

AWS Cloud Storage & DB Day 2014

2014/9/9

# 日経電子版におけるAWSと その周辺サービスの活用事例

日本経済新聞社

デジタル編成局

白銀 傑

(システム・サービス開発担当)



# Agenda

- 日経とウェブサービス
  - 日経電子版のご紹介
- 電子版モバイルでAWSを採用
  - AWSを採用した理由
  - システム構成
  - AWSを使ってみて
- サービス開始後の施策
- 今後の展開



# 日経とウェブサービス

- 1996年に日経ネットを開始して以来、自社環境を中心に配信
- マーケットやマネーの情報配信と機能提供など  
ニュースサイト以外も
- サービスインフラとして、データセンターを保有
  - 主な配信設備はオンプレで構築しサービス(これまで)
  - それなりの設備とそれなりのノウハウが社内にはある
- 各種アプリケーションは自社でも開発
- 2010年に日経電子版創刊

**NIKKEI NET**  
日本経済新聞社

殖やせる人の最強サイト  
**マネー&マーケット**

**日本経済新聞** 電子版



# 日経電子版のご紹介



300本の記事

紙面

+ 600本 =



900本の記事

電子版

日本経済新聞 電子版

電子版オリジナル記事

日経BP社  
Nikkei Business Publications, Inc.

NQN 日経QUICKニュース社

TV TOKYO

FT.com  
FINANCIAL TIMES

# 日経電子版のご紹介



携帯



Android & iPhone



Android & iPad iPhone



モバイル



電子版  
マーケット



Windows8



PC



新聞

# メディアとして

- 必ず最新の情報がいち早く届く
- いつでもどこでも、持ち歩いて読める

- マルチデバイスに対応
- コンテンツの更新感がある
- 応答が速い
- 耐障害性が高い



# 日経のサービス

- 紙面と電子版(PC向け)
- スマホとタブレット向けアプリ
  - 朝刊・夕刊＋最新情報を提供
  - 紙面そのまま、めくるように
- モバイル向けのサービスを強化



# 日経電子版モバイル(2013/5~) をAWSで

- PC以外で情報収取するライフスタイルに
  - 会社ではPCで 朝と夕方にはモバイル端末で
- モバイル端末のブラウザ向けにレイアウトした画面を配信
- 通信環境によらず高速なパフォーマンス
- AWS上に構築



# AWSを採用した理由

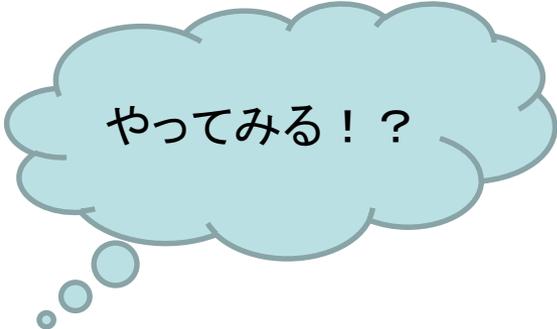
- ユーザーがシステム設計できる
- リージョンが選択できる
- サービスの成長に応じて、スケールアウト/イン・スケールアップ/ダウンが可能
- 構築及び運用管理コストのカット（人件費込み）
  - サービス開発に人的リソースを投入したい
  - インスタンスの起動が簡単で試しやすい



当然のメリット  
だけでなく・・・

# AWSを採用した理由

- 新システム、新サービスを作るなら技術者として・・・
  - 電子版としてモバイル端末向けプラットフォームを新たに構築したい
  - 単純に新しい技術を使ってみたい
  - 将来に向けて、新しいことにチャレンジしたい
  - 社内リソースで構築したい
    - 199x年：自分たちで作るしかなかった
- 少ない技術者、少ないコストで
- ブランドはしっかり守って
- クラウドを使えば安く簡単にできる??



やってみる!?

# リスクに対する対応

## ① AWSと自社DCとの使い分け

- どのデータをクラウドに保持するかはポリシーを持って定義

## ② 自社DCにある監視システムとの併用

- 既存のシステムは自社DCで集中監視している

## ③ AZやリージョンを分けて稼働させる

- 24H365D ニュースを報道

# 電子版モバイルのシステム構成

Cache  www.nikkei.com

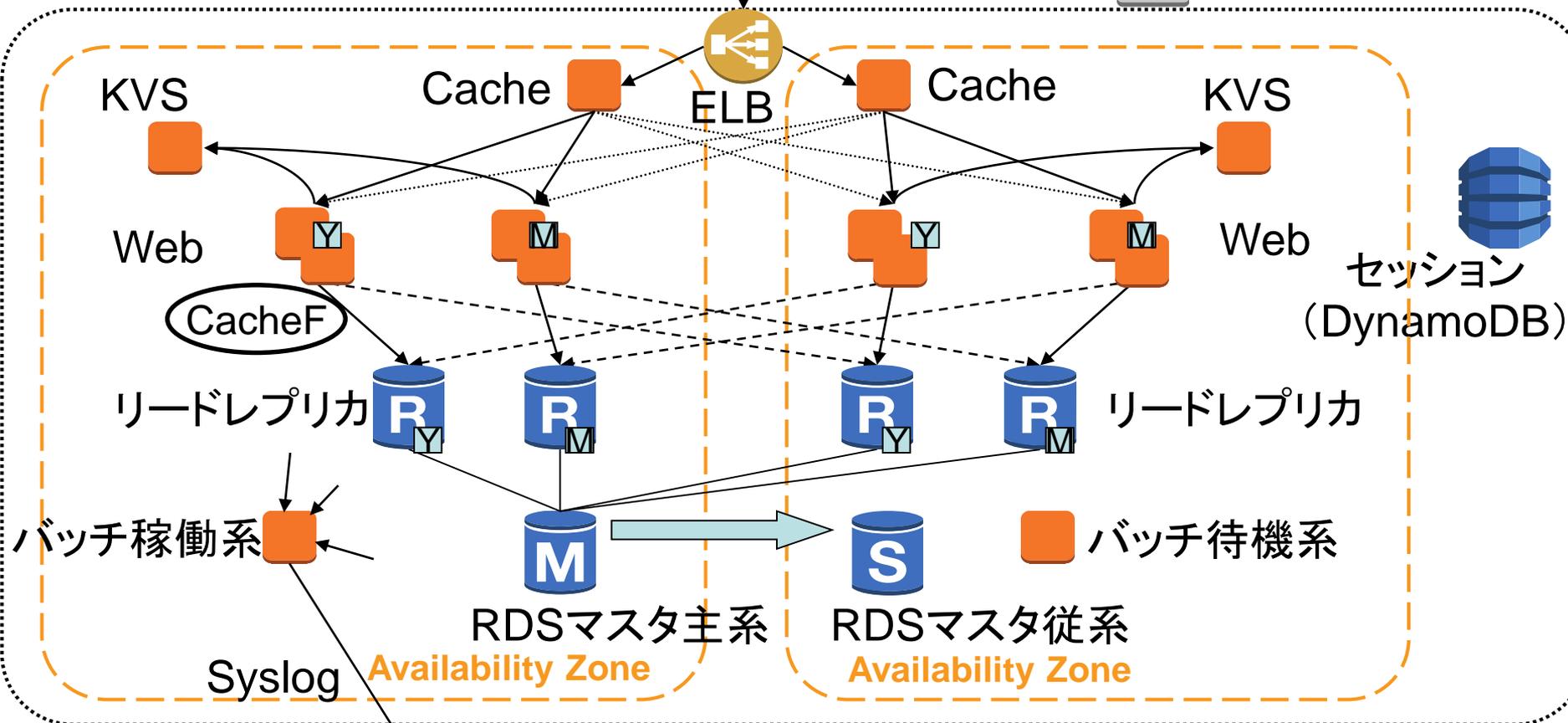
 EC2

 有料会員向け  
 無料会員向け

AWS

redirect  mw.nikkei.com

東京リージョン



自社DC

監視

Cache、Webサーバーから

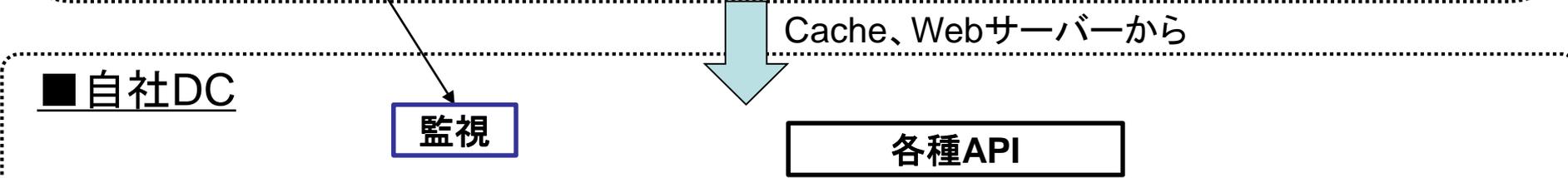
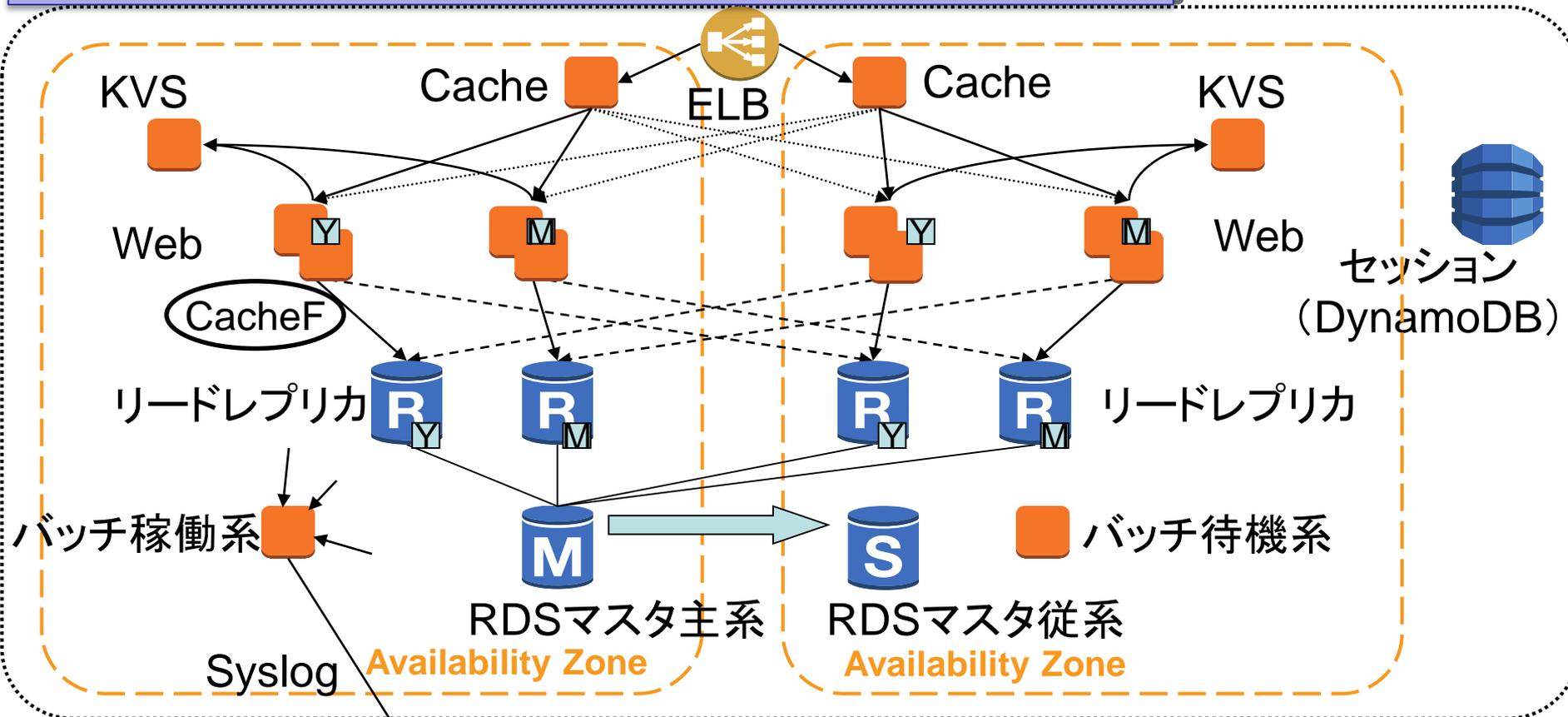
各種API



# キャッシュサーバー

- Cookieスイッチング
- ユーザー権限によるキャッシュオブジェクトの切り分け
- Webサーバーへのヘルスチェック

- Y 有料会員向け
- M 無料会員向け



# 監視：CloudWatch/SNSを利用し自社監視装置で

■ バッチサーバーにSNSがHTTP通知、自社DCにsyslog転送

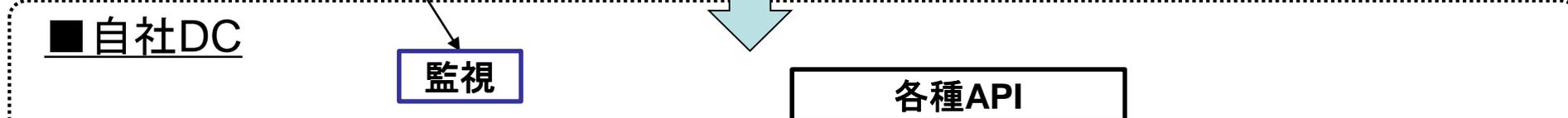
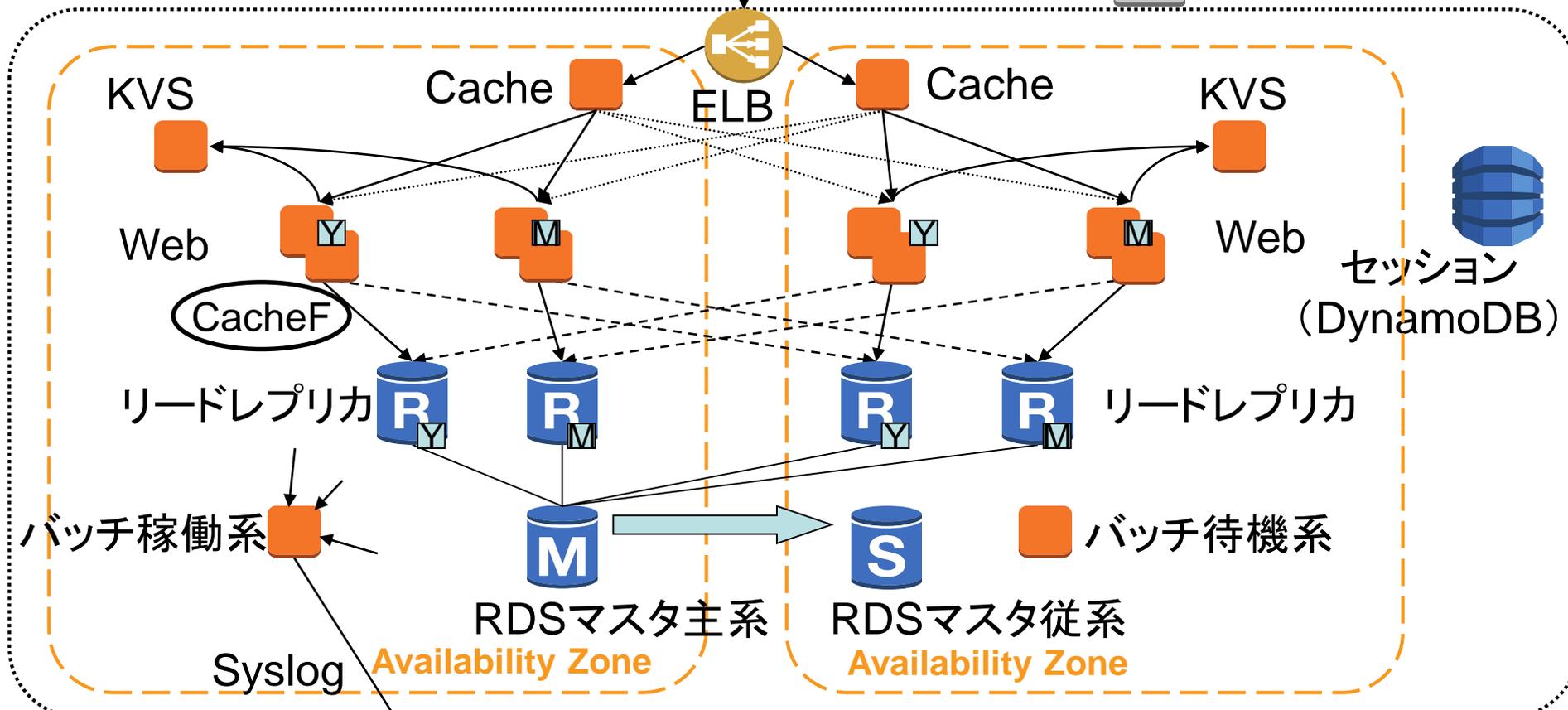
有料会員向け

■ 無料会員向け

AWS

redirect → mw.nikkei.com

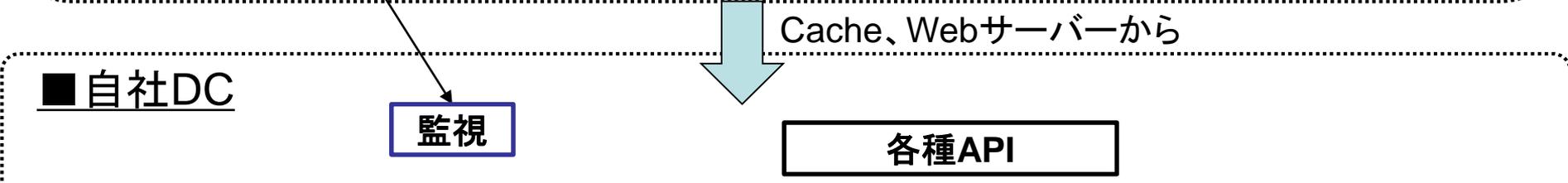
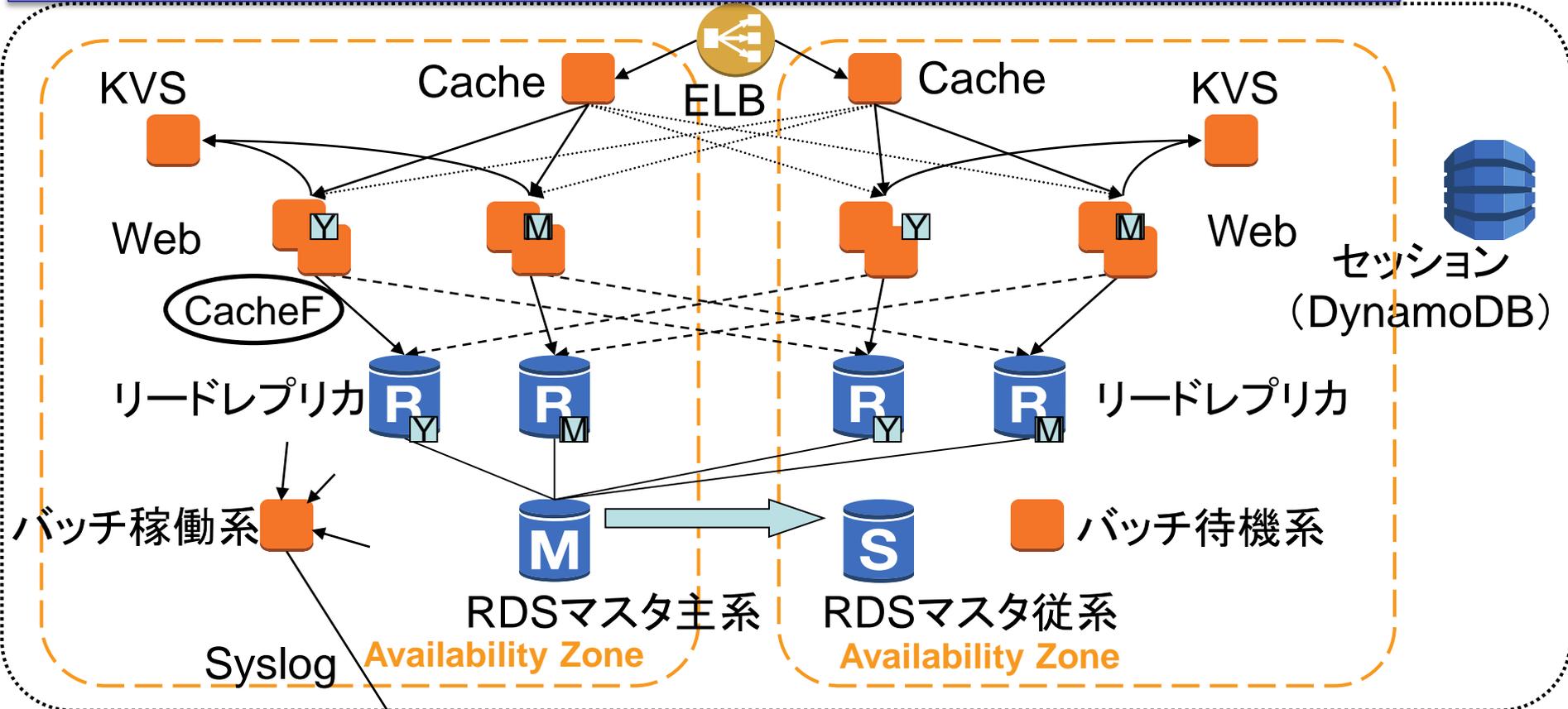
■ 東京リージョン



# コンテンツ

- バッチサーバーが自社DCのAPIを使って定期的に更新記事を取得  
+ WebサーバーがオンデマンドでAPIをキック
- キャッシュサーバーが直接APIをキックすることも

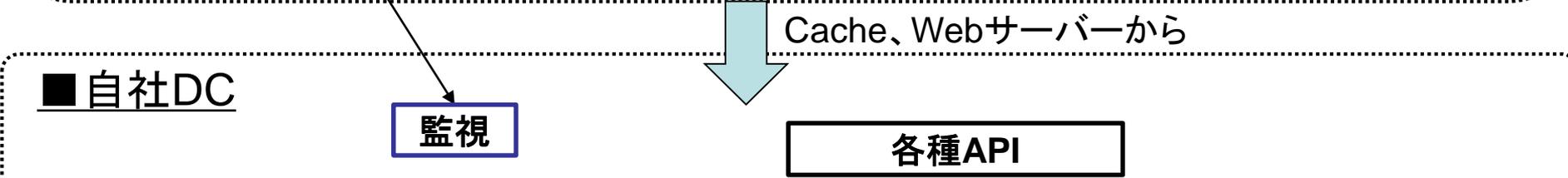
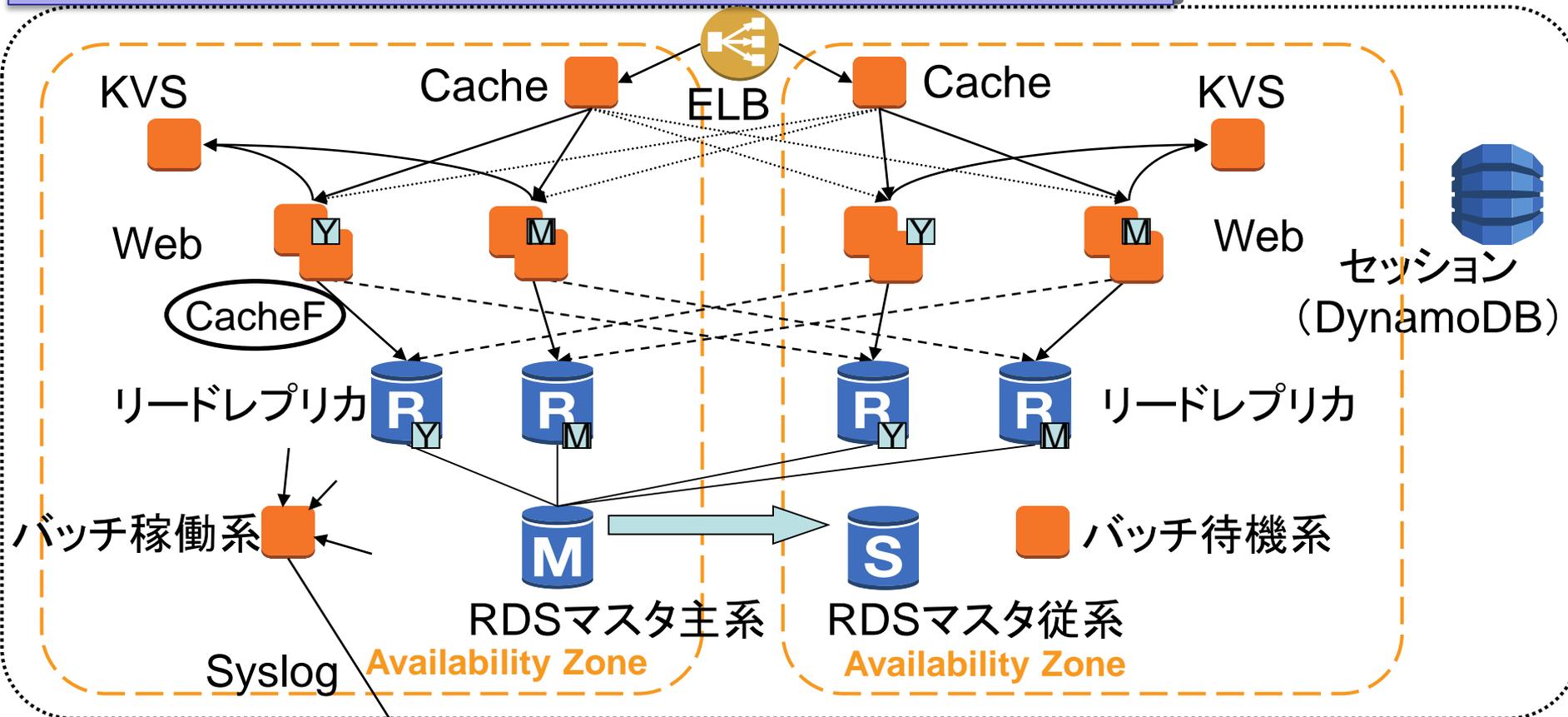
社員向け  
社員向け



# RDS: 記事データベース

- Multi AZ構成
- マスター主系従系 + リードレプリカ
- マスターへの書き込みはバッチサーバーからのみ

- ☑ 有料会員向け
- ☒ 無料会員向け



Cache  www.nikkei.com

 EC2

 有料会員向け  
 無料会員向け

redirect → mw.nikkei.com

AWS

東京リージョン

KVS

Cache 

ELB

Cache 

KVS

### クライアント技術

- API + JSON
- Web Storage
- backbone js
- オープンソースの有効活用

  
東京リージョン (Amazon RDS)

バッチ稼働系 

 バッチ待機系

RDSマスタ主系

RDSマスタ従系

Syslog **Availability Zone**

**Availability Zone**

Cache、Webサーバーから

自社DC

監視

各種API

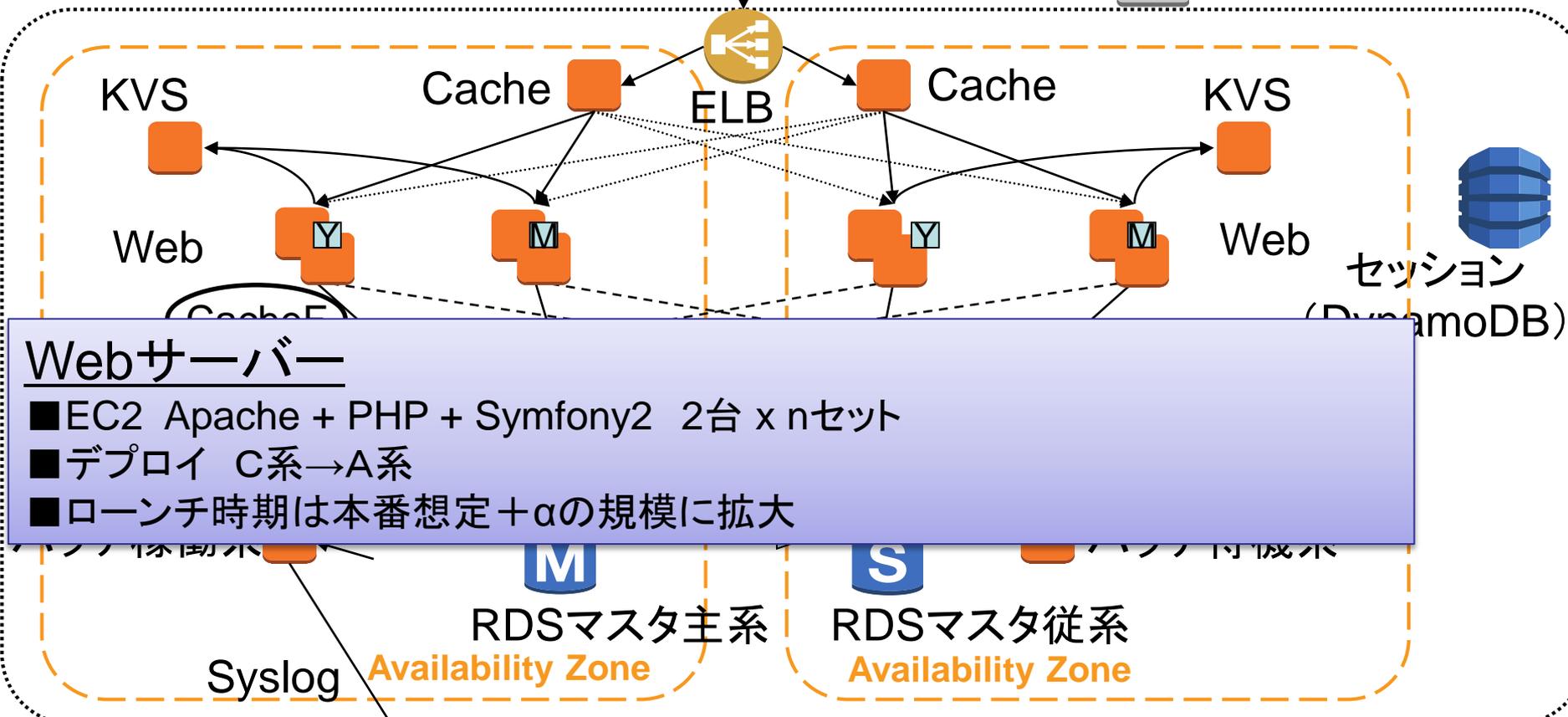
Cache  www.nikkei.com

 EC2

 有料会員向け  
 無料会員向け

redirect  mw.nikkei.com

■東京リージョン



### Webサーバー

- EC2 Apache + PHP + Symfony2 2台 x nセット
- デプロイ C系→A系
- ローンチ時期は本番想定+αの規模に拡大

■自社DC

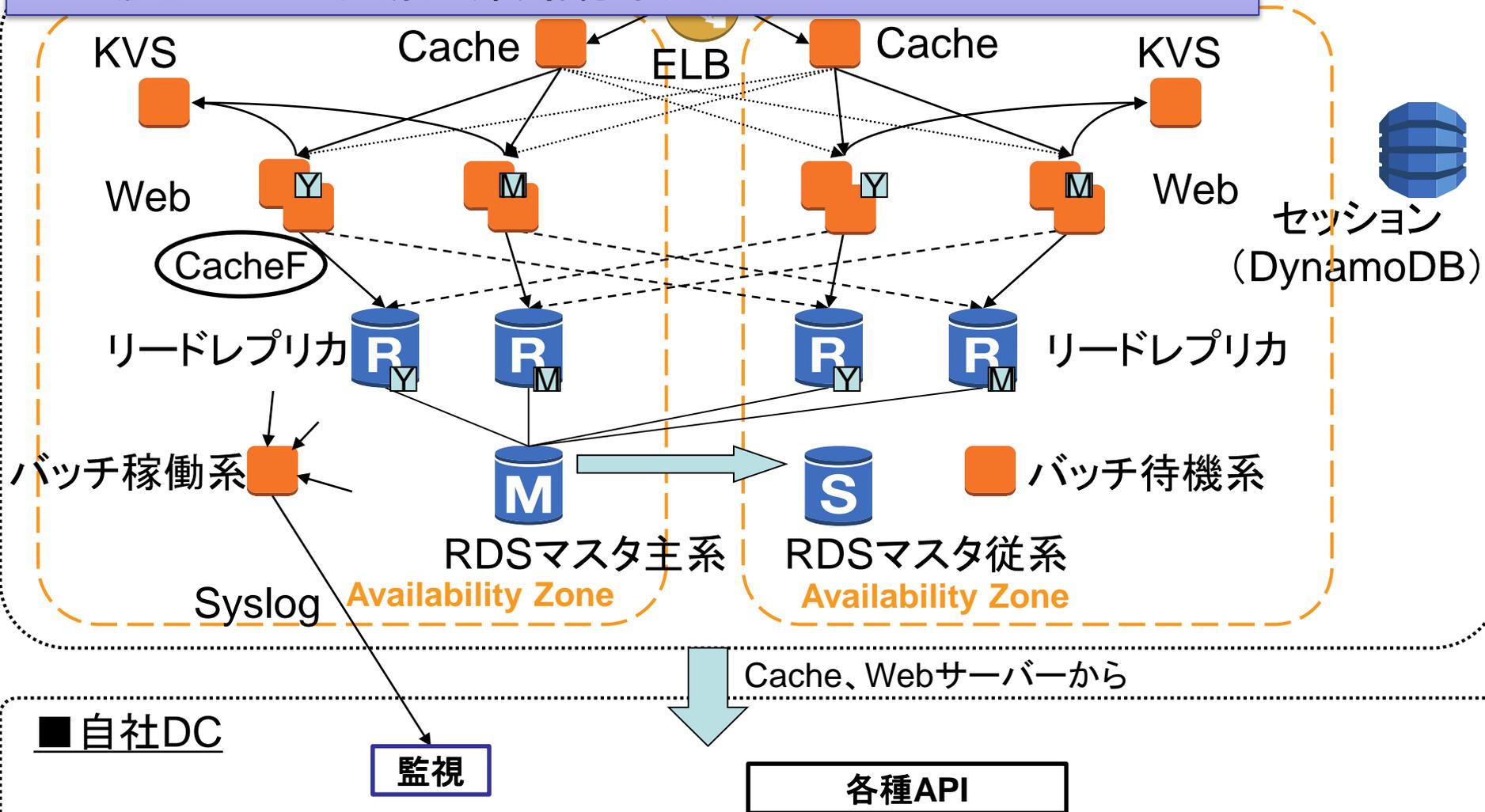
監視

各種API

# 認証、セッション管理

- セッションテーブルをDynamoDBに作成  
頻繁なデプロイをサービスを継続したまま実施できるように
- 日経IDAPI (OpenID Connect + シングルサインオン) を利用
- PC版とセッション同期 (登録、解約時など)

有料会員向け  
無料会員向け

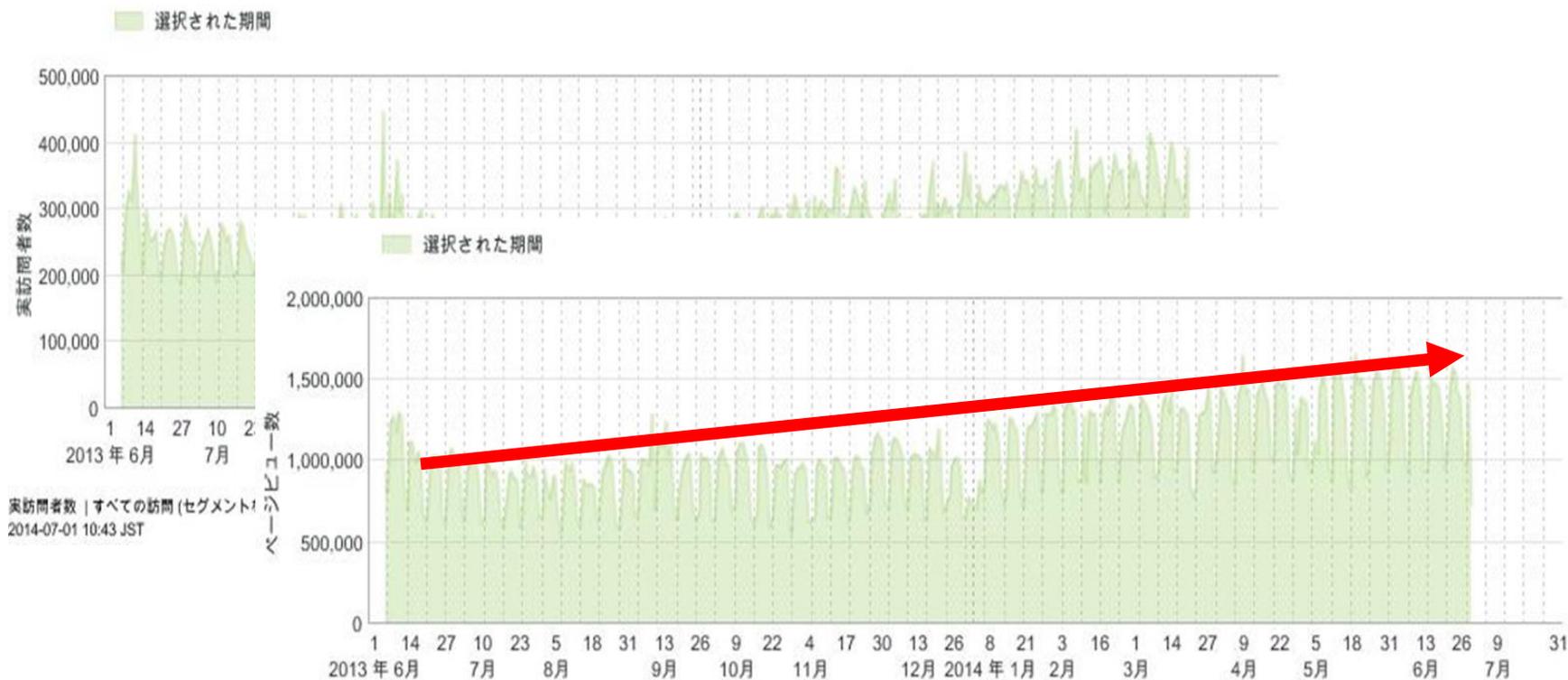


# DB構成

- RDSに記事データを格納しウェブサーバから参照
  - 高速に記事を返すため、ウェブサーバは以下の順で参照
    - ローカルキャッシュ(APC) → KVS(Memcached) → RDS
  - RDSはmulti-AZ構成
  - バッチからRDSのマスターを更新
- セッション管理にDynamoDBを採用
  - RDSへの書き込みがボトルネックになりやすいことを回避
  - 今回の構成では、ウェブからRDSへの書き込みはしない

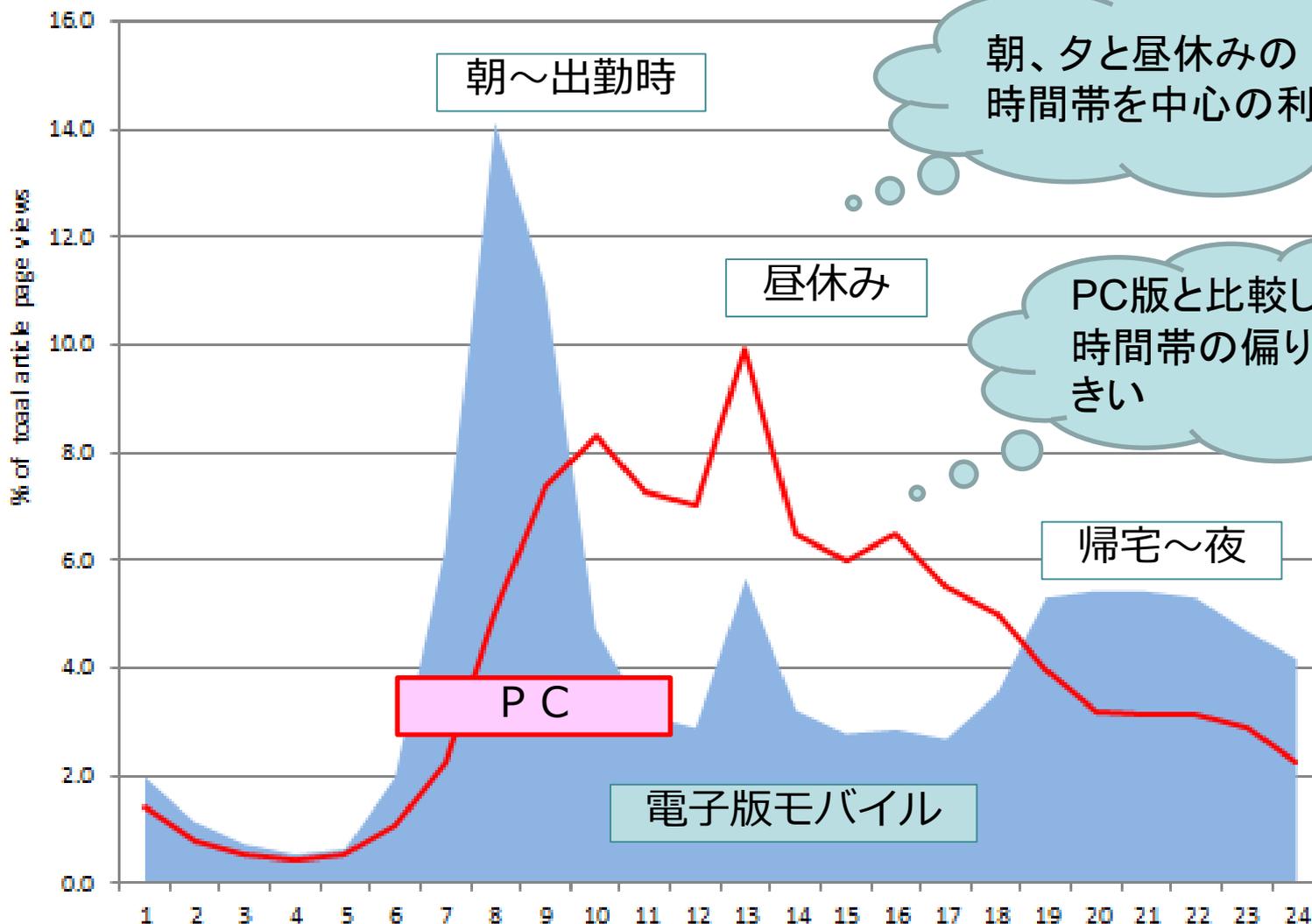
# サービスを開始してから

- PV、訪問者数の順調な増加



ページビュー数 | すべての訪問 (セグメントなし) | 2013-06-01 - 2014-07-31 | Adobe Analytics で生成されたグラフ 日時: 2014-07-01 10:42 JST

# 一日のPV遷移



朝、夕と昼休みの時間帯を中心の利用

PC版と比較し利用時間帯の偏りが大きい

朝～出勤時

昼休み

帰宅～夜

PC

電子版モバイル

# サービスを開始してから

- 大きなトラブルはほぼなし
- 監視機能を使いこなせていない
- クラウドのメリットをまだまだ活用できていない
  - アクセス状況にあわせたスケールアウト、イン
  - DynamoのThroughput設定
- ELB/EC2の障害(1時間強)
- たまに、自社DCとAWS間の通信が切れる
- DynamoDBの設定に悩まされる

徐々に進化中

こんなこともまれに

# DynamoDBを使ってみて

- ログイン時の同時接続数のチェック、セッション切断にscanを利用
- scanはテーブル内の全レコードを走査
  - 重い、Read Capacity Unitを多く消費する=コスト増
- Provisioned Read Capacityを超えてなくても、APIでProvisionedThroughputExceededException
  - 特定のパーティションでThroughputオーバー
  - scanを極力使わないよう、セッションテーブルを変更
  - キーやインデックスの設計が重要(キーに設定した項目以外での検索がスキャンになりコストがかかる)



最新の状態が分かりづらい  
reloadで上がったり下がったり

Amazon DynamoDB Explore Table: sessions

id	data	expires	nikkei_member_no
"NK_DSW_S_ppgtac5tkg7hn0ecu349n4f6d3"	"pGHFQXUba0JSJMKF5qSjuSWqWu00Dcdi"	1409735178	"a5b66cb8dc72aa6ae2225778"
"NK_DSW_S_550p0lqro47ace87onmp9i59p:"	"pGHFQXUba0JSJMKF5qSjuSWqWu00Dcdi"	1409736303	"5cbda5c564154a8eed36e9"
"NK_DSW_S_ob3hnrdoaaavs5kns5jg6a0447"	"pGHFQXUba0JSJMKF5qSjuSWqWu00Dcdi"	1409738339	"86f5ab304c1c32a9b70953dcfcf1210edboce"
"NK_DSW_S_lr43eatk82qjisd9ofgk7oc0"	"pGHFQXUba0JSJMKF5qSjuSWqWu00Dcdi"	1409733567	"6e9269309062f8c97dc4a804c8942484c40:"
"NK_DSW_S_qcn1tqunp9tuu5fdclllkv4590"	"pGHFQXUba0JSJMKF5qSjuSWqWu00Dcdi"		
"NK_DSW_S_tlkpvn7ef0avrve994svuhb2"	"pGHFQXUba0JSJMKF5qSjuSWqWu00Dcdi"		
"NK_DSW_S_1ea0nsgj3mriemtdrprvriqo4j5"	"pGHFQXUba0JSJMKF5qSjuSWqWu00Dcdi"		
"NK_DSW_S_us12c8ml1363q1uacaa9v3je"	"pGHFQXUba0JSJMKF5qSjuSWqWu00Dcdi"		

プライマリーキーを変えた  
別テーブルを作成  
→テーブル設計が重要

Amazon DynamoDB Explore Table: user\_sessions

nikkei_member_no	insdate	sessids	updatedate
"2cc8e4dc285a3845459287be4e9e630e22b:"	1409567140	"TNK_DSW_S_dj3j55agr5qeve41d11ddr8a3"	1409663020
"eceb7801fece77ce47cd7681059cfe18ef4."	1409661407	"TNK_DSW_S_id8fajuqfdl473a5ce9kr0pm3"	1409661407
"ce198b7eb20d0beb44bcaabe496bcf3631d6f"	1409528000	"TNK_DSW_S_6ila6337rabss0r54m358v3"	1409655513
"2d43a11aa4bc3422098f2a6ba7a05c12c3e3"	1409640636	"TNK_DSW_S_gf467j2l6f5cuojpsaodg41q3q4"	1409720656
"800fa9b5e6e2a24c12a822d0c61a9eb132ad"	1409712203	"TNK_DSW_S_56accu1abnekei839di6jy95"	1409712203
"055bbd27a8a8823584967984d87efcdf5f8c4"	1409720974	"TNK_DSW_S_6jhpacc15da9h212126sah2e6"	1409720974
"407aa1cda1d2643302624ecbae261627101:"	1409658851	"TNK_DSW_S_mtc88g01js995lonldda4693h4"	1409658851
"67945f6379ccf5408b7a647b3a20286e3bae"	1409627598	"TNK_DSW_S_ljo2q5b7896m5i0svrg7ghf3"	1409713761

# サービス開始後の施策

- DynamoDBの利用拡大
- アクセス状況にあわせたリソースの有効活用
  - Throughputの設定
- EMRと組み合わせたデータ抽出
- SQSとS3を活用した新サービス
- 電子版以外にも・・・

# 保存記事情報を自社DC (RDB) から DynamoDBに 2014/4~

## 1. 自分だけのニュースをデザインする「Myニュース」登場！

「Myニュース」では、自分が気になる記事をひとつの画面でチェックできます。

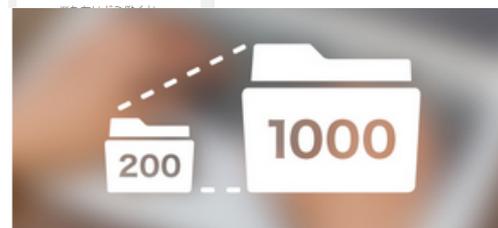
連載・コラム、気になるトピック、保存記事など、個人の設定にあわせてニュースをまとめてお届けします。また、過去に読んだ記事や、日経電子版があなたにおすすめするニュースも確認できます。

もちろん、PC・スマホ・タブレット間で個人設定は連動、その時読みたい環境に最適な形でニュースをチェックできます。



## 2. 記事の保存件数が1,000本に大幅増量！

記事を保存件数を従来の200本から1,000本に大幅増量。朝の忙しい時間は朝刊やWeb刊にさっと目を通して気になる記事を保存、昼休みや仕事帰りにじっくり読み返すといった使い方がしやすくなりました。保存記事に関連した新着記事は「Myニュース」でお知らせします。

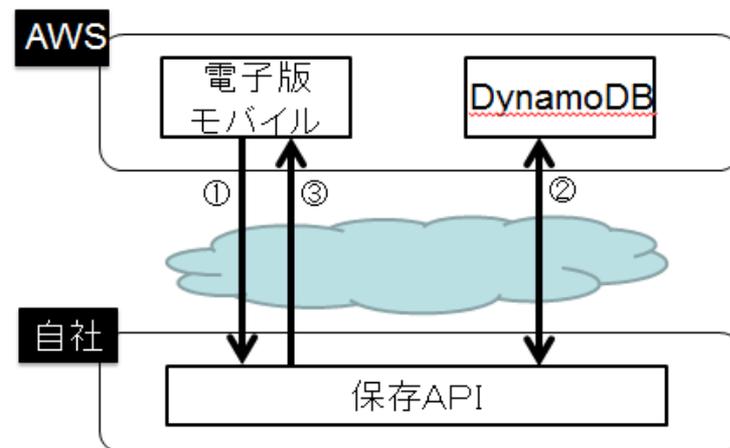


# 保存記事情報を自社DC (RDB) から DynamoDBに 2014/4～

＜要件＞保存記事をデバイス間連携可能にし、保存可能本数も拡大

- 保存記事情報を格納していたDBがスケールアップ構成
- リソースが固定化されたオンプレシステムでは扱いにくい

- DynamoDBはスループットの調整のみで対応できる
- 電子版モバイルで実績を積んだ
- 将来の電子版システムのAWS移行にも繋がる？



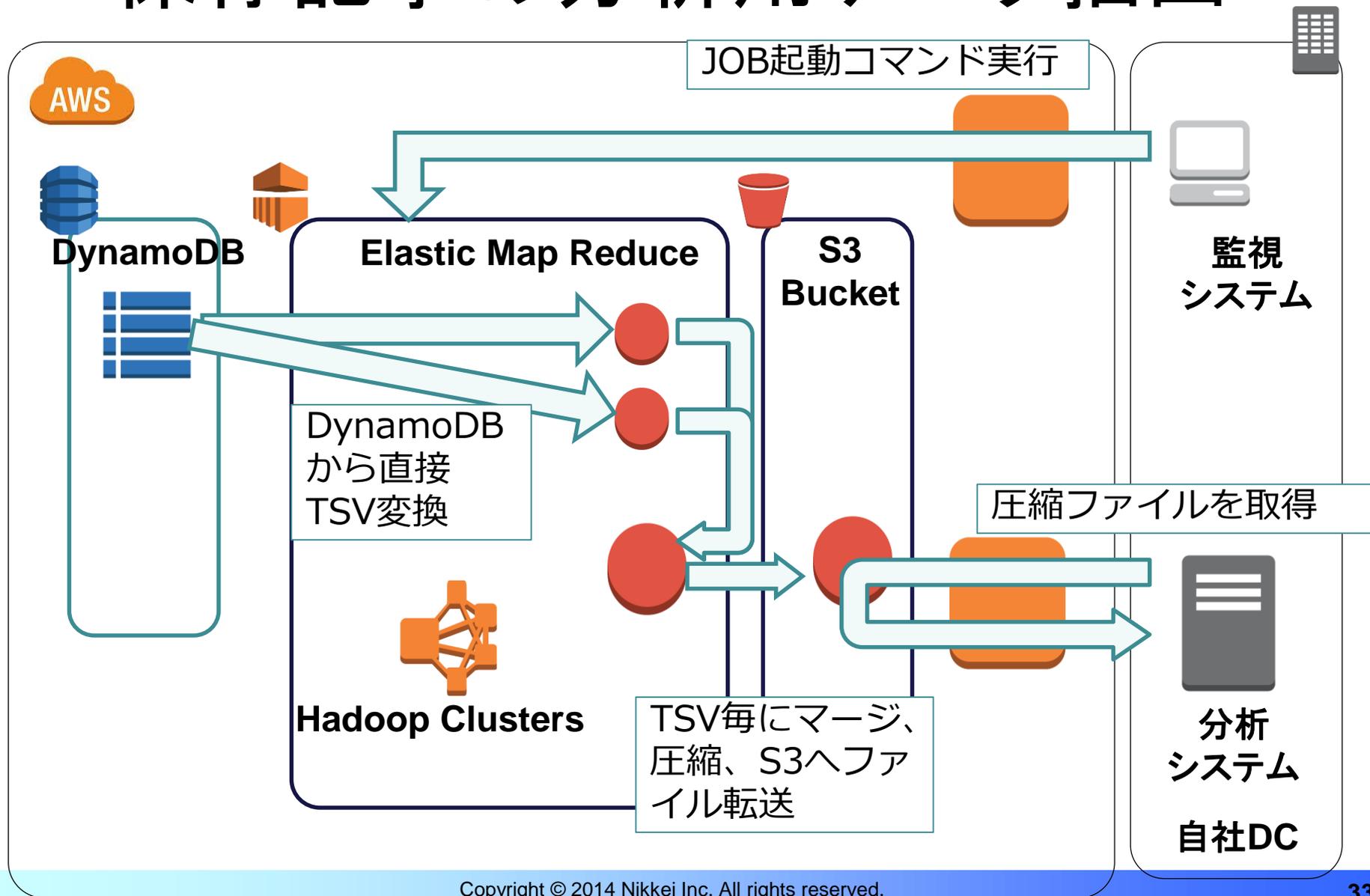
- 移行日に自社RDB内の記事保存情報を一気に移設

# 保存記事の分析用データ抽出

<要件> 電子版読者の記事保存データ(DynamoDBに格納)を自社DCにある分析システムに1日1回定時に提供

- EMRクラスタを立ち上げ、Hiveを使ってクエリを行い、S3に書き出す
  - SQLを書く感覚で必要な操作を実行できる。
  - データが増えても一定時間で処理を終えるためにはEMR内のインスタンス数を増やすだけで対応可能。
  - EMRは処理を実行している間だけの課金で済む(1時間単位)
  - 当時はまだDynamoDBからデータをExportする機能がリリースされていなかった

# 保存記事の分析用データ抽出



# 紙面の切り抜きイメージの表示

## 2014/6~

外形標準課税が広がれば、収益の高い企業は税負担が軽減される。赤字企業にとっては、できるだけ早く廃止されたいとみられる。

一方で、外形標準課税は給与総額が増えるほど課税額が増えるとの指摘もある。現在の仕組みを単純に見直すべきだとの声もある。

政府・与党は法人実効税率を数年で約6%下げ、改革を段階的に進めていく方針だ。



### 法人事業税、赤字企業の負担増

政府・与党検討 外形課税 倍以上に実効税率1.5%下げ

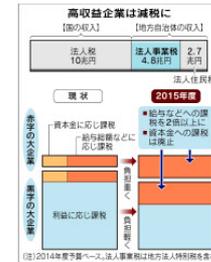
2014/8/29付 | 日本経済新聞 朝刊 [有料会員限定]

小 中 大 保存 印刷 リプリント

政府・与党は赤字の大企業の税負担を増やす検討に入る。給与の総額に基づく課税など、企業が黒字か赤字かに関係なく納める税金の割合を2015年度から2倍以上に増やす。代わりに黒字企業の税金を減らす。安倍政権は黒字企業の利益にどれだけ税金がかかるかを示す**法人実効税率（総合2面きょうのことば）**の引き下げを公約しており、これを1.5%分以上下げる。

対象になるのは地方自治体の収入になる法人事業税だ。約4.8兆円の税収のうち、大企業が納めている約2.8兆円の総額は変えず、負担の内訳を赤字企業は重く、黒字企業は軽くする。企業が利益を増やすと得をする仕組みにする。政府は29日に開く自民党税制調査会の幹部会合に、利益以外にかかる外形標準課税を増やす案を示す。

政府は法人税の実効税率を現在の約35%から数年の間20%台まで下げる方針だ。経済財政諮問会議の民間議員は15年度に2%（約1兆円）以上の引き下げを提言し、外形標準課税の割合を2倍以上に引き上げる案を示す。



小 中 大 保存 印刷 リプリント

外形標準課税とは (2014/8/29付)

# 紙面の切り抜きイメージの配信

＜要件＞紙面の切り抜きイメージを提供  
 イメージのURLは期限付きに

– S3に紙面の切り抜きイメージを格納

(サムネイルと拡大画像)

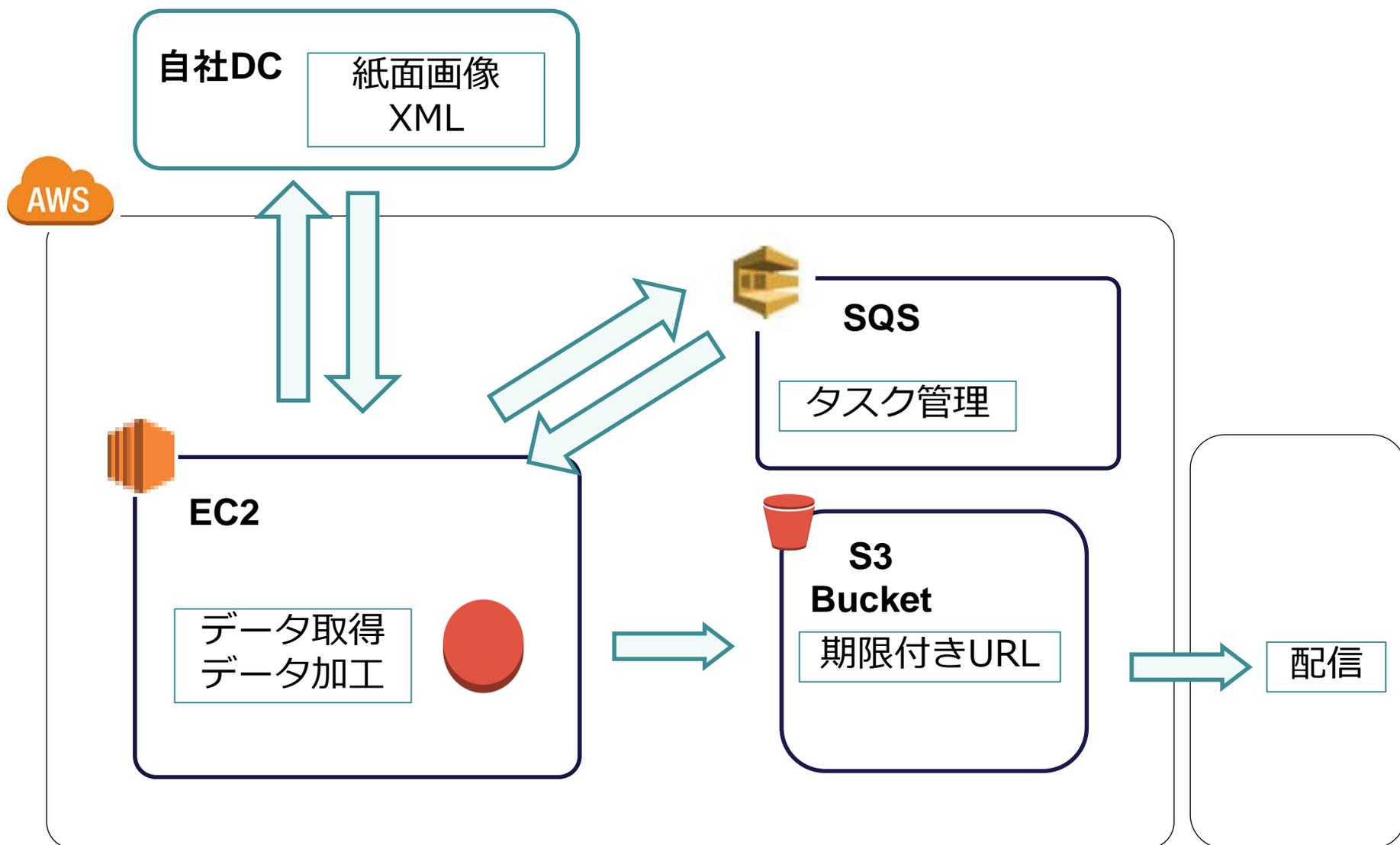
– 有効期限付きのURLを発行

– SQSでタスク管理



2014/8/29付 日本経済新聞 朝刊

# 紙面の切り抜きイメージの配信

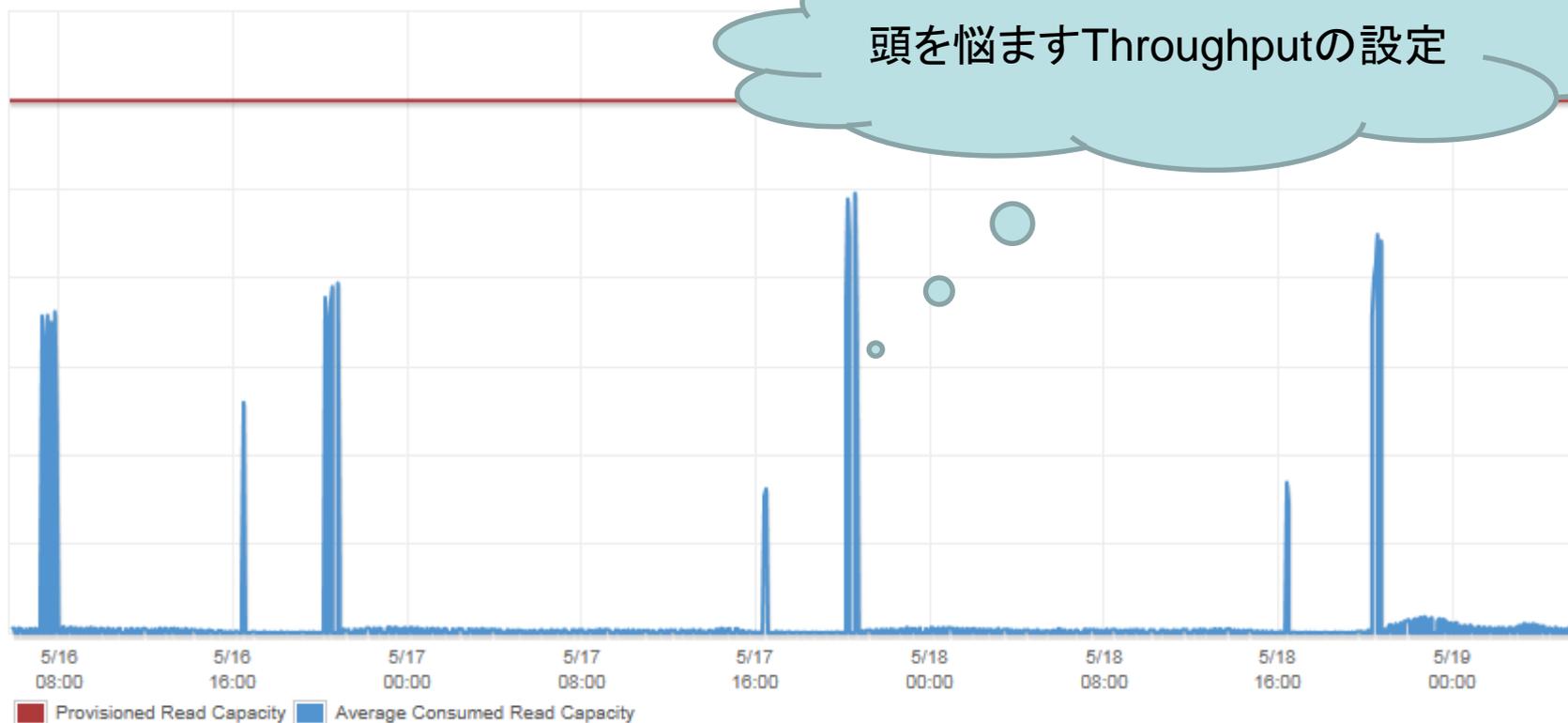


# DynamoDB Throughputの設定

Read Capacity (Count)

Statistic: Time Range: Last 3 Days

Period: 5 Minutes

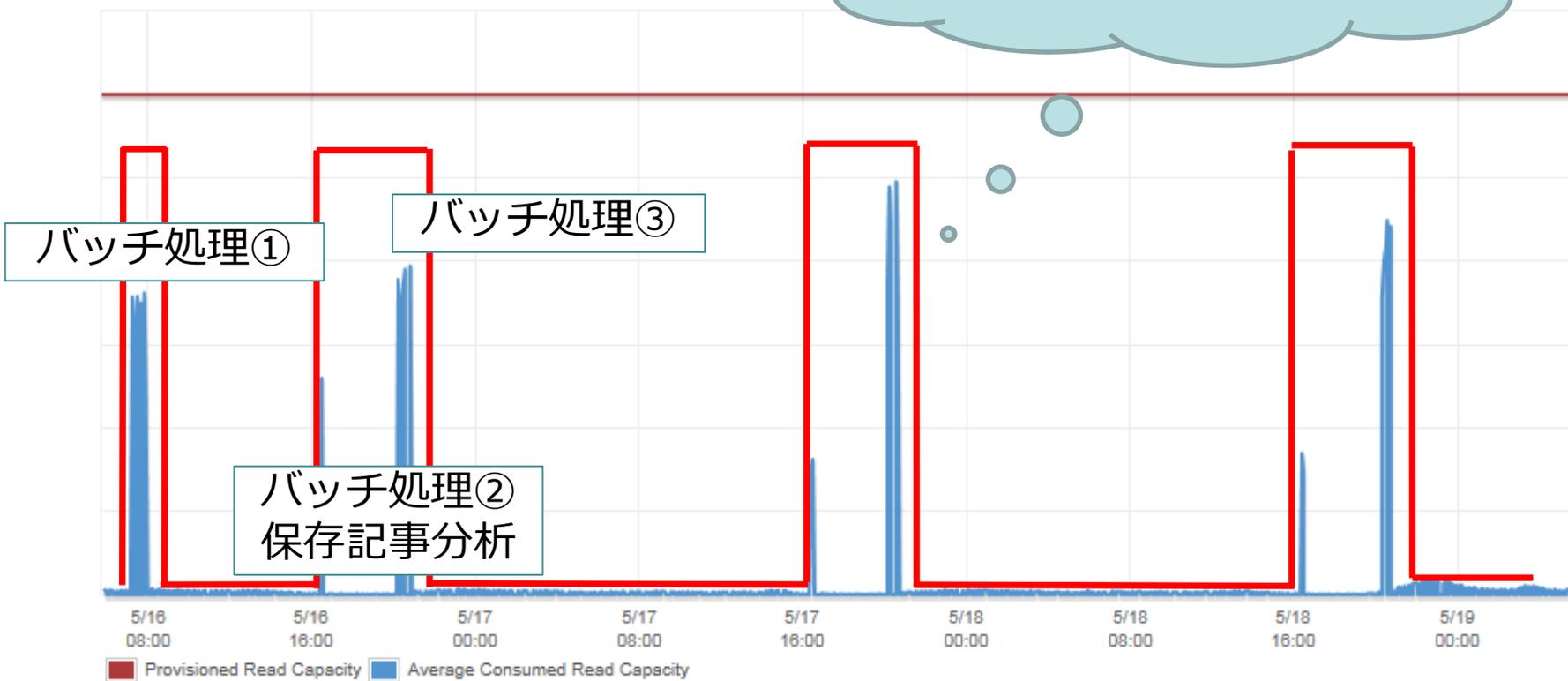


# DynamoDB Throughputの設定

Read Capacity (Count)

Statistic: Time Range:

利用状況に応じて上限値を変更しコスト削減

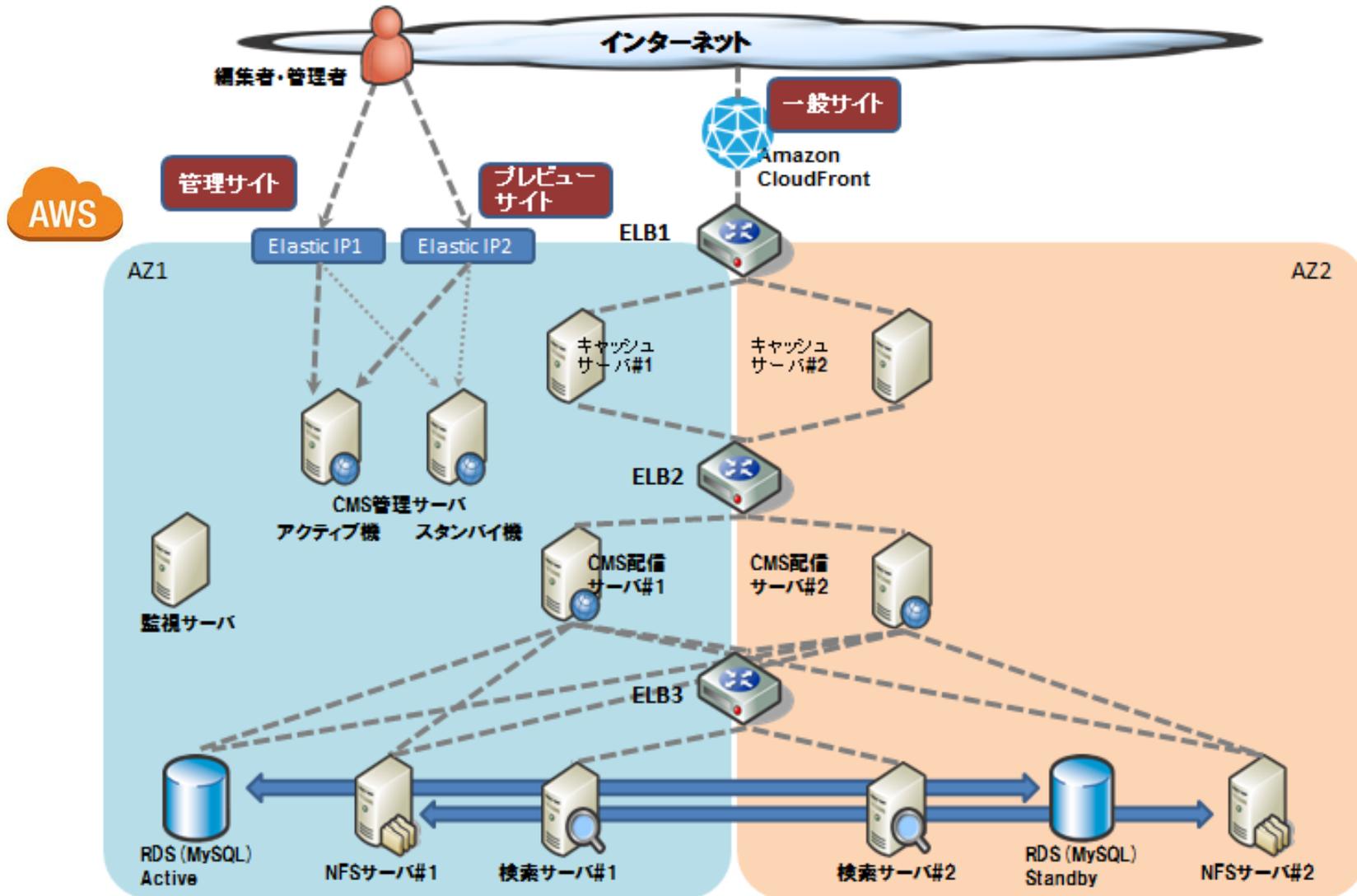


# 電子版以外にも活用1

- Nikkei Asian Review (NAR)
  - アジア経済圏の“今”を日経の視点で発信する英文ニュースメディア
  - 2013/11リリース
  - AWS上に編集システムと配信システムを構築
  - EC2, CloudFront, RDS, ELBで構成



# 電子版以外にも活用1

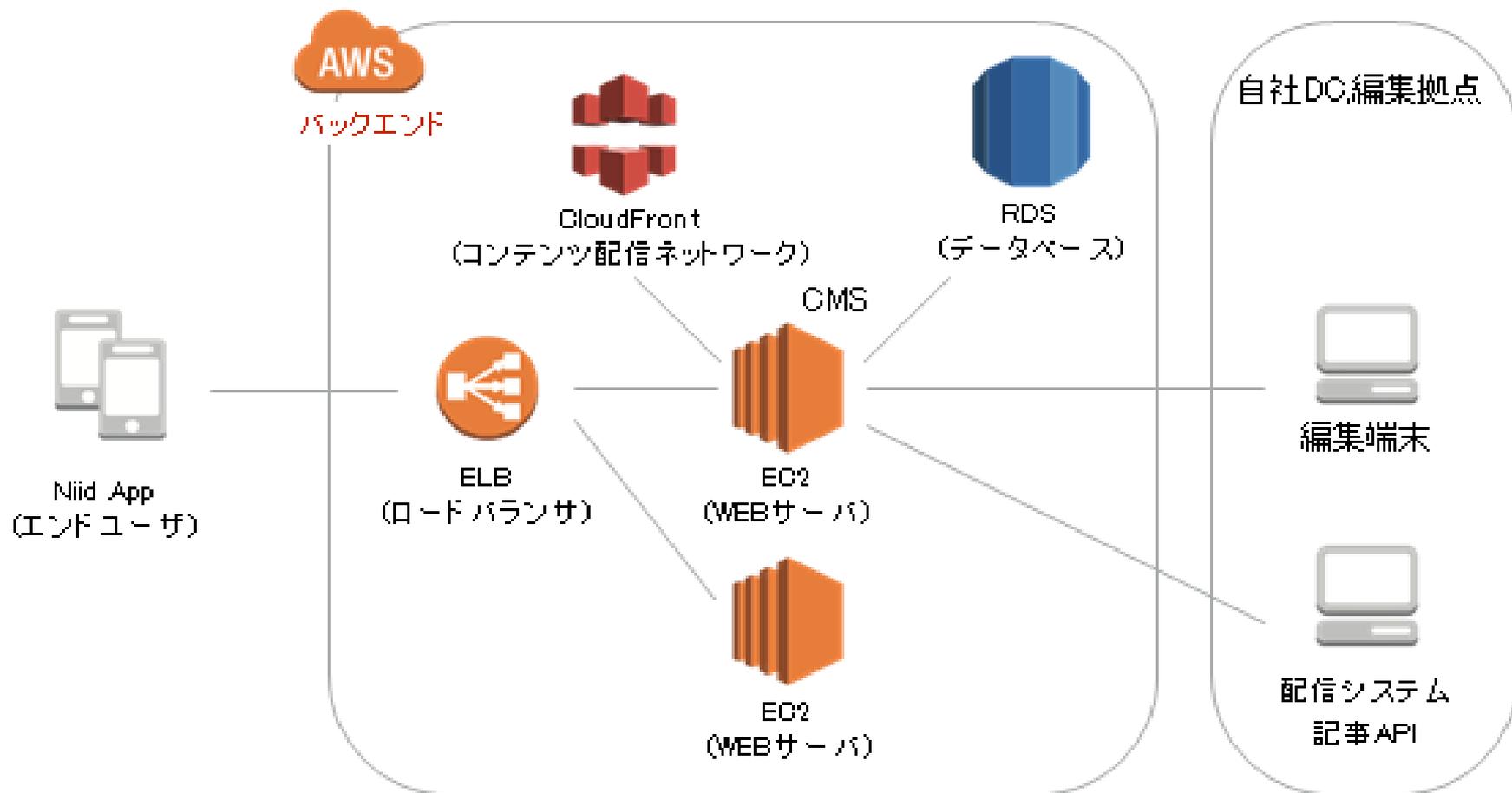


# 電子版以外にも活用2

- Niid(ニード)
  - 日経5紙から厳選したビジネスニュースを提供するアプリ (iPhone、iPad 向け)
  - 2014/9/1リリース
  - AWS上に専用編集システムとコンテンツ配信システムを構築
  - EC2,CloudFront,RDS,ELBで構成



# 電子版以外にも活用2



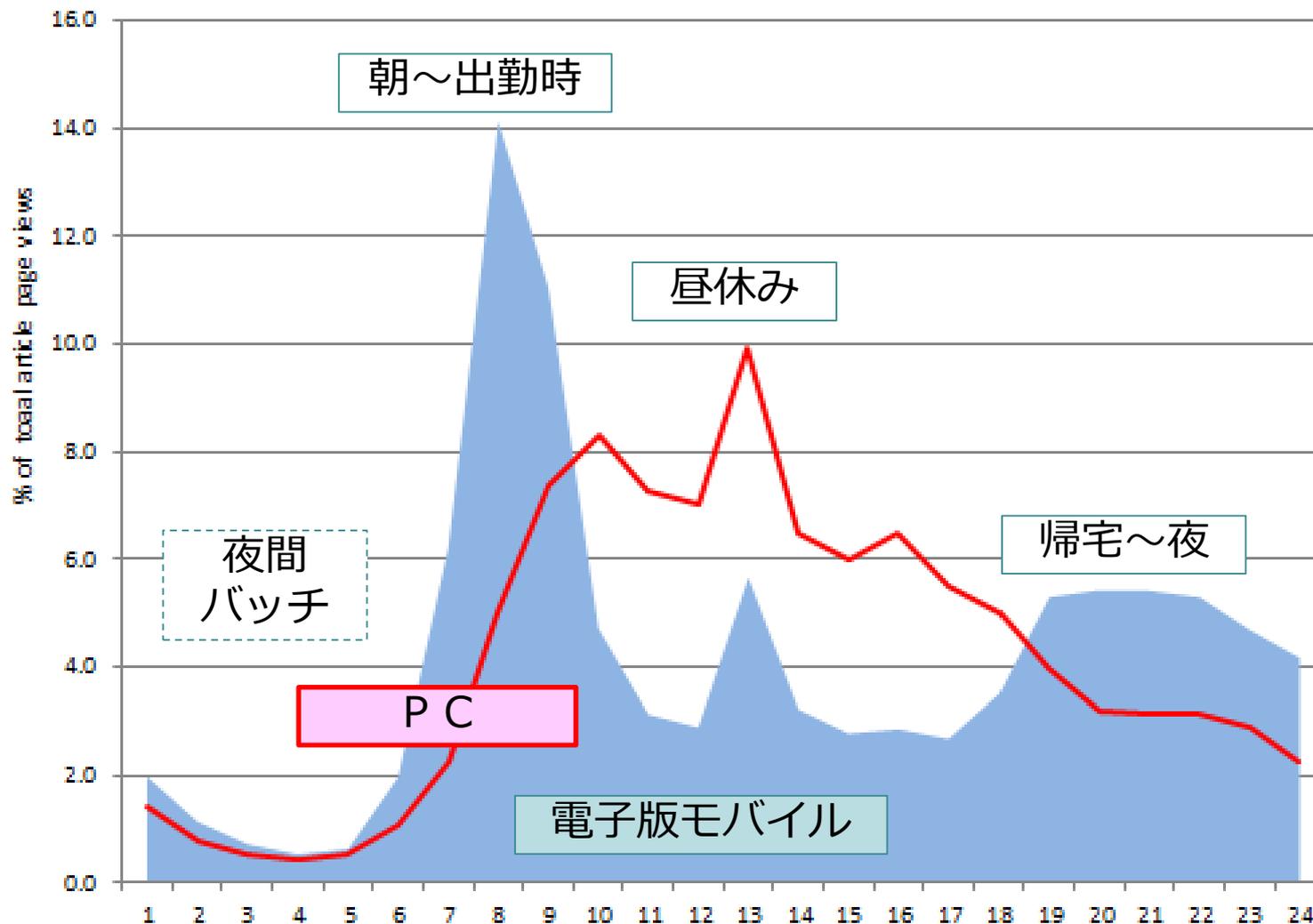
# 今後の展開

- システムの自動構成、自動起動
- 利用状況にあわせて適切なリソースを使用
  - DynamoDBのThroughputをより最適に
  - EC2インスタンスの起動数をより最適に
- データ分析
- AWS(クラウド)利用の拡大・・・

# システムの自動構成、自動起動

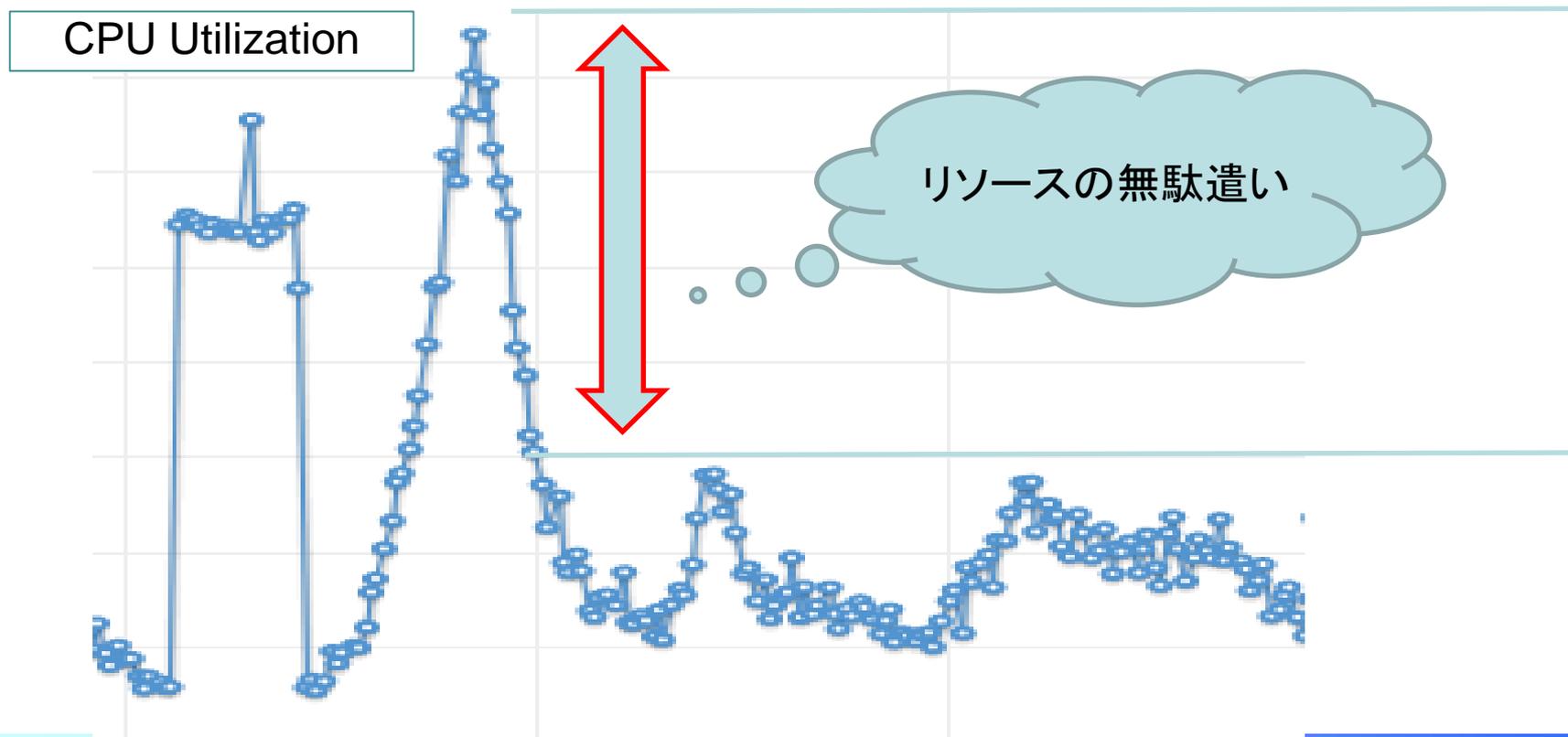
- EC2インスタンスの追加、起動は簡単だが、システム設定を常に最新にあわせる必要がある
  - システムを起動したいときに、自動的に最新の設定で起動できる
- ① Ansibleでミドルウェアのインストールと設定  
(AWS OpsWorksと同等の機能)
  - ② Serverspecでミドルウェアの自動テスト
  - ③ GithubとCapistranoで自動デプロイ
  - ④ CasperJS (UIとAPI)で自動テスト
  - ⑤ システム起動、組み込み

# システムの自動構成、自動起動



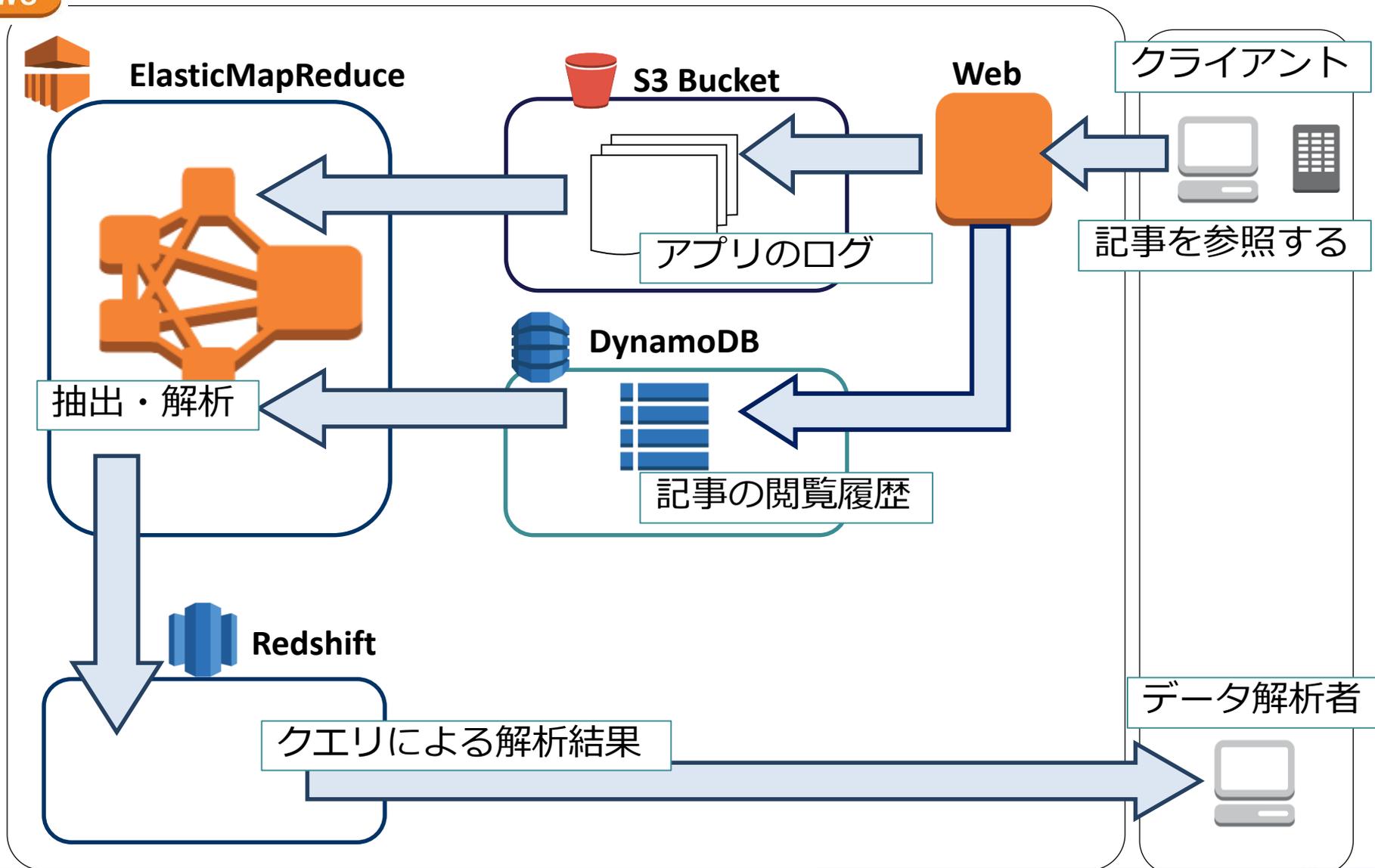
# EC2の自動起動

- 利用状況にあわせたサーバ起動台数の制御
- 突発的なアクセス増にも速やかに対応可能



# 記事閲覧履歴とデータ解析

AWS



# 記事閲覧履歴とデータ解析

- ユーザー毎の閲覧履歴をDynamoDBに格納する
- アプリケーションログと閲覧履歴をEMRで分析・集計してRedshift用に加工作る
- RedshiftにSQLを投げて結果を分析する。

# クラウド利用の拡大

- 電子版モバイルで培った知識を他システムや新サービスでもさらに活用
- 電子版(PC版)をAWS上に移設？
  - 自社DCとの使い分けにはポリシーを持って
  - 段階的にクラウド環境に移設

# 日経電子版は これからも 日々進化を続けます

AWSさんと共に・・・

他のパートナー様も歓迎です

ご清聴、ありがとうございました。

