Subject: [BUG][cryo] Create file on restart ?
Posted by Sukadev Bhattiprolu on Wed, 16 Jul 2008 18:50:27 GMT
View Forum Message <> Reply to Message

cryo does not (cannot ?) recreate files if the application created
a file before checkpoint and the file does not exist at the time
of restart.

Note that the 'flags' field in '/proc/$pid/fdinfo/$fd' will not
have the O_CREAT (or O_TRUNC, O_EXCL, O_NOCTTY) flags. These
are cleared in __dentry_open()).

At the time of restart, is there a way for cryo to know that the
file must be created ?

To reproduce:
 - run following program,
 - checkpoint after the first printf
 - rm /tmp/foo1
 - restart # fails to open file during restart

```
#include <stdio.h>
#include <unistd.h>
#include <errno.h>
#include <sys/fcntl.h>

main()
{
 int fd;
 int i;
 char *buf = "abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ";

 fd = open("/tmp/foo1", O_RDWR|O_CREAT|O_TRUNC, 0666);

 if (fd < 0) {
  perror("open");
  exit(1);
 }
 printf("%d: Opened '/tmp/foo1', fd %d\n", getpid(), fd);

 for (i = 0; i < strlen(buf); i++) {
  if (write(fd, &buf[i], 1) < 0) {
   printf("Error %d writing %c to file, i %d\n",
     errno, buf[i], i);
   exit(1);
  }
  printf("%d: i %d, wrote %c\n", getpid(), i, buf[i]);
```

```
  sleep(2);
 }
}
```

_____
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers

---

Subject: Re: [BUG][cryo] Create file on restart ?
Posted by serue on Wed, 16 Jul 2008 19:26:04 GMT
View Forum Message <> Reply to Message

Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
>
> cryo does not (cannot ?) recreate files if the application created

I think that's for the best.

Don't you?

The most we should do is make sure cryo has a clear enough error
message.

-serge

> a file before checkpoint and the file does not exist at the time
> of restart.
>
> Note that the 'flags' field in '/proc/$pid/fdinfo/$fd' will not
> have the O_CREAT (or O_TRUNC, O_EXCL, O_NOCTTY) flags. These
> are cleared in __dentry_open()).
>
> At the time of restart, is there a way for cryo to know that the
> file must be created ?
>
> To reproduce:
> - run following program,
> - checkpoint after the first printf
> - rm /tmp/foo1
> - restart # fails to open file during restart
>
> ---
> #include <stdio.h>
> #include <unistd.h>
> #include <errno.h>
> #include <sys/fcntl.h>
>

```
> main()
> {
>  int fd;
>  int i;
>  char *buf = "abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ";
>
>  fd = open("/tmp/foo1", O_RDWR|O_CREAT|O_TRUNC, 0666);
>
>  if (fd < 0) {
>   perror("open");
>   exit(1);
>  }
>  printf("%d: Opened '/tmp/foo1', fd %d\n", getpid(), fd);
>
>  for (i = 0; i < strlen(buf); i++) {
>   if (write(fd, &buf[i], 1) < 0) {
>    printf("Error %d writing %c to file, i %d\n",
>      errno, buf[i], i);
>    exit(1);
>   }
>   printf("%d: i %d, wrote %c\n", getpid(), i, buf[i]);
>   sleep(2);
>  }
> }
> _____________________________________________________
> Containers mailing list
> Containers@lists.linux-foundation.org
> https://lists.linux-foundation.org/mailman/listinfo/containers
```

_____

Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers

---

Subject: Re: [BUG][cryo] Create file on restart ?
Posted by Sukadev Bhattiprolu on Wed, 16 Jul 2008 20:45:29 GMT
View Forum Message <> Reply to Message

Serge E. Hallyn [serue@us.ibm.com] wrote:
| Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
| >
| > cryo does not (cannot ?) recreate files if the application created
|
| I think that's for the best.
|
| Don't you?

