

Microsoft Big Data и Hadoop

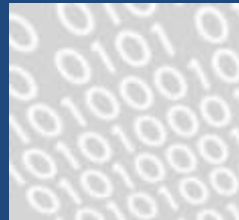
Вячеслав Архаров

Менеджер по развитию бизнеса платформы
приложений Microsoft



Мир меняющихся данных

Взрыв
данных !



10x
прирост
каждые
пять лет

85%
от новых
типов
данных

Volume
Velocity
Variety

Hadoop



Дешевая и
распределенная
обработка и
хранение

Легкая
доступность
данных



Cloud

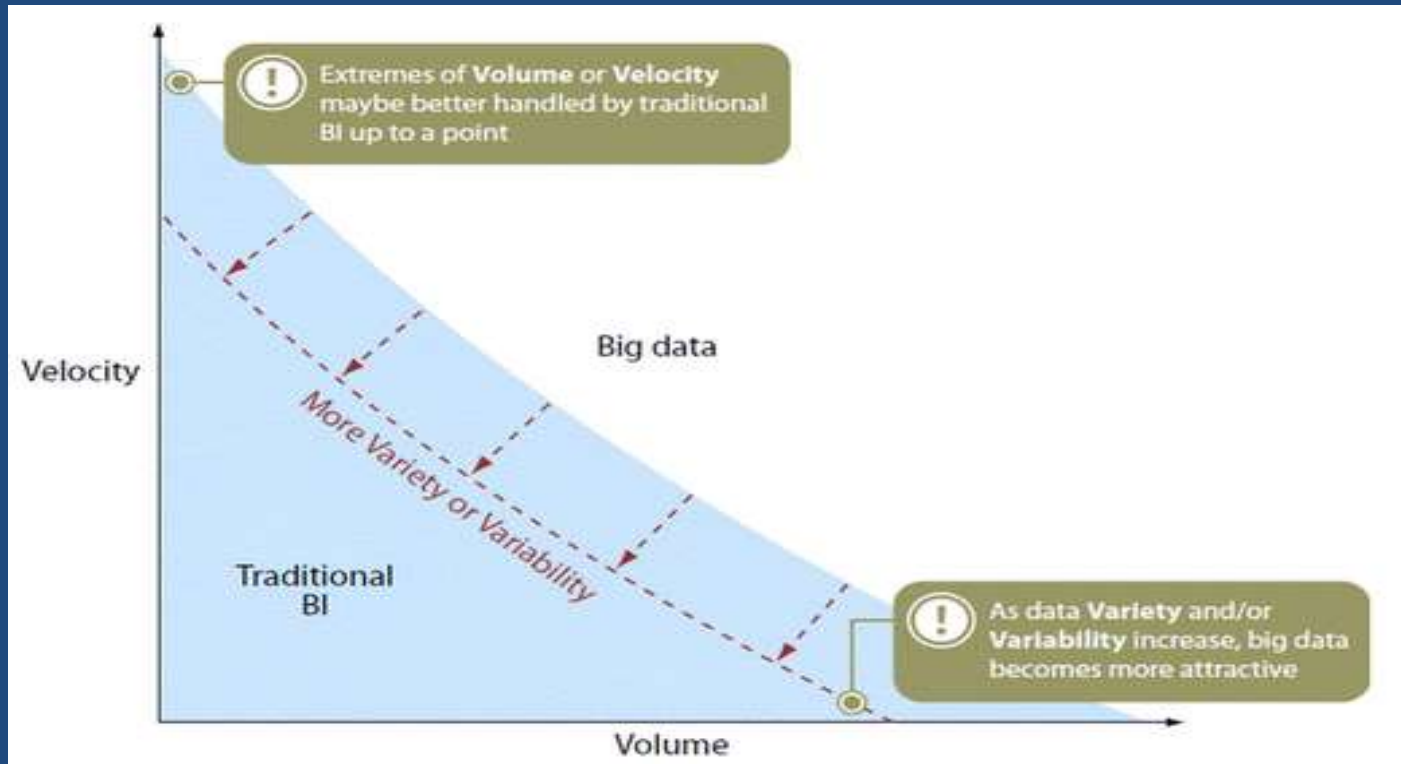
“

К 2015 году компании, которые строят современные системы управления данными, будут финансово опережать своих конкурентов на 20%

– Gartner, Mark Beyer, "Information Management in the 21st Century"

”

Четыре «V» для Big Data



Volume (Объем)

Превышает физические лимиты вертикального масштабирования

Velocity (Скорость)

«Окно принятия решения» мало по сравнению со скоростью изменения данных

Variety (Форматы)

Множество различных форматов данных усложняют интеграцию

Variability (Вариабельность)

Может существовать масса вариантов интерпретации и анализа одних и тех же данных

<http://nosql.mypopescu.com/post/9621746531/a-definition-of-big-data>

Scale Up!



