



OpenStack Sahara Road to Kilo

ミランティス・ジャパン

04月10日2015年

www.mirantis.com/jp

Copyright © 2015 Mirantis, Inc. All rights reserved

- Hadoop
 - ご紹介
 - データ
 - 分析
- Sahara
 - ご紹介と歴史
 - ユースケース
 - クラスタ管理
 - エラスチック・データ・プロセッシング
 - アーキテクチャー
 - Kilo ロードマップ
- 導入事例
- デモンストレーション
- Q&A

Hadoop

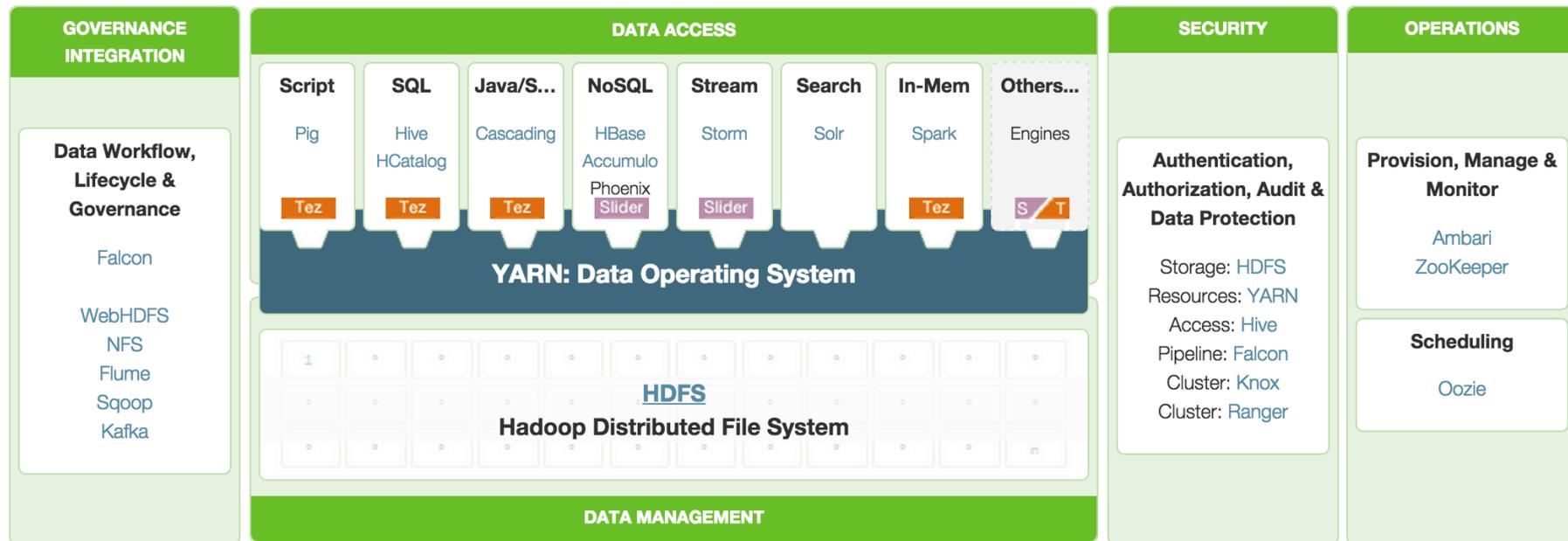
Open-source ecosystem for reliable, scalable, distributed computing

安定性、拡張性的の分散処理のオープンソース・エコシステム

ご紹介: Hadoopとは

- 元々:
 - Yahoo!のプロジェクト
 - Apache財団に入りました
 - Googleの発行したホワイトペーパー
- 現在:
 - プロジェクト・エコシステム
 - Apache財団から支援されている
 - 元のプロジェクトと実現されているエコシステム

