

最先端研究開発支援プログラム（FIRST）中間評価に係るヒアリング  
（超巨大データベース時代に向けた最高速データベースエンジンの開発と  
当該エンジンを核とする戦略的社会サービスの実証・評価）

1. 日時 平成24年9月19日（水）14:00～14:50

2. 場所 中央合同庁舎4号館12階 共用1214会議室

3. 出席者

相澤 益男 総合科学技術会議議員

奥村 直樹 総合科学技術会議議員

今榮東洋子 総合科学技術会議議員

青木 玲子 総合科学技術会議議員

大西 隆 総合科学技術会議議員

西尾章治郎 大阪大学情報科学研究科教授（外部有識者）

秋永 広幸 産業技術総合研究所ナノデバイスセンター長（外部有識者）

今井 浩 東京大学情報理工学研究科教授（外部有識者）

國枝 秀世 名古屋大学大学院理学研究科教授（外部有識者）

黒部 篤 株式会社東芝セミコンダクター&ストレージ社半導体研究開発センター長  
（外部有識者）

久間 和生 三菱電機株式会社常任顧問（外部有識者）

倉持 隆雄 政策統括官（科学技術政策・イノベーション担当）

川本 憲一 政策統括官（科学技術政策・イノベーション担当）付参事官（最先端研究  
開発支援プログラム担当）

4. 説明者

喜連川 優 東京大学生産技術研究所教授（中心研究者）

安達 淳 国立情報学研究所教授（研究支援統括者）

上田 修功 国立情報学研究所客員教授

豊田 正史 東京大学生産技術研究所准教授

## 5. 議事

### 【事務局】

それでは、皆さんお揃いですので、定刻より若干早いのですが、これより研究課題、「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的社会サービスの実証・評価」の中間評価に係るヒアリングを始めさせていただきたいと思っております。

本日の出席者は、お手元の座席表のとおりでございますが、喜連川先生中心研究者初め研究課題の皆様方にはお忙しい中、お集まりいただきましてありがとうございます。

本日の配付資料についてはお手元の一覧のとおりでございますので、ご確認をいただければと思います。その中で、研究課題側からお配りしました配布資料のうち資料2につきましては、機微な情報が含まれているということでヒアリング終了時に回収させていただきますので、あらかじめご了承ください。

このヒアリングは非公開で行いますが、後日、今後の研究発表、あるいは知的財産権等に支障が生じないことを確認した上で、議事の概要を公開させていただきます。

ヒアリングの時間配分につきましては、研究課題側から説明15分、質疑応答35分とさせていただきます。説明については時間厳守をお願いいたします。

説明に当たっては、あらかじめお願いしておりますが、課題全体の研究の進捗度合いと目標達成見通しについて、国際的な優位性、サブテーマの役割、相互関係を含めて簡潔で明瞭なご説明をお願いできればと思います。説明では、終了5分前に予鈴、終了時間に本鈴を鳴らさせていただきますので、時間が来ましたら説明の途中であっても中断していただければと思います。質疑応答では終了3分前に予鈴を鳴らさせていただきます。

それでは、ご説明をよろしくお願いいたします。

### 【説明者】

東京大学の喜連川でございます。超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的社会サービスの実証・評価というやや長いタイトルでございますけれども、この研究テーマの中間報告をさせていただきたく存じます。

超巨大データベースという言葉はもし今名前をつけるといたしますと、ご存じのようにビッ

クデータ時代ということになるかと思いますが、米国のホワイトハウスが、今年の3月29日に、ビッグデータはインターネットやパソコンと同じくらい大きく科学や産業、社会を変革するであろう、だからこそ我々は200ミリオンドルを長期にわたって投入する、という極めて強いメッセージを寄せられましたものですから、それ以降、米国のメディアも日本のメディアも大きく取り上げているということでございます。

まず最初に、こういうことが起こる2年半以上前に私どものこの研究の重要性をご認識いただきまして、最先端研究開発支援プログラムにご採択頂けましたことを心より感謝させていただきたいと思っております。おかげさまで、私どもは時代を先取りして骨太でしっかりとした研究を進めてくることができたと感じている次第です。

