楽天におけるビッグデータとその収集・解析基盤の構築

平手 勇宇（楽天株式会社）　王 永坤（楽天株式会社）

概要　楽天は 1997 年の創業以来，会員数・流通額・提供サービス数を急速に拡大させており，楽天市場を中心とする楽天のシステムでは，様々なサービスから生成される多種多様な大規模データが日々蓄積され続けている．ビッグデータと呼ばれるこれらのデータ群を，いかに統一的なプラットフォームで収集，解析し，ビジネス側に迅速かつ的確にフィードバックを行うかが，楽天にとって重要な課題となっている．この課題に対処するために，楽天では現在，楽天スーパーDB と呼ばれる大規模ユーザ属性 DB プラットフォーム，およびグローバルイベント解析プラットフォームを構築・運用を行っている．本論文では，楽天でのビッグデータの例を紹介するとともに，２つのプラットフォームの紹介を行う．

1. はじめに

楽天は，1997 年にオンラインショッピングモール「楽天市場[1]」のサービス提供を開始して以来，会員数，流通額を伸ばしており，2011 年は楽天市場における年間流通額が 1 兆円を突破した. また, 楽天市場だけではなく，オンライン旅行サイト「楽天トラベル[2]」や, ポータル・メディア事業，銀行やクレジットカード等の金融事業など, 提供するサービスを拡大させてきている. さらには，楽天はグローバル化を進めており, アメリカ, フランス，イギリス，ドイツ，タイなど，EC サービスを提供している地域を拡大させている．

事業の拡大に伴い，楽天グループのシステムでは，多種多様なフォーマットの大規模データが，至る所で日々生成され続けており，これらのデータをいかにして，横断的に収集・蓄積し，解析をすることでサービスに活用していくのかが重要な課題となっている．このような状況を踏まえ，現在楽天では，多種多様な大規模データの収集，蓄積，解析を実現するプラットフォームの構築・運用に取り組んでいる．本稿では，その中の取り組みの例として, 7,500 万人を超える会員のユーザ属性情報を一元的に管理する DB プラットフォーム，およびグローバルイベント解析プラットフォームの構築・運用を取り上げ, 楽天におけるビッグデータへの挑戦について述べる．

本論文では，以降，次のような構成をとる．第 2 章にて楽天で取り扱っているビッグデータの例を示し，第 3 章にて楽天スーパーDB について紹介を行う．第 4 章では，グローバルイベント解析プラットフォームについて紹介を行う．第 5 章にてまとめを行う．

2. 楽天におけるビッグデータ

楽天では様々な種類の大規模データを扱っているが，代表的な大規模データの例として，楽天市場の商品データ，および商品レビューデータが存在する．

図 1 に，2004 年 10 月以降の楽天市場にて取り扱っている商品数の推移を示す*．図 1 に示す通り，2004 年 10 月の段階では，取り扱い商品数は約 900 万商品であったのに対し，2012 年 8 月 21 日現在，取り扱い商品数は約 1 億 700 万商品と，爆発的に増加している．

同様の傾向は，楽天市場の商品レビューデータにも見られる．図 2 に，2003 年 8 月以降の楽天市場の商品に対して投稿されているレビュー[3]総数の推移を示す†. 2003 年 8 月よりスタートしたレビューサイトは, 2012 年現在, 約 8,680 万のレビューが存在し, 日々1 万程度のレビューが投稿されている．

図 1, 図 2 に示した商品データ, 商品レビューデータは楽天が取り扱っている大規模データの一部であり，これらのデータの他に，7,500 万人を超える会員情報，8,000 万を超える購買履歴データ，ユーザによる検索クエリログ，商品ページのクリックスルーログ，広告のクリックログ，クレジットカードの利用情報等の大規模データが存在する．

多種多様なサービスから得られるユーザ属性情報，および様々なフォーマットの大規模データを横断的に収集・解析し，サービスに役立てていくことは大変重要である．そこで楽天では，現在，全サービスが共通して利

