

Implantación de una solución en red para Backups en IRAM usando Bacula

Pablo Mellado
Instituto de RadioAstronomía Milimétrica

Jornadas Técnicas de RedIRIS 2010



Índice

- Quiénes somos
- Nuestro peculiar entorno
- Necesidades especiales de copias de seguridad
- Por qué Bacula
- Introducción a Bacula
- Nuestra experiencia con Bacula
- Ventajas e inconvenientes de esta solución

Quiénes somos

- Instituto de investigación fundado por el CNRS francés, el MPG alemán y el IGN español.
- Su sede principal se encuentra en Grenoble.
- Mantiene y desarrolla dos observatorios: el telescopio de 30m en pico veleta y el interferómetro en Plateau de Bure en los Alpes franceses.
- El personal está compuesto por más de 120 personas entre científicos, ingenieros, técnicos y personal administrativo. Alrededor de 35 trabajando en España.
- En España existen dos centros de trabajo: las oficinas en Granada y el observatorio en Pico Veleta.



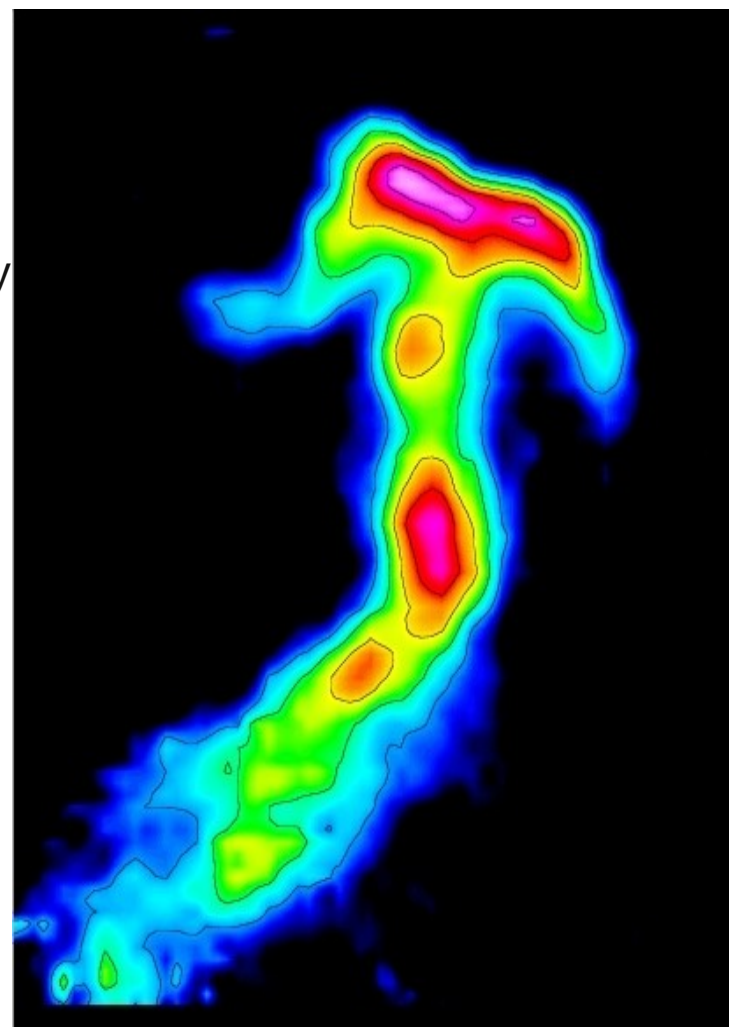
Nuestro peculiar entorno

- Nuestro telescopio se encuentra situado a 2850m de altitud, muy cercano al Pico Veleta.
- El acceso en invierno se realiza mediante máquinas ratracs.
- Estamos rodeados por la estación de esquí.
- La conexión de nuestras redes de datos se realiza mediante radioenlaces redundantes de hasta 50 Mbps full-duplex.
- Es bastante común que algunos días de invierno, el observatorio se quede completamente aislado.



Necesidades especiales de copia de seguridad

- Nuestro telescopio está observando 24 horas al día, sólo para por mal tiempo o por mantenimiento.
- Por lo tanto, el telescopio genera gran cantidad de datos durante las observaciones.
- Las observaciones realizadas son una información muy sensible que debe ser copiada para evitar cualquier pérdida.
- Las máquinas esenciales para la observación son repaldadas durante las horas de mantenimiento, que se planifican semanalmente.
- Una copia completa de todos los sistemas es almacenada en cinta “para siempre” a principios de año.
- En lo posible, debemos ser capaces de recuperar observaciones muy antiguas hechas por científicos.