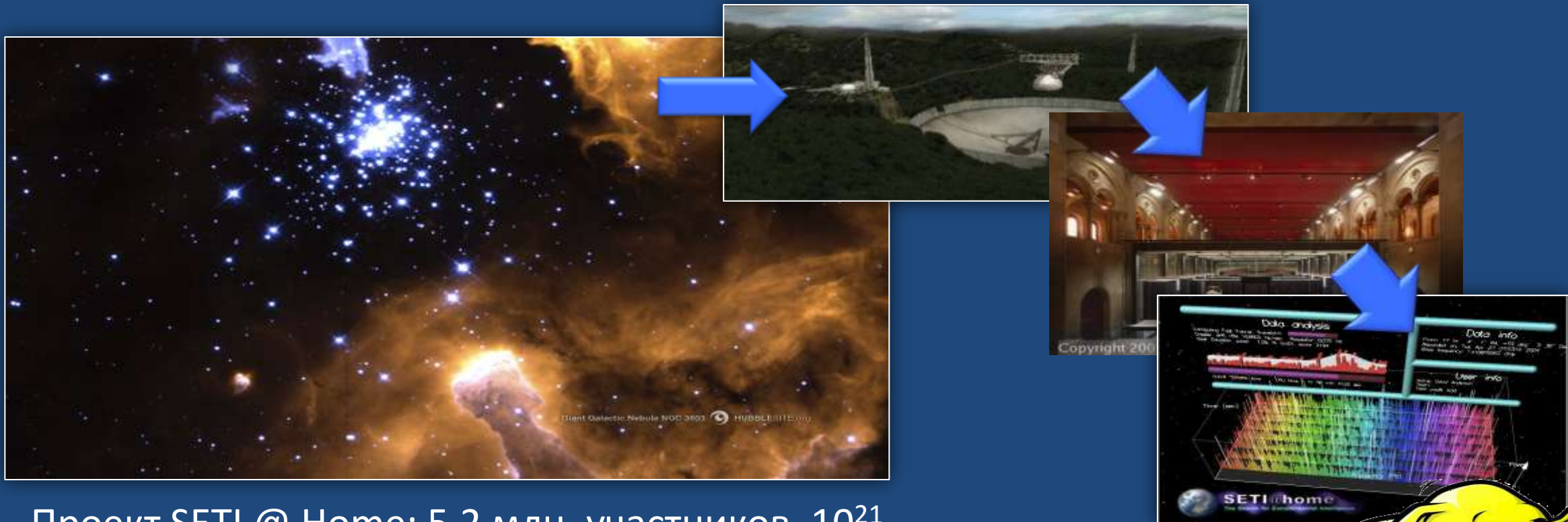
С помощью телескопа Хаббл мы можем получать потрясающие фотографии Вселенной с расстояния 45 млн. Световых лет

Аналогия со scale-up системой:

- Специализированное оборудование
- Заточенное под узкую задачу



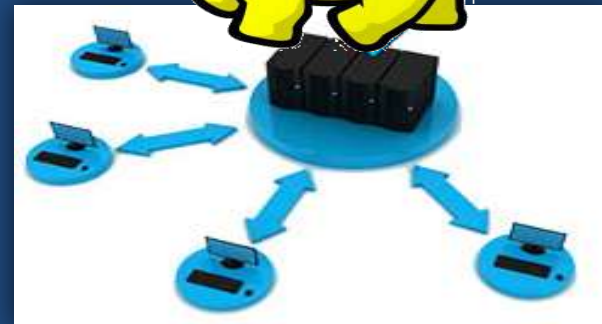
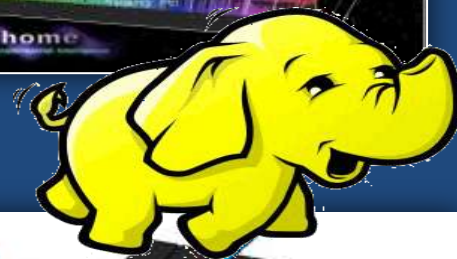
Scale Out | Распределение вычислений