* 最新の取扱い商品数は，http://www.rakuten.co.jp/にて公開されている．

† 最新のレビュー登録数は，http://review.rakuten.co.jp/にて公開されている．

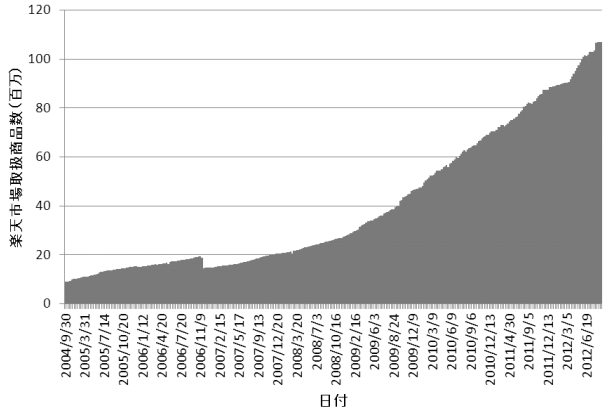

図 1. 楽天市場での取り扱い商品数の推移

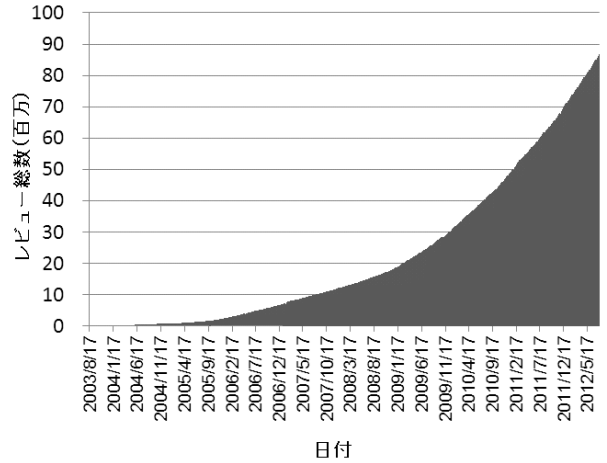

図 2. 楽天市場に投稿されているレビュー総数の推移

用するユーザ属性情報の DB プラットフォーム(=楽天スーパーDB)，および，多種多様な大規模ログを収集し解析するグローバルイベント解析プラットフォーム等の構築・運用を行うことで，これらの課題に立ち向かっている．楽天スーパーDB は，増加する楽天会員を一元的に管理する DB であり，会員に紐づく静的な情報を保持する DB プラットフォームである．対してグローバルイベント解析プラットフォームは，アプリケーションから発生する多種多様なログデータを収集する解析基盤で，動的な情報をリアルタイムに取り扱うプラットフォームである．以降では，これら 2 つのプラットフォームについて，具体的に述べていく．

3. 楽天スーパーDB

楽天スーパーDB とは，楽天グループが運用している多種多様なサービスから発生するユーザの属性及び行動に関する情報を，一元的に管理する DB プラットフォームである．図 3 に示す通り，楽天スーパーDB には楽天会員約 7,500 万人のユーザ属性情報，およびユーザの行動に関する情報として，購入履歴情報，アンケート情報，

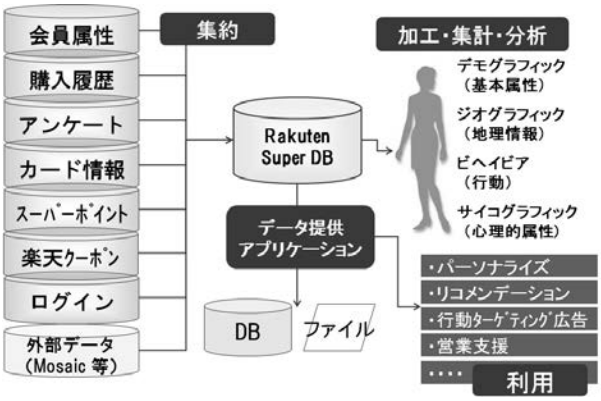

図 3. 楽天スーパーDB の概要

カード情報などの情報が蓄積され続けている．また，ユーザ属性情報・行動情報の一次データだけではなく，これらの情報を，デモグラフィック観点，ジオグラフィック観点，ビヘイビア観点，サイコグラフィック観点等で加工・集計・分析した結果についても同様に保持している．この DB に格納されている情報は，パーソナライズ機能，レコメンデーション機能，行動ターゲティング広告機能等に応用されている．

