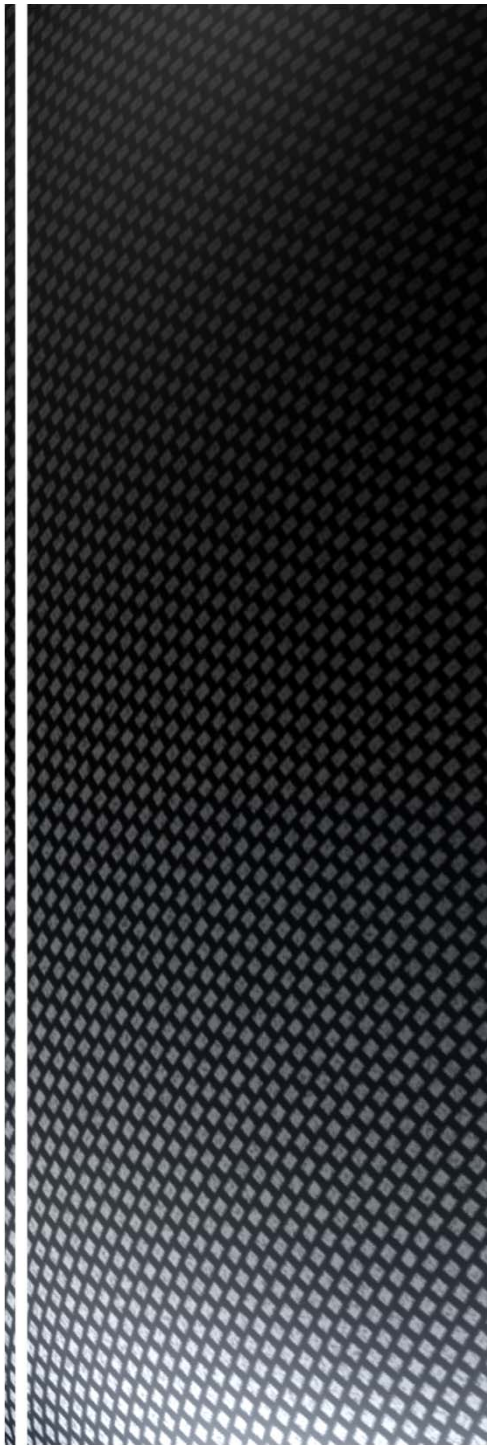
OS-25a

NOSQLデータベースでの機密データの管理と暗号化

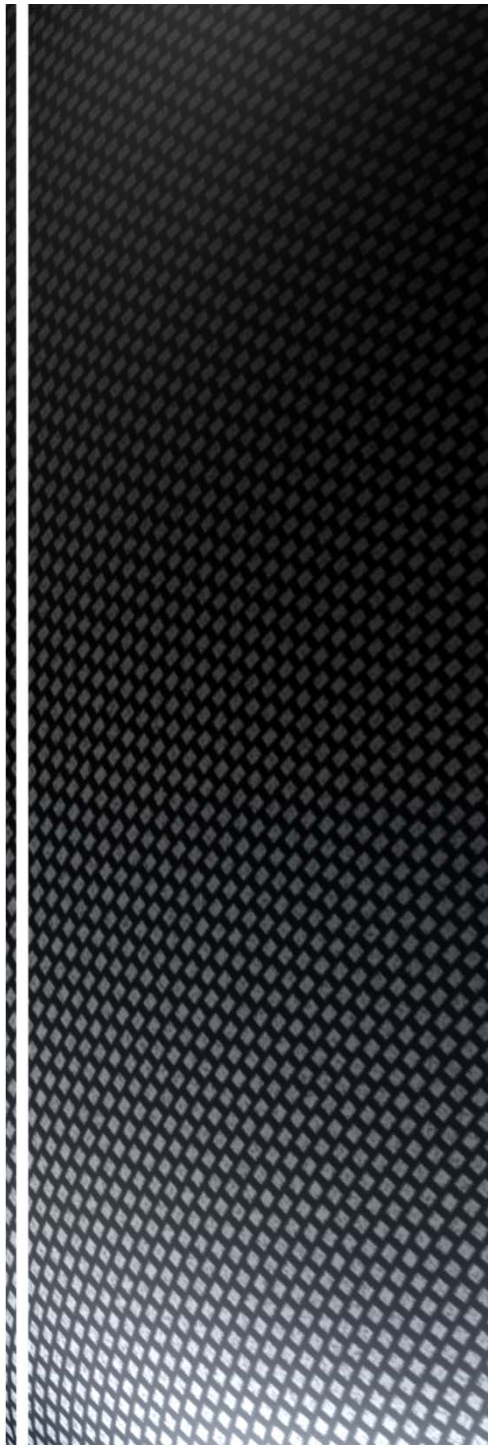


(株)神戸デジタル・ラボ　先端技術開発事業部
岩瀬　高博

アジェンダ

- 生み出されるデータについて
- NOSQLデータベース
- NOSQLデータベース「okuyama」について
- okuyamaによる暗号化実施の検証
- まとめ

