

Command completion time measurement tool Whitepaper extension

<Produced>

JVCKENWOOD Corporation

Yoshinobu Nagase

MIKASA SHOJI CO.,LTD.

Ryosuke Shimizu

Takayuki Yokoyama

Hirota Seisakusho Limited

Hisao Ichimura

Hiroshi Sato

The University of Tokyo

Kazuo Goda,Ph.D,

TOYO Corporation

Masumi Okada

<Sample HDD >

Filed Three Corporation

Fieldthree Sotokanda Bldg.,5-3-6,Sotokanda,Chiyoda-ku,

Tokyo,101-0021,JAPAN

Introduction

2008 年に発行された“Command completion time measurement tool Whitepaper”は、発行当時の 28bitLBA、1 セクター512Byte を前提とした GByte 単位の HDD のパフォーマンス評価を中心に考慮されてきた。

ところが、発行後 8 年経過した現在において、HDD は 48bitLBA への ATA 規格の拡張と 1 セクター4096Byte 等の大容量化技術の実用化により TByte 単位が主流になりつつある。この HDD の大容量化により、これまでの“Command completion time measurement tool Whitepaper”の特に Detailed benchmark をそのまま運用すると評価時間に非常に多くの時間を要することが分かっている。そこで、ここに“Command completion time measurement tool Whitepaper”の基本的な評価方法とその結果を維持しつつ、特に TByte 単位の HDD の評価時間の短縮を目的とした評価方法を“Command completion time measurement tool Whitepaper”の拡張として定める

このように、本“Command completion time measurement tool Whitepaper”の拡張は、基本的に元の“Command completion time measurement tool Whitepaper”の Detailed benchmark 評価方法についてのみ、その評価方法を変更すること無く、その運用方法を TByte 単位の HDD の評価時間短縮に対応して定めた物である。

◆従来の評価方法と同等の特性が得ることができ、かつ評価時間を短縮するために、以下に方法を定める。

CCTL 測定値時間の短縮(Long BLK 化)

CCTL では、Read(又は、Write)の Command 実行時間を先頭から最終まで測定し、Command 実行時間をプロットする事で、HDD 全体の時間バラツキを把握するものである。

HDD 全領域に対し高速のアクセスを実現するには、次の項目が必須である。

1. 転送ブロック数を増やす。(発行 Command の転送ブロック数を大きくする)

実験では、256(128KB) → 16,384(8MB) と 64 倍に拡大

※ 試験器では、Max 64MB の転送が可能であるが、未サポート HDD があることから 8MB とした。

2. シーケンシャルアクセスのみとする。

HDD の特性比較に適している。

3. キャッシュを有効にする。

キャッシュ機能を有効にすることで、Command 終了 → 次の Command 発行間に時間ロスが見えなくなる。

また、HDD 側は連続して動作するため、無駄な回転待ちを抑制できる。

● 拡張項目

1. アクセス単位は、基本4096セクター（1セクター512Byteで2MByte）を最小単位とし、そのn倍を拡張として用意する

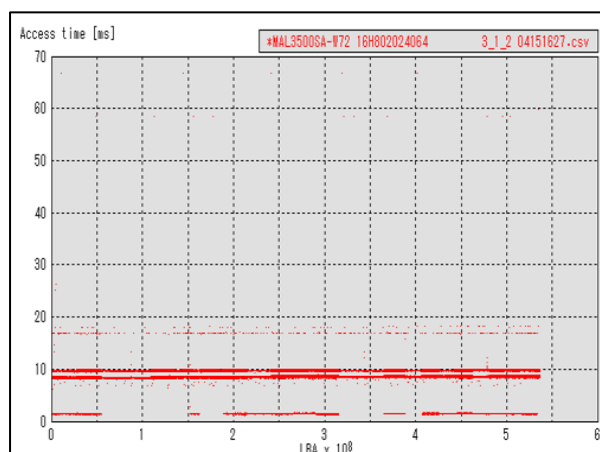
“Command completion time measurement tool Whitepaper”は28bitLBAで有ったことから、1回の転送量の最大値を0xFFの256セクターとし、1セクター512Byteとして128KByteを最大としてきた。