ご紹介: Hadoopのアーキテクチャー



© <http://hortonworks.com/hadoop/>

ご紹介: ユースケース

- 大規模データを格納
 - HDFS、 Data Locality、 冗長性..
- ETL (Extraction, Transformation, Load)
 - 抽出し 利用しやすい形に加工、 データベースへの書き出し
- データ分析
 - データを利益に変換する

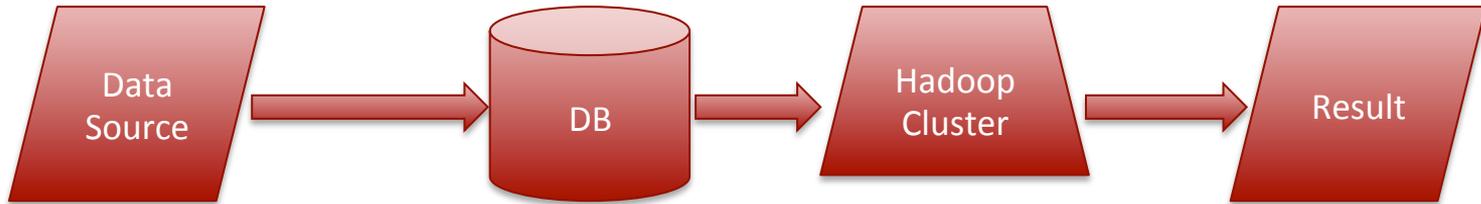
データ: 利用事例

- サーバーのログ
- ソーシャルメディアのデータ
- 患者の履歴と健康情報
- Webサイトからの集積情報
- HVAC(空調系) やセンサー情報
- その他

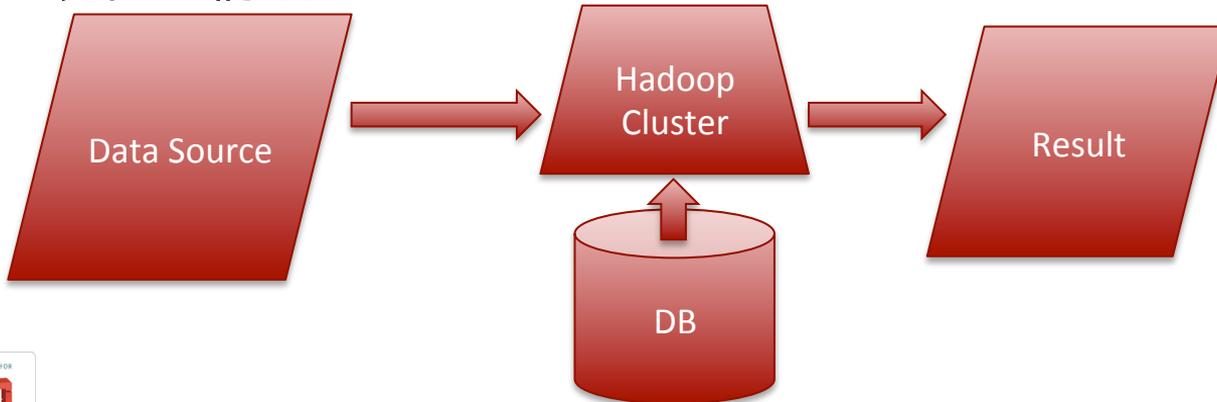
- 構造系/フォーマット vs. 非構造系/非フォーマット
 - IDCやEMCによれば 40 zettabyteデータ量の成長の見込み¹
 - しかし70 -80%が非構造系
- 画一性の無さ
 - Source A differs from Source B
- Dirty Data
 - 非正確、エラーを含んだデータ

¹ http://ikt.hia.no/perrep/eip_ind.pdf

- 元々の構成



- 現在の構成



<https://hadoop.apache.org/>

Copyright © 2015 Mirantis, Inc. All rights reserved

アナリシス: さらなる課題

- 費用的な問題
 - ハードウェア (Computation)
 - リソース (Network, Storage)
 - 人材
- データ解析における反復性
 - Ask the right question