生み出されるデータについて



昨今のデータを取り巻く環境

- 昨今ビッグデータというキーワードの台頭にもあるようにデータ(情報)へ注目が高まっている。

1. 企業(公共)活動によりつくられるデータ
2. インターネットメディアに生成されるデータ
3. 個人の活動によりつくられるデータ

昨今のデータを取り巻く環境

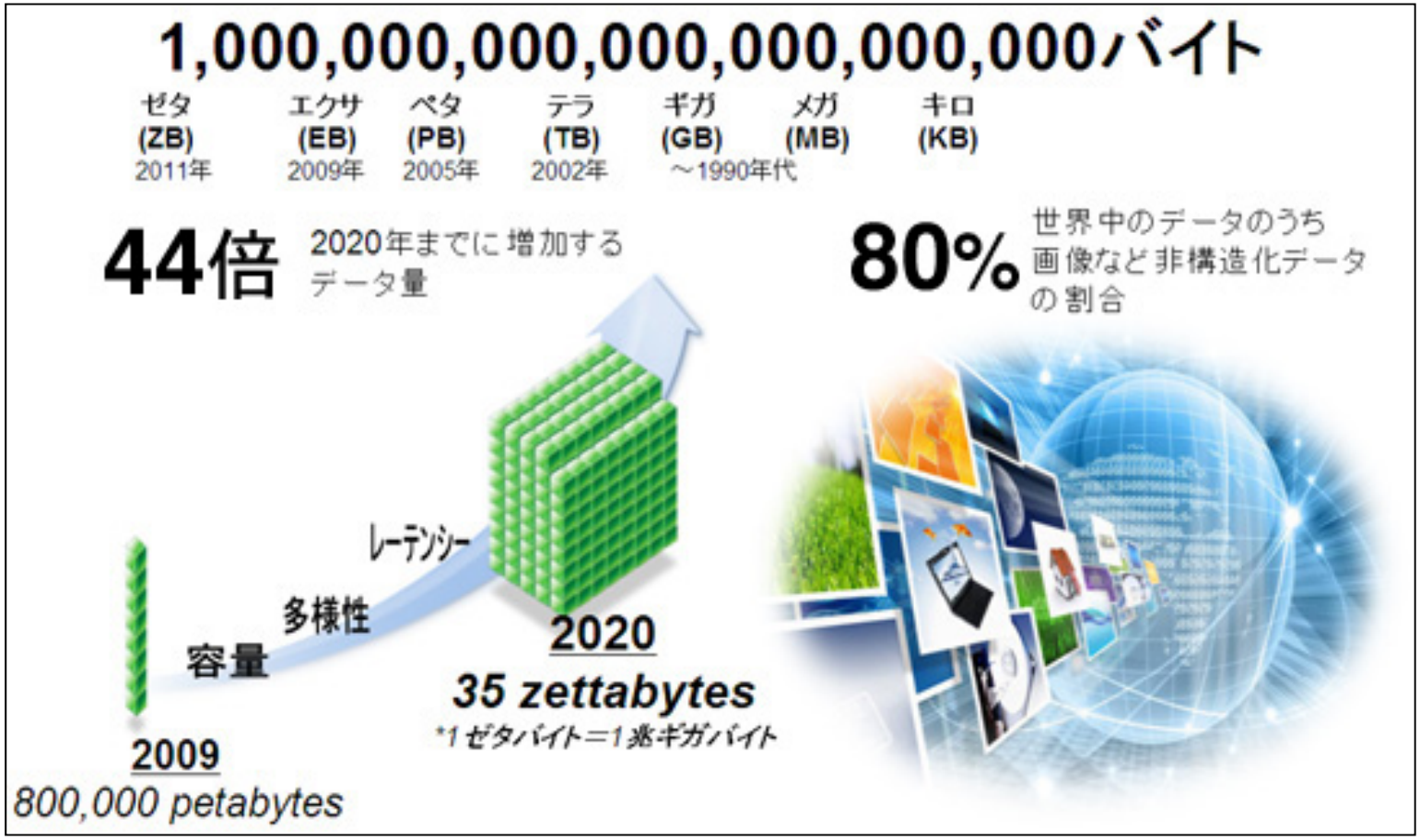
- どのような場所にデータは保管されているのか？

1.企業内サーバなどクローズドな環境
- 業務的情報
- 顧客情報
- 社内情報

2.インターネット等のオープンな環境
- 個人のHP、blog情報
- 情報サービス企業による情報(Facebook等)

