KEKB コンピューターシステムにおける分散システム技術とその運用

真鍋 篤*、佐々木 節、柴田 章博、川端 節弥、渡瀬 芳行
文部省 高エネルギー加速器研究機構（KEK） 共通研究施設 計算科学センター
*) Atsushi.Manabe@kek.jp

概要

KEKB コンピューターシステムは KEK の高エネルギー素粒子実験である Belle 実験の実験準備計算、データー収集、実験解析のためのミッションクリティカルシステムであり、1 日 24 時間、最大 6 ヶ月連続無停止運用される。年間 30TB のデーターを収集、蓄積し 1000SPEC int95 以上の計算処理能力をもち、常時数十人がログインしてプログラム開発、解析を実行する。本論文は、多数のサーバー群の上でどのような分散コンピューター技術がどのように適用されてこのシステムが実現されているか、負荷分散、高可用性にどのように配慮されているかを述べる。

Distributed computing technologies applied on the KEKB computer system.

Atsushi Manabe*、Takashi Sasaki、Akihiro Shibata、Setsuya Kawabata, Yoshiyuki Watase
High Energy Accelerator Research Organization (KEK)　Computing ResearchCenter

Abstract:

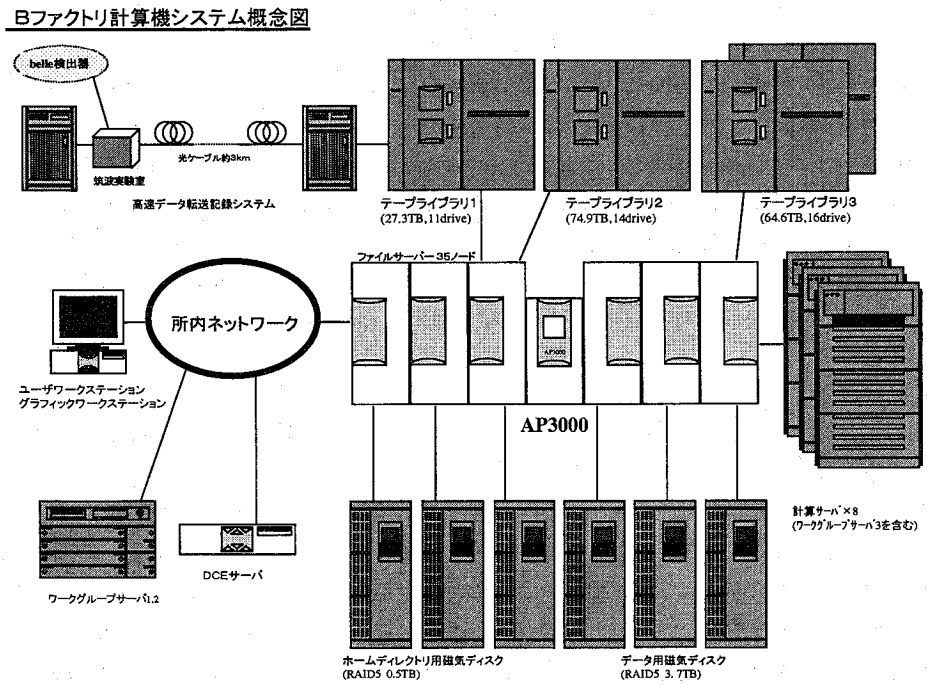
In present large High Energy Physics (HEP) experiments, a required computing power exceeds obviously that of a single computer, even with the highest end. The integration of a number of computers and storage devices with network requires high performance and robust distributed computing environment. The KEKB computer system is an example of such a computer system. It supplies ~1000 SPECint95 CPU power and 120TB file capacity and is operated under 6 months 24hous/day non stop policy. This paper describes how distributed computing technologies are applied to the system to realize the huge and robust computer system.

はじめに

最近の大規模、高エネルギー素粒子実験においては、必要とされる計算機資源を単一の計算機により実現するのは、困難あるいは可能であっても費用対効率が悪いために分散コンピューティングを利用したシステムを利用せざるをえない。分散コンピューティングにおいてはハードウエアの構成要素が多く、。また、要素間の複雑な相互依存関係が生じがちであるため、ユーザーにとって名も知らぬ一台の計算機に問題が生じただけで、システム全体に影響をおよぼすことがある。このため大規模なシステムにおいては、それぞれの要素ごとにみれば信頼性に問題がなくても、全体としてみると非常に信頼性が低いものとなってしまう。これを避けるためには構成要素の冗長度をあげ、障害時にバックアップ可能とすることにより耐障害性をあげることが有効な方法である。

