

Der NetWorker 19.2.x+ kann parallele Recover Streams von nur einem einzigen Save Set erzeugen

Anstatt mit diesem neuen Pfund zu wuchern, wird diese interessante neue Eigenschaft in den NetWorker 19.2 Release Notes nur so beiläufig erwähnt, daß sie selbst erfahrenen NetWorker Insidern gar nicht richtig bewußt wurde:

Improve restore performance with multiple threads

Recover workflow for file system backups has been enhanced to perform restores in parallel. This splits recover requests into multiple recover requests resulting in more than one recover stream. This has improved recover performance in comparison to earlier versions.



Das bedeutet, daß vor allem Restores von Save Sets mit sehr vielen kleinen Dateien z.T. wesentlich schneller durchgeführt werden können, als sie beim Backup gedauert haben.

Ursächlich hierfür ist eine neue Option für den NetWorker `recover` Befehl: `recover -z`

Leider ist die Beschreibung in der *NetWorker Command Line Reference* (wie gewohnt) sehr minimalistisch:

```
recover -z          Causes the recover to enable parallel restore.
```

Die eigentlichen Details dieser neuen Eigenschaft finden Sie schließlich im *NetWorker 19.2 Performance Optimization Planning Guide* und dort im Abschnitt *Parallel restore*. Bevor ich auf die praktische Durchführung eingehe, möchte ich die wenigen Zeilen übersetzen, weil in ihnen wichtige Details und Randbedingungen aufgeführt sind.

Parallel Restore

Ab dem NetWorker 19.2 wurde der Wiederherstellungsworkflow für Dateisystem-Sicherungen so erweitert, daß auch parallele Wiederherstellungen durchführbar sind.

Die neue verbesserte Logik versucht, Wiederherstellungsaufgaben in mehrere Wiederherstellungsprozesse aufzuteilen, wodurch eine schnellere Wiederherstellung als nur einem einzigen Stream möglich ist.

Diese neue Eigenschaften werden von Data Domain- und AFTD-Geräten unterstützt:

- *Bei der Wiederherstellung einzelner Dateien und Verzeichnisse (auf der Datei-Ebene)*
- *Bei der Wiederherstellung eines vollständigen Backups (eines Save Sets)*

File level Restore

- Werden Dateien aus einem einzigen Save Set ausgewählt, werden pro Wiederherstellungsprozess maximal vier Threads (Unterprozesse) eröffnet.
- Werden Dateien aus mehreren Sicherungssätzen ausgewählt (d.h. bei weniger als vier Save Sets), werden für jeden Recover Prozess bis zu vier Threads zugewiesen. Für jeden einzelnen Save Set wird genau ein Thread zugewiesen.

CR: Ich denke, diese Aussage bezieht sich auf die PSS Funktionalität des NetWorker.

- Werden Dateien aus mehreren Save Sets ausgewählt werden (d.h. bei mehr als vier Save Sets), wird diese Aufteilung nicht mehr angewendet, denn sie würde die Einstellung für den max. Parallelismus bei Wiederherstellungen überschreiten.
- Werden Dateien von mehreren Save Sets ausgewählt, die alle auf dem gleichen Volume gespeichert sind, wird jedem Save Set genau ein Wiederherstellungs-Thread zugewiesen.
- Werden Dateien von mehreren Save Sets ausgewählt, die auf verschiedenen Volumes gespeichert sind, wird versucht, die Sessions gemäß den eingestellten Sessions für das Laufwerk zu verteilen. Kann dies nicht gleichzeitig geschehen, wird der sog. „safe mode“ angewandt, bei dem jedem Volume nur eine einzige Session zugewiesen wird. So wird sichergestellt, daß die Wiederherstellung erfolgreich durchgeführt werden kann, wenngleich sie so nicht mit optimaler Geschwindigkeit durchgeführt werden kann.

Save set Restore

- Sind die File Index Informationen für einen bestimmten Sicherungssatz online verfügbar, wird versucht, die o.a. Logik der Aufteilung anzuwenden und für den Wiederherstellungsprozess maximal vier Treads zu öffnen.
- Sind die File Index Informationen für einen bestimmten Sicherungssatz jedoch nicht online verfügbar, kann der NetWorker nur die altbekannte Methode verwenden und die Wiederherstellung mit nur einem einzigen Thread durchführen.

CR: Wie Sie sehen, kommt das *Save Set Recover* diesmal nicht ohne Client File Index aus.

Bei einer Wiederherstellung auf Dateiebene werden die Wiederherstellung-Threads zuerst an der Grenze des Sicherungssatzes zugewiesen. Geben Sie also z.B. zwei Save Set IDs zur Wiederherstellung an, wird jedem Save Set zunächst ein Thread zugewiesen - danach erst werden die verbleibenden zwei Threads nach der Wiederherstellungslogik zugewiesen.

Die Funktion zur parallelen Wiederherstellung ist standardmäßig nicht aktiviert. Führen Sie einen der folgenden Schritte aus, um die parallele Wiederherstellung zu aktivieren:

- Führen Sie den Wiederherstellungsbefehl `recover` mit der Option `-z` aus.
- Aktivieren Sie in der NetWorker Management Console (NMC) bzw. in der NetWorker Admin GUI das Kontrollkästchen **Advanced Options** und tragen Sie im Edit Feld **Extra recover options** die Option `-z` ein (siehe Beispiel auf Seite 5).



Als erfahrener NetWorker Administrator wissen Sie natürlich, daß Dell EMC hiermit nicht das Fenster NMC (**NetWorker Management Console**), sondern vielmehr das Fenster **NetWorker Administration** meint. Die Fenstertitel unterstreichen diese Tatsache eindeutig.