楽天スーパーDB には，多種多様なサービスからの情報が記録されているため，単一サービスからの情報だけでは分かりえない，横断的なユーザの特長を把握することが可能である．この特性は極めて重要であり，たとえば，あるサービスの利用経験がないユーザであっても，他のサービスでの利用経験があれば，当該ユーザの特徴を把握することができ，レコメンデーションにおけるコールドスタート問題などの緩和に役立つ．

一方で，楽天スーパーDB には，多種多様なサービスからの個人情報が格納されているため，当該 DB は極めてセンシティブな情報の集合体とみなすことができる．したがって，セキュリティに関しては特に厳しく設計・運用されており，たとえば，データ取得の際には，当該 DB プラットフォームに直接接続するのではなく，別途データ取得アプリケーションを経由してデータを取得するようにしている．

4. グローバルイベント解析プラットフォーム

グローバルイベント解析プラットフォーム(=GEAP)とは，楽天が運用している様々なサービスから生成される多種多様な大規模ログを収集し，蓄積し，そして解析するプラットフォームである．

アプリケーションやサーバのログを収集・解析し，解析結果をアプリケーション側にフィードバックすること

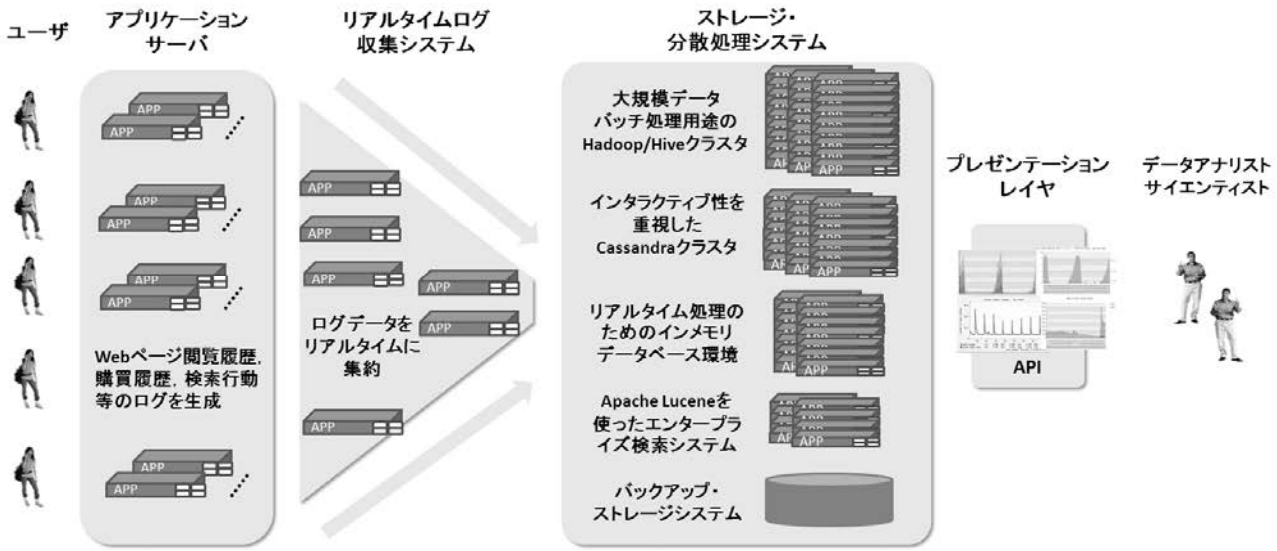

図 4 グローバルイベント解析プラットフォームのアーキテクチャ

で，サービスの最適化を図っていくことは，基本的な事である．楽天ではこれまで，このような解析は外部サービスを利用したり，サービス毎に独自で構築することで，最適化を実施してきている．しかし，外部サービスを利用する場合には解析のカスタマイズ性に限界があり，サービス毎に独自で構築する場合には，構築・運用コスト，メンテナンスコストが大きかったり，サービス間をまたぐ解析をすることができないという問題点がある．そこで，楽天では， GEAP を構築することによって，様々なサービスのログ収集，及び解析を，統一的なプラットフォームによって提供しようと試みている．