これに対し、48bitLBAでは、転送量の最大値が65536セクターと大幅に増えていることから、HDDのパフォーマンス評価に有効な設定値の検討を進めてきた結果、“Command completion time measurement tool Whitepaper”のアクセス最小単位を4096セクターとし、そのn倍でのアクセス単位で評価を行う物とする。

Sequential Write Cache-OFF (セクター長：256 ブロック長)

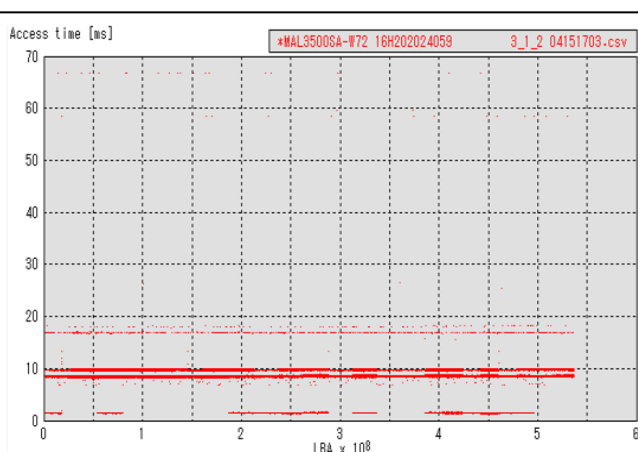
Sample-A (500GB HDD)

測定時間：5.5h / 平均転送レート：14.9MB/s



Sample-B (500GB HDD)

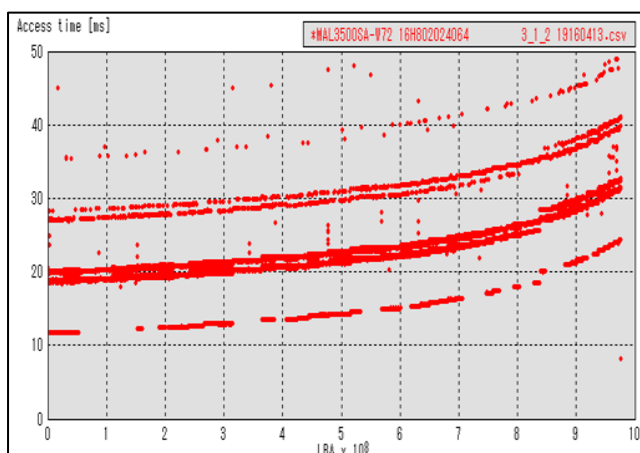
測定時間：5.5h / 平均転送レート：14.8MB/s



Sequential Write Cache-OFF (セクター長：4k ブロック長)

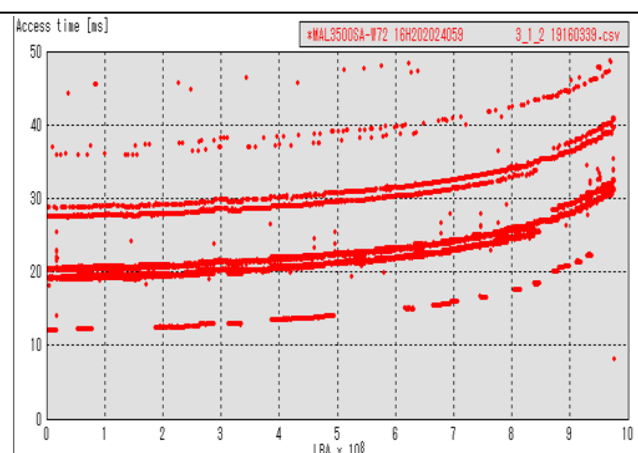
Sample-A (500GB HDD)

測定時間：1.5h / 平均転送レート：91.4MB/s



Sample-B (500GB HDD)

測定時間：1.5h / 平均転送レート：91.9MB/s



2. アクセスパターンはシーケンシャルアクセスに固定する

“Command completion time measurement tool Whitepaper”のDetailed benchmarkでは、アクセスパターンとして、シーケンシャルとジグザグアクセスを定義している。大容量HDDにおいて、ジグザグアクセスでは、HDDのパフォーマンスの比較評価をしにくいことから、シーケンシャルアクセスに限定する。