Leider hat sich dieser Kardinalfehler über Jahre in der NetWorker Literatur manifestiert und er behauptet sich hartnäckig. Doch warum sollten die Schreiber der NetWorker Literatur auch tatsächlich das beschreiben, was eindeutig zu sehen ist???

Parallel restore - Empfehlungen

Denken Sie daran:

- Die parallele Wiederherstellung ist am vorteilhaftesten, wenn ein Dateisystem mit hoher Dichte (ein Dateisystem bestehend aus einer großen Anzahl kleiner Dateien, wiederhergestellt wird..
- Sowohl die Wiederherstellung auf Dateiebene als auch die Wiederherstellung des Sicherungssatzes können bei paralleler Wiederherstellung eine erhebliche Wiederherstellungsleistung erzielen. Dell EMC empfiehlt, für eine schnellere Wiederherstellung die parallele Wiederherstellung zu verwenden.
- Die parallele Wiederherstellung ist bei der Wiederherstellung auf Dateiebene 2,5-mal schneller und bei der Wiederherstellung von Sicherungssätzen auf einer Windows-Plattform 3,3-mal schneller.
- Die parallele Wiederherstellung ist mit der Wiederherstellung auf Dateiebene 1,7-mal schneller und mit der Wiederherstellung von Sicherungssätzen auf einer Linux-Plattform 2,2-mal schneller.
- Dell EMC empfiehlt, dass Sie die parallele Wiederherstellung mit der Wiederherstellung auf Dateiebene auf einer Windows-Plattform verwenden, wenn Daten wiederhergestellt werden sollen, die weniger als 50% der gesamten gesicherten Daten betragen.
- Verwenden Sie auf einer Windows-Plattform keine parallele Wiederherstellung mit Wiederherstellung auf Dateiebene, wenn die zu erfassenden Daten weniger als 50% der Daten ausmachen Gesamtzahl der gesicherten Daten betragen.
- Um die parallele Wiederherstellung auf einer Windows-Plattform zu verwenden, benötigen Sie 1,5-mal mehr Speicher für die Wiederherstellung auf Dateiebene und 2,7-mal mehr für den Save Set Restore. Unter Linux beträgt diese Messung 1,2x für die Wiederherstellung auf Dateiebene und 2,6x für die Wiederherstellung von ganzen Save Sets.
- Sowohl für die Wiederherstellung auf Dateiebene als auch für die Wiederherstellung von Sicherungssätzen beträgt die CPU-Auslastung auf einer Windows-Plattform das 3,-fache und das 3,8-fache mehr auf einer Linux-Plattform.
- Verwenden Sie keine parallele Wiederherstellung, wenn Sie vier oder mehr Sicherungssätze wiederherstellen. Bei vier oder mehr Speichersätzen fällt die Wiederherstellung zurück zu der früheren Methode und verwendet einen Thread pro Sicherungssatz.
- Verwenden Sie keine parallele Wiederherstellung, wenn Daten mit DPSS gesichert werden.
(CR: Paralleler Speichersatz)
- Die parallele Wiederherstellung funktioniert am besten mit der komprimierten Wiederherstellungsfunktion, die beim Wiederherstellen von Daten aus der Data Domain aktiviert ist. Deshalb sollten Sie die Komprimierung nicht deaktivieren.
- Um die parallele Wiederherstellung effizient zu nutzen, sollte der NetWorker Client über mindestens 4GB RAM und 4 vCPUs verfügen.

Das folgende Beispiel wurde mit folgenden virtuellen Servern erstellt:

NetWorker Server
Windows 2019
NetWorker 19.2.1.0
2 virtuelle CPUs
8 GB RAM

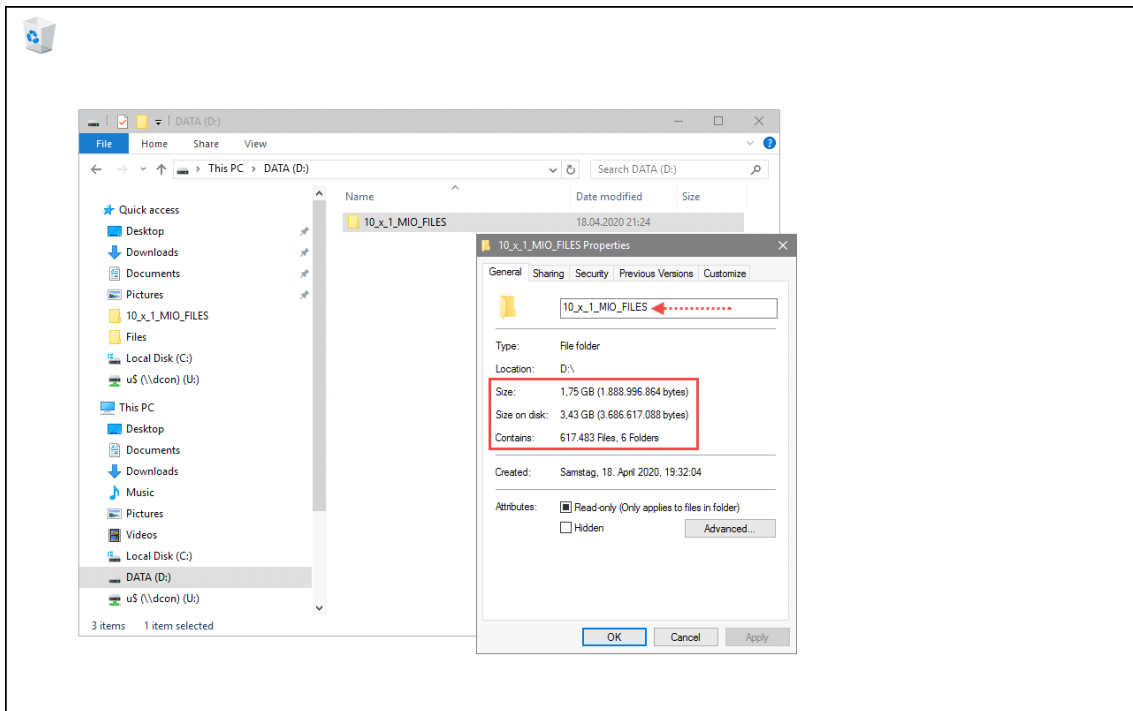
NetWorker Client
Windows 2016
NetWorker 19.2.1.0
2 virtuelle CPUs
8 GB RAM

