07	BackupCatalog
1944	Full	0	0	Error	01-Mar-12 13:03	BackupCatalog
1943	Incr	863	545.1 M	OK	01-Mar-12 12:55	win-backup
1942	Incr	0	0	Error	01-Mar-12 11:20	win-backup
1941	Incr	2,527	497.5 M	OK	01-Mar-12 08:04	one-backup-user
1940	Incr	1,144	199.6 M	OK	01-Mar-12 07:58	one-backup-misc
1939	Incr	0	0	Error	01-Mar-12 07:25	one-backup-misc

bat – Bacula Admin Tool

File Settings Help

Select Page

- two.kmpeterson.net-dir
  - Console
  - bRestore
  - Clients
  - FileSets
  - Jobs
  - Jobs Run
  - Pools
  - Media
  - Storage
  - Director Status

Command:

Console | Director Status

```
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/AppData/Local/History is a junction point or a different filesystem. Will not desce
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/AppData/Local/Temporary Internet Files is a junction point or a different filesyste
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/Application Data is a junction point or a different filesystem. Will not descend fr
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/Cookies is a junction point or a different filesystem. Will not descend from C:/ in
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/Documents/My Music is a junction point or a different filesystem. Will not descend
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/Documents/My Pictures is a junction point or a different filesystem. Will not desce
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/Documents/My Videos is a junction point or a different filesystem. Will not descend
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/Local Settings is a junction point or a different filesystem. Will not descend from
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/My Documents is a junction point or a different filesystem. Will not descend from C
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/NetHood is a junction point or a different filesystem. Will not descend from C:/ in
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/PrintHood is a junction point or a different filesystem. Will not descend from C:/
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/Recent is a junction point or a different filesystem. Will not descend from C:/ int
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/SendTo is a junction point or a different filesystem. Will not descend from C:/ int
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/Start Menu is a junction point or a different filesystem. Will not descend from C:/
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/KMP/Templates is a junction point or a different filesystem. Will not descend from C:/
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/Public/Documents/My Music is a junction point or a different filesystem. Will not desce
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/Public/Documents/My Pictures is a junction point or a different filesystem. Will not de
01-Mar 12:53 win.kmpeterson.net-fd JobId 1943:      C:/Users/Public/Documents/My Videos is a junction point or a different filesystem. Will not desc
01-Mar 12:55 win.kmpeterson.net-fd JobId 1943: VSS Writer (BackupComplete): "Task Scheduler Writer", State: 0x1 (VSS_WS_STABLE)
01-Mar 12:55 two.kmpeterson.net-sd JobId 1943: Job write elapsed time = 00:05:01, Transfer rate = 1.811 M Bytes/second
01-Mar 12:55 win.kmpeterson.net-fd JobId 1943: VSS Writer (BackupComplete): "VSS Metadata Store Writer", State: 0x1 (VSS_WS_STABLE)
01-Mar 12:55 win.kmpeterson.net-fd JobId 1943: VSS Writer (BackupComplete): "Performance Counters Writer", State: 0x1 (VSS_WS_STABLE)
01-Mar 12:55 win.kmpeterson.net-fd JobId 1943: VSS Writer (BackupComplete): "System Writer", State: 0x1 (VSS_WS_STABLE)
01-Mar 12:55 win.kmpeterson.net-fd JobId 1943: VSS Writer (BackupComplete): "MSSearch Service Writer", State: 0x1 (VSS_WS_STABLE)
01-Mar 12:55 win.kmpeterson.net-fd JobId 1943: VSS Writer (BackupComplete): "ASR Writer", State: 0x1 (VSS_WS_STABLE)
01-Mar 12:55 win.kmpeterson.net-fd JobId 1943: VSS Writer (BackupComplete): "Registry Writer", State: 0x1 (VSS_WS_STABLE)
01-Mar 12:55 win.kmpeterson.net-fd JobId 1943: VSS Writer (BackupComplete): "WMI Writer", State: 0x1 (VSS_WS_STABLE)
01-Mar 12:55 win.kmpeterson.net-fd JobId 1943: VSS Writer (BackupComplete): "COM+ REGDB Writer", State: 0x1 (VSS_WS_STABLE)
01-Mar 12:55 win.kmpeterson.net-fd JobId 1943: VSS Writer (BackupComplete): "Shadow Copy Optimization Writer", State: 0x1 (VSS_WS_STABLE)
01-Mar 12:55 two.kmpeterson.net-dir JobId 1943: Bacula two.kmpeterson.net-dir 5.2.6 (21Feb12):
  Build OS:               x86_64-unknown-linux-gnu redhat
  JobId:                  1943
  Job:                    win-backup.2012-03-01_12.50.27_13
  Backup Level:           Incremental, since=2012-02-29 12:42:29
  Client:                 "win.kmpeterson.net-fd" 5.0.3 (04Aug10) Linux,Cross-compile,Win32
  FileSet:                "Win-AllFiles" 2011-09-13 11:05:00
  Pool:                   "File-Incremental" (From Job IncPool override)
  Catalog:                "MyCatalog" (From Client resource)
  Storage:                "File" (From Job resource)
  Scheduled time:         01-Mar-2012 12:50:27
  Start time:             01-Mar-2012 12:50:30
  End time:               01-Mar-2012 12:55:33
  Elapsed time:           5 mins 3 secs
  Priority:               10
  FD Files Written:       863
  SD Files Written:       863
  FD Bytes Written:       545,130,348 (545.1 MB)
  SD Bytes Written:       545,307,464 (545.3 MB)
```

