

שאלות ותשובות בנושא שיחזור מידע והתקני אחסון.

מאוד יתכן שלא שמעת אפילו על שחזור מידע עד שהצורך בו פתאום עלה. בנוסף, יש הרבה בלבול בתחום הזה, עם הרבה פרסום סותר שם בחוץ. העובדה היא, שעקב פרסום ספאם של חברות מסוימות יש אולי מידע מוטעה יותר מאשר מידע אמין בחוץ על ענף זה. מדריך שאלות ותשובות זה נועד לעזור לענות על כמה שאלות עם תשובות פשוטות וישירות. אם יש לך שאלה שלא נדונה במאמר זה, אנא פרסם אותה בתגובות ונשתדל להרחיב עליה בהמשך. אם אינך מסכים עם משהו שנאמר בו, אנא פרסם גם את זה ונוכל לדון בו.

מהו שיחזור מידע?

שחזור מידע כרוך בשיטות שונות לקריאת מידע באמצעים ושיטות לא סטנדרטיות. בהתאם למקרה, זה עשוי לכלול שינויים באופן פעולת החומרה כדי לחלץ נתונים ממכשירי אחסון שנכשלו או לקרוא ולנתח נתונים מחוץ למערכת הקבצים הרגילה. יש אנשים המשתמשים בביטוי זה כדי להתייחס לאחזור נתונים המאוחסנים בגיבוי.

מהו שיחזור מידע לוגי?

מקרה בו אובדן נתונים נגרם, לא כתוצאה מכשל בחומרה, אלא בגלל תקלה פשוטה, מחיקה, חבלה וכו', מכונה שחזור מידע לוגי. כאשר תקלה כוללת חומרה כושלת, היא בדרך כלל נחשבת למקרה התאוששות ברמת החומרה ולא כלוגית. עם זאת, גם במקרה של התאוששות חומרה, ייתכן שעדיין נחוץ שחזור לוגי לאחר השלמת שחזור החומרה.

מהי מערכת קבצים או File System?

מערכת קבצים היא מערכת של מבנים המאוחסנים בכונן לוגי המשמשת לבקרה כיצד ואיפה מאוחסנים קבצים על ידי מערכת הפעלה. זה כולל טבלה או ספרייה של אילו מגזרים (סקטורים) לוגיים (LBA), קבצים בודדים מאוחסנים, כמו גם מבני תיקיות, אינדקסים ומיפוי מגזרים. באופן הזה מערכת ההפעלה עוקבת אחר כל הנתונים בכונן. אנלוגיה קלה היא לחשוב על זה כמו ספר טלפונים המשמש לחיפוש היכן נשמרים קבצים.

מהי גילוף קבצים (File Carving) או שחזור גולמי (Raw)?

גילוף קבצים כרוך בסריקת הנתונים הגולמיים של כונן המחפש חתימות פתיחה, קוד ייחודי לתחילת סוג קובץ ספציפי, ואז שימוש בחתימת סגירה או חתימת פתיחה של הקובץ הבא בכדי לאתר קירוב סוף כל קובץ.

השיטה המועדפת לשחזור מידע היא לנתח את מערכת הקבצים המקורית ולקבוע היכן אוחסנו קבצים שאבדו. עם זאת, אם יותר מדי ממערכת הקבצים אובדת, גילוף קבצים/שחזור RAW עשויים להיות האפשרות היחידה. החיסרון של גילוף קבצים הוא ששמות קבצים מקוריים ומבנה התיקיות יאבדו. כמו כן, לא לכל סוגי הקבצים חתימות פתיחה ניתנות לזיהוי.

מהו כונן/דיסק קשיח (HDD או Hard Disk Drive)?

כונן דיסק קשיח, כונן קשיח או דיסק קשיח מתייחס בדרך כלל למכשיר לאחסון נתונים המשתמש בפלטת מסתובבת (דיסק) עם ציפוי מצע המאפשר שמירה של אותות מגנטיים. לדיסק הקשיח יש מכלול של מערום ראשים, הכולל ראשי קריאה/כתיבה פרטניים אשר יכולים לקרוא ולכתוב מטענים מגנטיים לציפוי הפלטות.