なお、Outer-to-Inner、Inner-to-Outerのアクセス方向とジャンプパターンはこれまで通り、“Command completion time measurement tool Whitepaper”のDetailed benchmarkに準ずる

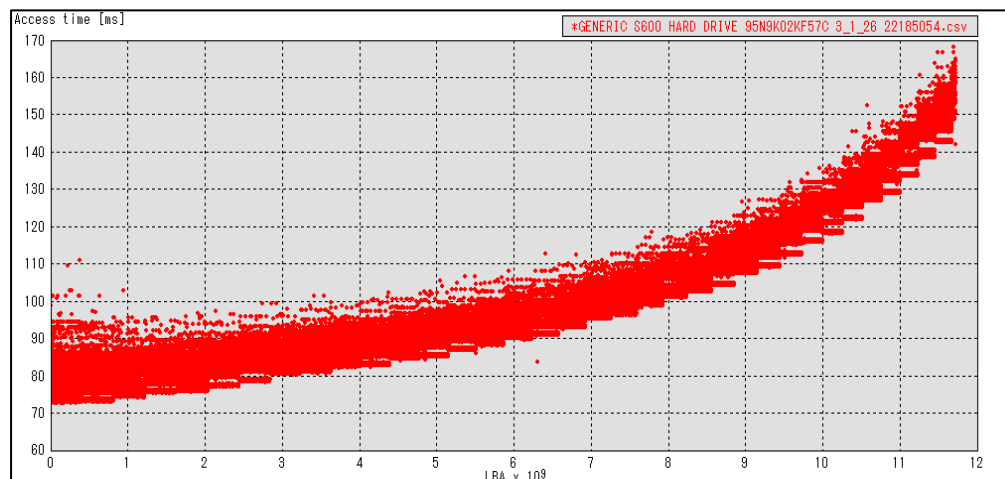
◆Long Block (16K以上) でCash-On状態で、 全領域 シーケンシャルに測定。

- 測定は、Read と Write で実施。
- 全域を最短時間で行うことが可能。
- この測定値のグラフ化で、 HDDのバラツキがある程度推測できる。
- バラツキのヒストグラムは、移動平均との差分で表示。

◆Sample 6TB HDD

Sequential Write Cache-ON (セクター長：32K ブロック長)

測定時間：約 10h / 平均転送レート：169.6 MB/s



測定データ出力 (CSV 形式)

測定Data出力Format (csv形式)	
File 名 形式: GENERIC S600 HARD DRIVE 95N9K02KF57C 3_1_26 22185054.csv	
Data内容	
<<Smart Tester Access Time Measure Log V4.2>>	
Log Format 999	
COMMENT : W/R=1 SEEK=1 Cache=3 Block=32768 Area= 0%-100%	
Device Model : GENERIC S600 HARD DRIVE	
Serial Number : 95N9K02KF57C	
Func=3 W/R=1 Exec-No.=26 Average count=1 Cache-mode=3	
Start LBA : 0	
Last LBA : 11721045168	
Access Blocks : 32768	
Skip Blocks : 0	
Start_LBA	Access_Time(ms)
0	78.333
32768	82.492
65536	81.986
98304	81.975
131072	82.5
163840	81.984
196608	81.966
229376	82.508
262144	81.984
.	.
11720916992	156.211
11720949760	156.211
11720982528	155.667
11721015296	142.16
End	End

3. HDDのキャッシュはONで行う

“Command completion time measurement tool Whitepaper”のDetailed benchmarkでは、HDDのキャッシュをOffにすることが推奨されていた。これを運用する場合、非常に多くの評価時間を要することが分かっている。