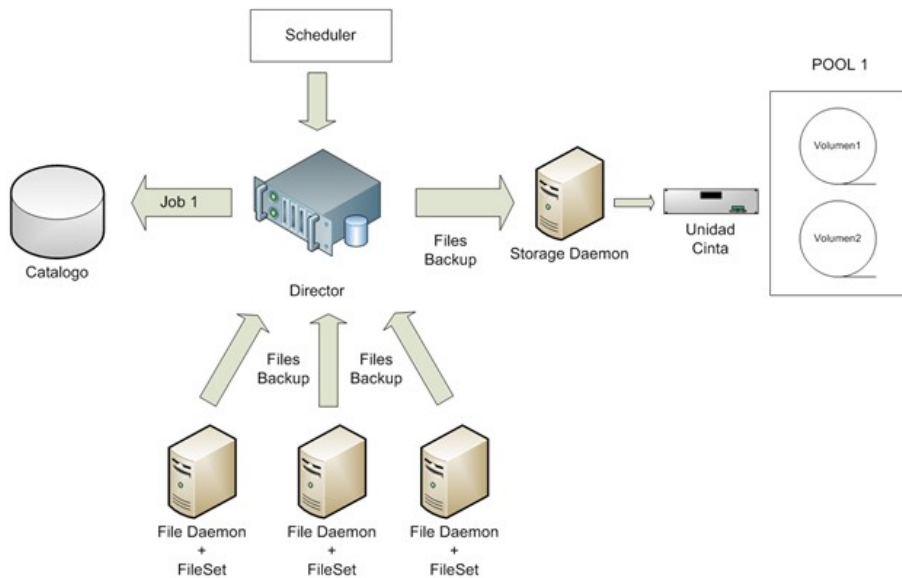
Por qué Bacula

- Es un sistema de backup en red, no es necesario el uso de NFS u otros.
- Está implementado en Python, como muchos de nuestros principales sistemas en el telescopio.
- Tenemos control sobre cuándo se guarda algo y en qué lugar.
- Soporta manejo de robots, y por lo tanto el cambio de cintas es automático.
- Almacena toda la “meta-información” en una base de datos, por lo que recuperar ficheros es muy “rápido” y fácil.
- Podemos verificar que el backup se ha realizado correctamente.
- Más fiable que nuestro anterior sistema basado en scripts que usaban el comando *tar*.

Introducción a Bacula

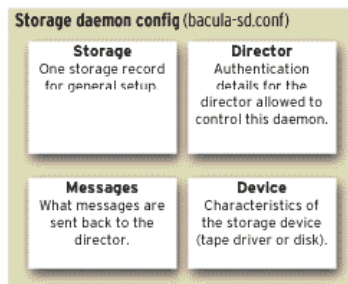
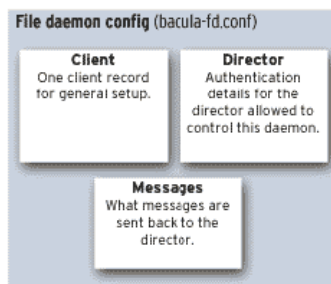
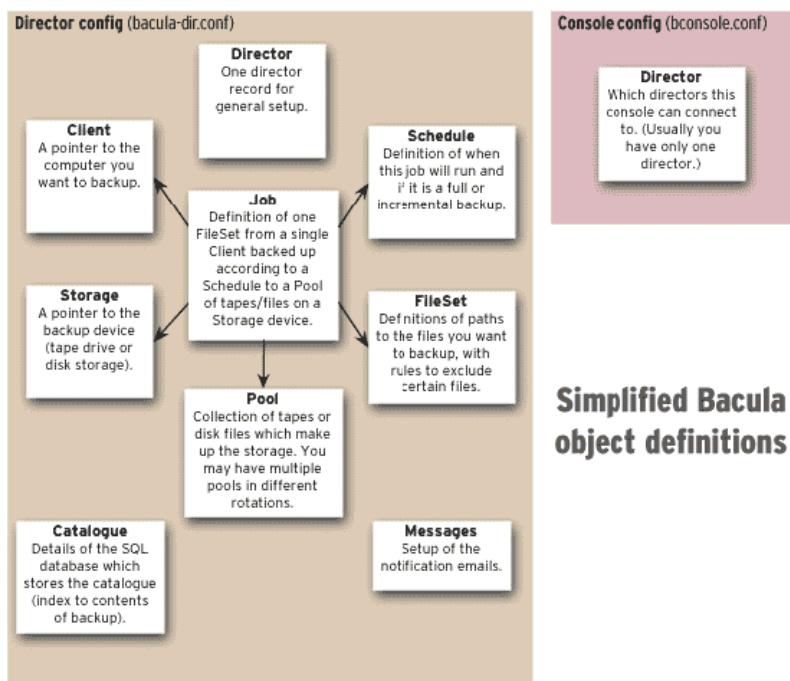
1/3 (Los Demonios)