KEKB コンピューターシステムの概要

KEKB コンピューターシステム（以下 本システム）は、B ファクトリー物理実験のデータ解析を目的として、1997 年 1 月に導入された UNIX(Solaris 2.5)を OS とするシステムである。13 台のワークグループサーバー、7 台の計算サーバー(1 台あたり 28CPU を有する)、最大容量 140TByte の高速デジタルカセットテープロボットライブラリー、総容量 4.2TBytes RAID ディスクおよびファイルサーバーとその他周辺装置を主な構成要素として持つ。周長約 2Km の KEKB 加速器で加速された素粒子を衝突させ、その現象を観測する検出装置が計算機センターより約 3km 離れた実験室にあり、実験データーは高速データー転送記録システムにより本システムに転送される。終日 30～５０人のユーザーが 40 台のワークステーション、70 台の X 端末を使用してワークグループサーバーにログインしてプログラム開発やインタラクティブ処理作業をおこなう。多量のデータや計算資源を使用するデータ解析やシミュレーションなどの仕事は、分散バッチシステム LSF により計算サーバーに処理をサブミットすることにより実行する。6 台の RAID ディスク装置は 100GB 毎にパーティション分けされ、それぞれがファイルサーバー専用のワークステーション 1 台に接続されサービスされる。 ファイルサーバーは FDDI ネットワークと 200MB/s の転送能力をもつ AP net と呼ばれるネットワークにより、計算サーバーなどのクライアント機に接続されている。システムの主なジョブと蓄積されるデーター量を次頁表に示す。



Job の種類	必要計算能力 (SPECint95)	必要 I/O 性能 (MB/s)	平均同時実行数	データー生成量	記憶媒体
実験データー収集	-	15MB/s	1	10TB	tape
実験データー解析 （再構成）	120	6MB/s	1	15TB/年	tape
実験データー解析（分類）	20	6MB/s	1	1TB/年	mag.disk
全体シミュレーション	120	5MB/s	1	8TB/年	tape & disk
ユーザーによる解析１（抽出）	8	5MB/s	2~3		disk
ユーザーによる解析２	2	2 MB/s	~20		disk
ユーザーによるシミュレーション	20	<1MB/s	~10		disk

分散コンピューティング技術の適用

ユーザーアカウント管理

分散環境において、ユーザーの認証と資源へのアクセス権の認可および uid などの属性の管理は基本的な要件である。このような目的には、NIS(Network Information system)が広く使われてきたが、本システムにおいては、セキュリティーの観点から kerberos 認証を実装する Transarc 社の DCE (Distributed Computer Environment)を使用している。DCE では、おもな要素サーバーとして、ディレクトリーサービス を提供する、Cell Directory サーバー、ユーザー認証と認可をおこなう セキュリティーさーばー、時間の同期をとるための 時間サーバー等があるが、いずれのサーバーも複数台（2~3 台）を用意し、負荷分散を行うと同時に障害耐久性を高くしている。例えばセキュリティーサーバーの場合、任意の一台のサーバーがマスターとなりセキュリティー情報に読み書きの権限をもつ。ほかのサーバーはクライアントからの情報の読み出し要求にのみに応じる。定期的に両者のデータベースは同期が図られ、クライアントが増えた場合サーバー間で負荷の分散がはかれる。 マスターサーバーに異常が起こった場合はオペレーターの介入によって、マスターサーバーを変更できる。自動に変更しないのは、マスターが２つできてデーターベースに矛盾が生じるのを避けるためであり、マスターサーバーに異常があったり、ネットワークに異常があってネットワークが分割された場合でもユーザー認証自体に支障はない。新規のユーザー追加に支障が出るだけである。

計算能力資源の負荷分散

本システムでは、長時間の計算処理はバッチによる処理を基本にしている。計算サーバーの負荷分散のために、platform comp.社の LSF（Load Sharing Facility）を使用している。LSF ではキュー毎に実行ジョブの資源制限、ジョブスケジューリングポリシーを設定することができ、ユーザー間、グループ間の公平なジョブの実行が可能である。LSF バッチシステムは主に、バッチキュー全体を管理するマスターバッチサーバー、情報を統合する負荷情報サーバー、ジョブが実行されるバッチサーバー、ユーザーがジョブをサブミットするクライアントからなる。負荷情報サーバーにより各バッチサーバーの情報が定期的に収集され、マスターバッチサーバーは次にジョブを実行するバッチサーバー機を選択する。バッチサーバーに異常が起こった場合、実行中のジョブは失われるが、バッチシステム自体がダウンすることはない。情報サーバーやマスターバッチサーバー機に異常が起こった場合は、他のバッチサーバー機が代わってマスターバッチサーバー、情報サーバーとして機能する。サービスはすべてのバッチサーバーがダウンしないかぎり継続することができる。バッチキューのパラメーターなどを、運転をつづ