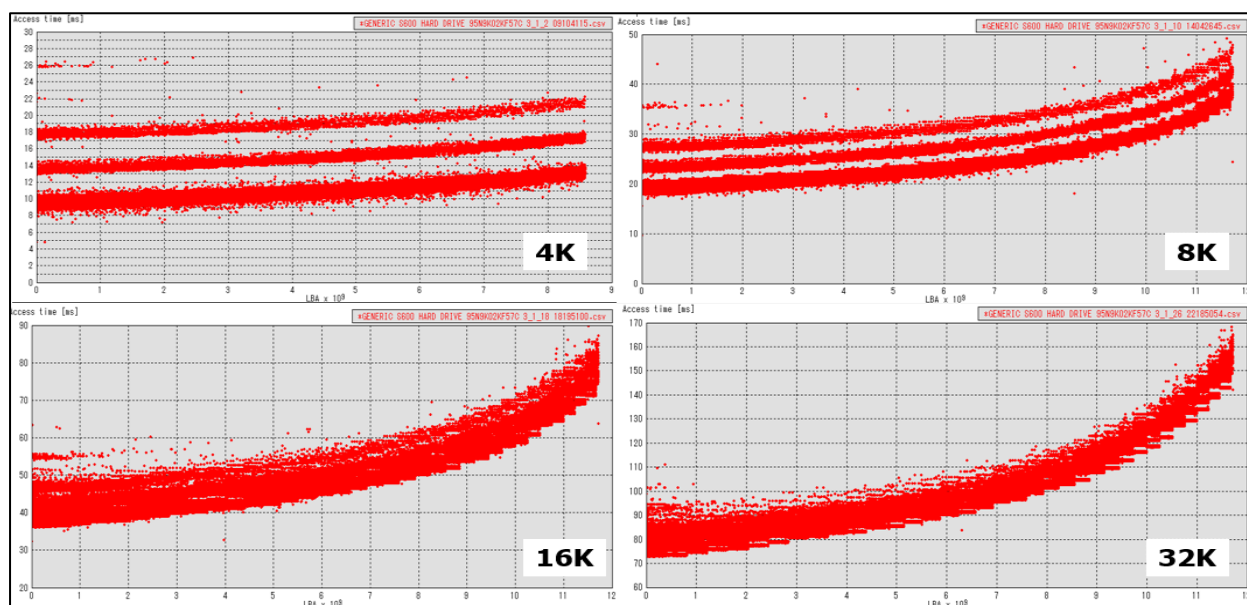
しかし、アクセスパターンをHDDの内部キャッシュよりも大きく取ることにより、キャッシュを超えた範囲でのHDDのパフォーマンスが評価出来ることが確認されたので、評価時間の短縮のため、キャッシュをONにすることを定める。

◆転送ブロック数(セクター長：4k/8k/16k/32k,ブロック長)を増やし、キャッシュをONにすることを定めたことにより、キャッシュOff 時と比較した結果を以下に示す。

■Sequential-Write Cache ON

測定時間：37h (全ブロック長のTotal)

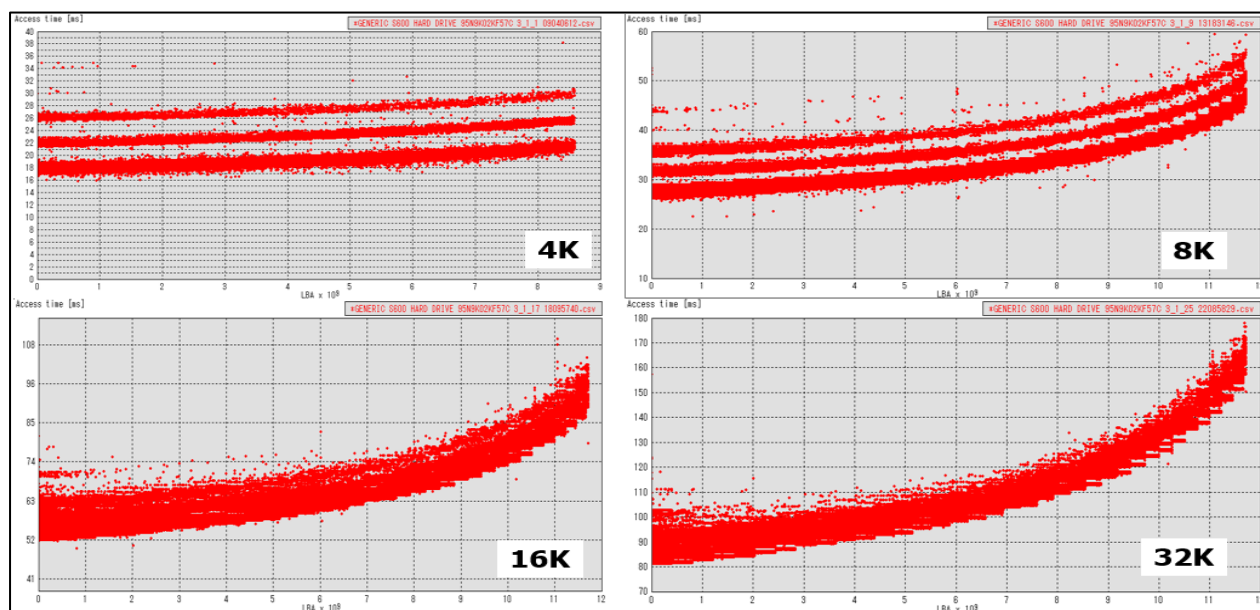
(Sample:6TB HDD)