昨今のデータを取り巻く環境

- どのような増加傾向にあるのか？



引用: http://www-06.ibm.com/software/jp/data/bigdata/

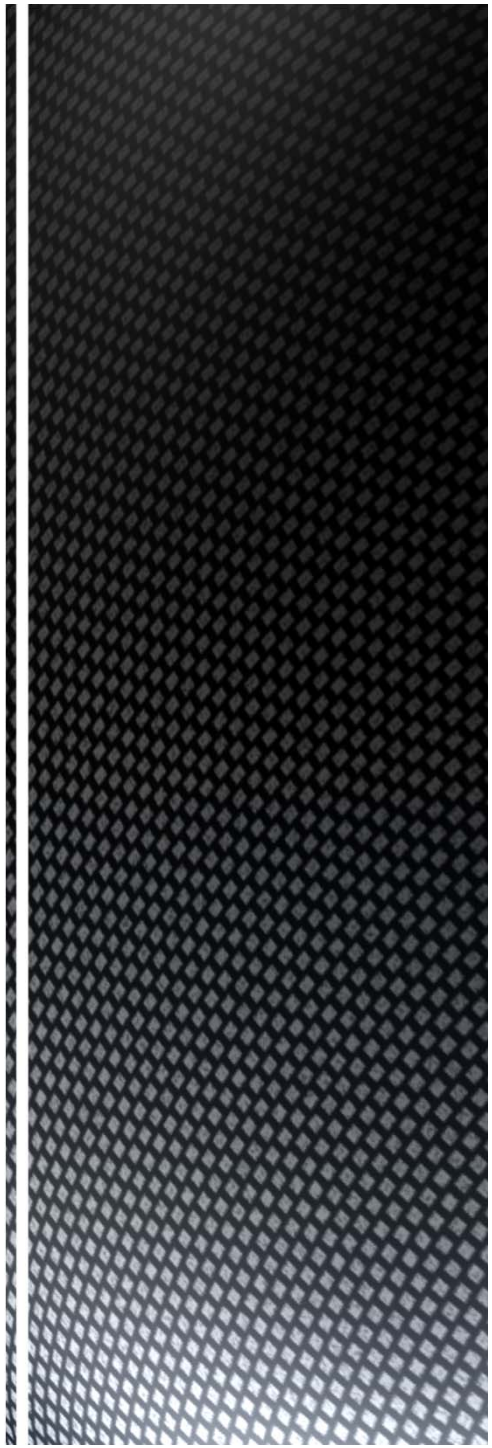
昨今のデータを取り巻く環境

- データの特性に関して

1.センサーデータ等の観測データ
　データ自体は数値等の羅列であり、
　意味を持たないデータである。

2.個人や企業等の機密データ
　データそのものに意味を持ち、管理に
　高いセキュリティーレベルを要求される。

NOSQLデータベース



データベースソフトウェア

- NOSQLデータベース

機能的特徴

1.非定型データの管理
　>柔軟な型管理によりあらゆるデータを格納

2.単一データの操作時の高速性
　>シンプルな操作方式により高速な参照、更新

3.高い水平スケーラビリティ
　>複数の計算機資源を統合管理

4.高い堅牢性
　>1データに対して複数個のコピーを作成

NOSQLデータベース

- NOSQLデータベース登場の背景

昨今ビッグデータに代表されるような非定型データの登場や、データ量の大規模化などにともない、注目され始めたデータベース

機能的にシンプルにすることで特定の用途での性能を追求したデータベース

データベースソフトウェア

- 様々なNOSQLデータベース
  取り扱うデータ構造から以下に大別される

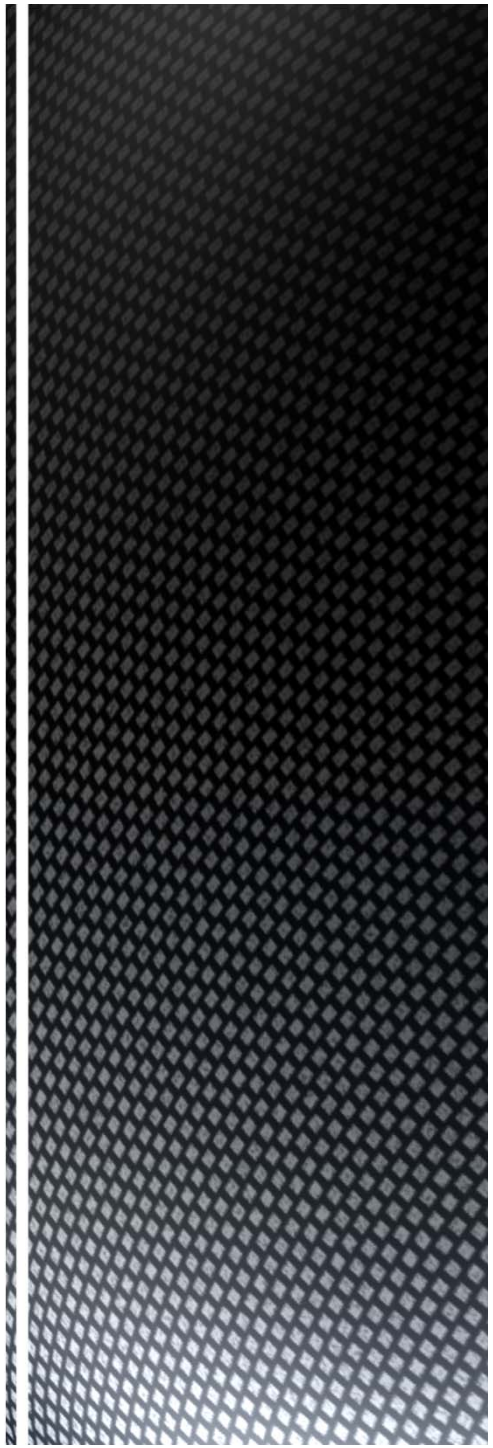
1.Key-Value Store
  >Key=Value関係のデータを管理