次の世代のITシステムの基盤というのは、ビッグデータを中心に非常に大きく変革がなされるであろうと考えております。つまりシフトしていくだろうと。この図にご覧いただきますように、いわゆるスマートグリッド、ITS、モバイル、ヘルスケア、農業、ロジスティクス、ありとあらゆる分野における高度なセンサー技術の進展によりまして、その活動が非常に精緻に可観測になり、観測された膨大なデータがクラウドの上の超巨大なデータベースに投入され、そこで高次の解析をし、従来には無い全く新しいサービスが生まれ、社会にフィードバックされる。こういう新しいIT基盤構造に、近未来において確実にシフトしていくだろうと考えております。

我々はこのリンクのアップリンクがビッグデータに相当すると考えているわけですが、ポイントとしましては、例えば現在アップリンクとダウンリンクの比は1：10でございますが、この構図ではアップリンクがダウンリンクよりも遥かに太くなることを意味しておりまして、根本的に大きくITシステムの構造が変わると考えております。その中でもビッグデータを入れる器であるこのデータベースエンジンというものが極めて重要であることは論を待ちませんし、ビッグデータを解析する技術も大切です。加えて、このような基盤で何ができるのか、社会がどう変革され得るのかという点に着目し、この研究を立案させていただきまして、今、申しあげました2つのサブテーマを設定させて頂いた次第でございます。

当初、自民党政権の時代に、1人平均90億円という予算額でございましたが、少しだけ遠慮致しまして、86億円ということでサブテーマ1とサブテーマ2の2つから成るテーマを申請致しましたが、その後民主党政権となり予算が大きく減額となりました。その中で本テーマではサブテーマ1のデータベースエンジンが無いとサブテーマ2のサービス実証が出来ないという依存関係がございますため、サブテーマ1を大きく縮減することは根本的に困難でした。した

が、いましてサブテーマ2を無しにしてしまうということもあり得ましたが、やはり納税者目線でプロジェクトの出口を大切にしたいという気持ちからサブテーマ1をやや縮減いたしまして、サブテーマ2を大幅に小さくするというような形で、この図のようにプロジェクトを再編成し、現在推進致している次第でございます。

まず、サブテーマ1のほうでございますけれども、このエンジン部分に関しましては、我々の技術を実際に社会に展開するためには、長らくデータベース管理システムを自社でコード開発されてきておられる日本で唯一の企業である日立さんとパートナーを組むことがベストという考えからで、日立さんをパートナーとして選ばせていただいた次第でございます。データベース管理システムは中にデータを入れ、長期間維持してゆかないといけないソフトウェアです。社会に該ソフトウェアを提供してゆくためには、大学がコードを開発するだけでは全くもって不十分で、コードを長期間にわたりメンテナンスする企業とのパートナーシップが不可欠とお考え頂ければと存じます。

次にこの図が私たちの研究・開発がターゲットとする領域を示しております。縦軸は実行頻度、横軸は実行あたりのデータのアクセス量を示しております。銀行、証券といえますのはこの辺のハイスループットのトランザクション領域に当たりまして、ほんの僅かのデータを、インデックス（索引構造）を用いて検索する領域でございます。ここはもうこれ以上高速化できないローアバウンドにあります。一方こちらはほとんど全部の情報を取り出して傾向分析するという、いわゆるデータウェアハウス系の応用に対応致しまして、これはハッシュクラスタリングなる手法が広く利用される様になっておりまして、ここも性能的にはローアバウンドになっております。

データが小さい場合にはこの2つの手法を切り替えるだけで済ませることが出来たのですが、データがビッグに（大きく）なってきますと少なくともない、多くもない、微妙な中間領域、この領域での仕事が現実には実質的にほとんどのケースになるだろうと想定されるわけですが、ここに対しましてはまだ決定打となる手法がまだないということから、我々はこの領域で圧倒的な性能を発揮する新しい方式を創案し、そのアイデアを実現した新しいデータベースエンジンの研究・開発を推進しております。

この図に示しますように、従来型の方式では、同期I/O、つまり赤いリクエストを出してその答えが返り、その後黄色を出して黄色が返るという順番にやっていくという決定的な処理手法が採用されております。この枠組みでは毎回実行順序が同じで、もちろん処理結果も同じになります。