- Tres demonios: *director*, *storage* & *file*.
- El demonio *director* se encarga de coordinar y planificar los backups.
- El demonio *storage* gestiona las unidades de backup y almacena los datos que se le envían. Típicamente se ejecuta en la máquina que tiene instalada la unidad de cinta.
- El demonio *file* provee los ficheros que se tienen que copiar. Este demonio se instala y ejecuta en las máquinas cliente.



Introducción a Bacula

2/3 (Los objetos)



- **Schedule**: Define planificaciones en el tiempo y si son incrementales o totales.
- **FileSet**: Define un conjunto de ficheros a ser copiado, con reglas para excluir.
- **Pool**: Conjunto de cintas para almacenaje.
- **Device**: Dispositivo de almacenaje, puede ser un robot.
- **Storage**: Define una conexión con un demonio Storage.
- **Job**: Definición de un trabajo que copia un FileSet de un determinado cliente siguiendo un Schedule que se almacena en un determinado Pool con un Device.
- **Catalog**: Detalles de la base de datos donde se almacena toda la *meta-información*.
- **Messages**: Define la información que se envía por email y a quien.

Introducción a Bacula

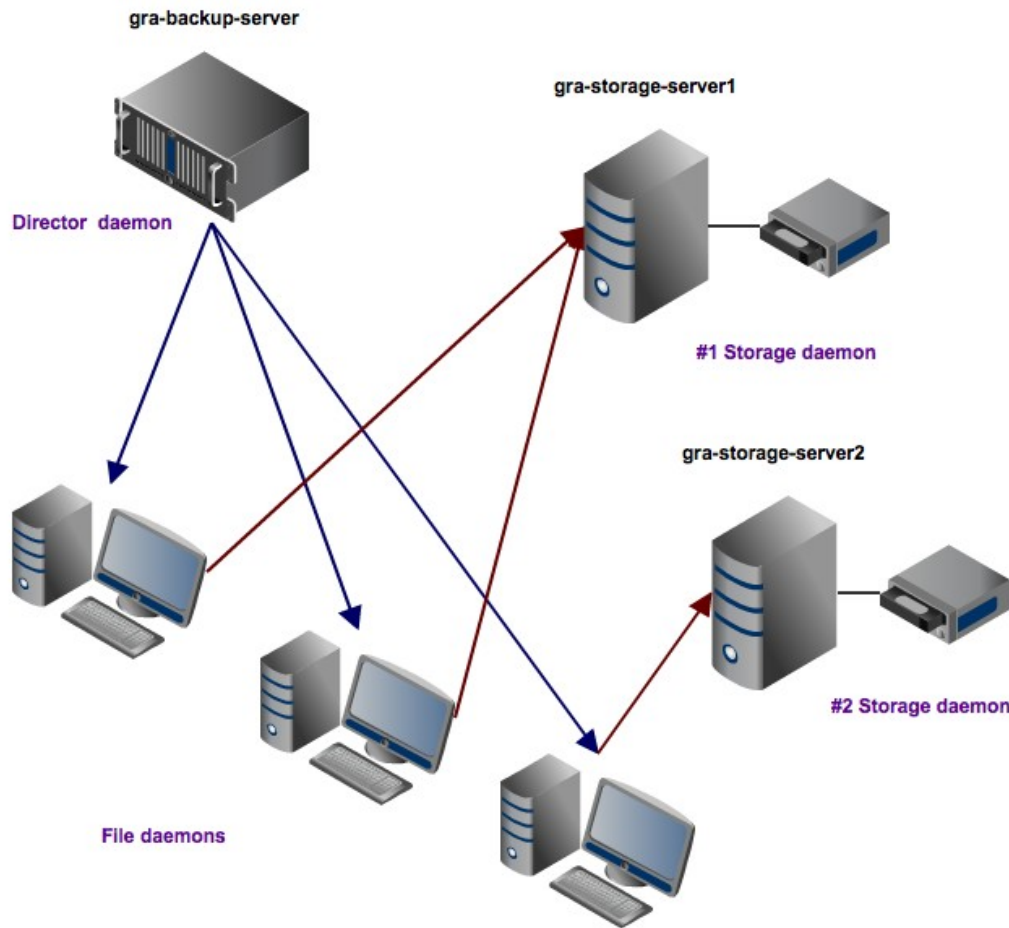
3/3 (Características destacables)