Sahara

OpenStack Big Data as a Service

クラスターの管理やデータの解析ツールを
より**シンプル**にし、**機動性**と
ユーザの**利便性**を高めるため

*To make cluster management and
data analysis tools **simpler**,
agile and **accessible** to users.*

Sahara: 歴史

- Savanna (0.3まで)というコードネームで開始
- ミランティスとHortonworksとRedHatが2013年(Portland)に開始
- 現在はミランティスがプロジェクトをリード (PTL Sergey Lukjanov)
- Juno (2014)でOpenStackに編入
- 多くのコミュニティメンバーが貢献中！



<https://launchpad.net/sahara>

Sahara: 基本構成

- イメージ
 - Sahara用のGlance VMテンプレート
- プラグイン
 - Distribution and version drivers
- テンプレート
 - Node Groups: Master nodes, Computation nodes..
 - Cluster: 複数のnode Groups
- クラスタ

Sahara: 利用例

- テンプレートを使った再現性
- ライフサイクル
 - 開発、QA、本番
- 自動プロビジョニング
 - ハイブリッド、仮想化、ベアメタル
- バースト状況などの負荷管理
- クラスターの管理
- EDPワークロード



Sahara: クラスターの運用

- Hadoopクラスターの運用とプロビジョニング
 - Hadoopクラスターコンフィギュレーションのテンプレート
 - REST APIを使ったクラスターの開始と運用
 - 手動でのクラスターの拡張・縮小 (ノードの追加・削減)
 - データノードのアンチアフィニティ
 - Swiftの統合
- OpenStackダッシュボード(Horizon)へのユーザインターフェースの統合
- プラグインのシステムが他のHadoopディストリビューション(Vanilla Apache, Hortonworks, Cloudera, Spark)との統合を容易に！