I can understand that configuration or data files should exist, but

not sure about temporary or log files that an application created
upon start-up and expects to be present. Should the admin find
out about them and create them by hand before restart ?

_____
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers

---

Subject: Re: [BUG][cryo] Create file on restart ?
Posted by serue on Wed, 16 Jul 2008 20:57:37 GMT
View Forum Message <> Reply to Message

Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
> Serge E. Hallyn [serue@us.ibm.com] wrote:
> | Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
> | >
> | > cryo does not (cannot ?) recreate files if the application created
> |
> | I think that's for the best.
> |
> | Don't you?
>
> I can understand that configuration or data files should exist, but
> not sure about temporary or log files that an application created
> upon start-up and expects to be present. Should the admin find
> out about them and create them by hand before restart ?

I think the admin should have set the destination environment such that
the task is restarted in the same network fs in the same directory, with
no files having been deleted.

Am I wrong?

-serge

_____
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers

---

Subject: Re: [BUG][cryo] Create file on restart ?
Posted by Matt Helsley on Wed, 16 Jul 2008 20:59:23 GMT
View Forum Message <> Reply to Message

On Wed, 2008-07-16 at 14:26 -0500, Serge E. Hallyn wrote:
> Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):

> >
> > cryo does not (cannot ?) recreate files if the application created
>
> I think that's for the best.
>
> Don't you?

I agree. I think drawing the line for process checkpoint/restart before preserving the contents of mounted filesystems is very reasonable since mounted filesystem(s) can already be preserved with your choice of tool(s). I think it also gives us more options as far as using checkpointed images for error recovery; if we take the concept of checkpoint too far we may limit ourselves to merely reproducing errors rather than also giving ourselves a means to recover from errors.

Cheers,
-Matt Helsley

> -serge
>
> > a file before checkpoint and the file does not exist at the time
> > of restart.
> >
> > Note that the 'flags' field in '/proc/$pid/fdinfo/$fd' will not
> > have the O_CREAT (or O_TRUNC, O_EXCL, O_NOCTTY) flags. These
> > are cleared in __dentry_open()).
> >
> > At the time of restart, is there a way for cryo to know that the
> > file must be created ?
> >
> > To reproduce:
> > - run following program,
> > - checkpoint after the first printf
> > - rm /tmp/foo1
> > - restart # fails to open file during restart
> >
> > ---
> > #include <stdio.h>
> > #include <unistd.h>
> > #include <errno.h>
> > #include <sys/fcntl.h>
> >
> > main()
> > {
> > int fd;
> > int i;
> > char *buf = "abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ";
> >

> > fd = open("/tmp/foo1", O_RDWR|O_CREAT|O_TRUNC, 0666);
> >
> > if (fd < 0) {
> >   perror("open");
> >   exit(1);
> > }
> > printf("%d: Opened '/tmp/foo1', fd %d\n", getpid(), fd);
> >
> > for (i = 0; i < strlen(buf); i++) {
> >   if (write(fd, &buf[i], 1) < 0) {
> >   printf("Error %d writing %c to file, i %d\n",
> >   errno, buf[i], i);
> >   exit(1);
> >   }
> >   printf("%d: i %d, wrote %c\n", getpid(), i, buf[i]);
> >   sleep(2);
> > }
> > }
> > _____
> > Containers mailing list
> > Containers@lists.linux-foundation.org
> > https://lists.linux-foundation.org/mailman/listinfo/containers
> _____
> Containers mailing list
> Containers@lists.linux-foundation.org
> https://lists.linux-foundation.org/mailman/listinfo/containers

_____
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers

---

Subject: Re: [BUG][cryo] Create file on restart ?
Posted by Sukadev Bhattiprolu on Wed, 16 Jul 2008 21:26:09 GMT
View Forum Message <> Reply to Message

