

(2) 代表的な個別事業の紹介

上で紹介した 49 件の個別事業の中から代表的な 4 つの個別事業を選んで、以下で紹介する。

紹介するのは、以下のものである。

- 大規模データストリーム処理基盤の研究開発（領域：基盤（プラットフォーム）、事業者：富士通株式会社）
- 行動情報活用型クラウドサービス振興のためのデータ匿名化プラットフォーム技術開発事業（領域：基盤（安全性）、事業者：株式会社日立コンサルティング）
- 本格研究による人起点のサービス工学基盤技術開発（平成 23 年度次世代高信頼・省エネ IT 基盤技術開発・実証事業（サービス工学研究開発分野））（領域：応用（サービス）、事業者：独立行政法人産業技術総合研究所）
- 3D 映像品質評価技術調査研究（平成 23 年度次世代高信頼・省エネ型 IT 基盤技術開発・実証事業）（領域：応用（コンテンツ）、事業者：財団法人デジタルコンテンツ協会）

1) 大規模データストリーム処理基盤の研究開発

①事業目的

今後のクラウド浸透に伴い、ICT リソース（サーバ、ストレージ、ネットワーク）やアプリケーションの集約による効率化とともに、常時発生する多様かつ大量系列データ（データストリーム）を高速に処理・分析し、意味ある情報知識抽出組織や地域を超えて共有することによって、企業や社会の大域的な変革や新たな知の創造が可能になると推察される。

本事業は、公募技術 5)「高速化に関する技術 クラウドコンピューティングの処理速度を向上させる、大量データ収集・処理・分析技術、並列分散処理技術、ストリーミング処理技術等」に関する事業であり、上記のような背景の下、多様・大量データストリーム処理基盤サービスの実現に向け、クラウドコンピューティングの処理速度を向上させる、大量データ収集・処理・分析技術、並列分散処理技術としてのクラウド上の大規模データストリーム処理基盤、具体的には、以下を研究開発・技術検証することを目的としている。

【A】データストリーム処理開発実行環境

- ・多様で大量のデータストリームを低コストかつリアルタイムに処理し、かつ、負荷状況に応じて処理を並列分散化して高速な応答性を保証する、処理方式の記述及び実行環境の開発

【B】 ストリームストレージ

- ・ログ／センサー情報収集解析サービスなどにおける大規模ストリームデータの配置・格納方式の開発

【C】 ストリーム処理対応仮想化基盤

- ・仮想化をベースとしたクラウドサービスにおいて、大量のストリームデータを保存するネットワークストレージから一定時間内にデータを読み書きすることを保証するための I/O 性能保証技術の開発

【D】 ストリーム処理におけるセキュリティ

- ・大量のストリームデータを収集・格納・共有する環境において必要となるデータセキュリティの強化

【E】 上記技術を統合した技術検証

- ・実データなどを使用して各技術及び全体を検証

(参考：【D】・【E】は当該年度事業の範囲外ではあるが、将来的な関連項目として参考までに記載)