データフォーマットが異なる様々なサービスに対してログ収集・解析環境を提供するため，GEAP は，様々なログフォーマットに対して適用可能であることが必須である[‡]．また，近年の楽天グループのグローバル化の施策により，海外で運用されているサービスが増加しているため[§]，データ発生源の地理的な分布・ネットワーク上の分布も分散していく傾向にある．この傾向は今後一層強くなると予測される．このような楽天の環境に合わせ，GEAP は，次の 3 つの特長を有するように設計されている．

- ●データの多種多様性を吸収することができる．
- ●様々な用途に柔軟に対応できるよう，様々な解析を実施することができる．
- ●国外で生成されるデータも問題なく解析対象とすることができる．

GEAP の構築にあたって，我々は，Apache Hadoop[6], Apache Cassandra[7], Apache Flume[8]等，様々な OSS を採用している．また，我々は GEAP を構築することで得られた知見やバグ Fix 等，これらの OSS への貢献も実施している．

本章では以降，4.1 節にて，GEAP の概要を示した後，4.2 節で，GEAP を構築するにあたって遭遇した課題点とその解決方法を述べ，4.3 節にて，GEAP の応用例を紹介していく．

4.1 GEAP の概要

図 4 に示すように，GEAP は，次に示す 3 つのコンポーネントで構成されている．

- ●高信頼で，高速なリアルタイムログ収集システム
- ●ストレージ・分散処理システム
- ●GUI や API 等で，可視化したデータ分析結果を提供するプレゼンテーションレイヤ

ログ収集システムは，ログを生成する多数のホスト上で動作するアプリケーションから，ストレージ・分散処理システムにデータを集約するシステムである．収集に当たっては，ログデータを集約しながらストリーム形式で送信し続け，最終的にストレージ・分散処理システムにデータストリームを到達させる．国内外に位置する多数のサーバから，高信頼で，高速にデータを収集する必要があるため，我々は，Apache Flume[8]をベースとし，信頼性を向上させる手法を適用することで，ログ収集システムを構築している[**]．

ストレージ・分散処理システムは，ログ収集システム

[‡]楽天では，2012 年 11 月現在，日本国内にて 72 種類のサービスを提供している[4]．また，データフォーマット統一化がされていないため，サービスによってデータフォーマットが異なる場合がある．