これに対しまして、我々の非順序型と称しております新しい方式の場合は、ワッと出せるだけリクエスト出します。膨大な非同期 I O を用いて実現致しますが、この図にございますように、赤は一番最初に出したのですけれども、2 番目に返ってくる。黄色は 2 番目に出したのですが最後に返ってくるというふうにならざるを得ない順序になっていることが判るかと思存じます。即ち非決定的な処理が大きな特徴となっております。これを別の角度から見るべく実行トレースをグラフ化したものがこの図にございまして、に示しますと、この図のように、通常のはソフトウェアスタックのレイヤー間で行ったり来たりゆっくりと処理が進んでいます。それに対しまして私どもの手法は、この図のように膨大な数の非同期入出力が一度に出ているということをご理解いただけたらと思います。

それでは、私どものデータベースエンジンのパワーを感じて頂くべくデモをご覧に入れたいと思います。ビデオでお示ししておりますのは普通の廉価なディスクシステムです。実はかなり旧式で、最近では前面に LED がついている製品は売っていないものですから旧式なものを使っているわけですが、このディスプレイ画面に実行の様子をご覧に入れます。少々プロジェクタの輝度が低いので見にくいかもしれませんが、縦軸が I O P S、入出力のレートで、従来方式では、ご覧のように、非常に小さい値で推移していることがご理解いただけたらと思います。ここにアクセスするアドレスに対応して赤い点を表示しておりますが、大変スペースなためちょっと見えにくい状況です。もう少しいたしまして、今度我々の開発しております新しいアウトオブオーダー、非順序型のソフトウェアの実行の様子を示します。これだけの短時間で終わってしまいます。先ほどの通常システムの実行を途中で止めずずっとやっていると、この辺りまで延々とかかるところが、非常に稠密な I O を出すことによって高い性能が実現できることがおわかり頂けるかと思存じます。

もう 1 つ、デモをご覧に入れたいと思います。こちら側が従来の方式、反対側が非順序型です。パッと見ていただきますお分かり頂けますように、非順序型のパターンはバラバラであることがお分かり頂けるかと思存じます。つまり、実行するたびに毎回パターンが違ってきます。これがいわゆる非決定性、ノンディターミニズムと呼ばれているものに相当致します。これに対しまして従来方式であります順序型の実行と言いますのは、ちょっとしか実行が進んでいない為まだ細いのですけれども、ここからずっと扇型に回りまして、全周まわりここまで来ると完了という実行パターンに相当致します。まだここまでしか進んでいないということです。そして、他のアクセス構造体、インデックスも非常に高頻度にアクセスされていることがこのデモでおわかり頂けるかと思存じます。このように、通常の実行方式では殆どアクセスされていない

ということもおわかりいただけるかと思います。

さて、元のスライドに戻りまして、今のデモでちょっと見づらかった部分をご説明いたしますと、従来型のものでは薄く長い時間をかけながらアクセスしていたものを、今回の提案方式では、強烈に稠密なアクセスが実現されております。ハードウェア自身は2つの実験で全く同一で、ごく普通にあるシステムです。

さて、今後大きな流れとして、マルチコア化がございますが、現時点でマルチコアをまともに使えるアプリケーションというのは皆無に等しいと感じております。私どものエンジンはマルチコアに非常にフィットしておりまして、64コアの2年前の最大規模のシステムを震災直後に導入しまして実験を進めて参りました。

日立さんが頑張っていて、10テラバイトのデータベースでベンチマークをかけていただいておりますが、横軸がベンチマークのいろいろなクエリセット、縦軸が性能です。概ね300から500の間、つまり大体400倍ぐらいの性能が出ているということでございます。

実は震災直後の生研公開のときに、我々も頑張るぞ！という気持ちを込めて、来年に製品化するぞ！と言いまして、今年の終わりぐらいに製品発表のつもりだったのですが、今年になってビックデータが出てきましたものですから、大きく前倒しを致しまして5月末に製品発表させていただきました。

さて、大きな企業に対しては、しっかりとした会社がコードを保守、メンテナンスすることが必須で、そこを日立にお願いするというプロジェクトの構成をとりましたが、一方、もう少し小さめの会社にも私共が開発しました技術を利用可能なようにして下さいということを最先端プログラムの当初のヒアリングの際にもお受けいたしまして、そのご希望を反映すべく、大学ではオープンソースへ取組を行っております。オープンソースのデータベースで最も有名なものはここに示しますMySQLとPostgreSQLです。日本ではPostgreSQLがよく利用されますが、世界的に見ますと、やはりMySQLの利用が多いことから、MySQLへの我々の技術適用にチャレンジいたしまして、この図のように、MySQLに対しましても先ほどと同様、400倍ぐらいの性能が現時点で予定どおり出ているということでございます。

