Infortrend®
www.infortrend.com

ストレージソリューション　Fibre-SATA 【ES A16F-G2422】導入事例
東京大学　素粒子物理　国際研究センター 様



INDEX | 導入背景 | 選定のポイント | 運用状況 | インフォトレンド社 | ES A16F-G2422 | お問い合わせ

世界最先端の粒子加速器を用いて素粒子の研究を行い、宇宙に存在する様々な謎を探求している東京大学 素粒子物理国際研究センター。同センターは、世界各国で 1800 人以上のスタッフが参加する国際的な研究プロジェクト「アトラス実験」に参加しています。「アトラス実験」とは、CERN(欧州原子核研究機構)が建設中の LHC 加速器で行われる予定の大規模な実験の一つ。この実験で生成される膨大なデータは、世界中の研究機関に設置される計算機システムをグリッドコンピューティングで繋ぎ、分散して解析されます。2006 年末、「アトラス実験」の日本における研究・解析の拠点として、東京大学 素粒子物理国際研究センターに大規模な計算機システムが構築されました。そのシステムで、大容量ストレージとして採用されたのが、インフォトレンド社の【ES A16F-G2422】です。16 台の SATA ドライブを搭載した【ES A16F-G2422】が 140 台、合計物理容量 1.1 ペタバイトという大容量のストレージとして導入されています。今回は、実際に「アトラス実験」の研究を進めておられる東京大学 素粒子物理国際研究センター 准教授の真下哲郎理学博士に、【ES A16F-G2422】導入の背景やストレージ選定のポイントなどについてお伺いしました。

■システム導入の背景とストレージの選定のポイントを教えてください。



東京大学 素粒子物理国際研究センター
准教授　真下哲郎理学博士

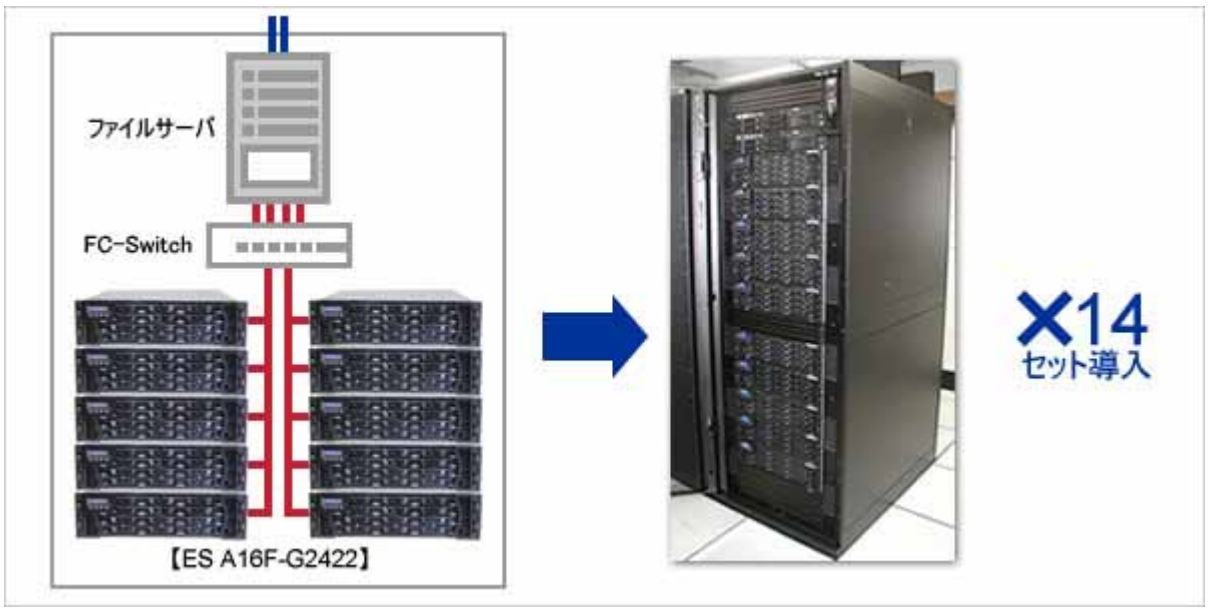
国際的な大規模プロジェクトである「アトラス実験」を行うための LHC 加速器が、今年（2007 年）に完成をします。この LHC 加速器からは、「アトラス実験」に限っただけでも、1 年で数ペタバイトという膨大なデータが生成されます。
その膨大なデータを解析するための日本国内の拠点には、処理能力の高い CPU と大量のデータを保存するディスクアレイを持つシステムの構築が必要でした。システムへの要件として、優先順位が高かったのは、CPU パワーとデータの保存。
データ保存のためのストレージに関しては、予算に制約があることから、コストを抑えながらも『大容量』、『耐久性』、そして『パフォーマンス』を求めました。