けながら動的に変更することが可能なためチューニングが容易である。

ファイルサービス

本システムにおいてはファイルの種類はユーザーの用途によりおおまかに以下の５種類に分類され、それぞれに異なる方法により提供されている。

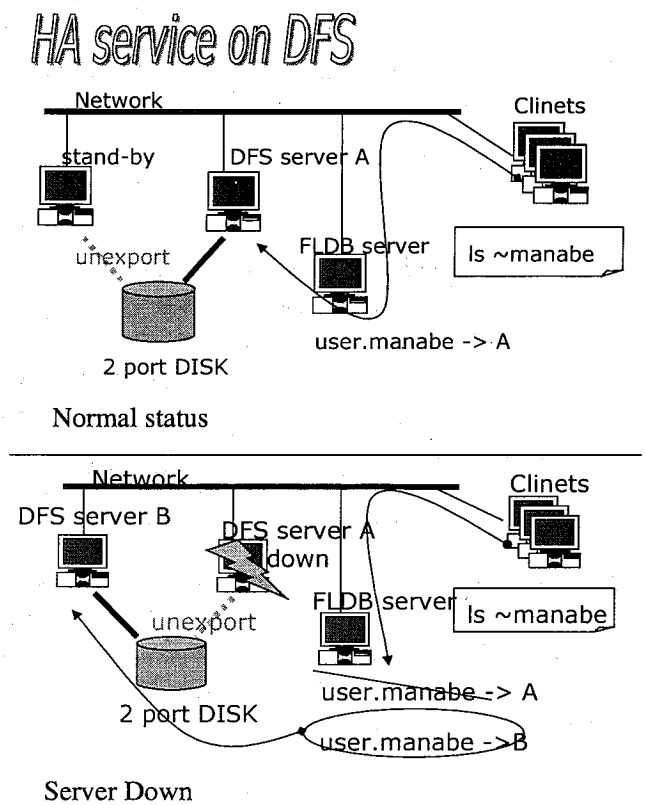
- システムファイル ：OS やシステムパラメーター設定ファイル、システムログなど。このファイルにアクセスできないと計算機が停止するため、本システムではシステムファイルは計算機に直接接続された独立な 2 つの安価な SCSI ディスクに二重化して保持している。２つのうち１つのディスクに異常が生じても運用続行が可能。ログ出力以外はほとんどが読み込みなので２重化によるオーバーヘッドはほとんどない。
- ホームディレクトリー ：ユーザー開発のプログラムソース、メイル、ドキュメントなどやいわゆるホームディレクトリーがこのファイルサービスにより提供される。容量が合計約 400GB。1 つのファイル容量は(4)に比較して小さく、読み出しと書き込み比率がほぼ１であり、同じファイルを複数のクライアント機から同時に使用されることから DFS ファイルシステムによりサービスを提供している。
- アプリケーション実行ファイル：ユーザー開発あるいは他で開発されたアプリケーションソフトウエアの実行ファイルであり総計約 40GB 程度。個別のファイルサイズは(4)と(2)の中間で読み出しの頻度が圧倒的に多く、多数のユーザーから同時に使われることがある。DFS ファイルシステムを利用して提供されている。DFS はファイルサーバーのレプリカ機能をもち、レプリカはクライアントに ReadOnly のサービスを提供する。ファイルサーバー間での負荷分散がおこなえ、また耐障害性も高くなる。
- データーファイル 1：解析結果のデーター等で複数のユーザーが使用するファイル。階層型ファイルシステムを含む NFS ver.3 により提供される。階層型ファイルシステムは使用頻度が下がったファイルが自動的にテープ媒体に移動(migrate)され、必要があるとディスクに移動される仕組みである。ファイル容量は、テープ階層を含めると最大約 60TB、ディスク階層で約 3.5TB である。それぞれのファイルは１GB 程度の大きさであり、複数のユーザーから同時にアクセスされることもある。読み出しと書き込みはほぼ同頻度。また、一人のユーザーが次々と一連のファイルをスキャンしていくことが通常であり、ファイルサイズが大きいこと共にクライアント側のキャッシュが有効に働かないばかりか邪魔になるため DFS は有効ではない。(2)(3)(4)のファイルシステムを提供しているディスクシステムは活線保守が可能なユニットを持つ RAID5 が使用されている。ディスクユニットの総数が 600 台と多いため、週一回程度のユニット故障が期待され必須の機能である。高速のスループットを必要とするので、ファイルサーバーと、計算サーバーの間は富士通 AP net と呼ばれる総計 200MB/s の転送速度をもつ特殊なネットワークにより転送が行われる。1 サーバー・クライアントあたりでは実測で 8MB/s 程度の性能をもつ。
- データーファイル 2：実験の生データーで、データー収集などの定型ジョブにより使用される。ファイル最大容量は約 80TB である。検出器からのデーターはすべてここに収集される。容量が大きいことと、最低転送速度を保証する必要があることから、テープの直接利用をつかっている。