bat – Bacula Admin Tool

File Settings Help

Select Page

- two.kmpeterson.net-dir
  - Console
  - bRestore
  - Clients
  - FileSets
  - Jobs
  - Jobs Run
  - Pools
  - Media
  - Storage
  - Director Status

Command:

Console | bRestore | Director Status

Select from Client list drop down then select from Job list drop down

Client list | Merge Jobs | View all Versions | Current Directory

File list -- drag below for restore

Type	File Name	Size	Date

Selection Range 0 - 500 File Filter

File revisions -- drag below for restore

InChanger	Volume	JobId

Drag and drop **File list** and/or **File revisions** items here for Restore

Clear | Estimate | Restore

Type	FileName	Size	Date	JobId	FileIndex	Nb Files

Backup Sets

- Disk
  - As-if tape volumes
  - Can be compressed
  - Limit size
  - Can require “mount” - so can defer location

Backup Sets

- Tape
  - Supports wide variety of media formats
  - Libraries/autoloaders, barcode labels &c.
  - No support for hardware encryption
  - Based on OS support (i.e., `mtx`, scsi)

Implementation Catalog

- Database-backed catalog
- Supports MySQL, postgres, Oracle, sqlite
- Generally requires some amount of tuning
- Schemas relatively accessible
- Console support for scripting

Implementation Devices

- Along with Disk Files, some support for optical media
- Bacula Enterprise includes NDMP
- FD can perform raw I/O (pipes)

Implementation Configuration

- File-based configuration for each daemon
- Director sections:

```
Director {}
JobDefs {}
Job {}
FileSet {}
Schedule {}
Client {} ...
```

Implementation

- Spooling
  - Buffering client data on disk
  - Drive tape drives at full speed
- Multiplexing
  - Multiple clients per output stream
  - Adverse impact on MTTR

Issues Complexity

- Configuration relatively clean, functions have grown organically
- so: lots of files
- duplication of or unclear parameters
- functionality is opaque.

Issues Complexity

- Documentation fair, pretty steep learning curve
- references difficult for beginners
- lots of “don’t forget your coat”
- mentions of “coming attractions” - not.

Issues Complexity

- Functionality partially developed then abandoned
- DR is dicey
- Project developed by a team, and is clearly the work of many hands.

Issues Performance

- Generally not inefficient, but:
  - recovery can be lengthy, depending on complexity of storage strategy
  - multiplexing clients onto backup media speeds up backup at the cost of recovery
  - recovery without a catalog possible, but also very slow

Issues Architecture

- Difficult to implement failover
- All functionality controlled from backup console, which must talk to single director
- Nomenclature confusion: difficult to understand certain limitations (“Pool” v. “Storage” v. “Device” WRT Director v. SD definitions).

Issues File Sets

- Until recently, selection of files by “last update” date.
- Modification of FileSet directive causes Full backup of target.
- “Accurate” backup introduces inefficiencies, still fairly new.

Example

- Simple Server + 1 client
  - Server: Director, SD, FD; disk-based set.
  - Configuration Overview
  - Example (incremental backup); restore