■ではなぜ、インフォトレンド社 ES A16F-G2422 を選定されたのですか？

今回の大規模システムの導入は 2006 年末でしたが、5 年ほど前から、計算処理のための CPU サーバやテープライブラリ装置も含め、本システムの規模を小さくした、実験的なシステムを作成して研究開発を行ってきました。
検証を行う中で、ディスクアレイに関してはインフォトレンド社製品を使用してきましたが、故障の頻度が少なく、大規模システムでも耐久性に問題がないことを実感していました。
また、データ転送性能に関しては、ホスト転送速度は 1 ジョブの最大で数十 M バイト/秒程度を予想していますが、多数のジョブからアクセスが集中しても、インフォトレンド社製品が、充分にそれをクリアできる性能を持っていることもわかりました。
インフォトレンド社製品は検証したどのモデルでも耐久性と性能に問題がないため信用できると感じていた上に、さらにコストパフォーマンスでも他製品を大きく上回っていると、考えていました。
今回のシステム構築の入札においても、落札したシステムインテグレータが提案した製品は、インフォトレンド社製品の【ES A16F-G2422】であり、我々が提示したデータ保存に関する必要条件と予算を、【ES A16F-G2422】が見事にクリアした、ということになります。

■実際のデータ保存の運用状況について教えてください。

2006 年度の本システムの RAID 構成には RAID6 を採用しました。RAID5 では、常時メディアスキャンを行うことが欠かせないのですが、RAID6 であればメディアスキャンを省略したとしても、データの損失などの被害が発生する可能性をかなり軽減できると考えています。また、RAID5 から RAID6 に移行しても、性能のダウンもあまりなく、非常に満足しています。



■今後のインフォトレンド社とのお付き合いについて

実は 10 年ほど前から、他の実験のデータ保存にインフォトレンド社の製品を使用していたんです。原因不明の故障が無く、とても耐久性に優れていることが評価のポイントです。
今後、新たに大規模なシステムを構築する際に、大量のデータの保存が必要になれば、個人的にはインフォトレンド社の製品を導入したいと考えています。



～真下哲郎理学博士へのインタビューを通して～

まさに世界規模、1800 人以上が関わる国際的な大プロジェクトとして行われるアトラス実験からは、宇宙の謎の深淵を覗くための貴重なデータが大量に生成されることに驚くと共に、その保存に使用されるストレージには、信頼性をはじめとする様々な条件が求められることを改めて実感しました。そして、その厳しい要求に【ES A16F-G2422】が応えられる製品であり、コストパフォーマンスが高かったからこそ、合計 1.1 ペタという大容量ストレージとして採用されたと感じました。

Fibre-SATA【ES A16F-G2422】

インフォトレンド社では、高性能・高機能な RAID ストレージを提供するため、データ・コントローラやファームウェアの設計・開発、そしてボードのマウントにいたるまで、すべてを一貫した開発・製造体制で行っています。だからこそ、信頼性が高く、常にテクノロジー・トレンドをリードする製品群をご提供できるのです。