2.Column Database
  >カラム単位でのデータ管理

3.Document Database
  >非構造テキストデータを管理

4.Graph Database
  >グラフ構造データを管理

NOSQLデータベース「okuyama」について



okuyamaについて

- okuyamaについて

1.NOSQLデータベース

・分散Key-Value Storeとして、2010年1月にオープンソースとして公開。
2011年9月には商用サポートを株式会社神戸デジタル・ラボにて開始。

・商用版があることもあり、企業での業務利用も行われている。現在、eコマースサイト、インフラ事業社向けプラットフォームなど、多様な場所で利用されている。

okuyamaについて

- okuyamaについて

  以下のような機能をもつデータベースである

  1.Key-Value型データ構造

  >Key-Value型でのデータ構造をもち、Tagなどの独自構造も併せ持ったデータ構造

  2.ストレージ構造

  >完全メモリ、ディスク+メモリ、完全ディスクなど多様なストレージ特性を持つ

okuyamaについて

- okuyamaについて

  以下のような機能をもつデータベースである

  3.分散環境に対応

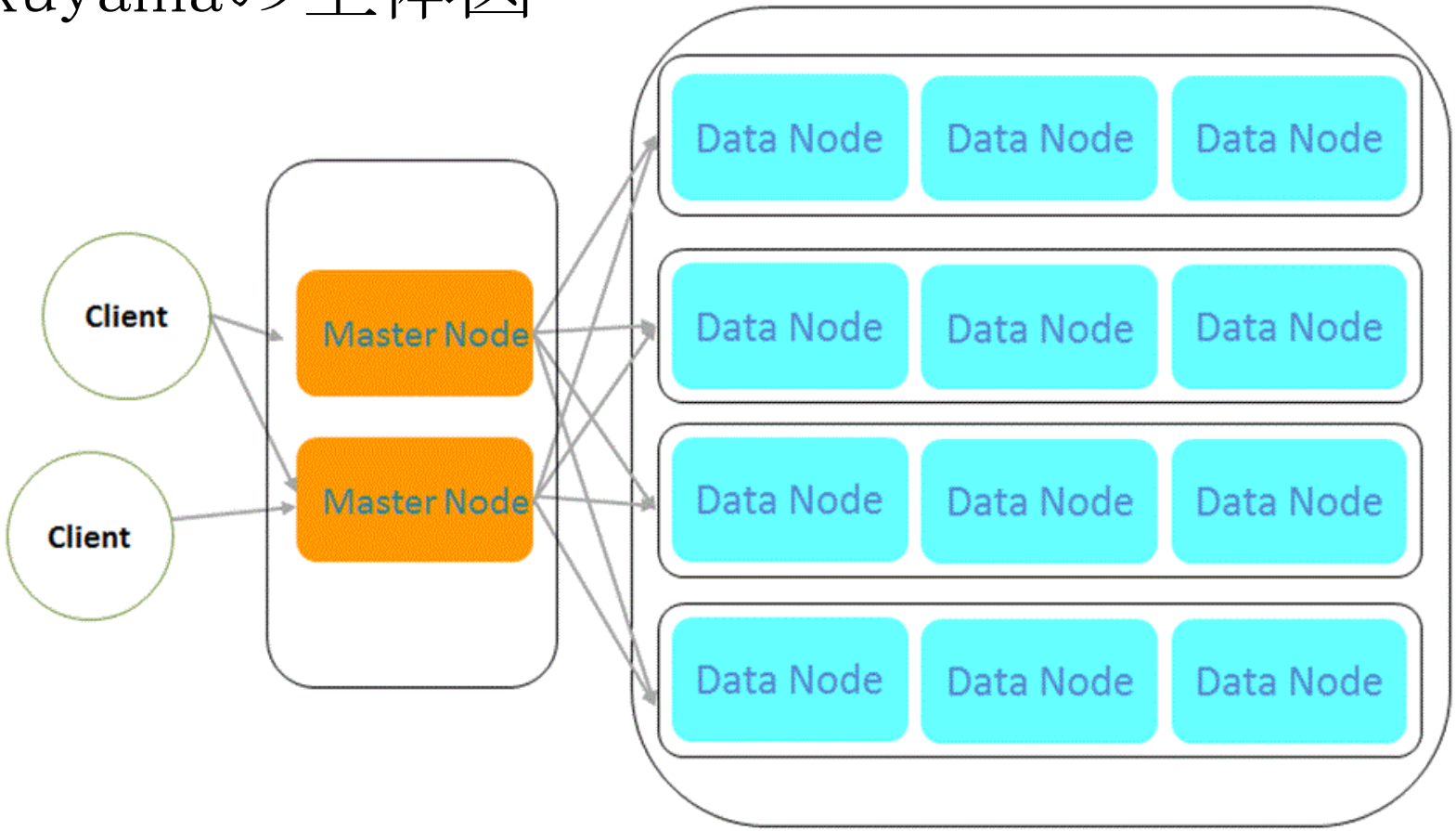
  >複数台の計算機を束ねて1つのストレージとして構築することが可能。
  また、全ての構成要素は多重化されるため、冗長化された環境を特別なハード、ソフトを利用せずに構築可能。