In diesem Fall weicht also meine Testumgebung ein wenig von den empfohlenen Einstellungen ab. Aber:

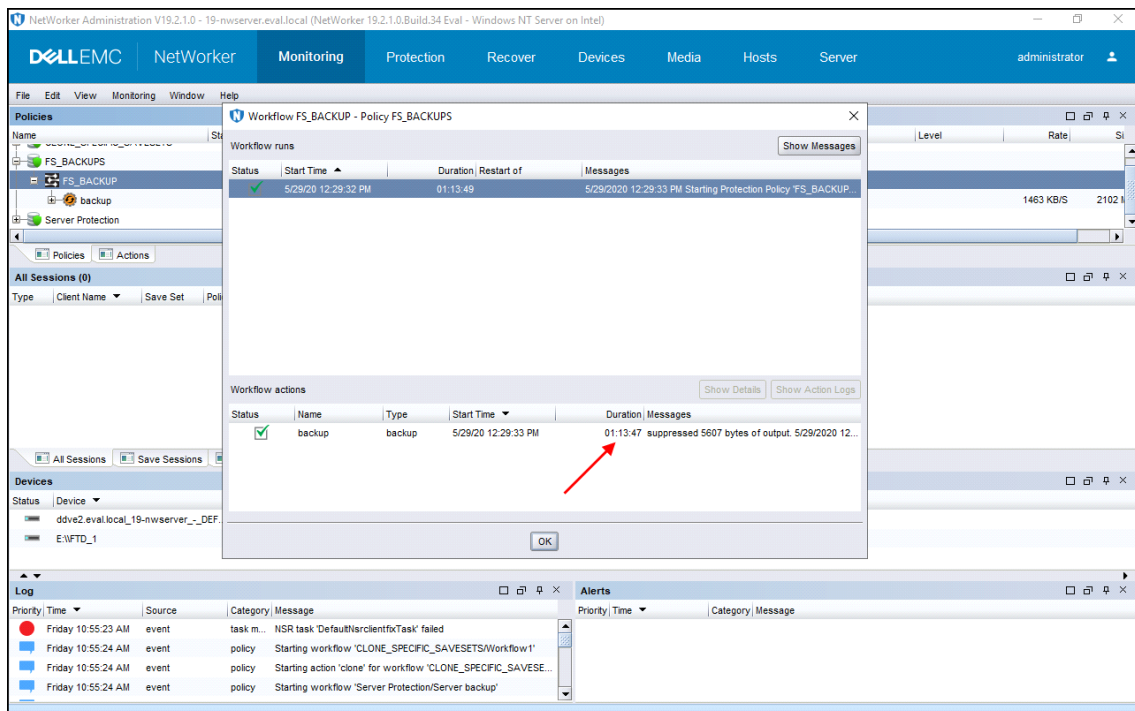


Mit ist klar, daß die erzielten Zeiten/Geschwindigkeiten ohnehin nur bedingt aussagekräftig ein können - mir geht es hauptsächlich darum, Ihnen das Funktionsprinzip zu erklären. Für wirkliche Benchmark-Tests ist meine Umgebung viel zu klein.

So sieht das Dateisystem des Volumes D: am NetWorker Client aus, das ich als Testobjekt verwenden möchte. Wie Sie sehen, enthält es sehr viele kleine Dateien:



Die ‚normale‘ Vollsicherung mit nur einem einzigen Stream auf eine virtuelle Data Domain braucht natürlich seine Zeit - vor allem dann, wenn sich die gesamte Testumgebung auf dem gleichen physikalischen Server befindet:



Dies sind die Informationen zum Save Set und zum Client File Index:

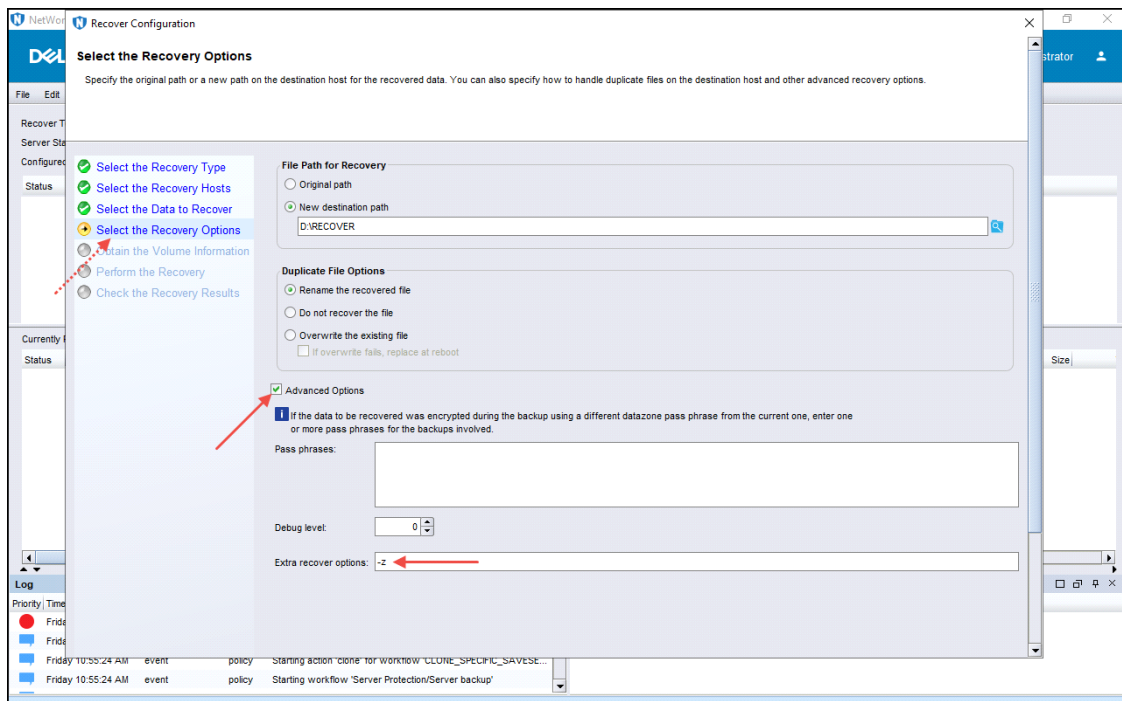
```

Administrator: Command Prompt
C:\>
C:\>
C:\>
C:\>wminfo -avot
volume      type      client      date      time      size  ssid      fl  lvl name
DEFAULT.001 Data Domain 16-client1 5/29/2020 12:30:11 2102 MB 4174439477 cb full D:\

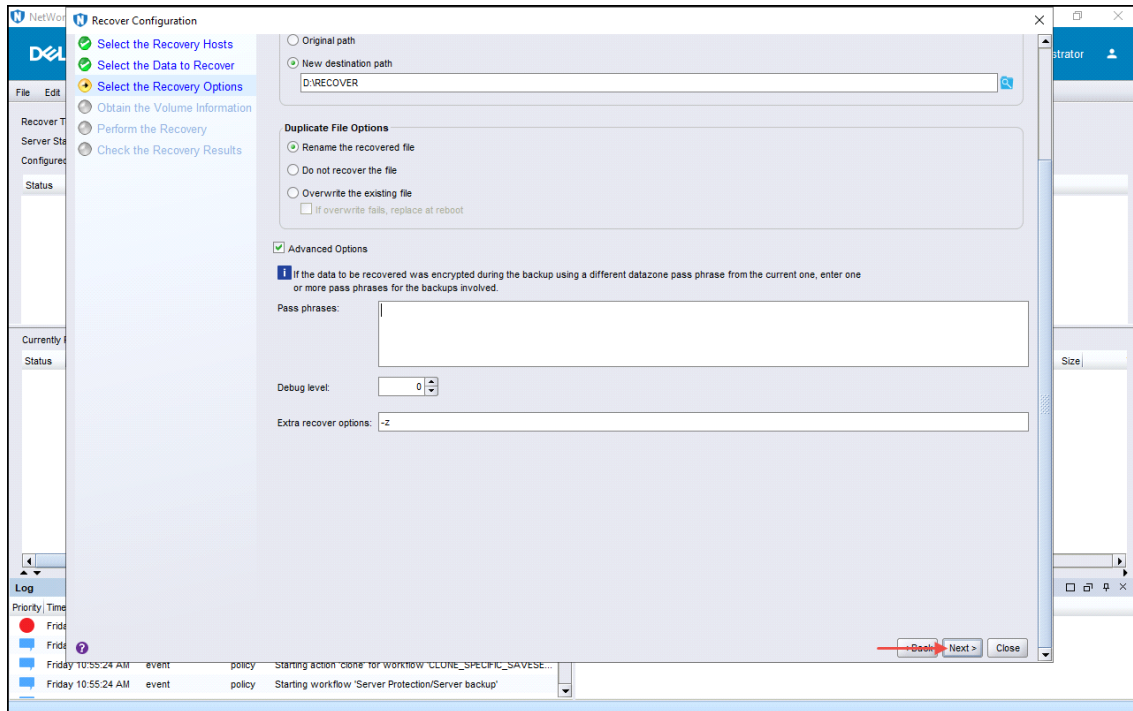
C:\>
C:\>
C:\>nsrsls 16-client1
E:\nsr\index\16-client1: 619448 records requiring 89 MB
E:\nsr\index\16-client1 is currently 100% utilized

C:\>
C:\>
  