■RAID コントローラー
750FX/600MHz CPU
ASIC266 offers Dual PCI-X 64/133 system bus
最大 2GB DDR266 キャッシュメモリサポート

■RAID5 ながら 最高の I/O パフォーマンスを発揮
Sequential Reads ：510 MB/second
Sequential Writes：350 MB/second

■2 ポートの 4Gbps Fibre ホスト・チャネル

■多彩な RAID レベルをサポート
0、1（＋1）、3、5、6、10、30、50、60

■16 台の SATA-II（3Gbps）ドライブ・ベイ

■3U シャーシで最大 8TB 容量をサポート(提案当時の容量)

■電源・冷却ファンを 2 重化

SATAドライブの信頼性を向上させるテクノロジー

インフォトレンド社のファームウェアは、RAID ストレージに関する 10 年以上の設計経験に基づく独自のテクノロジーによって、I/O プロセシング、ドライブ・ハンドリング、システム・マネジメントが強化されています。

I/O SMARTテクノロジー

●バックグラウンドでの論理ドライブの初期化が可能なため、最適な RAID 機能を素早く提供
●論理ドライブ、論理ボリューム、論理パーティションのフレキシブルな構築が可能
●I/O 入出力を保証
　イメージやストリーミング・アプリケーションに必要な連続データの継続性を保証 HDD にダメージがある場合でも遅延の発生を回避

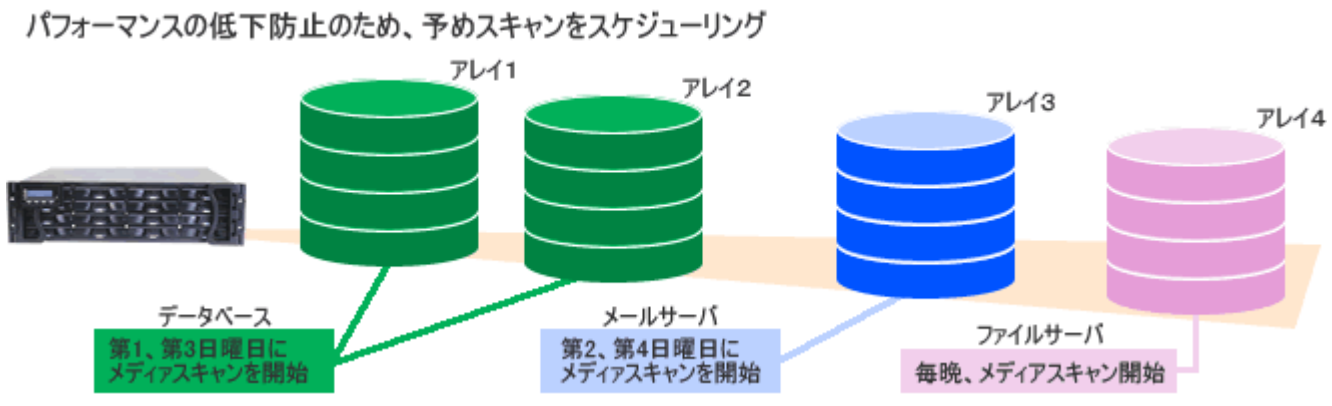
バッドブロックのハンドリング

【ES A16F-G2422】は、デグレード中やリビルド中にバッドブロックが発生した場合でも、システムに影響を与えないように、バッドブロックのハンドリング機能を強化しています。

●バッドブロックを発見した場合
1 バッドブロックを検出
2 他のメンバー・ディスクのストライプからデータを検索
3 同じブロック上にリライトを試行
4 リライトに失敗すると、そのディスクのセクターが不安定であると判定し、再構築されたデータを他のセクターへ転送

メディアスキャン

メディアスキャン機能が書き込みデータを検証し、不良セクターを検知・修正し、パリティエラーを修正します。隣接するストライプから再構築されたデータは、同じセクターにリライト、あるいは不良セクターから正常なセクターへと移行します。このメディアスキャンは、タスクスケジューラによって自動実行が可能。回復不可能なデータ・ブロックを無視し、影響を受けていないデータを守るといった設定も選べます。