Serge E. Hallyn [serue@us.ibm.com] wrote:
| Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
| > Serge E. Hallyn [serue@us.ibm.com] wrote:
| > | Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
| > | >
| > | > cryo does not (cannot ?) recreate files if the application created
| > |
| > | I think that's for the best.
| > |
| > | Don't you?
| >

| > I can understand that configuration or data files should exist, but
| > not sure about temporary or log files that an application created
| > upon start-up and expects to be present. Should the admin find
| > out about them and create them by hand before restart ?
|
| I think the admin should have set the destination environment such that
| the task is restarted in the same network fs in the same directory, with
| no files having been deleted.

or new files created ? For instance if the application was checkpointed
before it created a temporary file with O_EXCL flag, that temporary
file must not exist when restarting ?

|
| Am I wrong?

So we take a snapshot of the FS and checkpoint the application. Do they
need to be atomic ?

Eitherway, I withdraw the bug :-)

_____
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers

---

Subject: Re: [BUG][cryo] Create file on restart ?
Posted by Matt Helsley on Wed, 16 Jul 2008 22:31:00 GMT
View Forum Message <> Reply to Message

On Wed, 2008-07-16 at 14:26 -0700, sukadev@us.ibm.com wrote:
> Serge E. Hallyn [serue@us.ibm.com] wrote:
> | Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
> | > Serge E. Hallyn [serue@us.ibm.com] wrote:
> | > | Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
> | > | >
> | > | > cryo does not (cannot ?) recreate files if the application created
> | > |
> | > | I think that's for the best.
> | > |
> | > | Don't you?
> | >
> | > I can understand that configuration or data files should exist, but
> | > not sure about temporary or log files that an application created
> | > upon start-up and expects to be present. Should the admin find
> | > out about them and create them by hand before restart ?
> |
> | I think the admin should have set the destination environment such that

> | the task is restarted in the same network fs in the same directory, with
> | no files having been deleted.

[Assuming Serge meant: s/network fs/network, fs,/]

> or new files created ? For instance if the application was checkpointed
> before it created a temporary file with O_EXCL flag, that temporary
> file must not exist when restarting ?

I think that's not a problem given my assumptions above. The filesystem that the application restarts in would be the same because the admin should have set up the restart environment as Serge suggested. The admin can't rely on restart in an alternate environment. However, given knowledge of the application and environment, using an alternate environment may be a risk the admin is willing to take.

> |
> | Am I wrong?
>
> So we take a snapshot of the FS and checkpoint the application. Do they
> need to be atomic ?

If all the applications in a container are frozen then I think we can get fs snapshots consistent with checkpointed applications. Otherwise, yes, I think we'd be gambling that the checkpointed application isn't interacting with another, running, application via an intermittently-shared file.

Cheers,
-Matt

Subject: Re: [BUG][cryo] Create file on restart ?
Posted by Sukadev Bhattiprolu on Wed, 16 Jul 2008 23:20:24 GMT
View Forum Message <> Reply to Message

| If all the applications in a container are frozen then I think we can
| get fs snapshots consistent with checkpointed applications.

I agree in general, but cryo currently takes a checkpoint and the application resumes, which means the application could create the temp file. So, cryo should not resume the application until the FS snapshot is taken too I guess.

| Otherwise, yes, I think we'd be gambling that the checkpointed
| application isn't interacting with another, running, application via an
| intermittently-shared file.

Yes.

_____
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers

---

Subject: Re: [BUG][cryo] Create file on restart ?
Posted by serue on Thu, 17 Jul 2008 02:18:34 GMT
View Forum Message <> Reply to Message

Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
> Serge E. Hallyn [serue@us.ibm.com] wrote:
> | Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
> | > Serge E. Hallyn [serue@us.ibm.com] wrote:
> | > | Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
> | > | >
> | > | > cryo does not (cannot ?) recreate files if the application created
> | > |
> | > | I think that's for the best.
> | > |
> | > | Don't you?
> | >
> | > I can understand that configuration or data files should exist, but
> | > not sure about temporary or log files that an application created
> | > upon start-up and expects to be present. Should the admin find
> | > out about them and create them by hand before restart ?
> |
> | I think the admin should have set the destination environment such that
> | the task is restarted in the same network fs in the same directory, with
> | no files having been deleted.
>
> or new files created ? For instance if the application was checkpointed
> before it created a temporary file with O_EXCL flag, that temporary
> file must not exist when restarting ?
>
> |
> | Am I wrong?
>
> So we take a snapshot of the FS and checkpoint the application. Do they
> need to be atomic ?
>
> Eitherway, I withdraw the bug :-)