[§] 2012 年 11 月現在，12 カ国で EC サイトの運営を行っている[5]．

[**] 詳細については，4.2.1 にて示す．

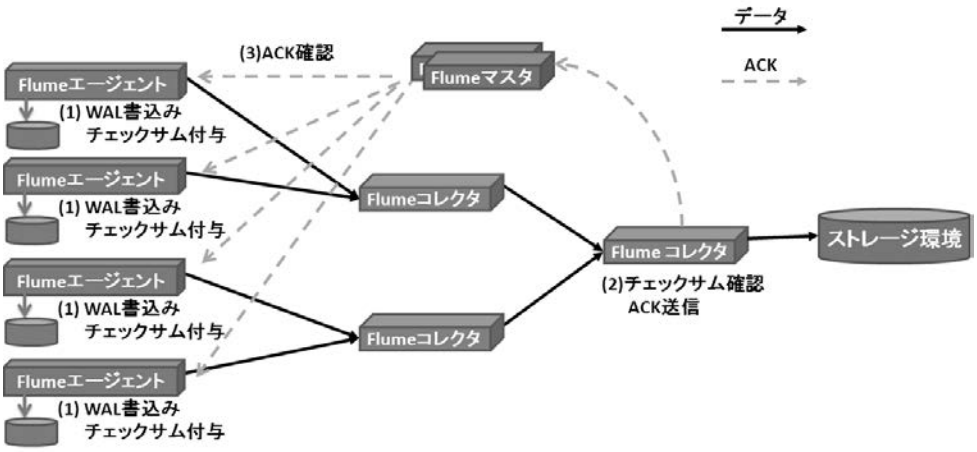

図 5 Apache Flume の E2E モード

によって集約されたデータを，保存・解析処理する役割を担っており，本プラットフォームのもっとも重要なシステムである．本システム上に存在するデータは，あらかじめ定義されたプロセスを通して処理されていき，処理結果はプレゼンテーションレイヤに向けて，継続的に送信されていく．ストレージ・分散処理システムは，多様なデータ規模・解析用途に対応しなくてはならないため，図 4 に示す通り，Hadoop/Hive クラスタ，Cassandra クラスタ，インメモリデータベース環境等の複数のコンポーネントで構成されている．これにより，データ規模・解析用途の違いによって，適用させるコンポーネントを変更することにより，あらゆる規模，用途の処理に対しても高速に処理する基盤を提供している††．

4.2 課題点と工夫点

第 4 章冒頭で示した楽天の環境において，GEAP を構築するに当たり，様々な課題が発生した．4.2 節ではこれらの課題点とその工夫点について，言及を行う．

4.2.1 データ収集の信頼性に関する課題と工夫

本プラットフォームを構築するに当たりまず考慮した点は，収集対象である多数のホストから，いかにしてリアルタイムに，効率的に，そして確実にデータを収集できるシステムを構築するのかという点である．この点を考慮し，我々はログ収集システムにおいて，End-to-End (E2E)モードを持つ Apache Flume[8]を採用した．

Apache Flume の E2E モードとは，データ到達性を保証するためのメカニズムであり，図 5 に示す通り，以下のような動作を行う．

(1) Flume エージェントは，先行書込みログ (WAL)を書き出し，チェックサムを付与したうえで，データを次のノードに送る．

(2) いくつかのノードを中継し，最終的な Flume コレクタに到達した際，Flume コレクタは，チェックサムを検証し，HDFS 等のストレージ環境にデータを蓄積する．この際，Flume コレクタは，ACK を生成し，全ての Flume ノードを管理する Flume マスタに生成した ACK を送信する．

(3) Flume エージェントは，定期的に Flume マスタと交信を行い，当該エージェントに関する ACK を取得する．

(4) Flume エージェントが指定時間以内に ACK を取得できない場合，当該エージェントは，データが目的地まで到達しなかったとみなし，データを再送信する．

上述の通り，E2E モードは，Flume マスタを介した ACK メッセージにより，データが目的地（最終到達ノード）に到達したか否かをチェックする．このメカニズムの利点は，ACK メッセージが，最終到達ノードから一律 2 ホップでオリジナルデータ送信ノードに届くということである．しかし，この E2E モードをそのまま我々の環境に適用させようとすると，以下のような問題点が発生する．

- Flume ノード集合が，異なったネットワーク環境に設置され，Flume マスタ，Flume エージェント間の通信ができない場合，E2E システムは機能しない．この問題は，たとえば海外拠点に設置された Flume ノードが，Flume マスタを参照しようとした際に発生するため，非常に重要な問題である．
- Flume マスタがクラッシュした場合，Flume エージェントは ACK を取得できなくなる．この場合，当該