引き続きまして、サブテーマ2の課題でございますけれども、分析というところに対しまして、少しご説明をさせていただきたいと思っております。

文部科学省がビックデータで概算要求をするということがつい先日科学新聞に報道されました。その中には大変すばらしい記載がございます。即ち「データ科学の成功事例を生み出す」努力をするということが書いてございます。この図で、私どもはこのデータベースエンジンの

レイヤで骨太の技術を構築しつつあるわけですが、この上の分析のレイヤにおきましても要素技術というのをもたくさんございまして、これを全部やるというのはさすがに時間的、あるいは資金的に無理と判断致しまして、むしろ成功事例として上から必要な技術をプルするというのを考えております。つまり社会的意義の高い成功事例を幾つか創出するという事で、この技術の有効性を世の中に問おうと考えた次第でございます。

最後に、BOP、グローバルという視点から実は今日もお越しいただいておりますが、バングラディッシュの先生と一緒にBOPの国のヘルスケアを考えようと発起致しまして、従来ノーベル賞の関係でグラミン・フォン・レディが有名かと存じますが、それを乗り越えるようなヘルスケアレディというものを考えようとしております。バングラディッシュではヘルスケアコールセンターに1日1万5,000のコールがかかってくるのことでございまして、地方には医師が少ないため、大きなニーズがございます。電話では当然のことながら相手の顔も見えず、血圧等も測らずに診療行為をしなくてはならず、そもそも非常に無理がございます。そこで、実際に医療機器を詰め込みましたアタッシュケースを安価に作りまして、ヘルスケアレディに持って頂き、患者を訪問し、種々の検査をして頂きつつ、遠隔の医療を行おうという試みでございます。10000人の大規模実証実験を計画致しております。中島先生は昨日まで実はバングラディッシュにおられ、今日お戻りになられこの評価会に参加して頂いております。この図のここの部分の患者が検診の結果要注意だったわけですが、このシステムを導入することによりこれだけの部分の患者の健康状態が大幅に改善していることが判ります。ビッグになると想定されますデータを今後精緻に解析することによりBOPに適した新しい医療というものを創出しようと試みている次第です。日本の外において根本的に世界を変える改革を目指し、本研究プロジェクトの社会還元を遂行してゆきたいと考えております。

このように、今、社会で何が根源的な問題であるのかということをしつかりと見据え、幾つかの抜本的に社会を変革するようなサービスの成功事例を見せていくということが、ビッグデータが世界を変えるということを世の中に示す1つのアプローチではないかと思っております。

最後ではございますけれども、現在、私ども一部製品化いたしましたけれども、最終的には1000倍の性能を持つデータベースエンジンを開発していきたいと考えております。

ご清聴を心より御礼申し上げます。

【事務局】

どうもありがとうございました。

それでは、これより質疑応答に移りたいと思います。

ここからの進行については、相澤先生のほうでよろしく願いいたします。

**【有識者議員】**

大変今日はわかりやすく、出口のところまでしっかりと示していただいたので、よく理解できたところでございます。昨年はかなりしつこく、これはどこまで行くのかということをお伺いさせていただきましたが、今日はシステムティックに出口のところを意識しておられて。

**【説明者】**

昨年は貴重なご示唆を御礼申し上げます。

**【有識者議員】**

今日は本当に社会が求めているところに出口を具体的に設定して、展開されていく、そのときに先ほど出口のところからプルするような形で中間層に位置づけられている要素技術でしょうか、こういうようなところをフィードバックして必要があればそういうところもというところなんですが、今日、そういう形で示されたことは、このファーストの期間内ではその部分についてはどの程度の研究がここで展開される必要があるのかと。あるいはこの段階ではもう直結で出口のところにもう直接出ているんだから、この段階でこのプロジェクト期間はいいんだというご判断もあり得るかとは思いますが、その辺についてのお考えをお伺いしたいんですけども。

**【説明者】**

出口からプルするというものの中間層の技術というものを今後どの程度、展開する必要があるのか、もしくは、今あるもので十分ではなかろうか、ということに関しましてご質問を頂戴していると理解してございます。やはり本当に社会に役に立つソリューションを出そうと取り組みを始めますと、多様な技術をコンパウンドしないと1つのソリューションに結実しないというのが実情でございます。この際、現行の技術だけで済むのかと申しますと、やはりそれだけでは済みません。本日ご紹介いたしました事例におきましても、新しいアルゴリズムをつくり出すという局面は出てきております。そのような試みをどんどんほかの分野でもやろうと考えておりますが、非常にダイバーシティが広いものですから、一概にこの分野にはこの技術