Well it's certainly beyond the scope of cryo. I'd prefer if we didn't have to snapshot the fs at each checkpoint (!) and I think any many or most cases (think one long-running scientific app or seti@home that just occasionally gets migrated or stopped for a reboot) it won't be an issue. But your point seems valid.

-serge

Subject: Re: [BUG][cryo] Create file on restart ?
Posted by serue on Thu, 17 Jul 2008 02:21:34 GMT
View Forum Message <> Reply to Message

Quoting Matt Helsley (matthltc@us.ibm.com):
>
> On Wed, 2008-07-16 at 14:26 -0700, sukadev@us.ibm.com wrote:
> > Serge E. Hallyn [serue@us.ibm.com] wrote:
> > | Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
> > | > Serge E. Hallyn [serue@us.ibm.com] wrote:
> > | > | Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
> > | > | >
> > | > | > cryo does not (cannot ?) recreate files if the application created
> > | > |
> > | > | I think that's for the best.
> > | > |
> > | > | Don't you?
> > | >
> > | > I can understand that configuration or data files should exist, but
> > | > not sure about temporary or log files that an application created
> > | > upon start-up and expects to be present. Should the admin find
> > | > out about them and create them by hand before restart ?
> > |
> > | I think the admin should have set the destination environment such that
> > | the task is restarted in the same network fs in the same directory, with
> > | no files having been deleted.
>
> [Assuming Serge meant: s/network fs/network, fs,/]

Well no I meant a network filesystem - at least if you're migrating apps around a cluster.

> > or new files created ? For instance if the application was checkpointed
> > before it created a temporary file with O_EXCL flag, that temporary

> > file must not exist when restarting ?
>
> I think that's not a problem given my assumptions above. The filesystem
> that the application restarts in would be the same because the admin
> should have set up the restart environment as Serge suggested. The admin
> can't rely on restart in an alternate environment. However, given
> knowledge of the application and environment, using an alternate
> environment may be a risk the admin is willing to take.

Yup. But Suka is right that in the case of the checkpointed app
continuing to run for a bit before being killed and restarted, it could
get out of whack with respect to the file system.

> > | Am I wrong?
> >
> > So we take a snapshot of the FS and checkpoint the application. Do they
> > need to be atomic ?
>
> If all the applications in a container are frozen then I think we can
> get fs snapshots consistent with checkpointed applications.
> Otherwise, yes, I think we'd be gambling that the checkpointed
> application isn't interacting with another, running, application via an
> intermittently-shared file.

What fun :)

I wonder whether the experience of users of c/r on sgi and cray could
teach us anything here.

-serge

_____
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers

---

Subject: Re: [BUG][cryo] Create file on restart ?
Posted by Oren Laadan on Thu, 17 Jul 2008 23:22:52 GMT
View Forum Message <> Reply to Message

sukadev@us.ibm.com wrote:
> Serge E. Hallyn [serue@us.ibm.com] wrote:
> | Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
> | > Serge E. Hallyn [serue@us.ibm.com] wrote:
> | > | Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
> | > | >
> | > | > cryo does not (cannot ?) recreate files if the application created
> | > |

> | > | I think that's for the best.
> | > |
> | > | Don't you?
> | >
> | > I can understand that configuration or data files should exist, but
> | > not sure about temporary or log files that an application created
> | > upon start-up and expects to be present. Should the admin find
> | > out about them and create them by hand before restart ?
> |
> | I think the admin should have set the destination environment such that
> | the task is restarted in the same network fs in the same directory, with
> | no files having been deleted.
>
> or new files created ? For instance if the application was checkpointed
> before it created a temporary file with O_EXCL flag, that temporary
> file must not exist when restarting ?
>
> |
> | Am I wrong?
>
> So we take a snapshot of the FS and checkpoint the application. Do they
> need to be atomic ?