```

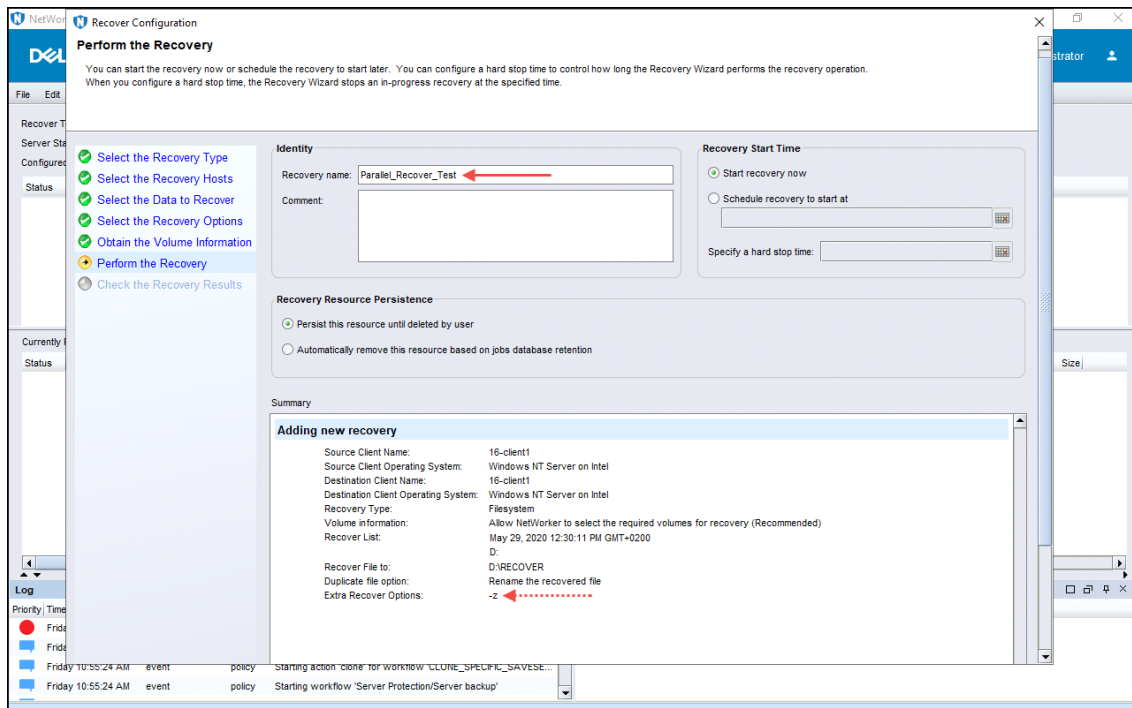
Und so aktivieren Sie im **NetWorker Administrator**, und dort im **Recover Wizard**, die neue Eigenschaft:



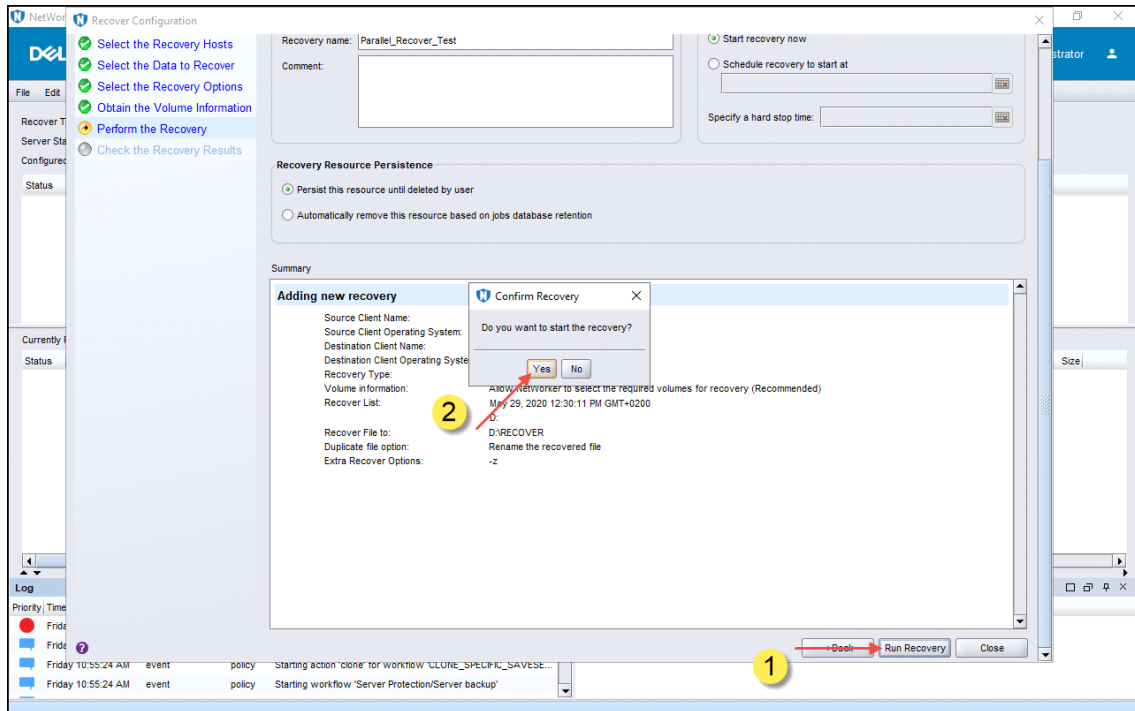
Fahren Sie mit **Next >** fort ...