が要る、あの分野にはあの技術が要るという整理が難しい状態であることをご理解いただけますと幸いです。先生から前回フォローアップ会議でご指摘をいただきまして、ほぼ全領域の分野に関して検討を行いまして、プロミシングな分野から進めてきているとうこととでございます。

ただ、社会に役立つ新しいサービスを実現しようとしますと、やっていく段階でどんどん必要な技術があらわれてくるのをまさに経験しているところでございます。完全なお答えになっていないかもしれませんが、N I Iも含めまして、この技術だったらこの研究者が世界で一番という情報は、少なくとも情報系では全部把握しておりますので、トップの研究者をダイナミックにお招きしながら新しい技術を創り上げる、それから、既存の技術もカスタマイズするという戦略で進めている次第です。

#### 【有識者議員】

そもそも検索速度1000倍向上というのは何の理由で決まっていたのでしたかね。当時、1000倍というのは非常にインパクトある目標で、30課題の中でも大変わかりやすい、しかもインクリメンタルではできないだろうという意味でも非常にわかりやすい目標であった。今日、お話を伺っても800倍とか、かなりのレベルに来られているんですが、果たして1000倍というのはどういう意味があって、決めていたんだろうか。お話伺っておりまして、場合によってはビックデータというこれから取り扱っていくことに日本でもなる。ですけれども、この取り扱っていく中で、技術が律速しているからこういう対象は取り扱えないんだという領域をなくしたいわけです。先生のこの成果というのは、私は非常にそういう意味では有力な業績になるのではないかと期待しているものですから、果たして1000倍は一体何だったのかという原点のお話を改めて伺えるとありがたい。

#### 【説明者】

コンピューターサイエンスにとりまして、1000倍という単位というのはシンボリックな意味を持ちます。例えばスーパーコンピュータの場合で言いますと、1ペタフロップスというのを超えたのが、実はIBMのロードランナというコンピュータですが、そのときは東芝、ソニーが作成したブロードバンドエンジンというものを、つまり、日本のゲームのマシンのプロセッサを、もうなりふり構わず投入して、1ペタフロップスの壁をアメリカは乗り越えました。

つまり1000倍単位のユニットがございまして、1000倍変わると世の中がふっと根本的に変わ

ったなという感じがするというのが、これは感覚的で恐縮でございますけれども、コンピューターサイエンスの技術者感として体感しているところでございます。1000倍変えるというのは、根本的に世の中が変わるというふうにご理解いただければありがたく存じます。

そういう意味で、我々はターゲットとして10倍とか50倍とかとは言っておりません。インテルのマイクロプロセッサが毎年どれくらい高性能化してきているかと言うと、1.5年で倍程度の性能しか上げておりませんので、それでも10倍、50倍というのは大変な数字なのですが、1000というのは非常にシンボリックなものですから、ここに挑戦しようと考えた次第でございます。ちょっと直接的なお答えになっているかどうかわかりませんが、ご回答になっておりますでしょうか？

#### 【外部有識者】

今の計算スピード1000倍の話ですけれども、実際に得られる性能は、このマシンに乗っけるアルゴリズムやプロセッサの数、あるいはデータ構造が、構造化されたものか画像のような非構造化データかによって、計算スピードは大分変わってくると思います。使う側からすると、どういう問題をどういう条件なら1000倍が出るのか、あるいは100倍になるのか、そういったところの指標を与えて頂けると、非常に使いやすくなると思います。チャンピオンデータの1000倍だけ言われても実用にはならないと思います。

#### 【説明者】

先ほど申し上げましたように、実は図のこの領域はローアバウンドです。ここは従来の技術に対して性能は上がりません。しかし図のこの領域では1000倍ぐらい上がります。逆な言い方をすると、この技術においては今のデータベースに比べて性能が落ちる領域は一切ないということを保証しています。