Проект SETI @ Home: 5.2 млн. участников, 10^{21} вычислений с плавающей точкой

Аналогия с распределенной commodity-системой

- Распределенная система с локальными вычислениями
- Работает с сотнями и тысячами узлов
- Множество точек резервирования позволяет избежать проблем





Вінг потребляет > 7 Петабайт в месяц

Сообщество Twitter генерирует более 1 терабайта сообщений ежедневно

Cisco предсказывает к 2013 году годовой поток интернет-траффика достигнет 667 Экзабайт

Практическое применение Big data



Финансы

Оценка рисков
Борьба с отмыванием денег; Анализ трендов;
Прогнозирование



Медиа и интернет

Анализ обращений; веб-аналитика; соц. медиа;
Анализ рекламы; Анализ изображений



Ритейл

Анализ поведения заказчиков; Анализ продаж;



Образование и госсектор

Национальная безопасность; научные исследования



Онлайн-игры

Анти-мошенничество



Фармацевтика

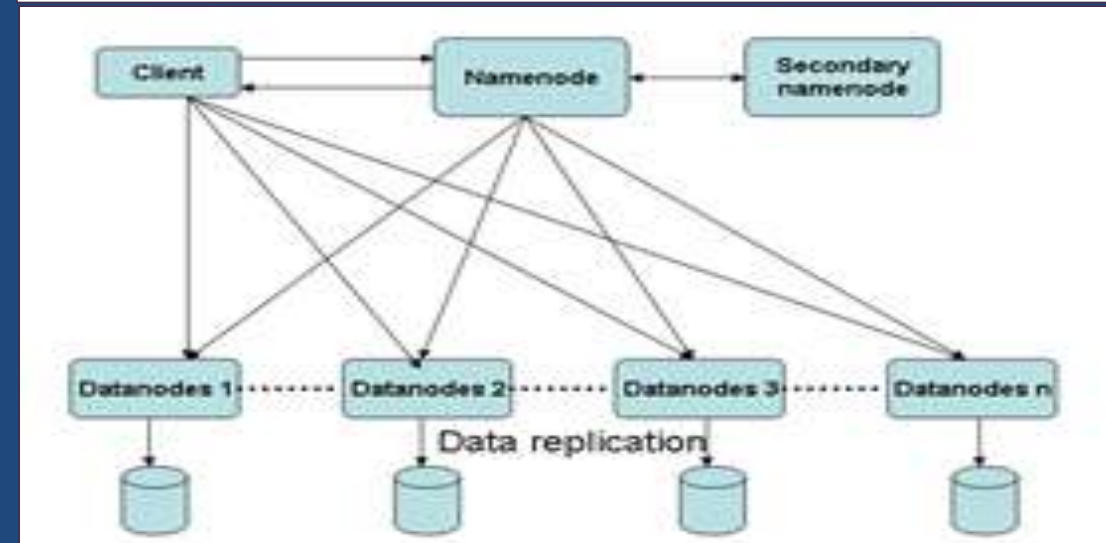
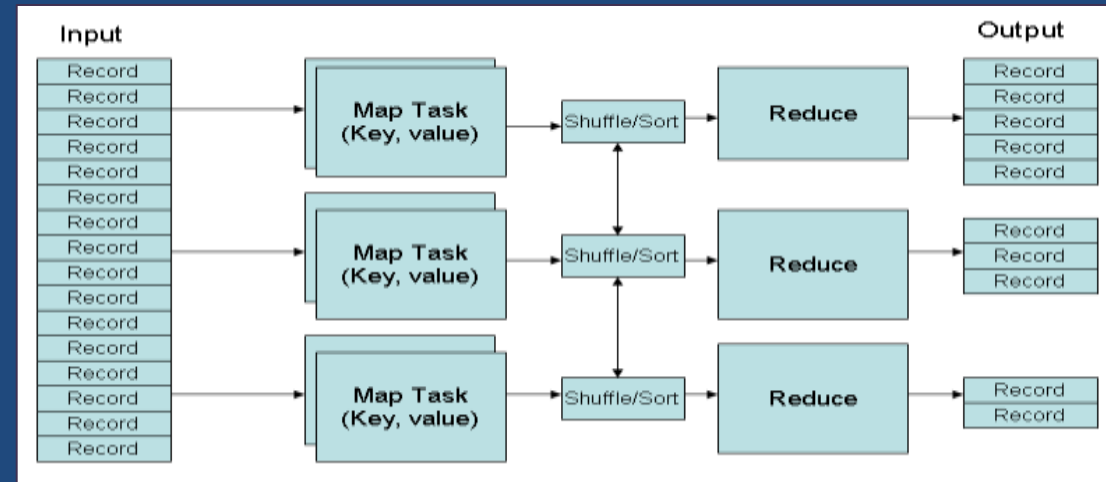
Генетические исследования