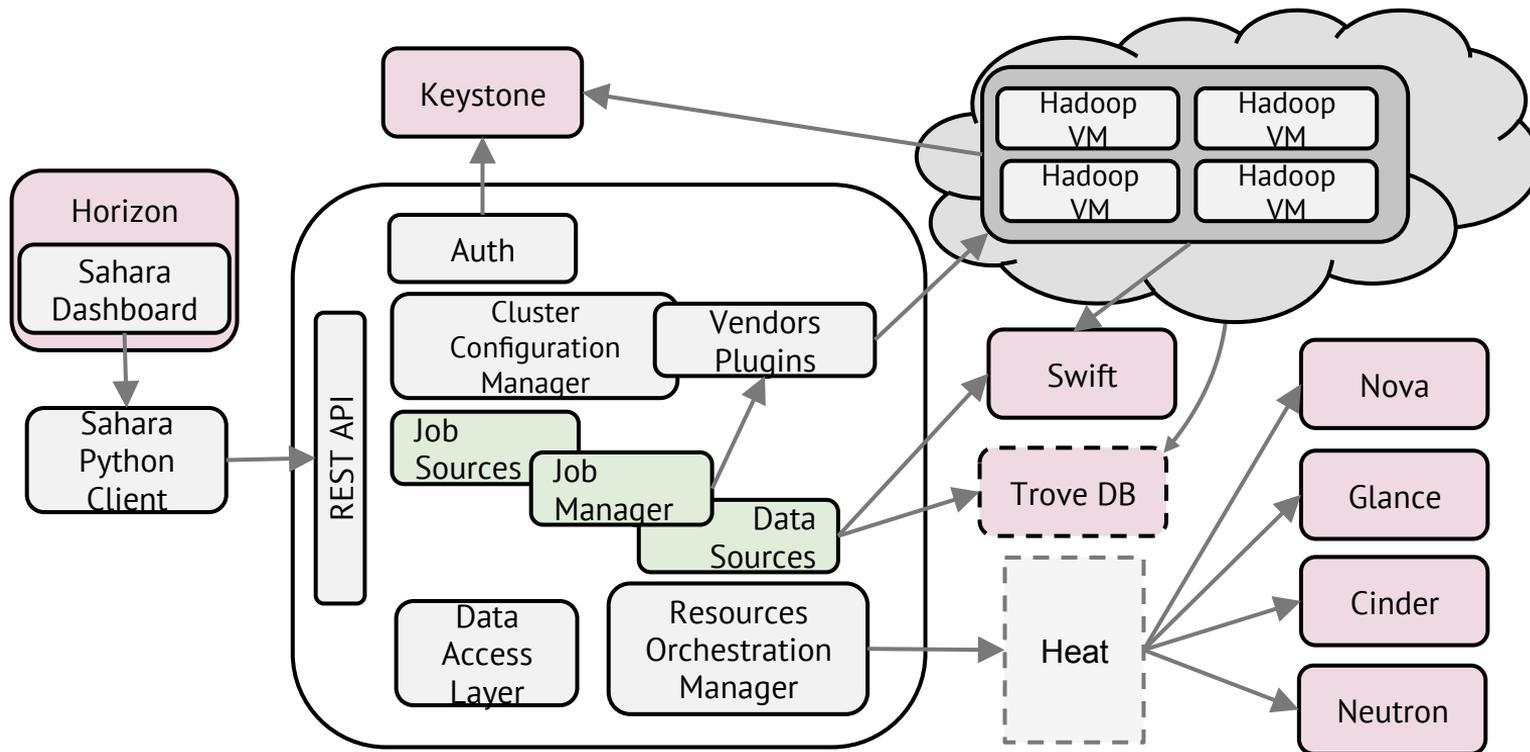
Sahara: Use Case : EDPとは

- EDP (Elastic Data Processing)
 - Saharaによって生成されたクラスター上でのJobの実行環境
- Data Source
 - 入出力元
- Job Binary
 - データ処理のスキ립ト
- Job
 - (パラメータを含んだ)Jobのテンプレート
- Job Execution
 - テンプレートに基づいたJobインスタンス

Sahara: EDP サポート

- インフラ環境へさらされないJob Execution (AWS EMRのような)
- プラグできるワークフローエンジン
 - Oozie, Spark
- Pluggable data sources
 - Swift API, HDFS, *Trove, *Ceph
- サポートされているjob タイプ
 - Jar, Pig, Hive
- 親しみやすいUIがHiveやPigに基づいたアドホックな解析リクエストに対応
- シングルJob用のトランジェントクラスターの生成

Sahara: アーキテクチャー



<https://wiki.openstack.org/wiki/Sahara>

Copyright © 2015 Mirantis, Inc. All rights reserved

Mirantis OpenStack環境でのSahara

- VM上でのHadoop cluster
 - Direct disk mapping increases IO performance (1-2% difference against BM)
- ベアメタルのサポート
 - Kilo/MOS 7.0でリリース、Fuelからの運用管理
- セルフヒーリング(自己回復)、プロビジョニング
 - HEATから導入しSaharaと連携
- データバックアップ、マイグレーション、アップグレード

Sahara: ロードマップ(KILO)

- 新しいディストリビューションのサポート
 - MapR plugin
 - Storm plugin
- Native Ceph サポート
- Ironic の統合(ベアメタルプロビジョニング)
- 分散Saharaエンジンの完成
- OpenStackダッシュボード内のSaharaページのユーザインタフェースの向上

<http://specs.openstack.org/openstack/sahara-specs/>

- お客様:
 - 米国のゲーム会社
 - 年間売り上げ 2,400億円
 - 社員数 5,000人
- 課題は?:
 - CDH (クラウデラ)ライフサイクル管理
 - 利用パターンで、バースト性のある負荷の管理
- どのように実現したか:
 - MOS 5.1 へのCDH pluginのポートし、QAステージも追加



デモンストレーション

MOS 6.0 Juno with Sahara

Q&A

How can we help you succeed with OpenStack?

We are looking for you

- 一緒にOpenStackを広めませんか？
- チームメンバーを募集中！
- info.jp@mirantis.com



- トレーニングサービスもご利用ください！