  4.クロスプラットフォーム

  >Java言語にて構築されているため、多様な環境での動作が可能

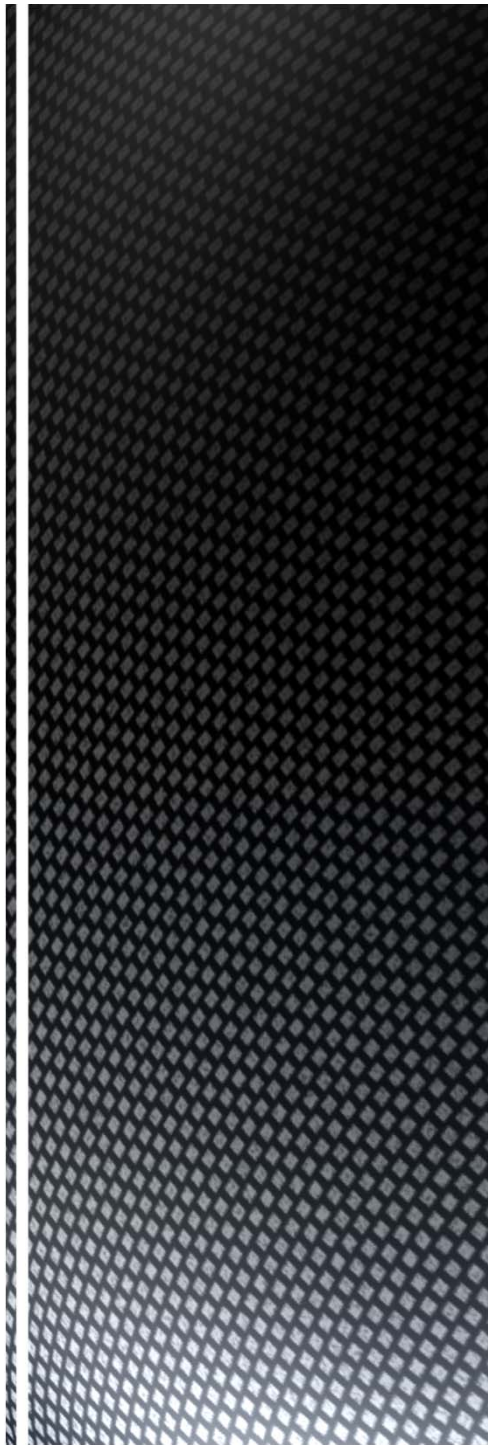
okuyamaについて

- okuyamaの全体図



・クライアント ⟶ マスターノード ⟶ データノード(×3)

okuyamaによる
暗号化実施の検証



機密データを扱うにあたって

- NOSQLデータベース上での機密データ管理

  - 性能を重視する上でokuyamaではデータ暗号化や通信経路暗号化などの機能は非搭載
  - その他のNOSQLに関してもこれらの機能を搭載していないことが多い
  - ビッグデータに含まれる機密データの増加