■Sequential-Write Cache OFF

測定時間：47h (全ブロック長のTotal)

(Sample:6TB HDD)



4. まとめ

◆大容量 HDD 対応、CCTL 測定値時間の短縮(Long BLK 化)結果について

元の“Command completion time measurement tool Whitepaper”を引用し同じ特性が得ることができ、かつ評価時間を短縮するために、以下の方法を定めた。

1. 転送ブロック数を増やす。
2. シーケンシャルアクセスのみとする。
3. キャッシュを有効にする。

結果

上記の方法を定めた結果、元の“Command completion time measurement tool Whitepaper”と比較し、評価精度を落とすこと無く、測定時間の短縮化が図れた。

● Appendix

大容量 HDD 対応 CCTL 測定に使用した機材と試験方法を以下に記載する。

■実施

株式会社 広田製作所

〒382-0005 長野県須坂市大字小河原 3954 番地 13

代表電話番号 026-245-1212

URL. <http://www.hirotass.co.jp>

E-mail. eigyo@hirotass.co.jp

■機材

- ・ SMART TESTER for 6G(本体)
- ・ ホスト PC (ホストプログラム用)

■試験方法

1) SMART TESTER に測定用 HDD を接続する。

(接続構成：図 1 参照)



(図 1)

SMART TESTER for 6G



-仕様-

項目	内容
名称	SMART TESTER for 6G
型格 ※1	S64PV1-F3
入力電源	AC95~245V 50/60 Hz
最大消費電流	8A
動作温度	0℃~40℃
試験インターフェイス	SAS/SATA 6Gbps対応 (6Gbps/3Gbps/1.5Gbps)
試験可能ドライブ数	4台
専用フィクチャ ※2	ドライブ4台分収容可能
ドライブ供給電圧 ※3	出力電圧 出力電圧1 : 5V 出力電圧2 : 12V 電圧可変範囲 出力電圧1 : 2V~6V 10mV刻みにて設定可能 出力電圧2 : 8V~14.5V 20mV刻みにて設定可能 電圧精度 出力電圧1 : 設定電圧±50mV 出力電圧2 : 設定電圧±100mV
ドライブ電源電圧測定	電圧測定範囲 : 0V~15V 電圧測定精度 出力電圧1 : ±1.5%FS 分解能 1mV 出力電圧2 : ±1.5%FS 分解能 1mV 125μsec~200msec間隔(125μsecステップ)
ドライブ電流測定機能	電流測定範囲 : 0A~4.0A 電流測定精度 出力電圧1 : ±1.5%FS 分解能 1mA 出力電圧2 : ±1.5%FS 分解能 1mA
対応OS※4	Windows 2000 / XP / 7
外部インターフェイス	LAN 10BASE-T / 100BASE-TX / 1000BASE-T
接続可能台数	64 (板)
外形寸法(W×D×H) (本体のみ)※5	258×258×58mm
重量 (本体のみ)	約 3Kg