Example Setup Background

- Centos 6.2
- Packages installed: `mysql-server bacula-director-mysql bacula-storage-mysql bacula-client bacula-console-bat bacula-console`
- MySQL: bacula user
- Bacula DB setup scripts
- create/chown `bacula` `/var/backup`

```
Director {                                  # define myself
  Name = bacula-dir
  DIRport = 9101                    # where we listen for UA connections
  QueryFile = "/usr/libexec/bacula/query.sql"
  WorkingDirectory = "/var/spool/bacula"
  PidDirectory = "/var/run"
  Maximum Concurrent Jobs = 1
  Password = "BaCuLaDiRpAsSwOrD"    # Console password
  Messages = Daemon
}
```

```
JobDefs {                                   # defaults - entries in Job definitions override these
  Name = "DefaultJob"
  Type = Backup
  Level = Incremental                 # will get forced to "Full" on first run
  Client = bacula-fd                        # by default, points to file daemon on this host
  FileSet = "Full Set"                      # pointer to list/specification of files for this backup
  Schedule = "WeeklyCycle"            # pointer to a scheduling directive
  Storage = File                      # references a "device" definition
  Messages = Standard                 # message logging/destination configuration
  Pool = File                               # media definition
  Priority = 10
  Write Bootstrap = "/var/spool/bacula/%c.bsr"          # recovery information
}
```

```
#
# Define the main nightly save backup job
#   By default, this job will back up to disk in /tmp
Job {
  Name = "BackupServerFiles"
  JobDefs = "DefaultJob"                    # simply use defaults
}
```

```
# List of files to be backed up
FileSet {
  Name = "Full Set"
  Include {
    Options {
      signature = MD5                 # catalog will store digest of file
      compression = GZIP                      # catalog will store digest of file
    }
    File = /boot                      # critical data
    File = /usr                             # and
    File = /root                      # system
    File = /var                             # files
    File = /home                      # and user directories
  }

#
  Exclude {
    File = /var/spool/bacula          # exclude bacula in-process dat
    File = /var/lib/mysql             # don't back up database files
    File = /var/backup                # don't back up backups et cetera et all ad infinitum
    File = /tmp
    File = /proc
    File = /tmp                       # exclude ... etc.
    File = /.journal
    File = /.fsck
  }
}
```

```
# Client (File Services) to backup
Client {
  Name = bacula-fd                                  # matches Name directive in FD
  Address = bacula-test-x                           # hostname

  FDPort = 9102
  Catalog = MyCatalog
  Password = "bAcUlAfDpAsSwOrDx@"                   # password for FileDaemon
  File Retention = 30 days                          # 30 days
  Job Retention = 6 months                          # six months
  AutoPrune = yes                                   # Prune expired Jobs/Files
}
```

```
# Definition of file storage device
Storage {
  Name = File
# Do not use "localhost" here
  Address = bacula-test-x                    # local host, in this example.
  SDPort = 9103
  Password = "bAcUlAsDpAsSwOrD"
  Device = FileStorage
  Media Type = File
}


# File Pool definition (This is pool used for jobs defined off JobDefs 'DefaultJob')
Pool {
  Name = File
  Pool Type = Backup
  Recycle = yes                              # Bacula can automatically recycle Volumes
  AutoPrune = yes                            # Prune expired volumes
  Volume Retention = 365 days                # one year
  Maximum Volume Bytes = 1G                  # Limit Volume size for ease of maintenance.
  Maximum Volumes = 100                      # Limit number of Volumes in Pool
  Label Format = "Test-"                       # Filename format (required to auto label!)
}
```

```
Schedule {
  Name = "WeeklyCycle"
  Run = Full 1st sun at 23:05
  Run = Differential 2nd-5th sun at 23:05
  Run = Incremental mon-sat at 23:05
}

# This schedule does the catalog. It starts after the WeeklyCycle
Schedule {
  Name = "WeeklyCycleAfterBackup"
  Run = Full sun-sat at 23:10
}

# This is the backup of the catalog
FileSet {
  Name = "Catalog"
  Include {
    Options {
      signature = MD5
    }
    File = "/var/spool/bacula/bacula.sql"    # this file created by 'RunBeforeJob'
  }                                          # directive in BackupCatalog job,
}                                            # then deleted by 'RunAfterJob' script.
```

Edge Cases VMware Backups

- vcbMounter: cycle through ESX server vmdk files
- scp to Bacula server
- script Bacula backups of VMs to tape

Edge Cases NetApp backups

- Problems retaining metadata from mixed NFS/CIFS environments
- Server access to shares not fast enough
- *Solution*: NetApp console, shell, dump to Bacula via pipe.
- Very fast, piece-recovery somewhat slow; not usable via ssh.

Edge Cases AWS Offsite

- Destination directory for backup sets amenable to synchronization
- If already encrypted, resource savings by not re-encrypting data
- Backup sets of reasonable size correlate well to AWS S3 “buckets”.
- But: Bacula assumes media is cheap!

Summary

- It's a complicated application, but what it needs to do is also complicated.
- Getting started not difficult, but concepts can be challenging.
- Open project, active = worth your time if these are problems you need to solve.

Summary Project

- Bacula: http://www.bacula.org
- Mailing Lists: https://sourceforge.net/projects/bacula/
- Latest release: 5.2.6, 26Feb12.

- Any (more) questions?
- Thank you!
- http://kmpeterson.com/special/bblisa-bacula

K. M. Peterson
kmp@kmpeterson.com
http://kmpeterson.com
@kmp