מהו כונן SSD או דיסק מצב מוצק?

כונן SSD הוא התקן לאחסון נתונים המיועד להחליף את הפונקציה של HDD, ושמשמש ברכיבי מצב מוצק ולא בפלטת מכאנות מסתובבות. בדרך כלל אלה משתמשים בזיכרון פלאש של NAND, עם זאת, ישנם כמה סוגים חדשים יותר המשתמשים בזיכרון 3D XPoint.

מדוע נדרש חדר נקי לשחזור מידע?

כוננים קשיחים רגישים ביותר לאבק או לזיהום אחר ובמידת הצורך, יש לפתוח אותם אך ורק בסביבה בעלת סינון מיוחד (HEPA). אם רכיבים פנימיים של דיסק קשיח כגון ראשי קריאה/כתיבה כשלו, יש צורך בסביבה נקייה במהלך ההחלפה.

לעתים קרובות מאוד שחזור מידע אינו מתבצע בפועל בחדר נקי כשלעצמו, אלא במתקן נקי שלטענת רבים הוא טוב יותר. ברדסים נקיים מסוגלים בדרך כלל לעמוד בניקיון גבוה יותר מאשר בחדרים נקיים שעלולים להיות מזוהמים ביתר קלות על ידי דייריהם האנושיים.

עבור סוגים רבים של שחזור מידע, חדר נקי אינו הכרחי. שחזור מידע הכרוך בתיקון קושחה בלבד, תיקון והחלפת PCB או שחזור לוגי אינו דורש חדר נקי כלל.

מדוע הכונן הקשיח שלי למעשה קטן יותר מאשר הנפח המוצהר שלו, האם יצרני הכונן הקשיח

אינם מסמנים את היכולת כנדרש?

התשובה הפשוטה לשאלה זו היא - לא. הם לא מפרסמים את היכולת של הכונן הקשיח שלך בצורה לא נכונה. אם קנית HDD 1TB, למעשה יש לך 1TB של אחסון נתונים במכשיר. התקלה אינה טמונה בתיוג שלהם. ונסביר מדוע.

1TB של נתונים הוא טריליון בייט או 1,000,000,000,000 בתים. כעת, אם לדוגמא אנו מורידים מהמדף דיסק קשיח Western Digital טיפוסי ובודקים אותו, יהיה לו איפשהו בסביבות 1,953,525,168 מגזרים.

גודלו של כל מגזר הוא 512 בתים, כך שעם קצת הכפל המהיר, אנו מגלים שלכונן יש למעשה:

1,000,204,886,016 בתים זמינים. וואו! הם למעשה נתנו לי 0.02% נוספים על הקיבולת הרשומה.

אז מדוע Windows מציג פחות יכולת ממה שיש בכונן הקשיח שלי בפועל?

אז היכן חוסר ההתאמה וההבדל בקיבולת HDD באמת שוכן?

הפער האמיתי אינו מגיע מיצרני הדיסק הקשיח ה"משקרים" לך. הם אומרים שמדובר בכונן של 1 מיליארד בתים וזה בעצם קצת יותר מהקיבולת הזו. הפער האמיתי טמון באופן שבו מערכת ההפעלה שלך (Windows) מחשבת את קיבולת הנתונים. למען האמת, חלונות בכלל לא נמדדים בגיגה-בייט ובטרה-בייט בכלל. במקום זאת, Windows מודד באמצעות Gibibytes ו-Tebibytes. אני יודע שהם מאויתים דומים, אבל יש הבדל עדין בכתיב ובמשמעות. מיקרוסופט, בחוכמתה האינוסופית, החליטה להשתמש עדיין בקיצור MB, GB ו-TB כפי שאנו משתמשים בדרך כלל עבור מגה-בייט, ג'יגה-בייט וטרה-בייט. עם זאת, ביתר דיוק הם היו צריכים להשתמש ב-GiB, MiB ו-TiB כדי לתאר את מה שהם מודדים.

אז מה ההבדל בין טרה-טייט לטבי-בייט?