†† 詳細は，4.2.3 にて示す．

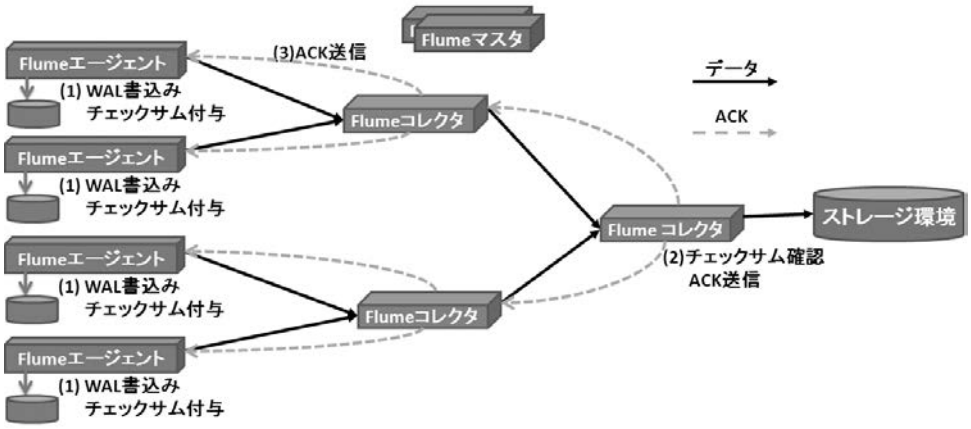

図 6　楽天の MA Flume

Flume エージェントはデータを再送信し続けることになる．Flume マスタを冗長化させることで，ある程度，この問題を解決することができるが，例えば Flume マスタのレプリカを作成している途中に Flume マスタがクラッシュした場合等，ACK が消失してしまう可能性が存在する．

- 多数の Flume エージェントが ACK を取得しようとした場合, Flume マスタがシステム全体のボトルネックとなる．Flume エージェントは，ACK の受信完了するまで，データを送ることができない．

これらの問題を解決するため，我々は，“Masterless ACK(MA) Flume[‡‡]”を構築した（図 6）．MA Flume では，図 6 で示す通り，データが送信されたルートとは逆のルートで ACK が送信される．つまり，

- 対象データがストレージ環境に記録された後，Flume コレクタはデータ送信元 Flume ノードに対し，ACK を送信する．
- ACK を受信した Flume ノードは，自身が ACK の宛先でない限り，データ送信元 Flume ノードに対し，ACK を送信する．
- オリジナルデータ送信元の Flume ノードは，データを送信した Flume ノードから ACK が返ってくるまで待機する．

このスキームを実現するために，MA Flume では，収集対象データに対し，通過してきた Flume ノードパスを付与していく．

我々は, 上述のメカニズムを持つ MA Flume を実装し，実験を行った．その結果，Flume マスターノードがクラッシュし，復旧しない状況下においても，このスキームを用いることでデータの送受信が継続できることを確認した．また，任意のノード間の通信量が多くなることが原因でパフォーマンスの低下が予測されるが, ACK の送受信部分の最適化を施すことで，オリジナルの Flume のパフォーマンスと遜色ない程度まで，MA Flume のパフォーマンスを向上することに成功している．現在，MA Flume モデルは GEAP に正式に採用され, 稼動している．

4.2.2 Hadoop の性能向上に関する課題と工夫

Hadoop クラスタにおいてリアルタイムに近い環境を実現するには，Flume で集約したデータを，可能な限り早く Hadoop クラスタに書き込む必要がある．これを実現するためには，データチャンクのサイズを小さくする必要があるが，これは Hadoop のパフォーマンス低下を併発することになる． なぜなら，Hadoop では，パフォーマンス向上のために，データチャンクのサイズを大きめに設定することが欠かせないからである. この問題は，楽天に限らず，Twitter 社等様々な企業が抱えている問題である[9].

この問題を解決するために，Lee らは，データチャンクのサイズが異なる複数の Hadoop クラスタを構築することで解決を図っている[9]．具体的には，データが到着した段階で，データチャンクサイズが小さい Hadoop クラスタに書き込みを行い，すぐに解析ができるような状況を作る．その後，複数のデータを集約し，データチャンクサイズが大きく設定されている Hadoop クラスタに配置しなおすことで，より高パフォーマンスな環境にデータを移動させる．このアプローチを繰り返していきながら，最終的に，大きなデータチャンクサイズが設定されている中央の Hadoop クラスタに集約することで，リアルタイム性を確保しつつ，パフォーマンスの向上を図っている[9].