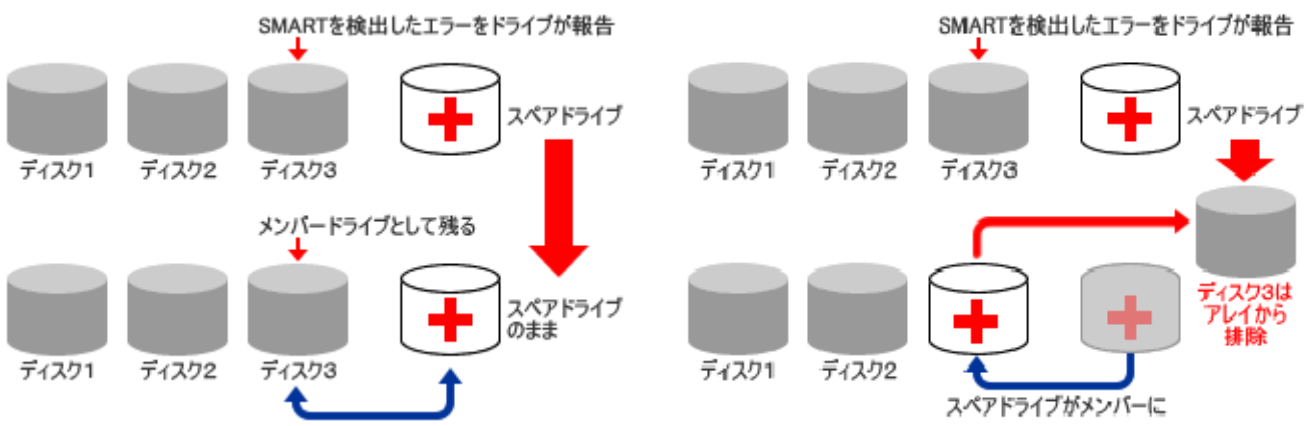


ディスク・クローニング

【ES A16F-G2422】で HDD の SMART 機能を有効にしておき、エラーを検出すると以下の 3 つの対処が行われます。対処方法は以下より選択することができます。
A 検出のみ: SMART エラーを管理者へ報告
B 検出/永久コピー: エラーの疑いがあるメンバー・ドライブをスペアドライブへコピーしておく
C 検出/コピー/置換:エラーの疑いがあるメンバー・ドライブをスペアドライブへコピーしてメンバー・ドライブと入換えます

●検出/永久コピー
ディスク 3 で障害が発生した場合、スペアがメンバーに変更
他のメンバー（図版内ではディスク 1）の障害時、スペアはクローニングを断念し、アレイの再構築を開始

●検出/コピー/置換
ディスク 3 で障害が発生した場合、スペアがメンバーに変更
ディスク 3 からスペアへのコピーが完了すると、ディスク 3 をアレイから排除、取り外します



【主な SMART 属性】
ヘッド高、データスループット性能、スピンアップ時間、セクターカウント再配置、シークエラー率、シーク時間性能、スピン試行の再カウント、ドライブ計測再試行カウント

ドライブ・スピン・ダウン

システムのアイドル時にドライブを止めるドライブ・スピン・ダウン機能を備えており、消費電力を削減。
また、ディスクの余計な熱の発生を防ぎ、ディスクの冷却コストも低減。無用なディスク稼動による劣化も最小限に抑えられます。

問い合わせ
株式会社インフォトレンドジャパン
住所：〒105-0023 東京都港区芝浦 1-7-14　岡家寿ビル
TEL ：03-5730-6551
FAX ：03-5730-6552
e-mail：sales@infortrend.co.jp
URL：http://www.infortrend.co.jp/

※この記事はキーマンズネットに 2007/6/27 に掲載されました。