2) ホスト PC のホストプログラムの起動から、被試験装置との接続チェック等を実施し、試験実施が継続可能な状態であることを確認する。

※SMART TESTER は、4port 同時試験可能であるが、用途に合わせて Port を選択し、試験可能。

3) 初期内部 SMART(Self-Monitoring, Analysis and Reporting Technology)の確認。
CCTL 測定を実施する前に、事前に初期内部 SMART 情報を取得し、現状の HDD の状態を確認し、各属性に異常値はないか等のチェックをする。

4) CCTL 測定用スクリプト(シーケンシャルアクセス&アクセス時間)で測定開始。

目的：シーケンシャルアクセスを行いアクセス異常アクセスの確認。

セクター長：4K、8K、16K、32K ブロック長で実験を実施。

測定範囲　： 全面

セクター長　： 4096 (2MB)

　　　　　　　： 8192 (4MB)

　　　　　　　： 16384 (8MB)

　　　　　　　： 32768 (16MB)

キャッシュ　　： On

Access Mode　： Write , Read , Write+Seek , Read+Seek

※機能：指定領域間を順方向にシーケンシャルアクセスを実施。この機能では、
R/W コマンドの実行時間を計測し、CSV 形式の File に出力。

5) CCTL(アクセスタイム時間)測定結果について

対象スクリプト : Sequential Write Cache(On) /セクタ長:4096(4KB)

出力された CSV File 名 : 例) GENERIC S600 HARD DRIVE 95N9K02KF57C 3_1_2 09104115.csv

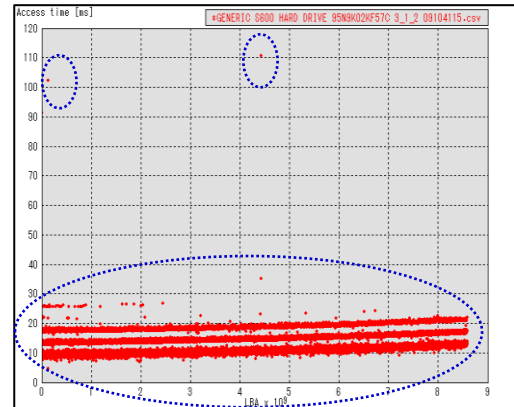
※CSV File を専用の(Measure Graph - HDD/SSD Tester)にてグラフ化に変換。

```
Log Format 999↓
COMMENT      : W/R=1 SEEK=1 Cache=3 Block=4096 Area= 0%-100%
Device Model : GENERIC S600 HARD DRIVE↓
Serial Number: 95N9K02KF57C↓
Func=3 W/R=1 Exec-No.=2 Average count=1 Cache-mode=3↓
Start LBA    : 0↓
Last LBA     : 11721045168↓
Access Blocks: 4096↓
Skip Blocks  : 0↓
↓
Start_LBA,Access_Time↓
0, 4.842↓
4096, 91.484↓
8192, 22.701↓
12288, 9.939↓
16384, 9.936↓
20480, 9.939↓
24576, 9.950↓
28672, 9.936↓
32768, 10.473↓
36864, 9.960↓
: ↓
: ↓
8589447168, 13.247↓
8589451264, 13.772↓
8589455360, 13.272↓
End, End↓
```

(CSV File)



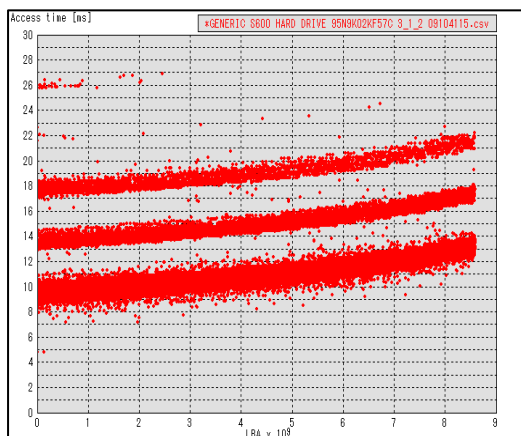
変換



(変換後のプロット図 1)