図表 3-9 にデータストリーム処理基盤の全体構成を示す。

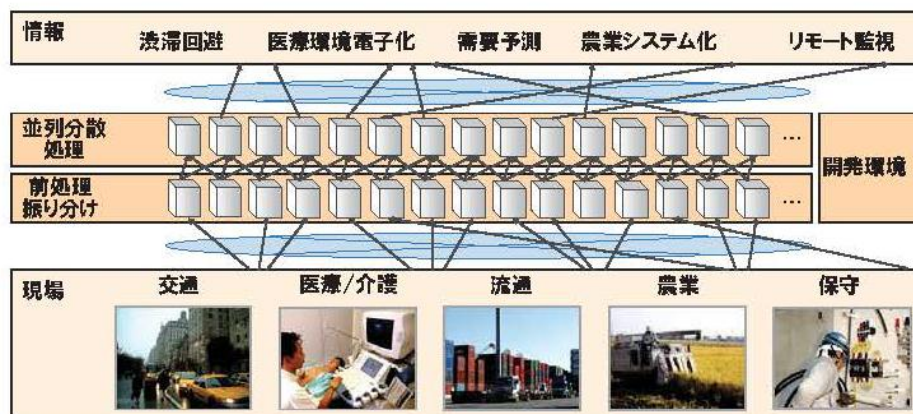


図表 3-9 データストリーム処理基盤

②データストリーム処理とは

ここで、データストリーム処理とは、実世界の活動や業務により常時発生する大量の時系列データをリアルタイムに処理・分析する技術のことである。このようなデータストリーム処理技術の利用によって、例えば、交通分野においては、時々刻々と変化する個々の車の位置情報を大量に収集し、道路ごとの渋滞状況を分析し、それを基に渋滞回避ルートを提示したり、信号制御したりすることによって、大域的に渋滞を解消し、個々人の移動時間を短縮し地域全体の交通の効率化するとともに、CO2 排出量を削減し環境負荷を低

減することが期待されている。図表 3-10 にデータストリーム処理の概観を示す。



図表 3-10 データストリーム処理の概観

③参考：データ量の一例

データ量の一例を以下に示す。

●農業分野

- 気象・土壌データ、GPS データ、画像データ、現場での気づきデータ、圃場データを日々蓄積するデータ容量は 5-10MB/件・日
- 独立行政法人農業研究機構の報告によると農業データは 10-30 年の蓄積が必要
- ビジネス展開時には 50-100 万件を目標
- 以上から、データ総容量は 100PB を超える

●医療分野

- 個人の生涯健康管理データ=PHR（Personal Health Record）が 2PB/人とされている

④事業概要

前記目的達成のためには、時々刻々と発生する膨大なデータストリームを負荷状況に応じていかに並列分散処理し、一定時間内に読み書きできるようにする、というような技術的課題がある。その解決のために、本事業においては、低コストかつ大量の ICT リソースが利用できるクラウド上で以下を実施する。

【A】処理ルールの分割や負荷に応じた再配置などを行うデータストリーム処理開発実行環境の開発

※H22 年度は、そのうちの並列分散方式の設計、及び基本方式の試作

【B】 ストリームデータを分割、データ関連付け、階層的配置、データ加工などを行うストリームストレージの開発

※H22 年度は、そのうちの要件／仕様の検討

【C】 一定時間内応答のための分散仮想スイッチ技術や多数VM集約時の I/O スケジューラ機能を強化したデータストリーム処理対応仮想化基盤の開発

※H22 年度は、そのうちの仮想スイッチ方式の調査、及び仮想スイッチ QoS 方式の設計

クラウド上での大規模なデータストリーム処理の動的な負荷分散[A]や無限に続くデータをリアルタイム処理に適しか形式で蓄積できる[B]、また、それを支える[C]は従来になく、以上により大規模データストリーム処理基盤サービスを実現し、企業や社会の大域的な変革や新たな知の創造に貢献する

⑤目標

技術的な目標指標は以下の通りである。

A) 現状

- ・スパコン上でスループット：500 万[件／秒]、レイテンシー：1[ミリ秒]

B) 目標

- ・IA サーバ(Intel Xeon、2 GHz)相当のプロセッサで構成されるクラウド上で、簡易なデータストリーム処理を想定して、データストリームを並列分散処理したり、並行処理したりする場合、以下の処理性能を目標とする。
 - 数 10 台以上のサーバでの並列分散処理によって、スループット:100 万[件／秒]以上、処理レイテンシー※:50[ミリ秒]以下
 - ※ネットワークによる遅延を除く。
- ・上記目標は 3 年目である H24 年度終了時の目標である。
- ・H22 年度の目標は、上記目標に向け、前記のようにデータストリーム処理基盤を構成し、個々の要素技術を設計・実装することによって、上記目標が設計上達成可能であるという見通しを得ることである。

⑥報告書目次

この事業の事業報告書の目次を、図表 3-11（5 枚）に示す。

目次

第1部	事業概要	7
1.	事業目的・概要・目標	7
1.1	事業目的	7
1.2	データストリーム処理とは	8
1.3	ご参考:データ量の一例	9
1.4	事業概要	9
1.5	目標	10
1.6	他社動向	11
1.7	今回差異	12
2.	実施内容	13
2.1	課題と提案技術	13
2.2	実施スケジュール	21
第2部	データストリーム処理開発実行環境	24
1.	データストリーム市場・技術動向	24
1.1	データストリーム処理システムの技術動向	24
1.2	クエリ記述言語の技術動向	28
2.	技術的課題	31
3.	基本設計	35
3.1	設計方針	35
3.2	全体構成	35
3.3	クエリの並列分散実行方式	36