機密データを扱うにあたって

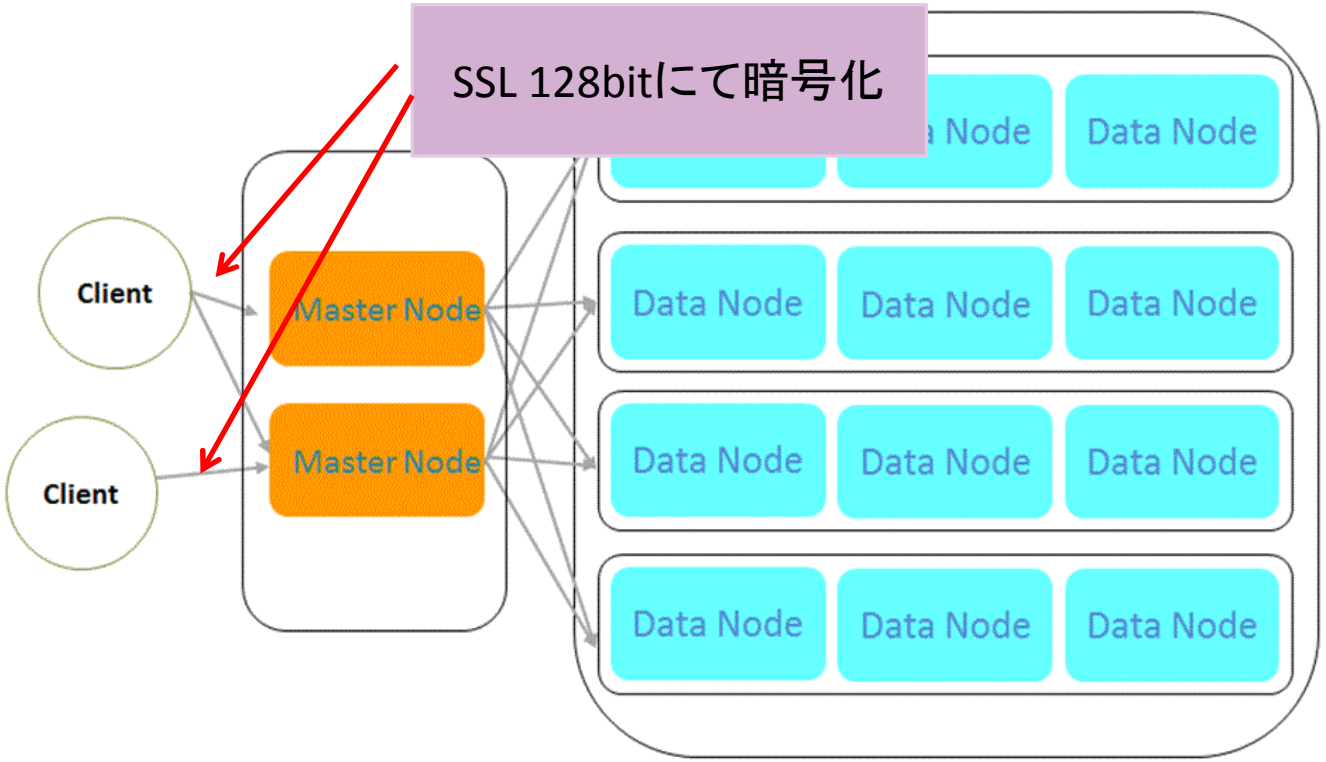
- NOSQLデータベースへの機能追加

・NOSQLデータベースを用いて機密データを管理する為にセキュア機能の基礎を追加

・追加後の性能を測定することで実利用に耐えうるものか検証

機密データを扱うにあたって

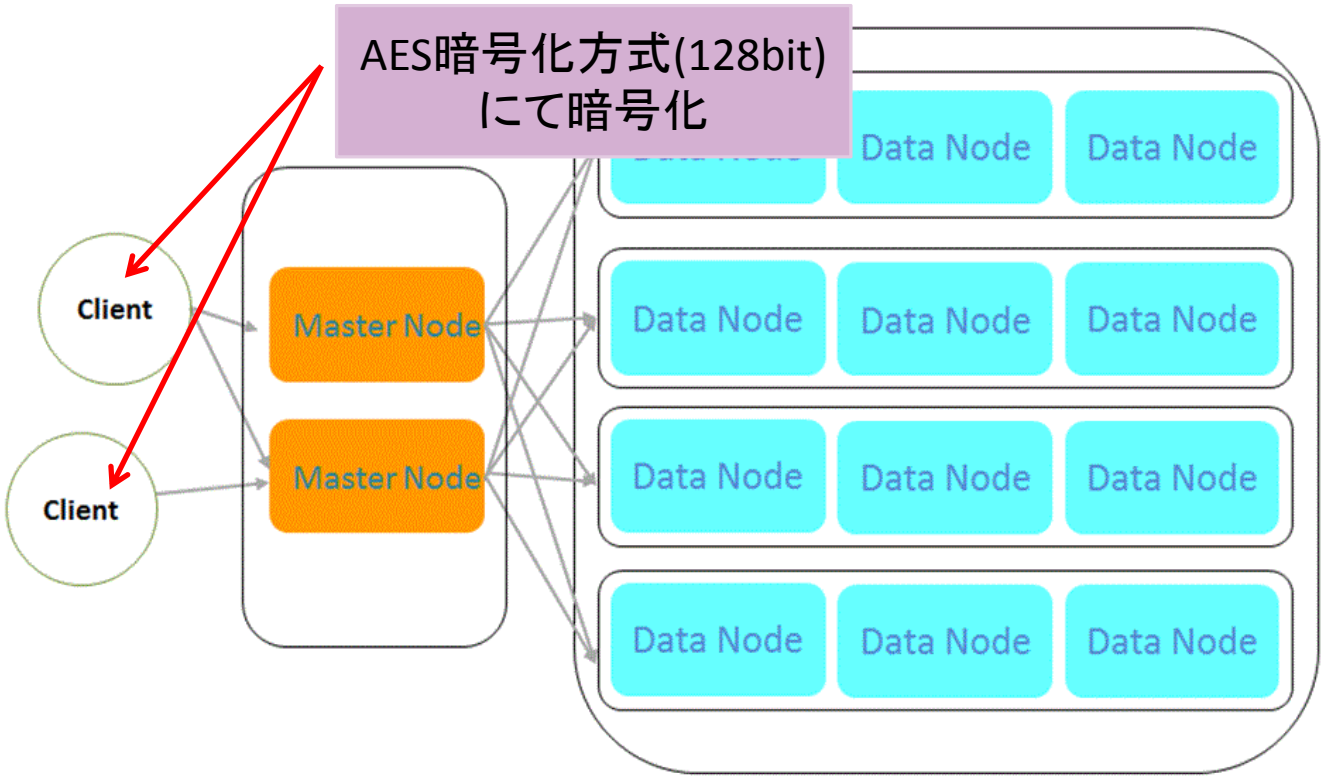
- 暗号化箇所
  - ・通信経路の暗号化



・クライアント ⟶ マスターノード ⟶ データノード(×3)