これに対し，我々は，HDFS の“append”関数を用いた

[‡‡]米国特許出願第 13/675, 352（米国特許出願日 2012 年 11 月 13 日），特願 2012-017514 号

アプローチ方法を実験している．つまり，複数のデータを集約したうえで，同一 Hadoop クラスタ内に格納するアプローチである．この同一クラスタ内にデータを集約するアプローチをとることで，データの集約に伴うデータの移動コストを低くすることができ，文献 9）で提案されている手法に比べ，ネットワーク負荷の軽減が実現できることが見込まれている．

4.2.3 多種多様なデータソースに関する課題と工夫

第 4 章冒頭で示した通り，GEAP が対象とするデータは，様々なサービスから生成される様々なフォーマットのデータである．また，今後も日本国内外で提供するサービス数が増加していくことが予測されるため，フォーマットが異なる様々なデータを,いかにして柔軟に扱い,単一のストレージ基盤に蓄積していくかが，非常に重要なポイントである．

そこで我々は，様々なデータに共通して存在する共通要素を抽出し，共通要素を利用して任意のデータを統一的な形式で表現できるようにすることで，この問題を解決している．具体的には，下記のような手法により，任意のデータを統一フォーマットに変換している．

- データを共通要素に関するデータと，それ以外のデータに分割する
- 共通要素以外のデータを，json 形式に変換する
- 任意のデータは，共通要素とそれに紐づく json 形式のフォーマットに変換する

共通要素は，Hadoop にデータをバケットする用途，Hive パーテションを作成する用途としても活用されており，分散処理・ストレージ基盤部のパフォーマンス向上にも寄与している．

4.2.4 多様な解析手法・処理要求に関する課題と工夫

本プラットフォームは多種多様なデータを集約・蓄積しているため，様々な解析手法のニーズが存在し，解析スピードに関する要求も様々である．たとえば，マネージメントに属する社員は,長期間のデータを対象とした,比較的長いスパンでの解析をする必要がある．このような解析の要求に対応するため，本プラットフォームは，大規模データに対して複雑なクエリを受け付けられるようにしなくてはならない．また，データ解析・マーケティング部門に属する社員は，インタラクティブな解析を実施する必要がある．したがって，ある過去数日・数週間のデータを対象にしたクエリを比較的短時間で返さなくてはならない．さらに，オペレーションに近い立場の社員は，リアルタイムに近い時間差で，データを集約した結果が必要である．

このような多種多様な解析要件を満たすために，我々は，Flume 上で流れるデータストリームの宛先を仕分けできるようにするモジュールを開発した．このモジュールを活用することで，あるストリームは，長期間のアーカイブが蓄積されている Hadoop クラスタに集約されることで長期間のデータに対する解析に利用され，あるストリームは，比較的短期間のデータを対象としたインタラクティブな解析を行うために Cassandra クラスタに集約され，さらに別のストリームは，リアルタイムに近い環境下で，ダッシュボードライクな GUI インタフェースを提供するために，インメモリデータベース環境に集約されるといったことが可能となり，様々な解析要件に対して応えうるプラットフォームになる．

4.3 GEAP の活用事例

ログ解析を通したサービスの最適化を実現するため，様々なサービスが GEAP を利用することを検討している．また，サービスへの適用の他，マーケティング部門，研究機関のメンバも，モデルや解析手法を構築するために GEAP に蓄積されているデータを参照しながら解析を行っている． 本節では，GEAP の利用方法の興味深い事例として，クラウド環境においての GEAP の活用事例について述べる．

4.3.1 クラウド環境のための Logging as a Service

楽天グループには，約 2,000 人のエンジニアがおり，多数のサービスのアプリケーションを開発し，運用を行っている．そのため，楽天グループでは，アプリケーションの開発・テスト・リリース環境プラットフォームをエンジニアに迅速に提供するために，自社内でのクラウド環境，Platform as a Service(PaaS)環境の提供を行っている．この結果，クラウド環境にて動作するアプリケーションが急増しているが，ログ解析のためには，クラウド環境上で生成されるアプリケーションログを，ログ解析環境に集約し，解析するモジュールをサービス毎に開発せねばならず，開発コストの増加につながる．