Yes they do, in the sense that the FS must be snapshotted when the
container is quiescent to ensure consistency.

Oren.

>
> Eitherway, I withdraw the bug :-)
> _____
> Containers mailing list
> Containers@lists.linux-foundation.org
> https://lists.linux-foundation.org/mailman/listinfo/containers

_____
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers

---

Subject: Re: [BUG][cryo] Create file on restart ?
Posted by Oren Laadan on Thu, 17 Jul 2008 23:35:19 GMT
View Forum Message <> Reply to Message

Serge E. Hallyn wrote:
> Quoting Matt Helsley (matthltc@us.ibm.com):
>> On Wed, 2008-07-16 at 14:26 -0700, sukadev@us.ibm.com wrote:
>>> Serge E. Hallyn [serue@us.ibm.com] wrote:

>>> | Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
>>> | > Serge E. Hallyn [serue@us.ibm.com] wrote:
>>> | > | Quoting sukadev@us.ibm.com (sukadev@us.ibm.com):
>>> | > | >
>>> | > | > cryo does not (cannot ?) recreate files if the application created
>>> | > |
>>> | > | I think that's for the best.
>>> | > |
>>> | > | Don't you?
>>> | >
>>> | > I can understand that configuration or data files should exist, but
>>> | > not sure about temporary or log files that an application created
>>> | > upon start-up and expects to be present. Should the admin find
>>> | > out about them and create them by hand before restart ?
>>> |
>>> | I think the admin should have set the destination environment such that
>>> | the task is restarted in the same network fs in the same directory, with
>>> | no files having been deleted.
>> [Assuming Serge meant: s/network fs/network, fs,/]
>
> Well no I meant a network filesystem - at least if you're migrating apps
> around a cluster.
>
>>> or new files created ? For instance if the application was checkpointed
>>> before it created a temporary file with O_EXCL flag, that temporary
>>> file must not exist when restarting ?
>> I think that's not a problem given my assumptions above. The filesystem
>> that the application restarts in would be the same because the admin
>> should have set up the restart environment as Serge suggested. The admin
>> can't rely on restart in an alternate environment. However, given
>> knowledge of the application and environment, using an alternate
>> environment may be a risk the admin is willing to take.
>
> Yup. But Suka is right that in the case of the checkpointed app
> continuing to run for a bit before being killed and restarted, it could
> get out of whack with respect to the file system.
>
>>> | Am I wrong?
>>>
>>> So we take a snapshot of the FS and checkpoint the application. Do they
>>> need to be atomic ?
>> If all the applications in a container are frozen then I think we can
>> get fs snapshots consistent with checkpointed applications.
>> Otherwise, yes, I think we'd be gambling that the checkpointed
>> application isn't interacting with another, running, application via an
>> intermittently-shared file.
>
> What fun :)

>
> I wonder whether the experience of users of c/r on sgi and cray could
> teach us anything here.

if you are checkpointing to migrate the application - you need not worry
about the file system, as it may not change while you migrate.

if you are checkpointing to be able to be able to recover from an error
later, you need to snapshot the file system, but you may get away with
it in some cases.

if you are checkpointing to be able to travel back in time (return to
older than last checkpoint), you certainly need to snapshot the file system.

in any event, I think this is something we may want to discuss in the
mini-summit.

Oren.

>
> -serge
> _____
> Containers mailing list
> Containers@lists.linux-foundation.org
> https://lists.linux-foundation.org/mailman/listinfo/containers

_____
Containers mailing list
Containers@lists.linux-foundation.org
https://lists.linux-foundation.org/mailman/listinfo/containers