機密データを扱うにあたって

- 暗号化箇所
  - データ自体の暗号化



・クライアント ⟶ マスターノード ⟶ データノード(×3)

実験環境について

➢ 実験環境

計算機スペック

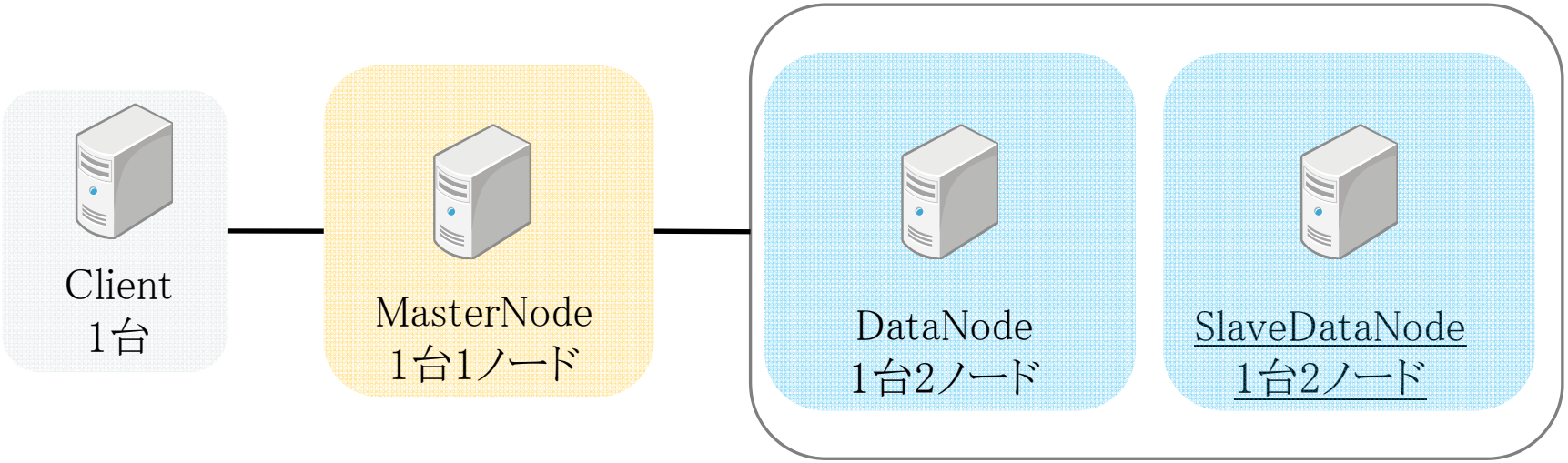
CPU	Memory	Disk	NIC	OS	Java
Intel(R) Core(TM) i5-2400 CPU @ 3.10GHz	4GB	200GB (7200rpm)	1Gbps	CentOS 64bit	Oracle JDK1.7(64bit)