ＤＦＳ と ＮＦＳ

ネットワーク分散ファイルシステムとしては、NFS が広く使用されているが、本システムでは、ホームディレクトリーサービスに DFS (Distributed File System) を使用している。ＤＦＳとＮＦＳの主な相違は、(1) クライアント側ファイルキャッシュ (2) ファイルディレクトリー情報 (実際のファイルのネットワーク上での物理的位置を含む) をファイルサーバー、クライアント機にもつか (ＮＦＳ), ネットワーク上のファイル情報サーバーにもつか (ＤＦＳ) という点、 (3) レプリカサーバーの有無である。

ＤＦＳでは各クライアントに必要に応じた量のキャッシュをもつことができ、複数クライアントのキャッシュと実体ファイルとの間の無矛盾、同期についてもよく考慮されている。クライアント数が増えた際にも複数クライアントからの同時アクセスによるサーバーが過負荷がおこりにくい。また、キャッシュにヒットしたデーターの読みだしについてはネットワークを介さないために平均アクセス時間が短縮される。一方、キャッシュにヒットしなかったデーターについては、キャッシュ操作のオーバーヘッドが生じるために一般にＮＦＳよりも転送速度があがらない。キャッシュヒット率が低い場合、頻繁にキャッシュの中身が更新されるため、これを管理するデーターベースやキャッシュ自体の更新、ガベージコレクションのオーバーヘッドが大きくなる。したがって、キャッシュのサイズやサーバーとクライアント間でやり取りされるデーターの単位であるチャンクのサイズ等のチューニングや、用途にあった使用が重要である。

上記(2)の特徴のために、DFS はファイルの物理的位置を動的に変更することができる (ただし移動中のファイルセットへのアクセスは待たされる)。例えば、あるユーザーのファイルセットを満杯になったサーバーから空いているサーバーへ移動することを、全体のサービスを停止することなく実行できる。また、左図のように二つの入出力口をもつディスクを使用し、コールドスタンバイ(cold stand by)サーバーを作ることができる。通常 DISK は server A がサービスしているが、A をメインテナンスする場合、スタンバイしていた server B に切り替えサービスを最短の休止時間で再開できる。ファイルの位置情報をもつ FLDB サーバーを変更することにより、クライアント側には透過的に切り替えが可能である。現在この仕組みをテストしている。



DFS ではファイルの複写をもつ複数のファイルサーバーをもつことができる。一台の読み書き可能マスターに対して複数の読み出し専用レプリカをも持つことができ

る。これは読み出しが主なアプリケーション等の供給に非常に有用であり、今後不可欠な技術であると考えられる。読み書き可能なレプリカサーバーは開発が待たれる技術であるが、ファイルの同期に困難があり、いまだ実現されていない。

システムの使用状況およびトラブル発生状況

ＫＥＫＢ実験は 1999 年の 2 月頃に本格的な実験開始がされる。したがって本システムにおいては、現在はテスト的なデーター収集と、シミュレーションデーターによる解析が行われている段階であり、昨年１２月現在の使用状況は計算サーバーの CPU 使用率が月平均で 70~90%、使用ファイル量 6TB(内４ＴＢが階層型ファイルシステムのテープ階層)で、ファイルサーバーの月平均ＣＰＵ使用率は 3~20%程度である。

述べてきたような各種の分散コンピューティングを適用した運用の結果はどうであったろうか。下図に現在までのトラブルの発生状況を描いたグラフをしめす。図中、ハード障害とある中に運用にまったく支障のないドライブユニットの交換は含まれていない。システムを運転してから一年半ほどは非常にトラブルが多かったことがわかる。主なトラブルの理由はソフトウエアおよびハードウエアのバグであり、特にマルチプロセッサー機でのＤＦＳクライアントソフトウエアの不具合、RAID コントローラーの不具合、ファイルサーバー機の CPU ファンの不具合によるトラブルが深刻であった。しかしながら、これらの障害がユーザー全体に影響を及ぼした回数は 1/10 以下であることがわかる。これらの障害は開発元の努力により約 1 年ほどで改善された。 図中システムダウンとあるのは、フォルトトレランス機能が存在しないところや機能しないことにより、ユーザーがまったく仕事ができなくなってしまう状況であり、当初は半月に 1 回程度もあった。また、これを解決するために頻繁なメインテナンスダウンが必要であった。しかしながら、最近はほぼ仕様をみたすダウン回数となっている。最後まで残っているハードウエア障害は CPU の故障による障害と、キャッシュメモリーのソフト放射線によると考えられるデーターエラーによるダウンである。