Hadoop: наиболее часто встречающийся вариант Big Data

Позволяет анализировать неструктурированные данные, распределенные по кластеру из commodity-оборудования

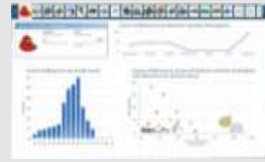
Это инфраструктура, которая **автоматически распределяет и реплицирует данные на множество узлов**, а также **выполняет операции поиска и вывода** по всем этим узлам

- Основан на алгоритме Google's MapReduce и Google File system (GFS)



Решение Microsoft для Big Data

ПРИВЫЧНЫЕ И
УДОБНЫЕ
ИНСТРУМЕНТЫ
АНАЛИЗА



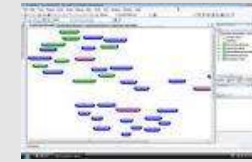
Power View



Excel with PowerPivot



Predictive Analytics



Embedded BI

BI-ПЛАТФОРМА



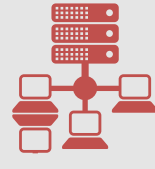
SSAS



SSRS



Нadoop на
Windows Azure



Нadoop на Windows
Server

Коннекторы



Microsoft PDW

НЕСТРУКТУРИРОВАННЫ
Е И
СТРУКТУРИРОВАННЫЕ
ДААННЫЕ



Sensors



Devices



Bots



Crawlers



ERP



CRM



LOB



APPs

МАРКЕТПЛАСЕ

Внешние данные и сервисы

Возможности решения Microsoft для Big Data



SELF-SERVICE

OPERATIONAL

PREDICTIVE

АНАЛИТИКА

MOBILE

REAL-TIME

COLLABORATIVE



ОБРАБОТКА ДАННЫХ

ПОИСК
КЛАССИФИКАЦИЯ

ТРАНСФОРМАЦИЯ
ОЧИСТКА

ОБЩИЙ ДОСТУП
МАСТЕР-ДАННЫЕ

УПРАВЛЕНИЕ ДАННЫМИ



РЕЛЯЦИОННЫЕ



НЕРЕЛЯЦИОННЫЕ



МНОГОМЕРНЫЕ



ПОТОКОВЫЕ

МАРКЕТПЛЕЙС

Внешние данные и
сервисы

Стратегия Microsoft в области Hadoop

Сделать Hadoop Enterprise-решением

Наша собственная реализация HADOOP

- Согласовать изменения с Apache Foundation
- Скачивается бесплатно

Оптимизирован для Windows & Azure

- Интеграция с AD и Systems Center
- Hadoop-as-a-service на Azure

Фокус на .NET разработчиков

- Интеграция с Visual Studio
- Поддержка C#

Обеспечение потребностей крупных компаний

- Производительность и масштабирование
- Высокая доступность
- Простота использования

Приглашаем вас посетить мероприятие посвященное запуску SQL Server 2012 в России !



Дата: 3 апреля 2012 г.

Место:

Отель Редиссон Украина

Начало: 10:00

Краткая программа мероприятия

9:30	-	10:00	Регистрация, приветственный кофе
10:00	-	10:15	Приветственное слово - Eugene Saburi, директор продуктовой группы SQL Server
10:15	-	11:15	Стратегия Microsoft в развитии платформы приложений, Eugene Saburi, директор продуктовой группы SQL Server; Nick King, руководитель группы маркетинга SQL Server 2012
11:15	-	11:30	Перерыв
11:30	-	12:00	Основные тренды в области развития бизнес-систем и платформы данных (Gartner, Accenture)
12:00	-	12:30	Аппаратные решения для построения платформы работы с данными (Hewlett Packard)
12:30	-	13:00	Приложения, необходимые для вашего бизнеса (SAP)
13:00	-	14:00	Обед
14:00	-	14:40	Обзор ключевого функционала SQL Server 2012 (Microsoft)
14:40	-	15:00	Перерыв
15:00	-	16:30	Успешные проекты и решения на базе SQL Server 2012 (компании Связной, Балтика, Digia)
16:30	-	18:00	Завершение мероприятия и развлекательная программа

Пожалуйста заполните анкету и мы с вами свяжемся для приглашения на SQL Server 2012 Symposium !

СПАСИБО ЗА ВНИМАНИЕ !

Microsoft[®]