- Conjunto de programas con arquitectura cliente-servidor.
- Funciona en *todos* los sistemas operativos, al menos el cliente.
- Reciclaje automático de cintas.
- Etiquetaje interno de cintas para evitar sobre-escritura.
- Compatible con lectores de códigos de barras para etiquetas de cintas.
- Restauración de ficheros por fecha.
- Guardado de backup en múltiples cintas, es decir cuando llena una, pide otra y continua.
- Posibilidad de habilitar un buffer en disco para hacer backup de forma que la cinta no sufra de “shoe shine”.
- Puede almacenar los backups en DVDs.
- Todos los bloques de datos guardados contienen un checksum.
- Es posible consultar la base de datos de ficheros almacenados por Bacula.
- Permite la ejecución de scripts antes y después de los trabajos de almacenamiento.
- Gestión de errores.

Nuestra experiencia con Bacula

1/5 (Arquitectura)

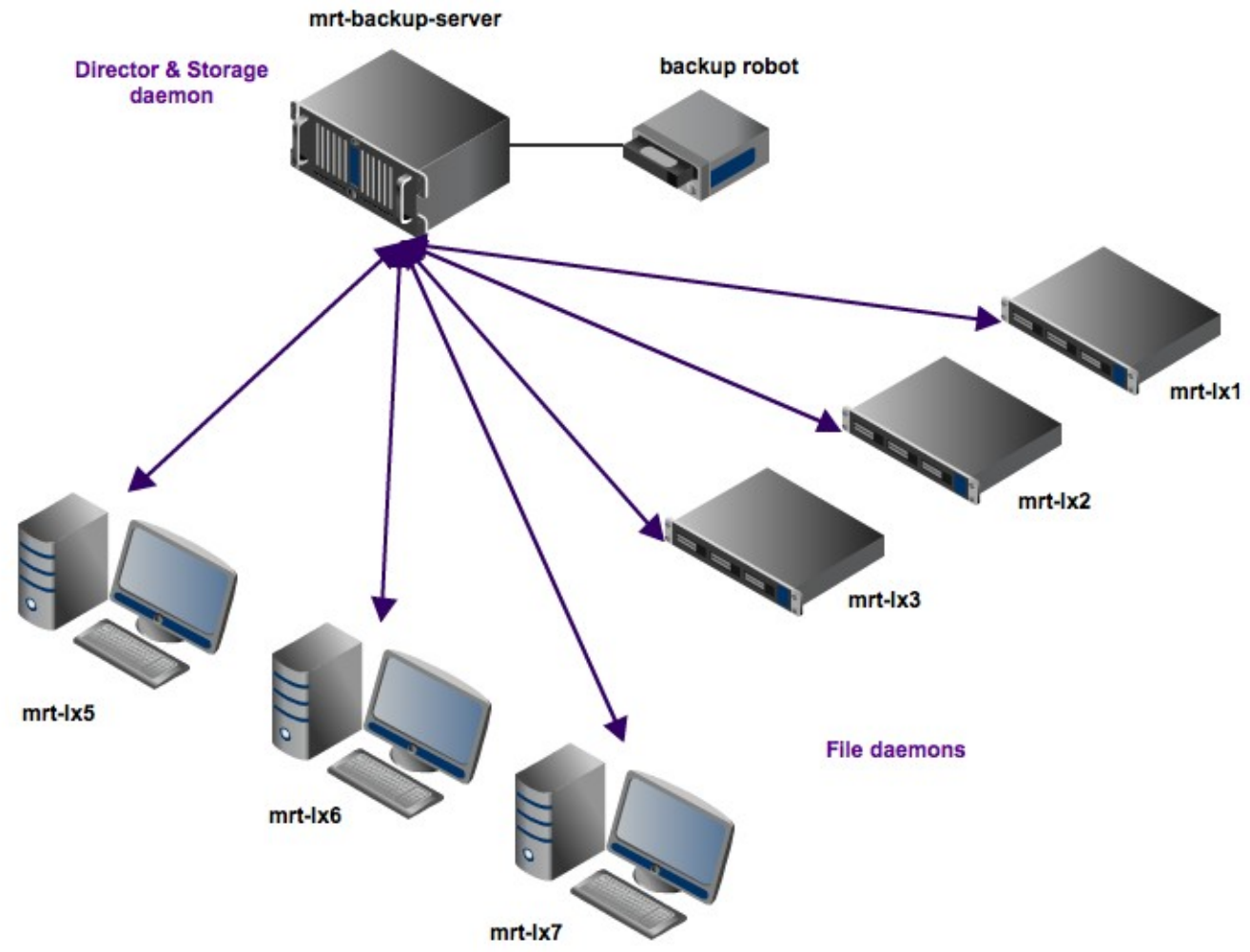
- Nuestra configuración en la oficina de Granada:



Nuestra Experiencia con Bacula

2/5 Arquitectura

Nuestra configuración en el telescopio:



Nuestra Experiencia con Bacula

3/5 (Nuestros objetos)

Objetos definidos en la oficina:

- **Schedule:**

- MonthlySave: Backup total segundo viernes de cada mes a las 1:10.
- NightlySaveUsers:
 - Backup total el primer sábado de cada mes a las 1:05
 - Backup incremental de lunes a sábado a las 2:05
- WeeklySaveProjectArchive:
 - Backup total el segundo sábado de enero y julio a las 1:05.
 - Backup incremental los miércoles a las 1:05.

- **FileSet:**

- VarEtcHome: /var, /etc & /home
- VarEtcHomeUsrlocal: /var, /etc, /home & /usr/local
- ProjectArchive: /prArch
- Users: /users

Nuestra Experiencia con Bacula

4/5 (Nuestros objetos)

- **Pool:**
 - Daily: sus cintas son GRIN01L3 y GRIN02L3, pueden ser recicladas después de 60 días.
 - Monthly: sus cintas son GRFU01L3 y GRFU02L3, pueden ser recicladas después de 1 año.
 - ProjectArchive: sus cintas son PRA01L3 y PRA02L3, pueden ser recicladas después de 4 meses.
- **Device:**
 - StorageServer1Drive: el dispositivo físico es /dev/nst0 de tipo LTO-3 y con directorio de spool en /var/lib/bacula/spool.
 - StorageServer2Drive: el dispositivo físico es /dev/nst1 de tipo LTO-3 y con directorio de spool en /backup/bacula-spool.
- **Storage:**
 - StorageServer1: su FQDN es gra-storage-server1.iram.es y usa el **Device** StorageServer1Drive.
 - StorageServer2: su FQDN es gra-storage-server2.iram.es y usa el **Device** StorageServer2Drive.

Nuestra Experiencia con Bacula

5/5 (Nuestros objetos)

- **Job:**
 - Users: se ejecuta con el **Schedule** *NightlySaveUsers*, copia los ficheros definidos en el **FileSet** *Users* en *StorageServer1* y usando el **Pool** *Daily* cuando el backup es incremental, y el **Pool** *Monthly* cuando el backup es total.
 - Cliente1: se ejecuta con el **Schedule** *MonthlySave*, copia los ficheros definidos en el **FileSet** *VarEtcHome* en *StorageServer1* y usando el **Pool** *Monthly*.
 - ProjectArchive: se ejecuta con el **Schedule** *WeeklySaveProjectArchive*, copia los ficheros definidos en el **FileSet** *ProjectArchive* en *StorageServer2* y usando el **Pool** *ProjectArchive*.
- **Catalog:** se almacena en gra-backup-server y en una base de datos llamada *bacula*.
- **Messages:** enviará los emails al alias bacula@iram.es y con prefijo [IRAM Granada Backup] incluido en el subject.

Ventajas e Inconvenientes

- **Ventajas:**

- Muy buena gestión de errores y feedback por parte del sistema.
- Gestión eficiente de las cintas, no se desperdicia espacio.
- Excelente documentación con un manual muy completo de más de 700 páginas.
- Etiquetado automatico de cintas usando un robot con lector de códigos de barras.

- **Inconvenientes:**

- Formato de escritura en cinta propio, no es tan accesible como un tar a una cinta.
- Configuración a base de ficheros de texto, no existe una GUI que ayude a realizar esta tarea.
- Curva de aprendizaje bastante pronunciada.

Gracias por su atención