結果、変換後のプロット図 1 では、100～110ms 近いアクセス時間が発生したため、帯状の部分が縮小し、分析しにくい。

拡大



(拡大変換後のプロット図 2)

変換後のプロット図 2 は、Access Time 値を、0～30ms に拡大表示することにより、各 HD 毎の記録密度の差、リトライ(まばらな点)が分析し易くなる。

- 6) 5 項で作成したプロット図から、各リード/ライトコマンドのアクセス時間のバラツキ(太い帯状になった部分や、その周りにまばらな点からのバラツキ)から、HDD にどのような挙動が占めているかが把握できる。又、各転送 Block Size の測定時間や平均転送レートからの要因の分析、さらに、各アクセス毎に出力された SMART 情報の差分から比較し起こりうる障害の早期発見と故障予測を分析する。
 <特に SMART 情報で重要視される項目>

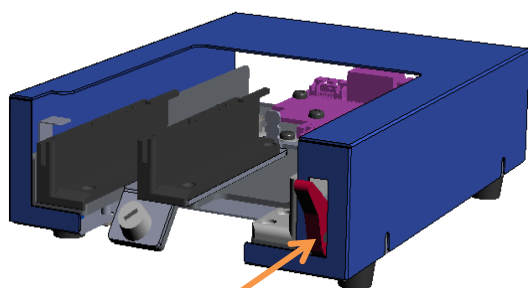
項目ID	項目名	詳細な説明
1	Raw Read Error Rate	この項目はハードディスクからデータを読み込むときに発生したエラーの割合を表す。数値が閾値より低い場合、ハードディスク内の磁気ディスクまたは磁気ヘッドに異常がある。
5	Reallocated Sectors Count	代替処置 (データを特別に予約した予備エリアに移動する) を施された不良セクタの数。
0A	Spin Retry Count	ディスクを規定の速度までスピニングしようとして再試行を試みた回数。
0D	Soft Read Error Rate	オフトラックの数。数値が0でなければバックアップを取る。
C4	Reallocation Event Count	セクタの代替処理が発生した回数。仮に処理に失敗しても回数に加算される。
C5	Current Pending Sector Count	現在異常があり、代替処理を待つセクタの総数。もし後で読み込みに成功したセクタがあれば、この値は減少する。
C6	Off-Line Scan Uncorrectable Sector Count	オフラインスキャン時に発見された、回復不可能なセクタの総数。この値が増加する場合は、磁気ディスクの表面に明確な問題がある。
DC	Disk Shift	ディスク (プラッタ) が衝撃などにより当初の固定位置よりズレた距離。

■追加評価実施例

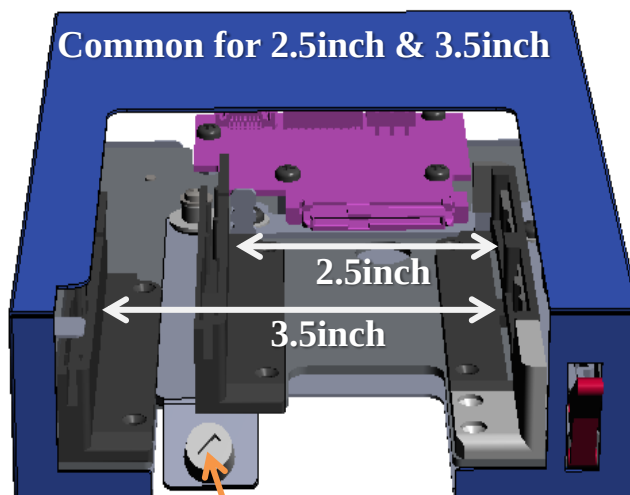
今回実施した CCTL 測定では、測定環境、特に HDD を外部に設置しての測定では、外部の振動が HDD に大きな影響を与える可能性があります。

そこで、株式会社広田製作所で独自開発した Drive Fixture は、免振構造で高い Drive Performance の効果を引き出します。

Drive Fixture



このレバーで HDD の挿抜が可能です。
※挿抜耐久回数：5,000 回 (弊社実証確認済み)



このレールを外すことで、
3.5inch Drive が対応することができます。