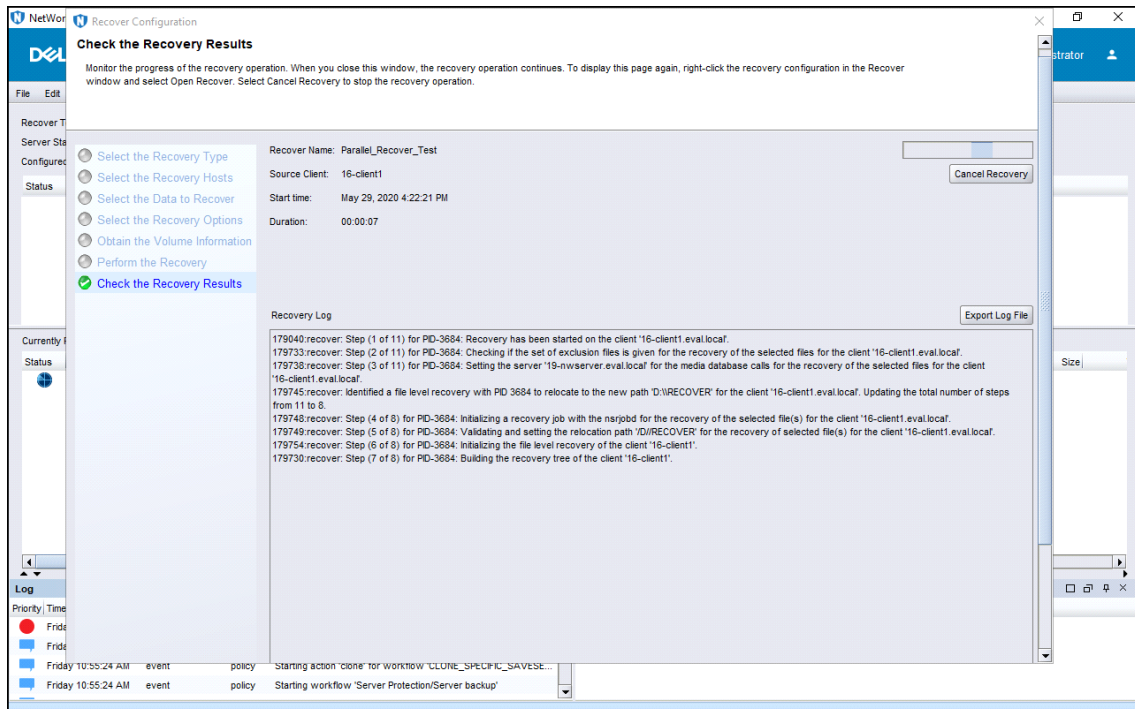
... und geben Sie schließlich diesem Vorgang einen eindeutigen Namen:



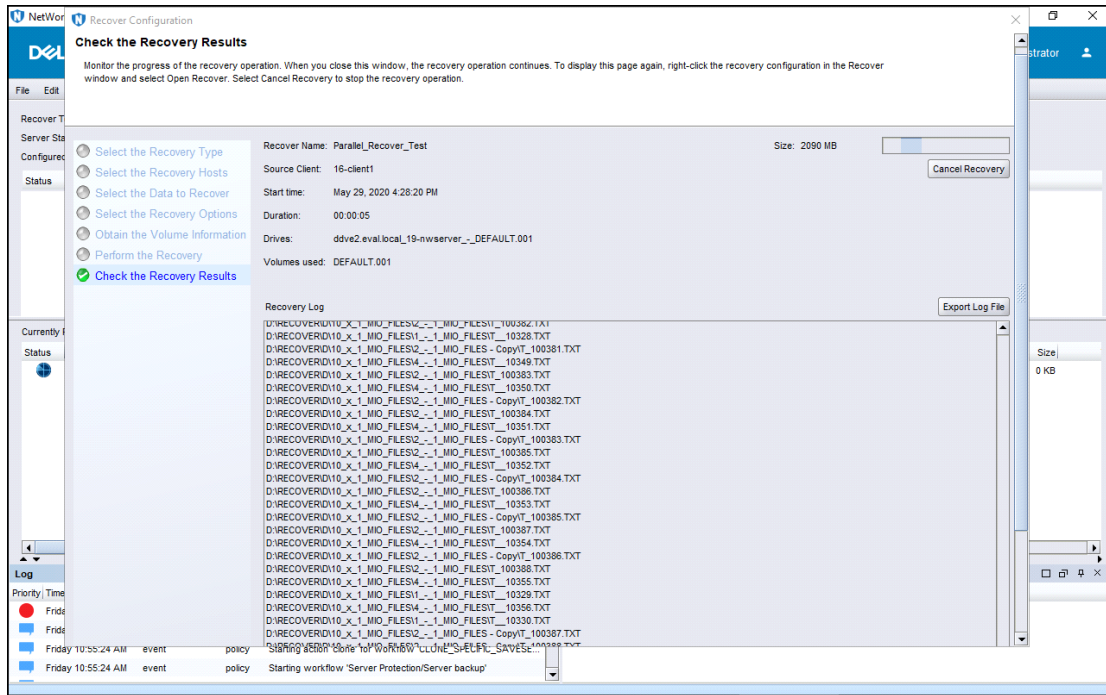
Starten Sie schließlich den Prozess und bestätigen Sie Ihren Wunsch wie üblich:



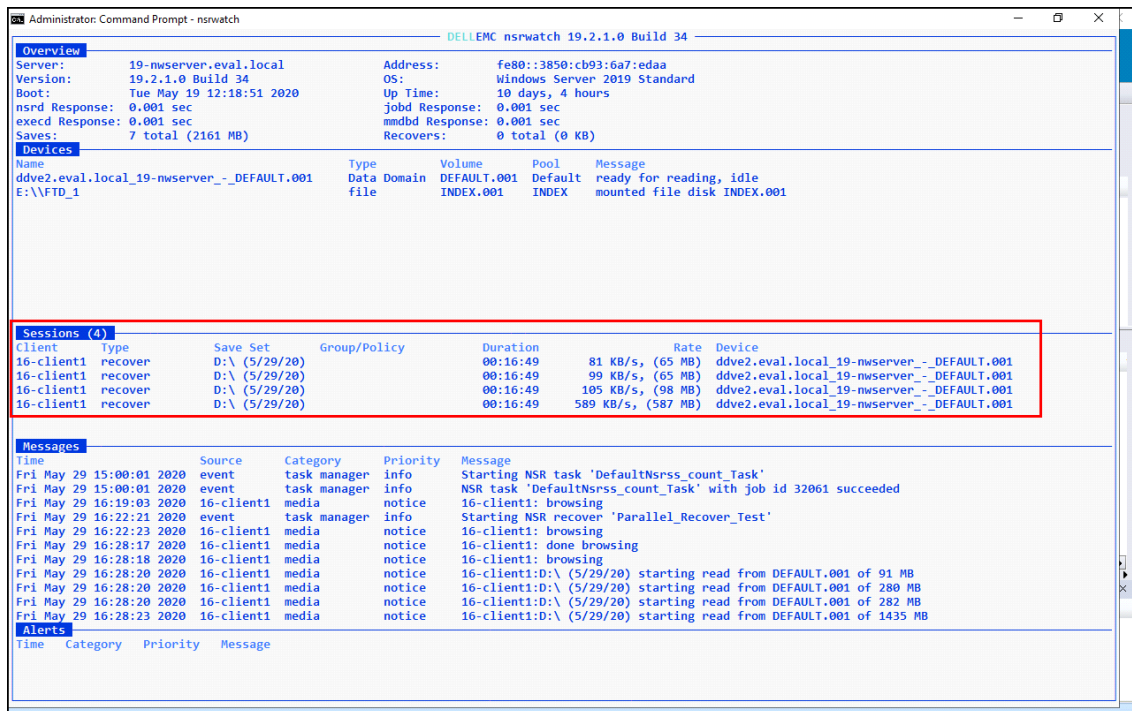
Jetzt findet eine etwas längere Phase der ‚Einleitung‘ statt. Ich denke, daß während dieser Zeit die Aufteilung der Dateien auf die einzelnen Threads erfolgt:



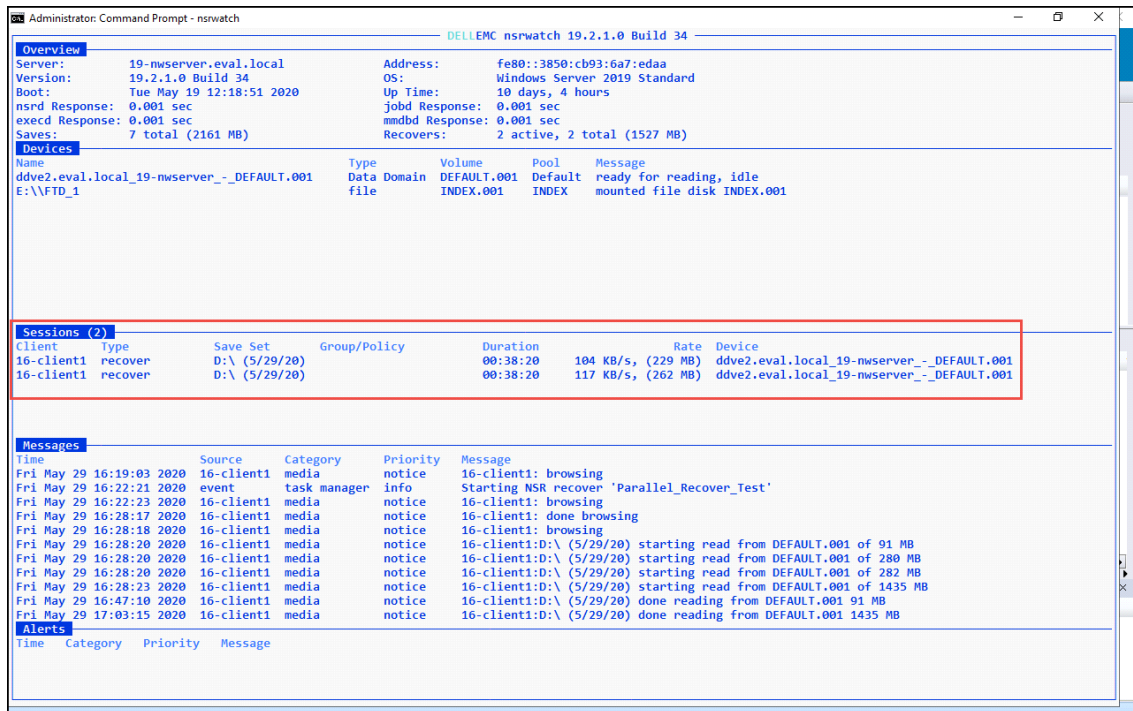
Schließlich beginnt die Wiederherstellung. Allerdings kann sich die Fortschrittsanzeige immer nur auf einen einzelnen Thread beziehen:



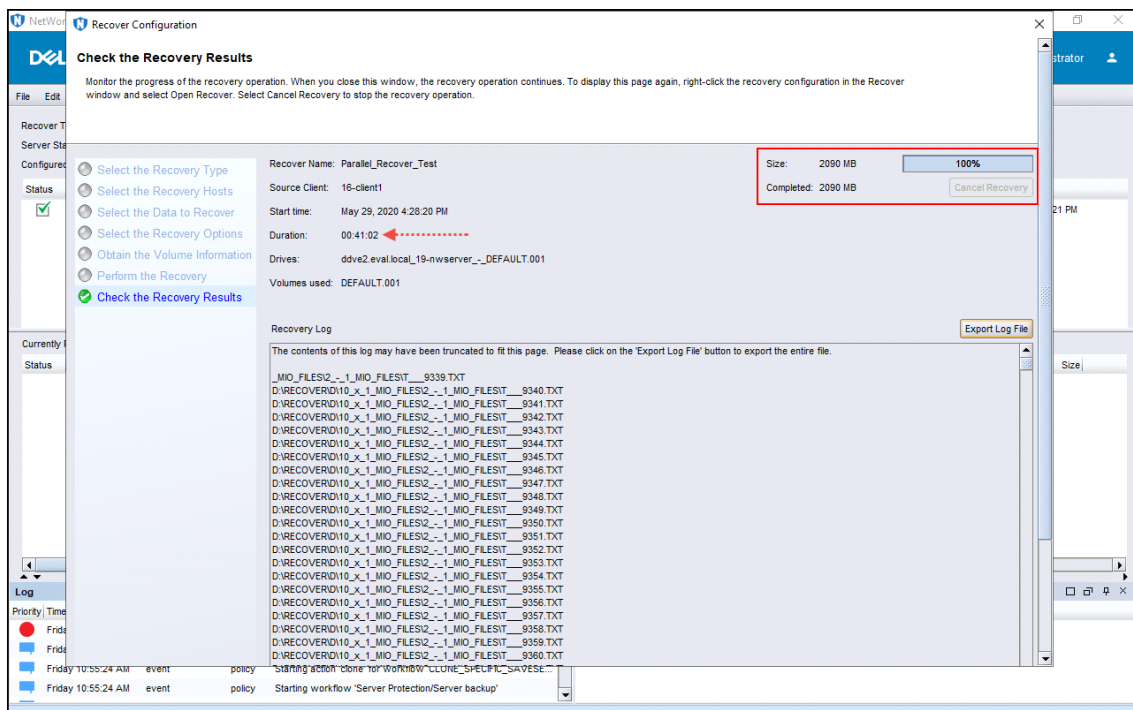
Mit dem Programm `nsrwatch` können Sie jedoch leicht verifizieren, daß auch wirklich bis zu 4 Threads gestartet wurden und auch aktiv sind:



Schließlich neigt sich der gesamte Vorgang seinem Ende zu und die Threads werden nacheinander abgebaut:

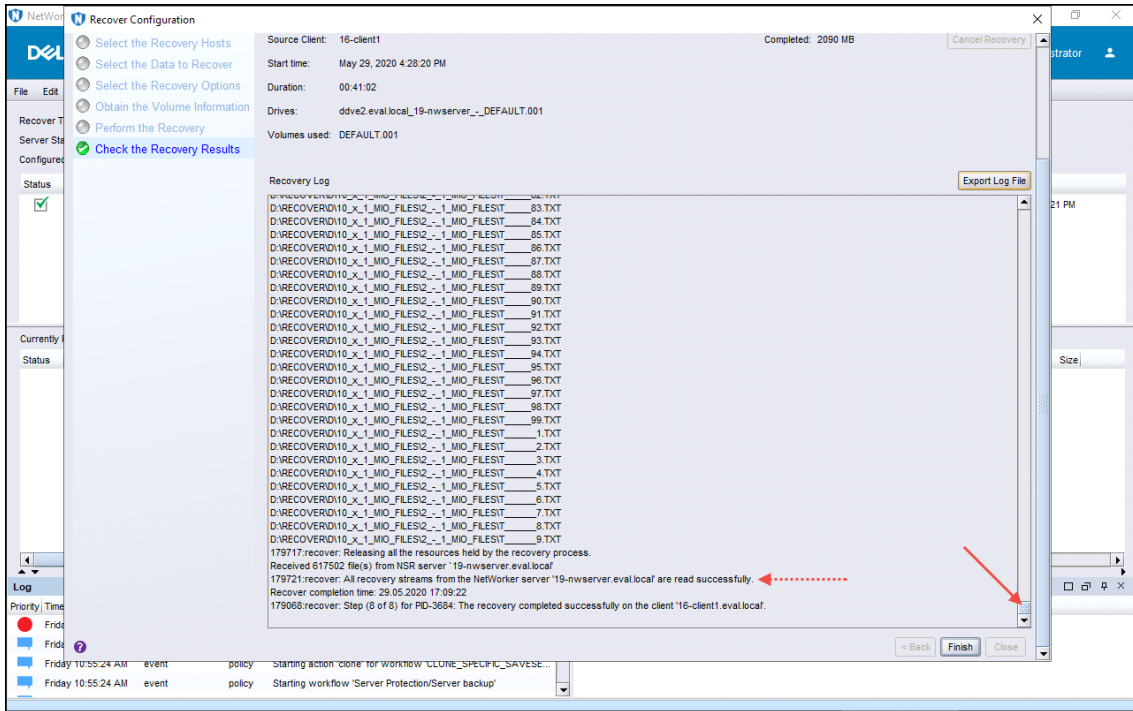


Zuletzt stellt sich das Ergebnis im *Recovery Wizard* wie gewohnt für den ganzen Prozess dar:



Jetzt können Sie die Zeiten vergleichen - in diesem Fall war die Wiederherstellung sehr viel schneller als die Sicherung (siehe Seite 5).

Und trotzdem ist der Vorgang erfolgreich abgeschlossen worden.



In der Liste aller Recovery Prozesse sehen Sie noch einmal die einzelnen Zeiten:

