【AWS Cloud Storage & DB Day 2014】

NTTドコモのデータ分析環境

2014年9月9日

株式会社NTTドコモ
サービスイノベーション部
森谷 優貴

自己紹介

- 森谷 優貴（Yuki Moritani, Ph.D）

- R&Dイノベーション本部 サービスイノベーション部 所属
  - サービス開発部門
    - ✓イノベーションサービス創造
    - ✓サービス・レコメンド・マイニング技術開発
    - ✓インターネット，セキュリティに関わる共通基盤技術

- 内部システムのアーキテクト，インフラエンジニア
  - AWS/オンプレのシステム/セキュリティ設計，構築，運用
  - AWSアカウント管理，ガバナンス担当
  - 共通ツールの作成，運用

- 本日は，ドコモがAWS上にデータ分析環境を構築するに至った背景と現在構築・検証中のシステムについてご説明します

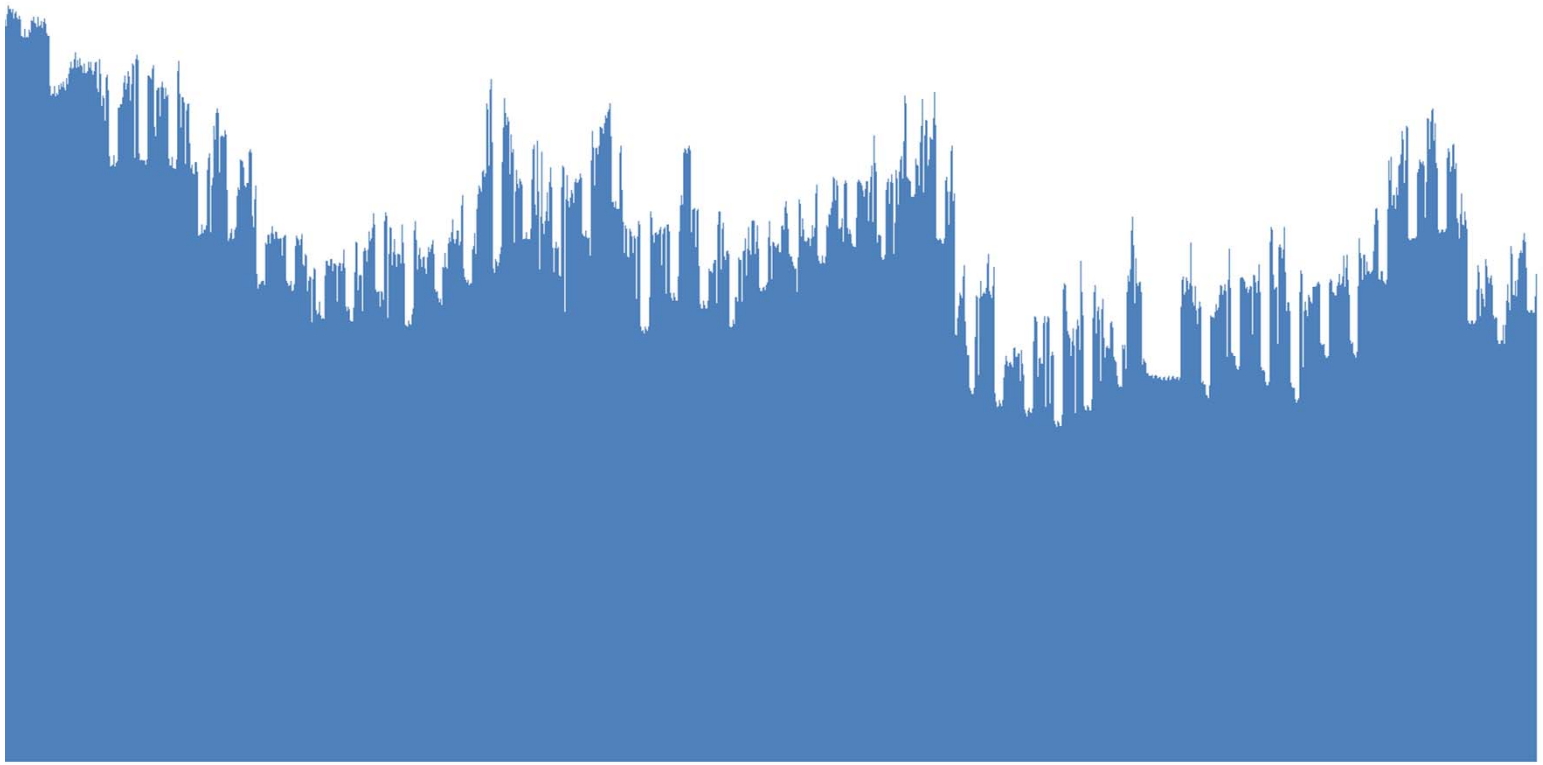
本日のAgenda

- NTTドコモのAWS利用
  - 利用・運用状況のご紹介
  - しゃべってコンシェル
- AWS利用経験の蓄積
  - 社内の意識変化
  - ガバナンスの必要性と取り組みのご紹介
- 業務系システムでのAWS利用：データ分析環境
  - 構築の背景
  - システムアーキテクチャ
- まとめ

NTTドコモのAWS利用

利用・運用状況のご紹介

- ４桁のEC2インスタンスが稼働中
  - Auto-scalingやカスタムチェックによって日々増減
    - ✓各システムでのコスト最適化も継続的に実施
  - Consolidated BillingでRIをShare
    - ✓登録アカウント数は継続的に増加中



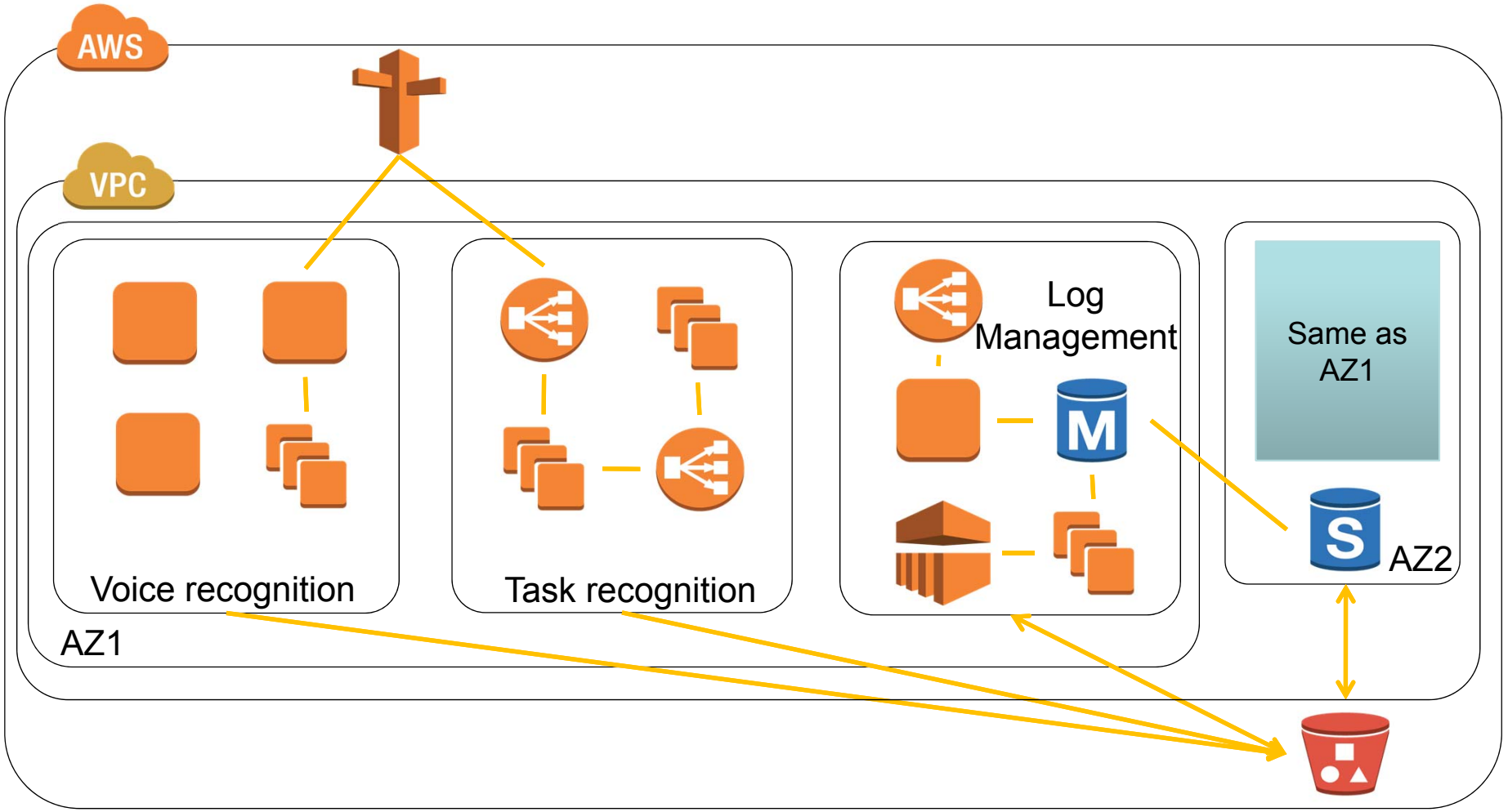
しゃべってコンシェル

- ドコモの音声エージェントサービス
  - ユーザの入力に応じて適切な答えを返す

入力＝音声
→ 音声認識
→ 意図解釈
→ 出力（文字＆音声）

しゃべってコンシェル

- システムアーキテクチャ（２０１２）
  - 音声認識，意図解釈，ログ管理をコンポーネントとしたシステム構成

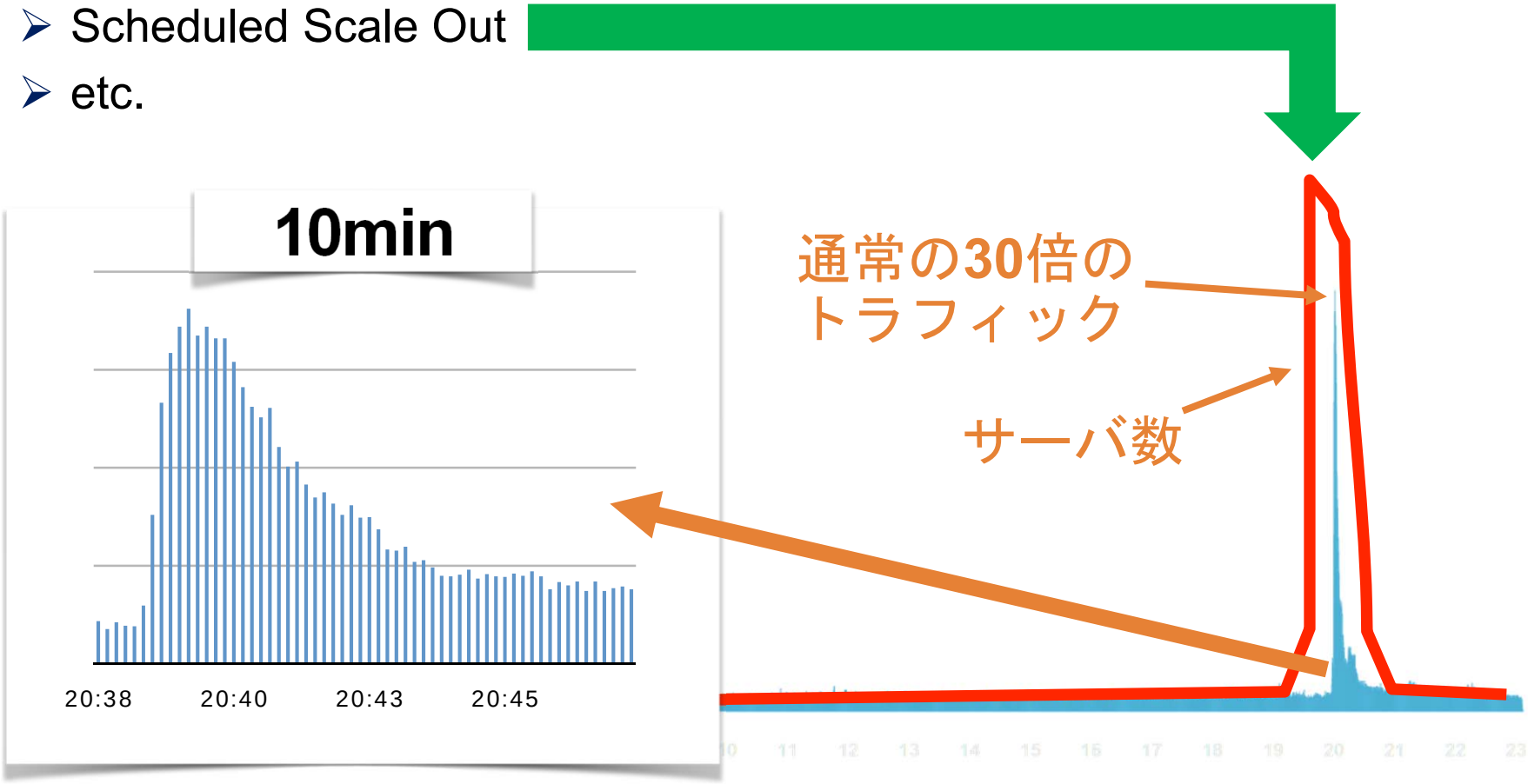


しゃべってコンシェル

- 2012年のサービス開始と共にトラフィックが急増
  - 当初は別のクラウドを利用していたがリソース問題が発生
  → AWSを使い始めた最大の理由
- 2012年4月 AWSに移行
  - トラフィック急増は続く
- 2012年7月 北カリフォルニアリージョンへ移行
  - リソースが潤沢なリージョンの利用
- 2012年9月 東京リージョンへ再移行
  - 遅延の改善
- AWSのリージョン間差異がないこと，CDPを利用したCloud Nativeな構成としたことの恩恵で，上記が実現できた

しゃべってコンシェル

- 使っているCDP
  - Multi-Datacenter
  - Clone Server
  - Scale Out
  - Scheduled Scale Out
  - etc.

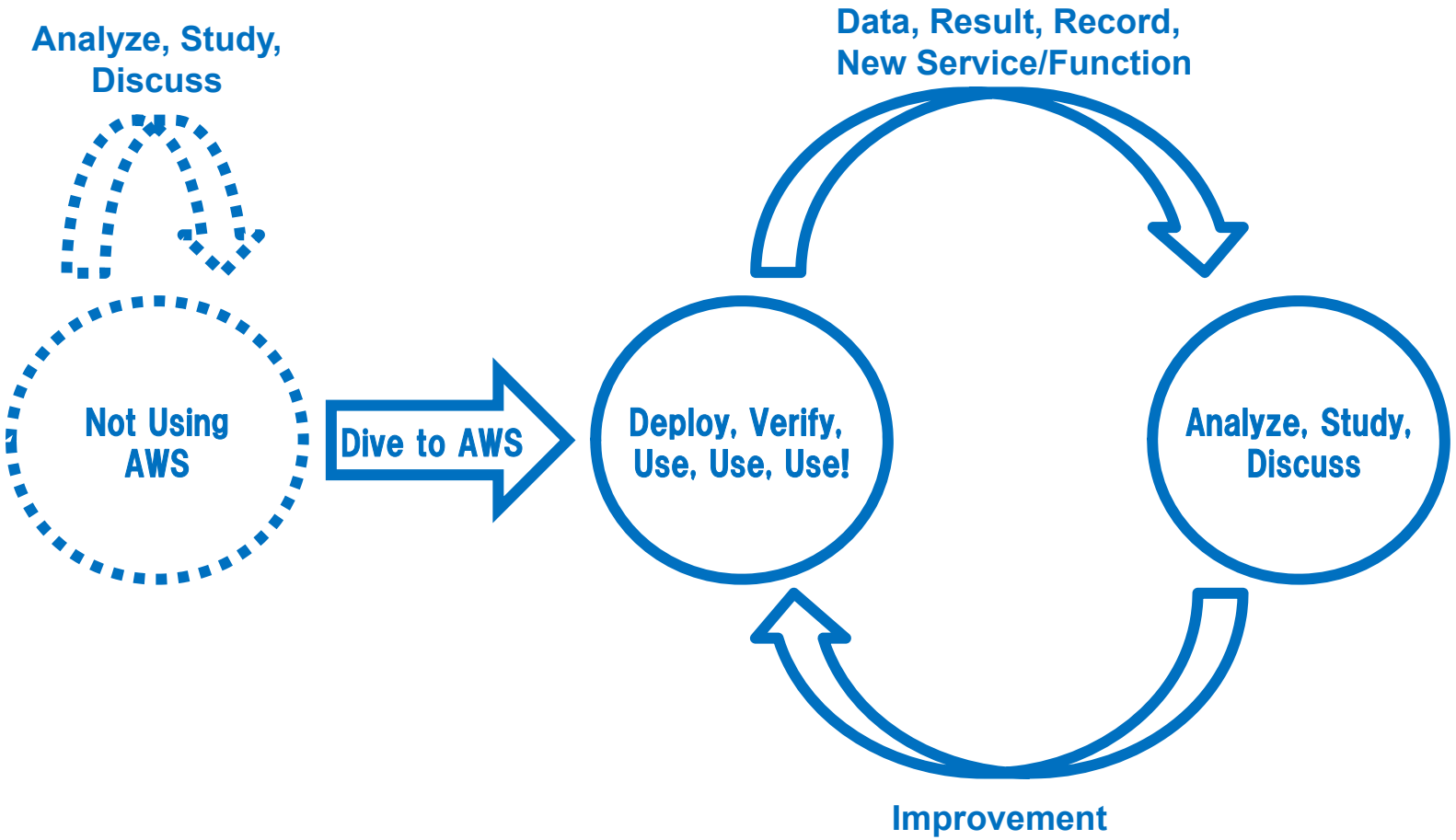


AWS利用経験の蓄積

社内の意識変化

- クラウドに対する意識
  - クラウドはよくわからない → クラウドも使い方次第
- セキュリティに対する意識，対策
  - 物理的境界 ＞ 論理的境界 → 物理的境界 ≧ 論理的境界
  - AWSの豊富なセキュリティ機能を活用
    - ✓ IAM，VPC，NACL，MFA，SSE/CSE，VPN，CloudTrail，etc.
- 使いながら改善していくことが大事
  - 実運用データを分析してScaling
  - 機能集約，機能分割
  - AWSの新機能リリース
    - ✓ 独自構築から機能活用へ移行

AWS利用スパイラル



- それでも全てが上手くまわるわけではない

特殊なケース

- 諸事情で小さく始められないプロジェクト
  - 例えば，新サービスながら最初から多くのユーザが存在

- Cloud Non-Nativeなデザインパターン： Try fast，think later
  - とりあえず既存のコンポーネントの組み合わせで非効率だとわかっていながらもサービスを作ってしまう　→　後で徐々に最適化する
  - サービス提供，品質，パフォーマンスを優先　→　後でコストを最適化する
  - 初版はセキュリティ，トラフィック対策を過剰に　→　後で改善
    - ✓ ウィルス対策，IPS/IDSを多くのノードに　→　実績見合いで減らす
    - ✓ Auto Scaleできないものは多めに　→　実績見合いで減らす
    - ✓ デザインはTraditionalで実績の多いものを利用　→　徐々に新しい方法に移行
    - ✓ 出来るだけバックアップは多めかつ頻繁に　→　後で不要なものを精査，除去
  - ただし，最適化に備えてRIは買わない

- 大きく作っても後から最適化できるのがクラウドの強み

ガバナンスの必要性

- Consolidated Billingは利用量に対する請求額を最適化
  - 個々のシステムが最適化されているかはBillingでは分からない

- クラウドを上手く使えば効果も大きいが，使い方を間違えると大変
  - セキュリティ的に問題のある構成，運用
  - 小さく作って大きく伸ばすことも，大きく作って最適化することもできない構成
  - オンプレでの開発に慣れていると陥りやすいこともある
  - アカウント数が増えてると全システムを細かくコンサルティングするのは困難

- 間違った使い方をしないためにはガバナンスが重要
  - クラウドを使う場合の考え方，お作法
  - 気をつけるべきセキュリティ観点

  等を利用者，開発者に正しく理解してもらう

ドコモにおけるガバナンス強化の取り組み

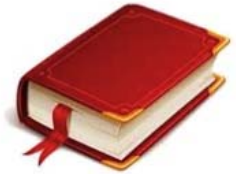
- クラウド開発ガイドラインの作成，展開
  - AWSを使う場合の考え方やお作法，ドコモの開発フローにおける各フェーズで考慮・実施すべき指針を記載（現在150ページ弱のボリューム）
  - 特に構成・セキュリティ等は重点的に網羅し，間違った使い方を抑止

クラウド開発ガイドライン

NTTドコモ
サービスイノベーション部
第3サービス開発担当

本ガイドラインについて

□ 対象
クラウド（特にAWS）を利用して開発を行う方



□ 本ガイドラインのカバーしている内容
- クラウドの特性の解説
- クラウドに開発における前提となる考え方
- 開発時に必ず実施すべきことと注意点
- ドコモ開発フローに沿ったクラウド利用ノウハウ

□ 本ガイドラインの利用方法
本ガイドラインはボリュームが非常に多いですが，クラウド利用時に分かっておかなければいけないことを記載しております．開発フローごとにまとめているのでの各フェーズに差し掛かった際に必要なところを読むといった利用をしてください．

※本ガイドラインではクラウド利用の場合に限定して作成しており，
オンプレミス利用時と共通となる考え方などは極力省いて解説しております．

3

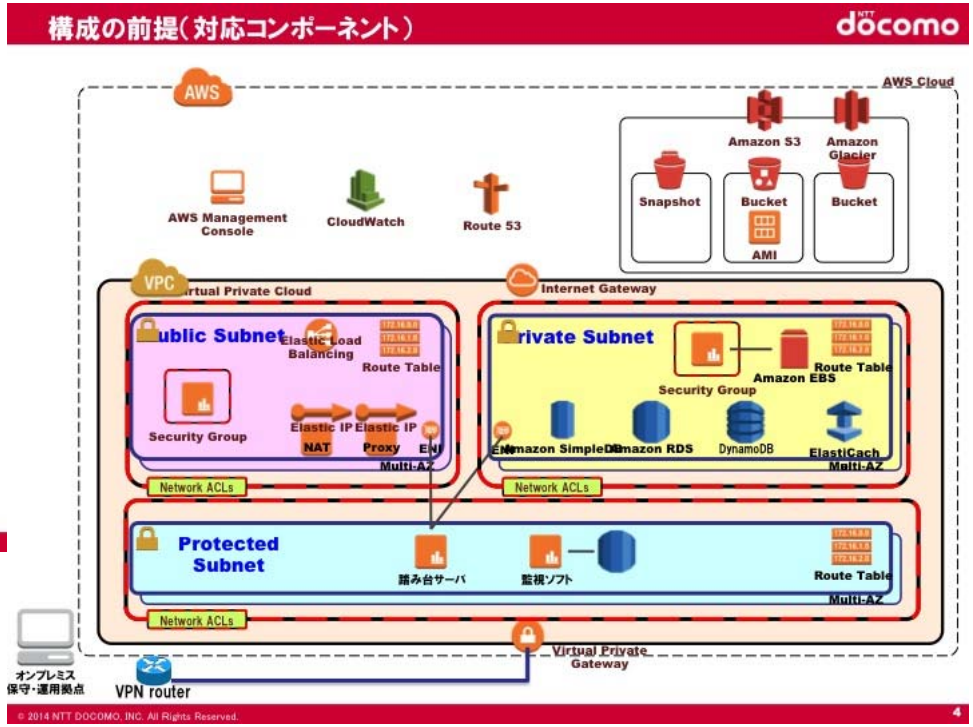
ドコモにおけるガバナンス強化の取り組み

- AWSデザインパターン（セキュリティ対策）の作成，展開
  - AWSを利用する際に必要となるセキュリティ要件を記載
  - 考慮漏れを抑制し，ドコモのセキュリティ基準に対する準拠性を高める

手のひらに、明日をのせて。 docomo

AWSデザインパターン
～SS対策基準対応要求仕様編～

サービスイノベーション部



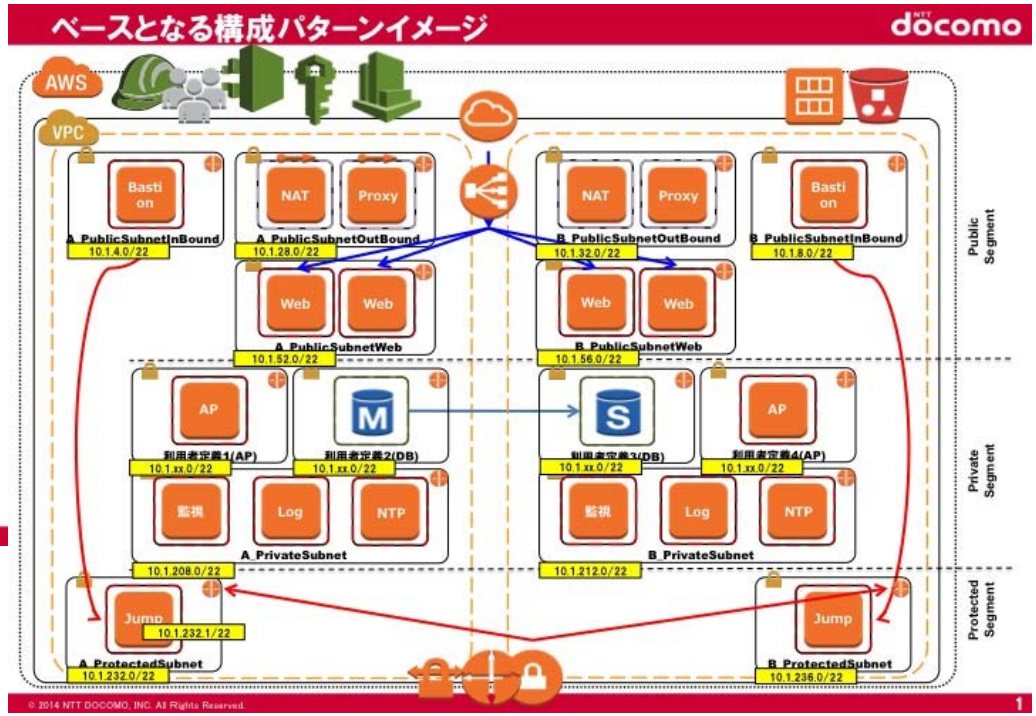
ドコモにおけるガバナンス強化の取り組み

- セキュリティ対策を満足するテンプレート作成，展開
  - マニュアルと共にCloud Formationのテンプレートを提供
  - デプロイのベースとして利用してもらい，意識と実環境のずれを抑制

手のひらに、明日をのせて。 docomo

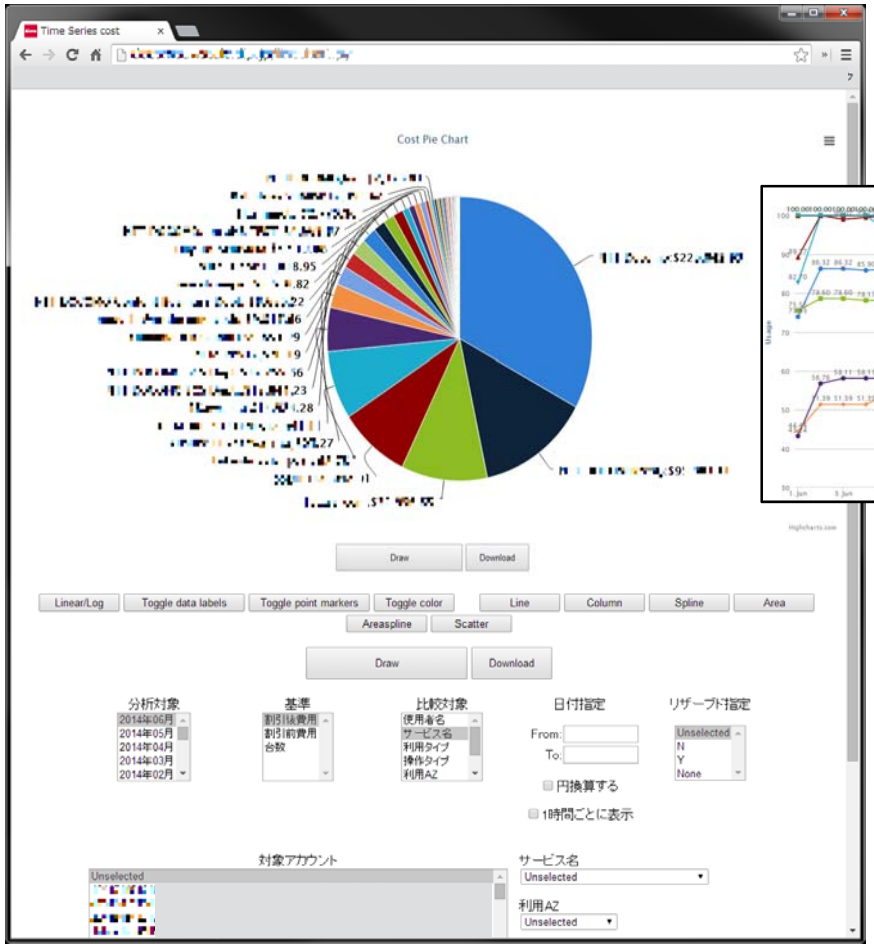
AWSデザインパターン
実装編

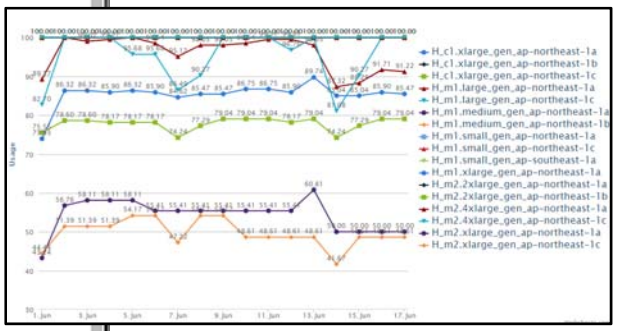
サービスイノベーション部
開発推進担当



ドコモにおけるガバナンス強化の取り組み

- コスト管理ツールの作成，展開
  - アカウント，サービス，日時，等に応じたコスト変遷，利用率等を表示





<u>表示形式</u>
- 利用アカウント別
- 利用サービス別
- 日付指定
- 1時間/1日毎
- リザーブド(RI)指定
- AZ
- インスタンスタイプ

<u>時系列でのコスト表示</u>
- 利用料表示
- 利用台数表示
- 利用アカウント別
- 利用サービス別
- 日付指定
- 円換算（為替反映）
- 1時間毎/1日毎
- リザーブド(RI)指定
- 利用サービス絞込
- 利用AZ絞込

<u>RIの有効利用度表示</u>
- 有効なRI数
- RI利用数
- RI利用率
- RI余剰数

【閑話休題】AWS Trusted Advisor

- AWS環境の基本的なチェックを実施してくれるツール

- 以下のセキュリティチェック結果を表示（一部はサポートプラン依存）
  - ✓ Security GroupのRuleチェック
  - ✓ AWSアカウントのMFA利用チェック
  - ✓ IAMの利用チェック，Password Policyチェック
  - ✓ S3のBucket Policyチェック
  - ✓ CloudTrail利用確認
  - ✓ Route 53のRecordチェック
  - ✓ RDSのSecurity Groupチェック

AWSの利用拡大

- Webサービスシステム

  2012年～

- 業務系システム

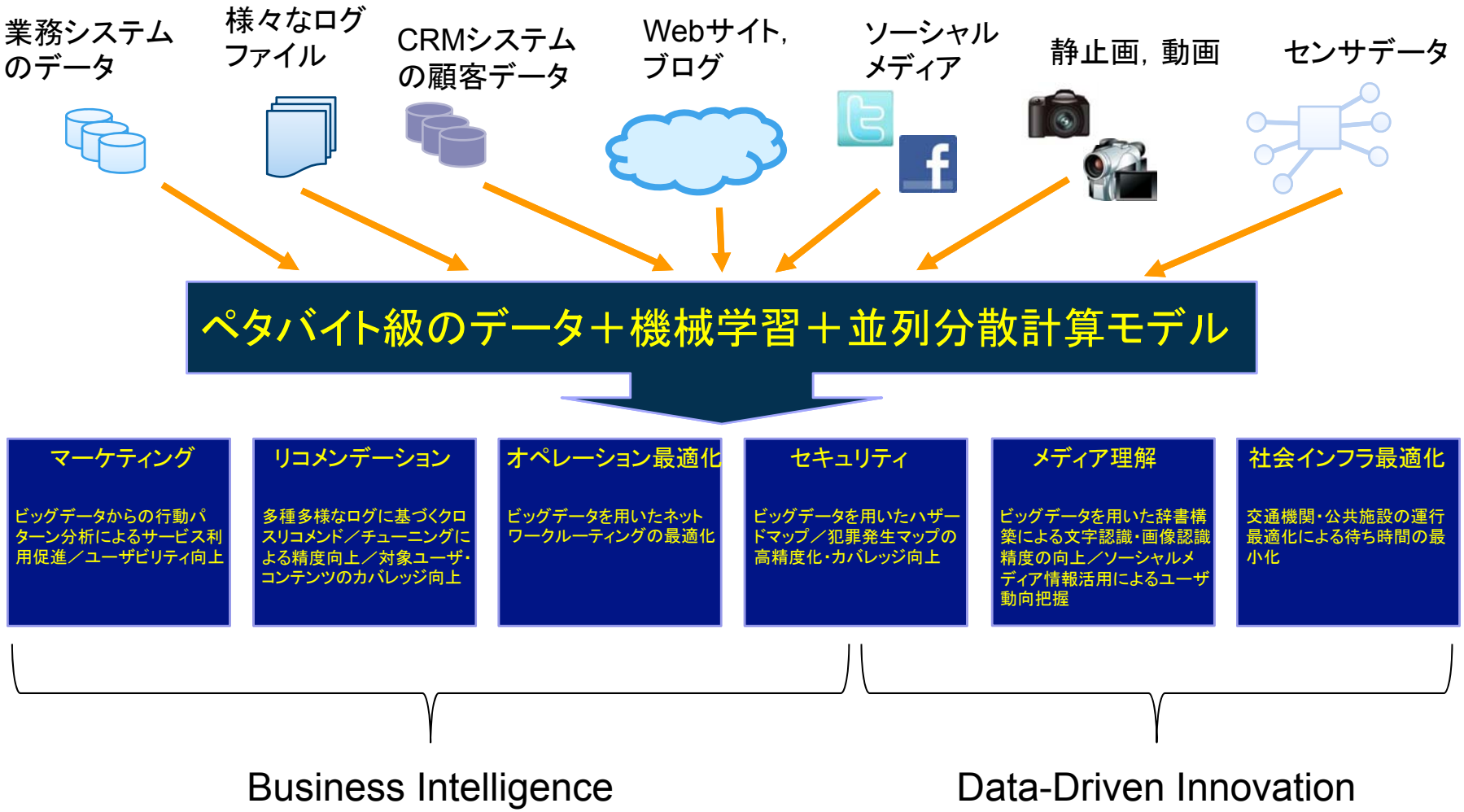
  2014年～

- ミッションクリティカルシステム

業務系システムでのAWS利用
データ分析環境

構築の背景

- ドコモにおけるビックデータ活用

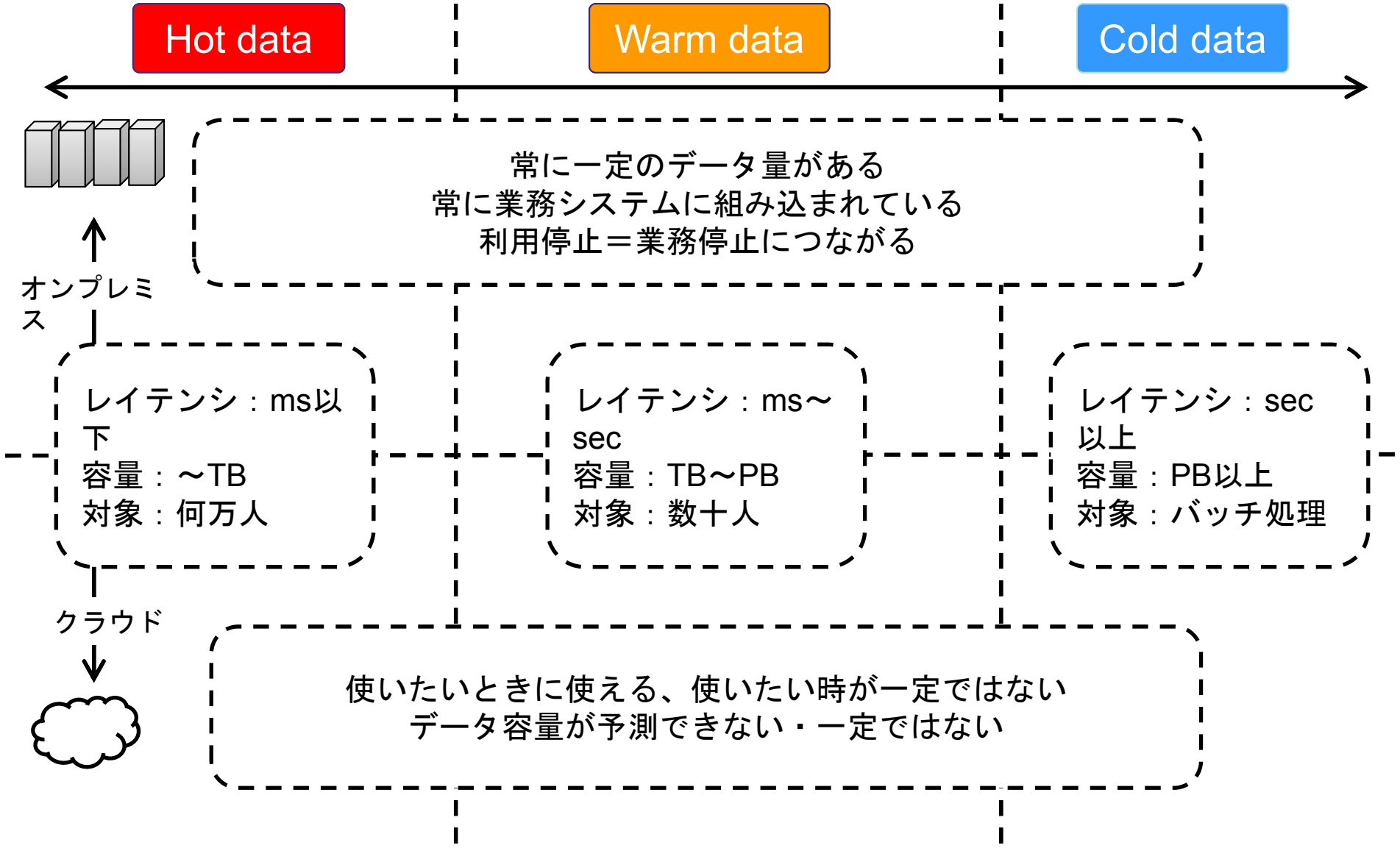


構築の背景

- データ分析環境の更改
  - PBクラスの基盤を新たに構築

構築の背景

o システム特性

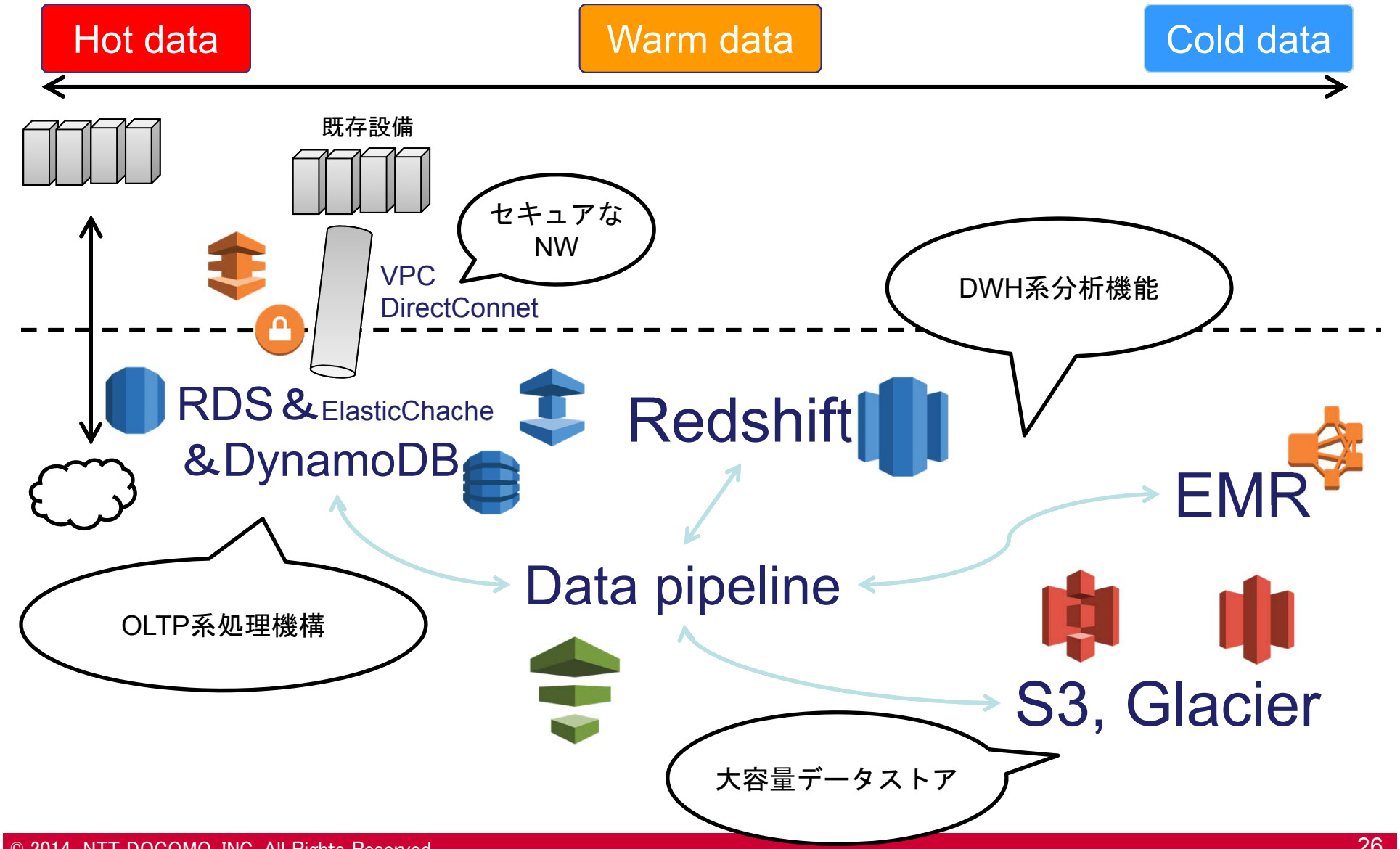


構築の背景

- データ分析のラインナップ

構築の背景

- AWSのデータ分析ラインナップ



構築の背景

- システム要件の整理
  - データ取得から分析開始までの時間
  - システムのパフォーマンス，可用性，信頼性，保守性
  - 将来の拡張性
  - 満足すべきセキュリティ・コンプライアンス
  - etc.

- 要件を満たす候補のピックアップ，性能検証
  - Software，Appliance，Amazon RedShift

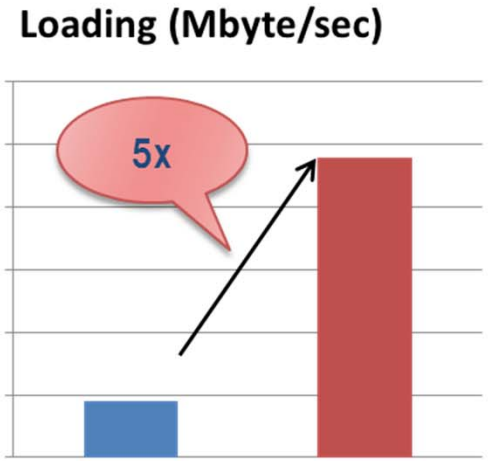
- セキュリティ評価
  - AWSのセキュリティ再評価
  - 業務系システムのセキュリティ要件との突合

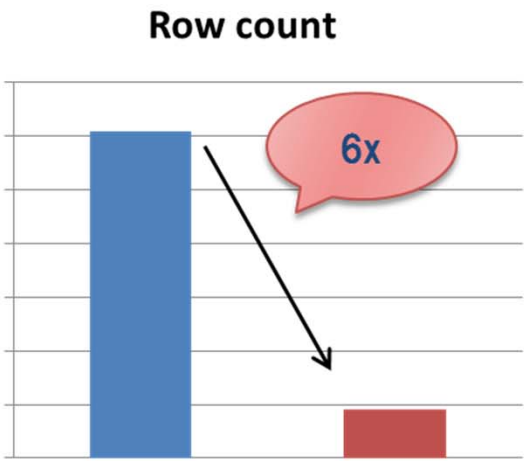
- 上記を踏まえて，最終的にRedShiftの採用を決定

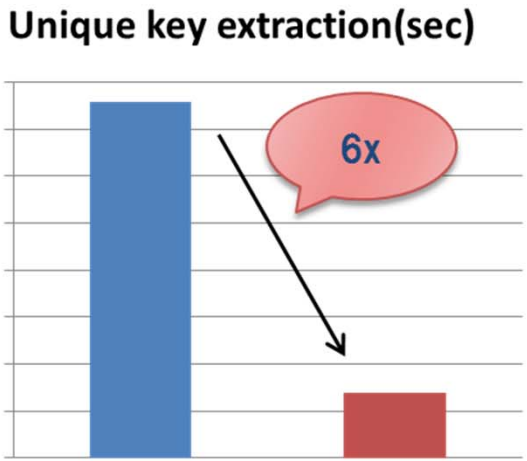
構築の背景

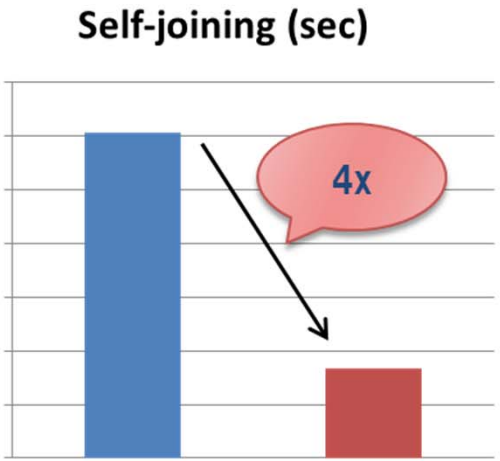
o RedShiftの基本性能検証

- データロード性能，行計算性能，Join性能，等を従来のSoftware製品の性能と比較
- 同一環境でないため，正確な比較ではないが，RedShiftの基本性能には問題がないことを確認





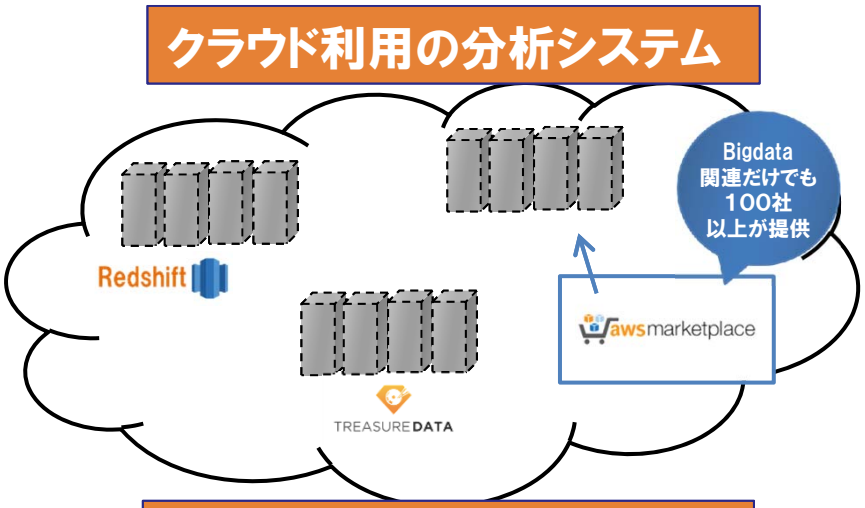


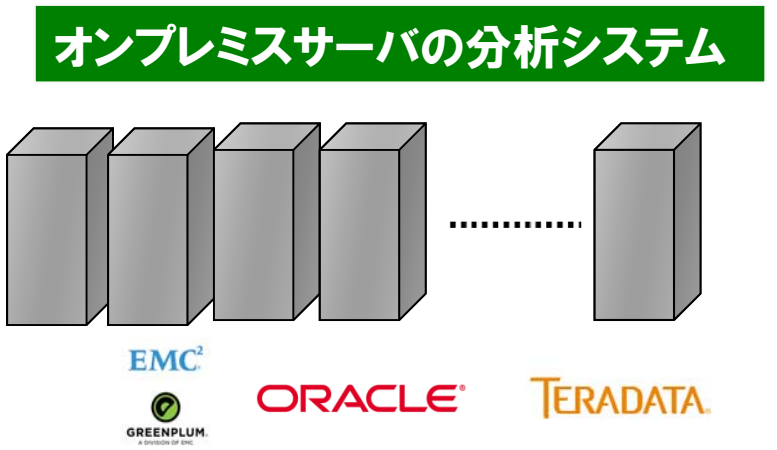


※ グラフの数値は参考

構築の背景

o セキュリティ評価…の前に，オンプレミスとクラウドの違い





構築・設計変更・機能追加の柔軟性	現状の機能は充実、新規機能追加は高コスト
Market Placeの製品を自由に使え，**常に最新の様々な会社のソフトウェアが利用可能**．構成も柔軟に変更可能． ◎	事前に性能評価し，**将来性も含めた検証が必要** 大きな単一システムが初期から必要で構成変更は困難 ×
新しいデータへの対応に対して，最適な分析システムの構築がPoCを含めても**数週間で可能**	新しいデータへの対応には，ハード購入，論理設計，製品選択などのため，**3～6ヶ月以上必要**
HW**更改不要**．常に最新のHWが利用できる．（AWSが実施） メンテナンス業務が**不要**（AWSが実施 ※独自構築部分除く）	HWが陳腐化した場合は**新規HWで再構築**する必要あり メンテナンス業務のための要員が**別途必要**
一時的な分析力増強、不要時の休止（費用削減）が可能	能力の柔軟な変更は困難で，**24時間365日の運用が必要**
クラウドのソフトウェアとして機能が未成熟な事がある ×	**製品の歴史が長く，機能は成熟している．** ◎
セキュリティはAWSと共同で責任を持つ	**セキュリティはユーザが全責任を持つ**

構築の背景

- セキュリティ評価
  - 物理的対策，運用・監視・体制，ログ・情報保護，個別対策，等のセキュリティ対策要件について以下の点を確認
    - ✓ AWS自体で満足しているか？
    - ✓ AWSの機能を利用して対策可能か？
    - ✓ ドコモが独自に対策可能か？
    - → ドコモが手を出せない部分で未対策の項目がないか？

構築の背景

- セキュリティ評価
  - AWSのIaaSとしてのセキュリティ
    - ✓ 各種認証取得要件の確認
    - ✓ AWSセキュリティチームからのヒアリング
  - 利用可能なセキュリティオプション
    - ✓ IAM，SG/NACL，CloudTrail，等
    - ✓ 暗号化オプション
    - ✓ サービス毎のログ
    - ✓ etc.

  - ドコモ側で適切に対応することで実現できるセキュリティ
    - ✓ 暗号危殆化対策
    - ✓ etc.

- 結論
  - AWS上でドコモが適切な対策をすればセキュリティ要件を全て満足可能

構築の背景

- 検証を踏まえて...
  - AWS上で構築可能なことを確認
  - しかしながら，オンプレでも構築は可能

- では，RedShiftを採用した最終的な決め手は？
  - AWSの機能拡充
    - ✓ 例：東京リージョンへのCloudTrail導入（2014年6月）
  - RedShiftの拡張性，機能拡充
    - ✓ 例：Query Flexibilityの拡大（2014年5月）
  - 更にAWSでは以下のメリットが享受できる

0. Transparency

- VPCとの直接接続でオンプレミスと透過な環境を実現

1. No provisioning

- データ量の増減に対応可能な柔軟性，コスト効率化

2. Sandbox

- AWSやMarketplace等で提供される様々な分析ツールを利用可能

構築の背景

- Sandboxの例

- Hadoop
  - Trying new analytics(e.g. machine learning, unsupervised clustering) with programming language

- Graph database
  - Trying social influence analysis

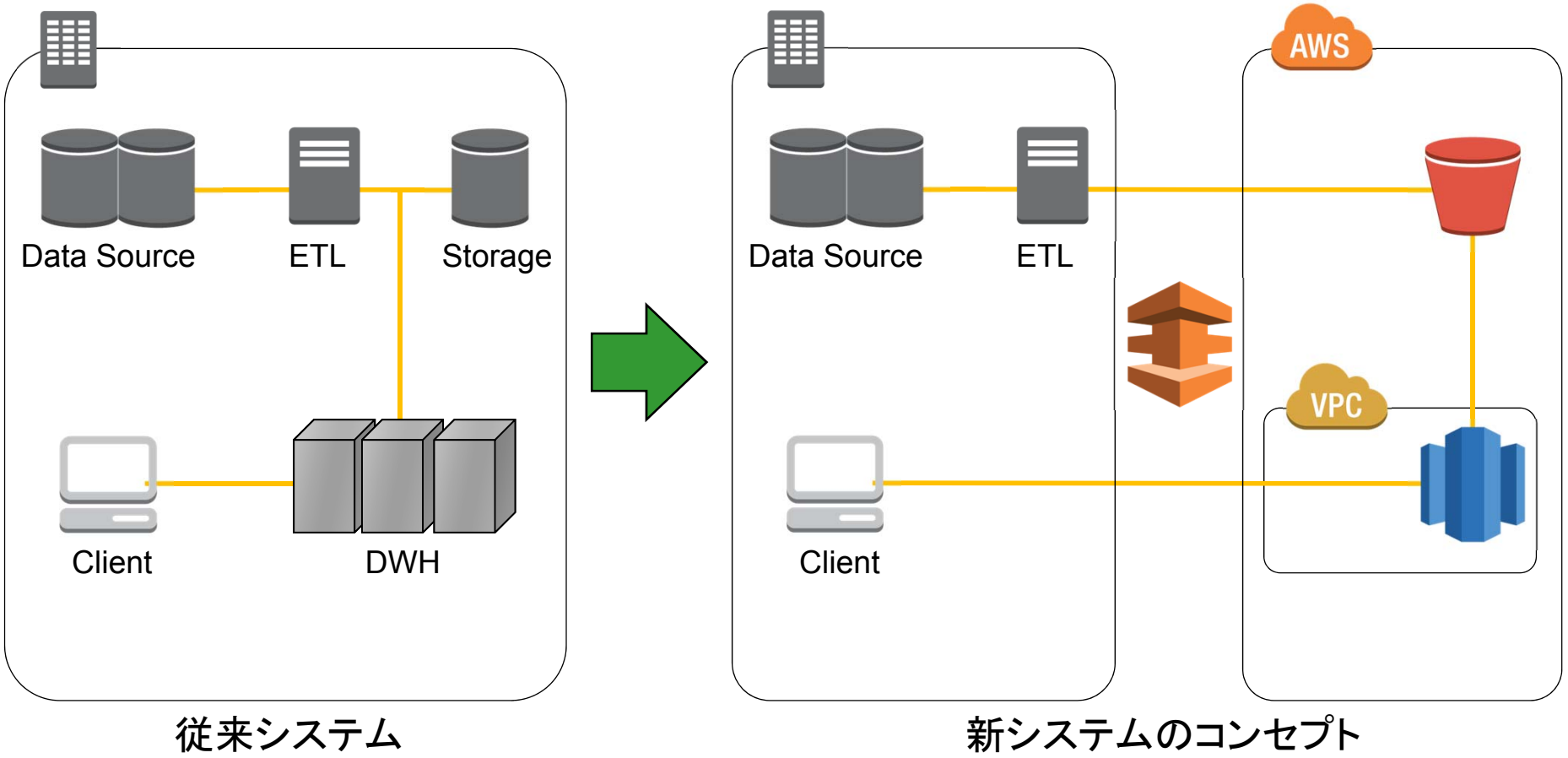
- Streaming process, In-memory database
  - Enabling real-time recommendation of EC sites

- ETL
  - Testing new data and injection to DWH

- BI and analytics platform (e.g. R)
  - Providing dashboard to internal customer temporary
  - Enabling to use favorite language of data scientists

システムアーキテクチャ

- 基本コンセプト
  - オンプレ環境と比較してユーザビリティを下げない
  - セキュリティを担保しつつ，クラウドのメリットと既存資産を活用可能に

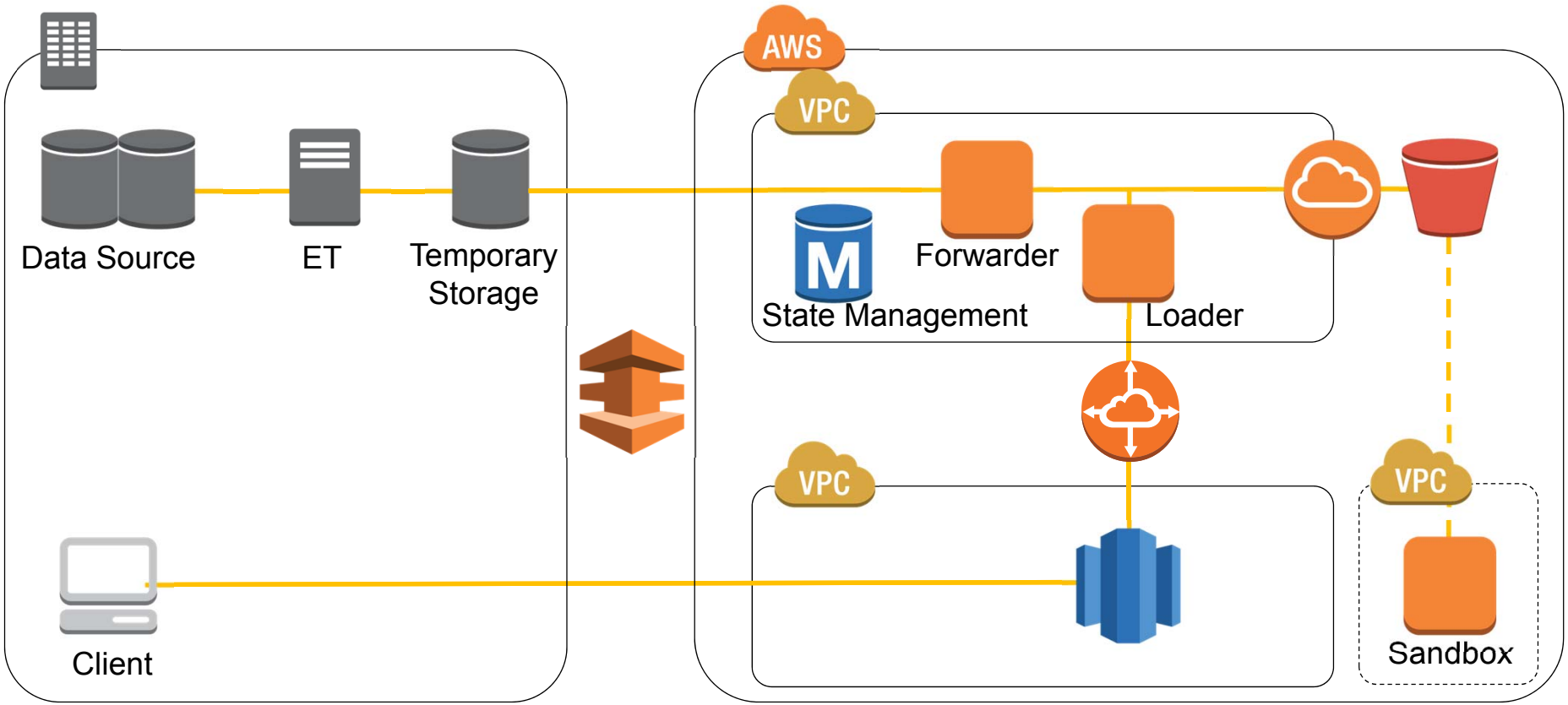


従来システム　　新システムのコンセプト

システムアーキテクチャ

検証中の構成

- コンセプトを実現するためのデータフロー設計
  - ✓ 管理者環境とユーザ環境を分離
  - ✓ S3をデータハブとしたSandbox拡張性



システムアーキテクチャ

- セキュリティ対策
  - 暗号化
    - ✓ オンプレミス
      - 従来通りに暗号化利用
    - ✓ S3
      - SSEに加えてCSEを利用
    - ✓ RedShift
      - Cluster暗号化オプションを利用
    - ✓ EBS
      - 暗号化オプションを利用
    - ✓ 通信路
      - SSL利用

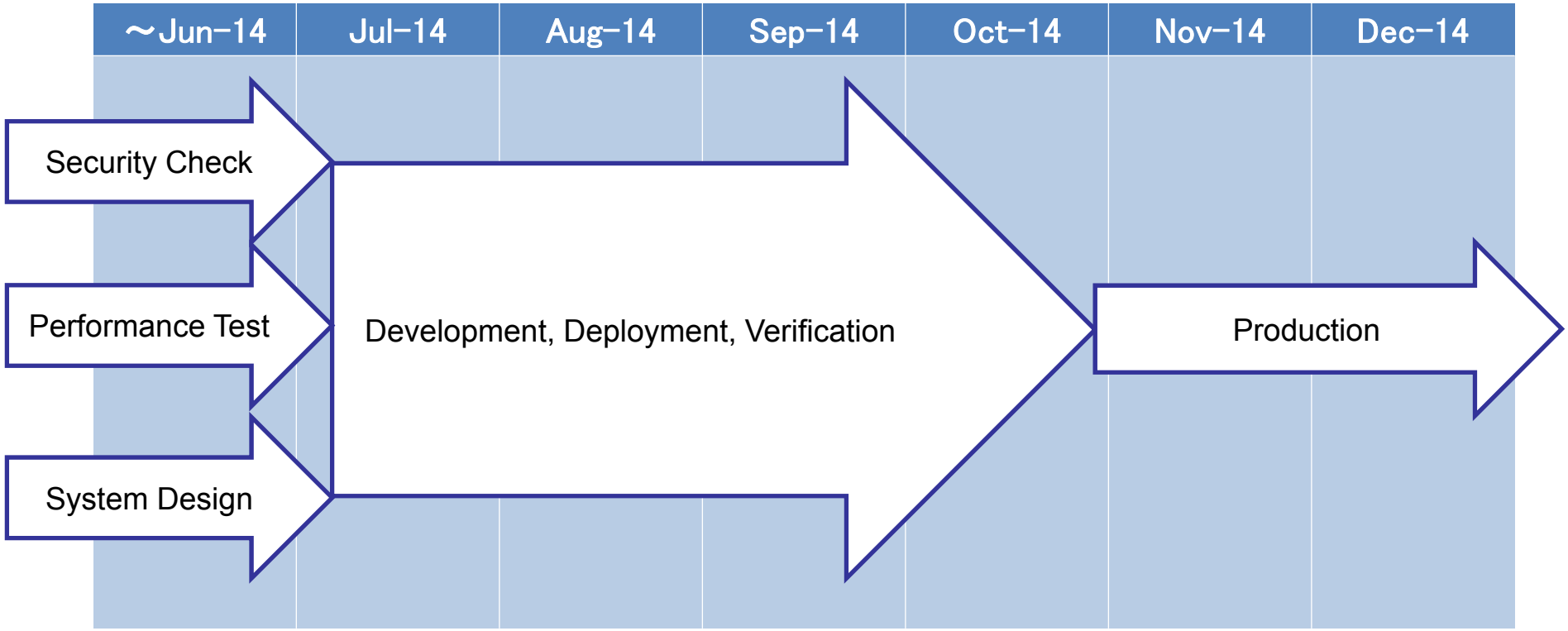
※利用する暗号化方式は最も堅牢なものを選択

システムアーキテクチャ

- セキュリティ対策
  - アクセス制御：AWS機能をフル活用
    - ✓ VPCを利用
    - ✓ Direct Connectを利用
    - ✓ IAM，MFAを適切に設定
      - Group/Roleの活用
    - ✓ SG/NACLを適切に設定
    - ✓ S3のACL/Policyを適切に設定
  - ユーザ制御
    - ✓ RedShiftのユーザを適切に設定
    - ✓ RDSのユーザを適切に設定
  - ログ管理
    - ✓ 全ての操作ログを取得
      - CloudTrail，S3/RedShift/RDSログ，等
  - OS以上のユーザ責任部分についても適切なセキュリティ対策を実施
  - オンプレミス側でも適切な制御・管理・対策を実施

スケジュール

- 2014年10月より運用開始予定



まとめ

- NTTドコモのAWS利用
  - 利用・運用状況
  - しゃべってコンシェル
- AWS利用が進むにつれて
  - 社内の意識変化
  - ガバナンスの必要性と取り組みのご紹介
- 業務系システムでのAWS利用：データ分析環境
  - 構築の背景
  - システムアーキテクチャ

まとめ

- AWS利用のステップ
- 1st Step：使ってみること
- 2nd Step：経験を蓄積し使いこなすこと
- 3rd Step：要件に合わせて柔軟に使うこと
- 4th Step： ドコモの今後にご期待ください

ご清聴ありがとうございました

連絡先
森谷 優貴（Yuki Moritani）
moritaniy@nttdocomo.com