この問題を解決するため，楽天の PassS 環境は，クラウド上で発生する様々なアプリケーションログ・及びシステムログの収集をグローバルデータ解析プラットフォームを利用して行うようにしている．本プラットフォームに API を介してアクセスすることにより，エンジニアはログ収集，解析モジュールを独自に構築するコストを削減できると共に，リアルタイムに，かつ迅速に，構築したアプリケーションの状況を把握することができる．

また，アプリケーションログだけではなく，システムログも同様に集約することで，クラウド環境を提供する側にも PasS 環境全体がどのようなステータスであるのかを簡単に確認することができ，PasS 環境の向上に有効な指標となっている．

5. おわりに

本論文では，多種多様なサービスから生成される大規模データを統一的に蓄積・活用していくために開発・運用されている 2 つのプラットフォームについて，紹介を行った．

楽天スーパーDB は，7,500 万人を超える楽天の会員属性情報を一元的に管理するプラットフォームであり，様々なサービスがこの DB プラットフォームに対してアクセスを行っている．また GEAP は，多数のサービスにて運用されている膨大な数のサーバから，データを収集・蓄積・解析するプラットフォームであり，現在社内の PaaS 環境や，大規模データ解析等に活用されている．

GEAP は，今後多くのサービスによって適用されていくことを予定しており，様々な種類のデータがリアルタイムで GEAP に集約されていくことになる．これらのデータを解析し，サービスに適用させていくことで，楽天のさまざまなサービスを，ユーザにとって使いやすいものにしていくよう努めていく所存である．

謝辞 本論文を執筆するに当たり，楽天株式会社 DU ビッグデータ部の Mr. Marthinussen Terje，杉野順清氏，GEAP チーム一同，また楽天株式会社 DU アーキテクチャコミッティ運営室の Mr. Waldemar Quevedo に，多大なる助言とサポートを頂きました．ここに感謝の意を表します．

参考文献

1) 【楽天市場】Shopping is Entertainment! ： インターネット最大級の通信販売，通販オンラインショッピングコミュニティ: http://www.rakuten.co.jp
2) 楽天トラベル:宿・ホテル予約 国内旅行・海外旅行 総合旅行サイト: http://travel.rakuten.co.jp
3) 【楽天市場】口コミ・レビューで人気商品を探そう！ みんなのレビュー・口コミ: http://review.rakuten.co.jp/
4) 楽天グループサービス一覧: http://www.rakuten.co.jp/sitemap/
5) All Rakuten Services: http://global.rakuten.com/group/
6) Apache Hadoop: http://hadoop.apache.org/
7) The Apache Cassandra Project : http://cassandra.apache.org/
8) Apache Flume : http://flume.apache.org/
9) G. Lee, J. Lin, C. Liu, A. Lorek and D. V. Ryaboy: The Unified Logging Infrastructure for Data Analytics at Twitter. PVLDB 5(12): 1771-1780 (2012)
10) D. Borthakur, J. Gray, J. S. Sarma, K. Muthukkaruppan, N. Spiegelberg, H. Kuang, K. Ranganathan, D. Molkov, A. Menon, S. Rash, R. Schmidt, A. S. Aiyer: Apache hadoop goes realtime at Facebook. SIGMOD Conference 2011: 1071-1080

平手 勇宇（正会員）
E-mail: yu.hirate@mail.rakuten.com
2008 年早稲田大学大学院理工学研究科博士課程修了．博士（工学）．早稲田大学メディアネットワークセンター助手を経て，2009 年より楽天株式会社楽天技術研究所．現在に至る．ACM, IPSJ, DBSJ 各会員

王 永坤（非会員）
E-mail: yongkun.wang@mail.rakuten.com
2011 年東京大学大学院情報理工学系研究科博士課程修了．博士（情報理工学）．現在，楽天株式会社ビッグデータ部に所属．ビッグデータ処理のためのデータベースシステム，ファイルシステム，分散処理システムに従事．ACM SIGMOD 会員．

投稿受付：2012 年 09 月 19 日
採録決定：2012 年 11 月 26 日
編集担当：中野美由紀（東京大学）