それから、先ほどいろいろな事例をお示ししておりますけれども、これはいわゆるビジネス系の基幹部分でございます。データベースの価値観をどこに置くかということに関係してくると思いますが、所謂非常に大手のインターネットの企業は廉価なシステムを利用しているというような印象がありますが、実は課金等の基幹部分は、本研究で開発しておりますような関係データベースを利用しております。即ち、ビジネス系のデータベースは非常に広く利用されているということを申し上げたく存じます。そしてまずは、このビジネスを対象とした業界のベンチマークで我々は勝負をしようと考えた次第でございます。ただ、今、久間先生がおっしゃ

られましたように、データは非構造の部分、つまりイメージのようなものもたくさんございます。こういうビジネスデータではない部分に対してどうしてくれるのかということのご質問も含まれていると理解します。

そのような非構造メディアデータ関しましては、現時点ではデータベースレイヤで大きなブレイクスルーが得られるような領域ではないと考えております。つまり大きな画像をとにかく読んで、そこを解析するという時にデータベースの役割はとにかくデータを取得するということになってしまいますので、そこで大きな工夫を入れる部分はあまりなく、画像の場合にはむしろ解析レイヤがより重要と考えております。一方で、先ほどご紹介しましたセンサーストリームのようなもの、これは株価のストリームと原則変わりませんが、そういうものは小さい粒の流れと見ることができまして、この種のデータを対象としたデータベース処理に対しましては劇的なブースティングができます。

このように、データベースにはいろいろな種類がある中で、どの部分が一番得意なのかというご質問をいただきますと、我々はまず世の中でビジネス規模が最も大きく、その効用が広くご納得いただけるような領域、別の言い方を致しますと、現状で多くの方々がビッグデータ時代を迎えるにあたってとにかく困っている領域を1000倍速くしたいと考えております。

非構造データに関しても、挑戦をしないと申し上げていることではありませんが、あまりに多様性が高いものですから、それに対応可能な一般的な方法論が世界的に見ましても現時点では解が得られておりませんことから、先ほども申し上げましたように、大きな社会的課題を解くという視点から、これはやらなければいけないということに対して1歩1歩チャレンジをしていきたいと考えております。その中で新しいデータベース、ちょっと今日はご紹介する時間がございませんでしたけれども、プロバブルスティックなデータベース、確率論的なデータベース、これらの要素技術の研究も進めておりますが、いろいろな新しい技術を創意工夫しながら取り組んでいきたいと考えております。

#### 【有識者議員】

今の幾つかのご議論の続きですが、1,000倍の高速化ということにはこだわりがあるというご説明で、今いろいろなデータがありますけれども、約300倍程度を達成したと。残り、ここでいただいたものでは800倍の高速性を実証するのを目指すと書いてあります。1,000倍までは開きがある。その1,000倍まで上げるところは、今まで開発されてこられた技術をそのまま延長していけばいいのか、それとも何か新しい発想を入れて、今までのとはまた違う質のもの

で1,000倍というものを達成しようとしているのか。そのあたりの展望についてお話しいただけますか。

**【説明者】**

頂戴致しましたご質問は事前の質問リストにもございまして、ご回答させていただきましたが、非常に経緯が複雑になっておりますことから、説明が込み入っておりご寛恕下さい。当初の申請予算に対しまして、大きく減額になりましたため、購入するハードウェアの規模を小さくせざるを得なくなりました。開発するソフトウェアそのものは当初想定していた規模のハードウェア環境を持ってくれば1,000倍の性能を発揮することができます。ハードウェアの導入に関しましては、2段階でやろうとしておりまして、現段階で大体400倍から500倍程度、もう1年後ぐらいの導入マシンでは800倍程度と計画しておりました。

ところが昨年、タイで大規模な洪水が起きたことをご存じかと思いますが、この影響でディスクの工場がタイに集中しておりましたことから、ワールドワイドの市場からディスクが急に品薄となりまして、ほとんど供給が途絶される状況になってしまいました。私共も買えなくなり困り果てておりましたが、ここ極めて最近の動きでございまして、ビッグデータというブームが生まれましてに起因してか、必ずしもその理由は明確に理解出来ておりませんが、ディスクの量産体制が急速に強化されて参りまして、この図に見えますように価格が下がるという事態までに至っております。このように、計画当初では購入不能と考えておりましたものが、現時点におきましては、1,000倍程度の性能向上を実現可能なシステム環境を購入することが可能となりまして、800倍ではなく、概ね1,000倍程度は達成可能と現時点で考えている次第でございます。ちょっとややこしいのですが、そういうふうにご理解いただければありがたいと思います。

**【有識者議員】**

研究的内容的には今の延長でハードウェア、ディスクを追加すればいいという、そういうことですか。

**【説明者】**

ハードウェアといいますかプロセッサの多コアへの対応が中心となりますけれども、もう一段大きなソフトウェアの工夫が必要となります。ここはかなり技術的に深い為詳細を簡単に一

言でご説明することは困難ですが、膨大な努力を重ねております。ITは1.5年ですべてが倍になるという世界でございますので、ハードウェアは、今の延長ですむかといいますと、それはそれほど簡単なものでもございません。現在のシステムは数多い筐体で構成されておりますが、これよりもはるかに少ない筐体数で、フットプリントも非常に小さくしつつも、コア数を増やし十分1,000倍の性能を出すということを考えている次第でございます。

「こだわり」という言葉を先ほどご指摘頂きましたが、誤解を招きますといけませんので、再度申し上げますと、奥村先生からおっしゃっていただきましたのでこだわりを熱っぽく語ってしまいましたが、極度にそこを目指しているということよりも、先ほど久間先生におっしゃっていただきましたように、むしろ安定して性能を出せるとか、技術的にはもっといろいろな課題がございまして、それらの一つずつ解決してゆく必要がございます。必ずしもピーク性能の1,000倍というものだけが全てのように強く申し上げているつもりでないことをご理解いただければと存じます。

#### 【外部有識者】

商品化が前倒しで進んでいるということで非常に素晴らしいことだと思います。先ほどの久間先生の質問とも絡みますが、例えば今までのあるデータベースシステムのあれでつくっていた大量のデータがあったとします。そこへ例えば日立さんからこのエンジンを買って導入すれば、そこでもともとつくったデータのあれと有効に使いながらこのエンジンを生かしていけるという、そういうポータビリティ的なものが備わっているのかどうか

#### 【説明者】

ようやくデータベースエンジンが動き出すようになりましたことから、現実に大きなデータをお持ちでありそうな方々とお話をしてみますと、日本では大きなデータを本当に運用しているところは実はあまり多くないことが判ってまいりました。この事実は昨年マッキンゼーのレポートにも出ているところでもあります。日立さんが製品化され、今幾つかの企業で現実にお使い頂き始めつつあります。それは先生に今まさにおっしゃっていただきましたように、今もちろんデータベースはありますが、圧倒的に速い環境の味見をして頂くべく、まずはデータ全部ではなくて、実証段階として、そのサブセットを切り抜いて当方のエンジンに入れて頂き、100倍程度の性能ゲインが得られるというようなことをまさに確認しつつあるという状況でございます。現業のシステムと新しい味見システムが併存してポータブルに稼働しております。

#### 【外部有識者】

今後、マシン to マシンのデータのやりとりが発生する中で非常にいい成果が出ているのではないかと今思ったところです。とはいえプロジェクトの評価でもあるので細かいところをご質問させていただきたい。

日立の方々のご貢献はどのような形になっているのか。これを教えていただきたい。

#### 【説明者】

多くのユーザがデータベースを長い期間利用されておられるわけですが、あまり気に留めておられないかと存じますが、OSが変わってもデータベースは変わらないということにお気づきになられるのではないかと思います。それぐらい長い期間利用されるソフトウェアというのは珍しいと言えます。又、OSのコードボリュームよりもデータベースのコードボリュームの大きい状況になっているとも聞きます。

このような長い歴史のあるデータベースソフトウェアに対しまして、今回、私どもが、従来には無い新しい実行原理を提案させて頂いているわけですが、それはエンジンと呼ばれる核部位に相当致しまして、現実にはその周辺部位に非常に多様な技術が過去延々と積み重なってきていることも事実です。数多くの周辺機能を我々がひとつひとつ全て作り込んでゆくことは非現実的で、短期間に利用可能なプロダクトとするには、現状有るものを可能な限り流用するなどの策が不可欠と考えました。日立のデータベースソフトウェアは大きな開発の蓄積があり、ユーテリティ、コンカレンシーコントロール、オブティマイザ等、そのままというわけにはゆきませんが、修正を加えることで流用可能な極めて多くの周辺モジュールがございます。従いまして、日立の現行データベースソフトウェアの最も核となる部分を我々のアイディアに基づく新エンジン、この新エンジンは共同で開発してきたわけですが、それで取り替えられると同時に、その周辺モジュールとのインタラクション、この部位は実際にそのコードを持っておられる日立しか実施出来ないわけですが、そういうところを日立さんが精力的にご開発になられているということがございます。

#### 【外部有識者】

応用で今、省エネであるとか新エネでデマンドオンレスポンスであるとかスマートグリッドとか電力系統、こういったところで膨大なデータ処理が必要ですね。これは日本全体の問題で

もありますね、このエネルギーの問題というのは。そういったところに今からアプリケーションの1つとして、金がないとおっしゃるかもわかりませんがやっておく価値が私はあると思います。

**【説明者】**

実はJSTのCRESTで今EMSというプロジェクトが立ち上がっておりまして私はそのアドバイザーメンバーを拝命しております。これはまさにスマートグリッドの理論的なフレームワークをしっかりとしようというプロジェクトでございます。先日も経産省が実施している日本で4か所の実証実験の一つである北九州に見学に出向いて参りました。三菱電機さんも非常に強くコミットされています。

そこで私が申し上げたのは、何とかデータを研究者にアベイラブルにしてくださいということです。現時点ではまだそれほどデータがあるわけではないようなのですが、色々とデータが蓄積されてまいりますと、ぜひ私どもの中で研究開発ができればと思っています。実はアメリカにエプシロンという組織がありまして、米国のスマートグリッドのデータを有しており、そこからデータを取得することも検討致しましたが、アメリカの電力環境と日本のインフラがあまりにも違うものですから、断念した経緯もございます。

**【外部有識者】**

だから日本独自のデータベースをつくっていかないといけないですね。

**【説明者】**

その通りでございます。特に関西は進んでおり、関電さんが数万世帯もう既にスマートメータを導入されておられますので、本日は非常に良いご示唆を賜りましたので対象領域として検討させて頂きたく存じます。

**【有識者議員】**

時間が参りましたので終わりますが、確認させていただきたいのは、今出口として事例を示されたのは、現在の300倍ぐらいを、この系で実現できることだと理解してよろしいわけですね。

このプロジェクトとしてのまとめ方ですが、先ほどの1,000倍規模に拡大して行って、それ

で展開していくというのはどういうまとめ方になるでしょうか。

**【説明者】**

先生のご質問と私の回答がピッタリ合っていないかもしれませんが、最先端でビッグデータの研究を進めて参りましたところ、非常に多くの方々からご興味を持って頂きまして、最近では1週間に一度以上講演をしたり、あるいはいろいろなジャーナリストの方からインタビューを受けたりするようになりまして、更に、どんどんと私共の取り組みが知られるようになり、「沢山のデータがあるのですが何とかありませんか」というリクエストが来るようになっております。その中で最先端としてやるべきことというクライテリアを丁寧に考えながら、即ち、本プロジェクトでやることの意義をきっちりと見据えながら社会的に価値の高い実証を実現すべく研究を進めている所存でございます。根源的に社会的に大きな課題であるとどの国民の方もご理解頂けるような問題に対して、やはりすごいことをやったねと何年か後に言っても頂けるようなチャレンジをしようとメンバ全員で日々努力致している次第です。

**【有識者議員】**

今のお答えはそれはそれで非常に結構だと思いますが、もう1つは1,000倍でなければできないというところはどこに焦点を当てられるかということをお伺いしたいということがあります。

**【説明者】**

ピッタリと1,000倍でなければいけないところがどこかというのは、時代とともに課題も、他の情報技術も変化しますので、特定が容易ではございませんが、少なくとも1,000倍速くなりますと、何時間とかかっていたものが本当に実時間で処理可能な状況となって参ります。これは非常に大きな思考の変革、すなわち、即座に返事を返せるシステムと、後から返事をするというシステムではシステムの作り方、サービスの在り方そのものが根底から変わるわけございまして、そのようなドラスティックな変化を可能とする技術をどのように社会で活用してゆくか、或いは、従来不可能であった全く新しい社会サービスの創出を検討するコミュニティをつくりたいと考えておりまして、今日もお越しいただいておりますがN I Iにそういうセンターをつくり、産学連携の議論の場を正に今醸成してゆこうとしているところでございます。



**【有識者議員】**

大変充実した議論が展開できたのではないかと思いますので、どうぞよろしく願いいたします。ありがとうございました。

**【事務局】**

どうもありがとうございました。それではこれでヒアリングを終了させていただきたいと思っております。ありがとうございました。