okuyamaスペック

Version	役割	実行サーバ台数	Process数	割り当てメモリ
OSSバージョン 0.9.5ベース	MasterNode	1台	1サーバ 1Process	3GB
OSSバージョン 0.9.5ベース	DataNode	1台	1サーバ 1Process	3GB
OSSバージョン 0.9.5ベース	Slave DataNode	1台	1サーバ 1Process	3GB
OSSバージョン 0.9.5ベース	負荷クライアント	1台	1サーバ 1Process (10Thread)	2GB

実験環境について

➢ 実験環境



◆ okuyamaのストレージ設定はKey=メモリ(圧縮なし)、Value=メモリ(圧縮あり)
◆ 全てのデータはコピーが作成される設定となる

実験内容について

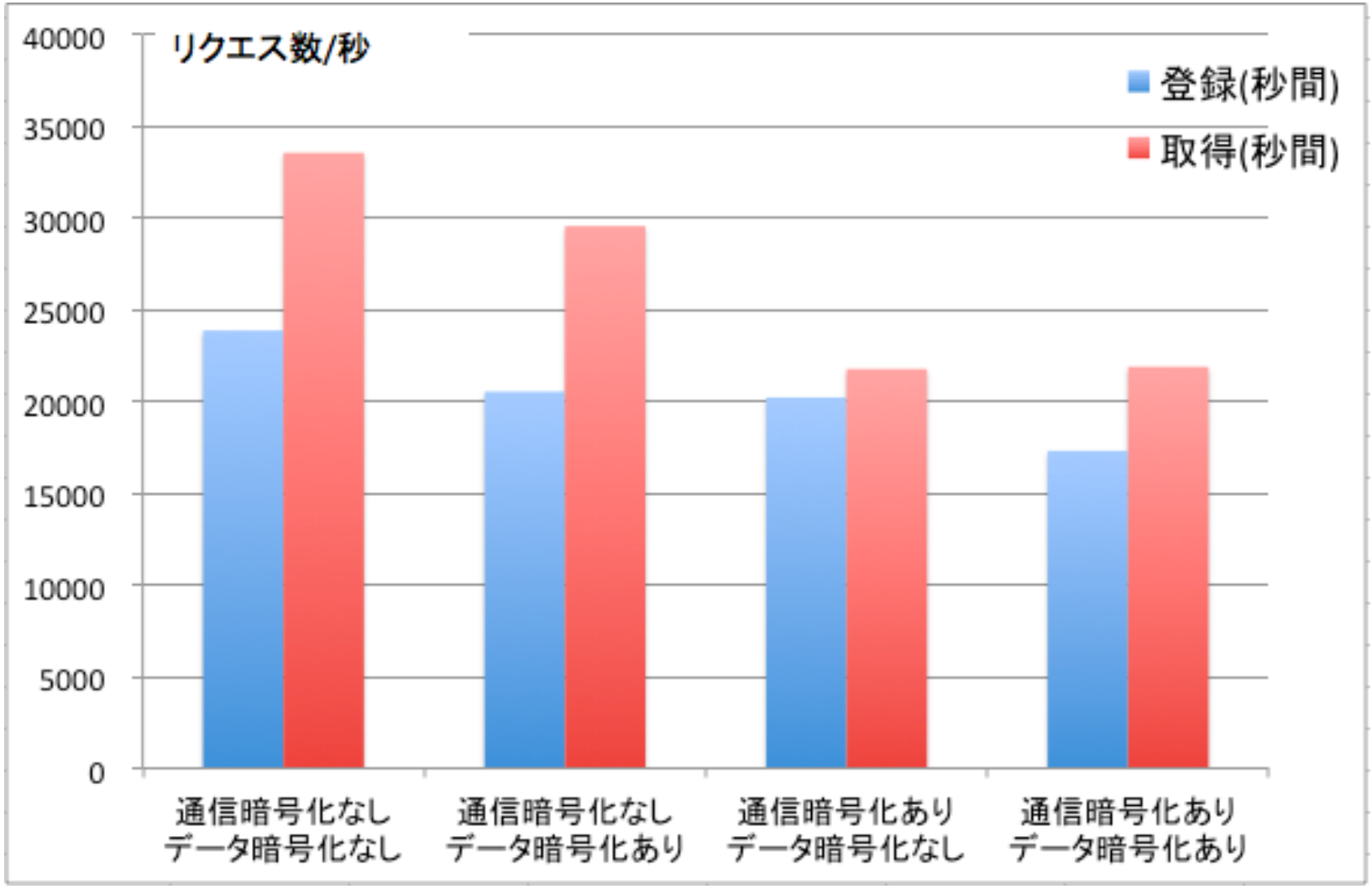
➢ 実験内容

- 1.新規登録テストパターン
  ユニークなデータを10個の独立したクライアントプログラムにより
  同時に1クライアントプログラムあたり10万件のデータ登録を行い、
  登録に要した時間を測定。Keyを20byte、Valueを50byteとした。
- 2.データ取得テストパターン
  10 の独立したクライアントプログラムより同時に取得処理を行う。
  取得対象のデータは「1」のテストパターンにて登録したデータとなり、
  1クライアント当たり10万件のデータを対象とする。
  全てのデータを取得するまでの時間を測定する。

上記テストパターンを以下の暗号化実施状態にて行う

	通信経路暗号化	データ暗号化
テストパターン1	無	無
テストパターン2	無	有
テストパターン3	有	無
テストパターン4	有	有

実験結果

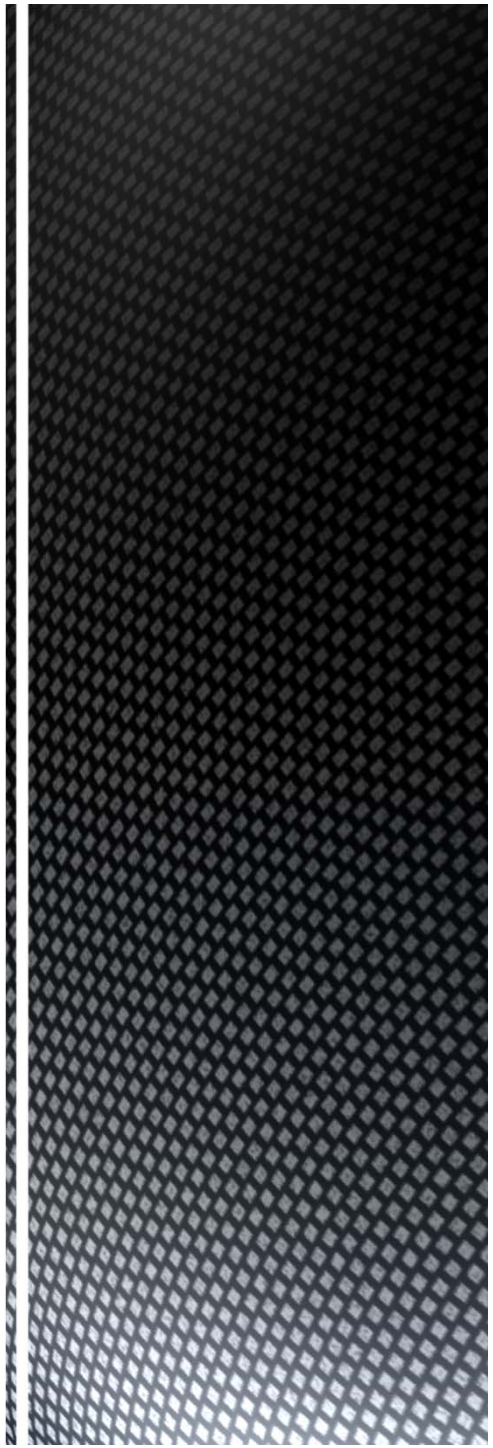

リクエス数/秒
登録(秒間)
取得(秒間)
40000
35000
30000
25000
20000
15000
10000
5000
0
通信暗号化なし
データ暗号化なし
通信暗号化なし
データ暗号化あり
通信暗号化あり
データ暗号化なし
通信暗号化あり
データ暗号化あり

結果考察

	登録処理数(秒間)	取得処理数(秒間)
テストパターン1 (通信経路:無、データ:無)	23,877	33,568
テストパターン2 (通信経路:無、データ:有)	20,539	29,570
テストパターン3 (通信経路:有、データ:無)	20,223	21,779
テストパターン4 (通信経路:有、データ:有)	17,313	21,899

- 登録処理に関しては通常時に比べデータ暗号化時で88%、経路暗号化時で64%の処理能力となった。
- 取得処理に関しては、データ、通信経路の両方の暗号化時において処理能力の低下は抑えられており、通常時の80%以上の性能が出せていることがわかった。
- 両方の暗号化を行なった場合の処理能力は経路暗号化のみの場合よりもわずかに劣る程度であり、このことからデータ暗号化の負荷は低いものと考えられる。

まとめ



まとめ

- まとめと今後の発展性について

  - まとめ
    本実験により暗号化をおこなったことによる
    性能への影響を把握することができた。
    両暗号化を行なった場合でも60%以上の性能を
    維持できることが分かった。また登録、取得
    両方において秒間15,000回以上の処理を確認
    できた。

  - 今後について
    通信経路暗号化はokuyama側にて実施される
    ため、ノード数(サーバ)増加による性能の変化の
    傾向を測定する。

ありがとうございました