כפי שציינו לעיל, טרה-בייט (הפרמטר שמייצרים HDD מודדים לפיו) הוא טריליון בייט. עם זאת, מספרים עגולים כאלה, שהם בסדר ברמה של חומרה, אינם מסתדרים יפה ברמה התכנותית. המגזרים הינם בני 512 בתים כל אחד. אשכולות (Clusters), שהם גוש הנתונים הקטן ביותר שכל תוכנית יכולה לכתוב גם נתח נתונים, הם תמיד מתחלקים במספר זה 512. אז אשכול יכול להיות 512-בתים, 1024-בייטים, 2048-בתים וכו', אך לעולם לא יהיה מספר עגול כמו 1000 או 2000.

על פי מרבית מערכות ההפעלה, קילובייט (קילו כלומר אלף) הוא למעשה 1024 בתים, ולא 1000 בייטים. הם נושאים את אותו העיקרון עם כל גודל גדול אחר כך. אז MB הוא GB1024, Kb הוא MB1024, ו-TB הוא GB1024.

אז מבחינת Windows, TB (למעשה טבי-בייט ולא טרה-בייט) היא $1024 \times 1024 \times 1024 \times 1024$ או 1,099,511,627,776 בתים. מספר שמתחלק יפה מאוד לפי כל גודל אשכול שאתה זורק אליו במקרה. וכך, תהליך חישוב הגדלים מפושט עבור המתכנתים.

מהו Parity במערכת RAID?

זוהי למעשה מערכת של סיכויים ושינויים שנוצרים כאשר נתונים נכתבים לראשונה למערך, המאפשר להשתמש בהמשך כדי לבדוק את עקביות הנתונים ולקבוע אילו נתונים אבדו במקרה של כשל בדיסק. מערכי RAID כותבים נתונים בפסים או בלוקים. בלוקים אלה של נתונים בינאריים מורכבים בסופו של דבר מאחדים ואפסים. אם אתה מכניס את כל הנתונים לטורים ואז סופר אותם, אתה מקבל מספר אי-זוגי של אלה או עם מספר שווה של כאלה. פשוט על ידי הוספת 1 למספר אחיד של אלה או 0 עבור מספר אי-זוגי מהם תוכלו ליצור בלוק די בסיסי של נתוני זוגיות. ובמערכות RAID ישנות, זהו כל העניין. המערכת בפועל של זוגיות RAID כיום מורכבת מעט יותר, במיוחד כאשר מתמודדים עם RAID 6 שיש לו יותר ממחסום זוגיות אחד. אך זה מתאר את עקרון היסוד של אופן פעולתו. ב-RAID 5 ו-6 נתוני הזוגיות מפוזרים (צפים) על פני כל הדיסקים במערך, ומחליפים אותם לדיסק אחר בכל כך הרבה בתים של נתונים (נקבעים בדרך כלל על פי גודל החסימה של המערך). התאמה זו נעשתה לאחר שהתגלה כי לדיסקים של זוגיות קבועה במערכי RAID 3 ו-4 הייתה שיעור כישלון גבוה בהרבה כתוצאה מגישה מוגברת מהמערכת.

מדוע מערך ה-RAID שלי קרס ואינו מצליח להיבנות מחדש?

לפעמים נתונים פגומים יכולים להשפיע על כל הכוננים במערך. למרות נוכחות נתוני הזוגיות, המערך נשבר במקום שאינו יכול לבנות אותו מחדש. לפעמים זה נובע מכך שיש יותר כוננים במצב לא מקוון מאשר המערך יכול להתמודד. פעמים אחרות זה יכול להיות פשוט מנתונים פגומים כאשר הנתונים הקיימים בכוננים והנתונים שצריכים להיות שם (לפי נתוני הזוגיות) שונים. לרוב זה נגרם כתוצאה

מהפרעות כוח פתאומיות בזמן שנכתבים נתונים למערך. סיבה נוספת לכך שמכשיר UPS הינו חובה לשרתים, וכן שרטיסי בקר RAID מתוחכמים מגיעים עם סוללה מובנית.