図表 3-11(その1) 報告書の目次(その1)

3.4	動的構成変更の方式	38
3.5	今回の設計方式におけるアーキテクチャの制約	40
4.	機能設計	41
4.1	システム構成	41
4.2	実現方式	47
4.3	実現機能	51
4.4	処理概要	51
4.5	複数ストリームの並列処理	53
4.6	アプリケーション固有処理インタフェース	54
4.7	データストリーム処理実行環境インタフェース	61
4.8	データストリーム処理開発環境インタフェース	73
5.	構成設計	78
5.1	システム・ユースケース一覧	78
5.2	システム・ユースケース詳細	79
5.3	クラス構成	96
6.	評価	103
6.1	目的	103
6.2	評価方法	103
6.3	測定結果	105
7.	まとめと今後の課題	167
7.1	成果	167
付録 A	デモシステム内容	169
A.1	動作概要	169
A.2	システム構成	172

図表 3-11(その 2) 報告書の目次(その 2)

A.3 構成設計概要	174
A.4 状態表示項目	184
第3部 ストリームストレージ	187
1. 技術的課題と開発すべき技術	187
1.1 技術的課題	187
1.2 開発すべき技術	188
1.3 ストリームストレージ性能要件	189
1.4 性能基礎値	191
1.5 ストリームストレージ機能要件	204
1.6 開発すべき技術項目	205
1.7 市場・技術動向	206
2. 基本設計	209
2.1 ユースケース	209
2.2 全体構成	212
2.3 ストレージストリームの動作概観	214
3. 機能設計	225
3.1 ストリームストレージクライアント	225
3.2 用語の定義	226
3.3 時刻公差の導入	227
3.4 イベント記録	228
3.5 イベント再生	230
3.6 イベント振り分け関数	231
3.7 加速再生	232
3.8 追っかけ再生	233

図表 3-11(その3) 報告書の目次(その3)

3.9	ストリーム-オブジェクト変換	236
3.10	ストリーム向け階層間制御	237
3.11	通信量削減方式	238
3.12	データ量削減方式	239
3.13	ストリームデータ用アクセス API	239
4.	まとめと今後の課題	261
4.1	成果	261
4.2	今後の課題	263
付録 A	性能基礎値測定データ	264
第 4 部	ストリーム処理対応仮想化基盤	277
1.	技術的課題と開発すべき技術	277
1.1	技術的課題	277
1.2	開発すべき技術	277
2.	調査・分析	282
2.1	I/O アクセス強化のための技術と評価	282
2.2	I/O 性能安定化のための技術と評価	291
2.3	スケーラビリティ向上のための技術と評価	293
2.4	トラフィック・負荷状況の見える化のための技術と評価	299
3.	ストリーム処理対応仮想化基盤の基本設計	300
3.1	分配器、ストレージ、処理エンジン構成	300
3.2	仮想化基盤の構成	304
4.	まとめと課題	348
4.1	成果	348
4.2	プロトタイプ評価から抽出された各構成要素の課題とその解決策	348

図表 3-11(その 4) 報告書の目次(その 4)

4.3 新規開発項目	351
付録 A. 調査分析項目一覧	354
A.1 vhost-net	354
A.2 Cgroup	355
A.3 Datapath	371
A.4 10GbE	371
A.5 VLAN (タグ拡張)	371
A.6 Link Aggregation (bonding)	372
A.7 VEPA, VEB	373
A.8 Reliable Multicast	374
A.9 L7 SW 連携	374
A.10 TRILL (TRansparent Interconnection of Lots of Links)	375
第 5 部 まとめ	376
1. 今回の成果	376
2. 今後の課題	382
3. 適用領域	384
4. 事業の普及	386

図表 3-11(その 5) 報告書の目次(その 5)