

علم البيانات والبيانات الضخمة في الإحصاء الرسمي واهداف التنمية المستدامة



ازدهار البلدان كرامة الإنسان



الأمم المتحدة
الإسكوا
ESCWA

ما هي البيانات الكبيرة / الضخمة

لا يوجد تعريف دقيق للبيانات الكبيرة :

- بيانات متعددة الأنواع والمصادر والأحجام
- مجموعة البيانات ذات حجم يفوق قدرة معالجتها باستخدام أدوات قواعد البيانات التقليدية من جميع، ونقل، وتخزين، وإدارة وتحليل تلك البيانات في غضون فترة زمنية مقبولة
- مجموعة أو مجموعات من البيانات التي لها خصائصها (مثل الحجم، السرعة، التنوع، التباين، صحة البيانات... إلخ)، ولا يمكن معالجتها بكفاءة باستخدام التكنولوجيا الحالية والتقليدية لتحقيق الاستفادة منها
- مجموعات البيانات التي تتميز بكونها فائقة حجم أو سرعة أو تنوعاً، بالقياس إلى أنواع مجموعات البيانات المعهودة الاستخدام

خصائص البيانات الكبيرة

حجم	سرعة	تنوع
GB	دورية	قواعد بيانات
TB	قريب من الوقت اللحظي	صور-صفحات الويب- تسجيلات
PB	الوقت اللحظي	الشبكات الاجتماعية-مقاطع الفيديو-الهواتف المتنقلة-البيانات الغير مهيكلة

فرص كبيرة للأجهزة الإحصائية

إحصاءات متاحة في
الوقت المناسب،
وبصور آنية

اعتماد اقل على
موافقة المستهدفين
لتقديم البيانات من
خلال المسوح

عبء اقل على
المستجيب

عدم الاستجابة اقل

زيادة دقة الإحصاءات
(أخطاء العينات اقل)

حماية العناصر
الأساسية
للخصوصية

التفاوت والتحييز
(من يملك ولا
يملك البيانات)



كيف يمكن لعلم البيانات والتحليلات المساهم في التنمية المستدامة

1 القضاء على الفقر
يمكن لأنماط الأنفاق على خدمات الهاتف الخليوي مؤشرات على مستويات الدخل

2 إنهاء الجوع تماما
يمكن للتعهيد الجماعي أو تتبع أسعار المواد الغذائية على الإنترنت أن يساعد في رصد الأمن الغذائي رسداً آتياً

3 الصحة الجيدة والرفاة
يمكن أن يساعد رسم خرائط حركة مستخدمي الهواتف الخليوية على التنبؤ بانتشار الأمراض المعدية

4 التعليم الجيد
يمكن لإبلاغ المواطنين أن يكشف عن أسباب تسرب التلاميذ من الدراسة

5 المساواة بين الجنسين
يمكن أن يكشف تحليل المعاملات المالية أنماط الإنفاق والآثار المختلفة للخدمات الاقتصادية على الرجل والمرأة

6 المياه الصحية والمرافق الصحية
يمكن لأجهزة الاستشعار المتصلة بمضخات المياه تتبع مواضع الوصول إلى المياه النظيفة

7 طاقة نظيفة وبأسعار معقولة
يسمح القياس الذكي لشركات المرافق بزيادة دقة الكهرباء أو الغاز أو المياه للحد أو تقييدها للحد من النفايات وضمان إمدادات كافية في فترات الذروة

8 العمل اللائق والنمو الاقتصادي
يمكن أن تقدم أنماط الحركة البريدية العالمية مؤشرات من مثل النمو الاقتصادي والتحويلات المالية والتجارة والناج المحلي الإجمالي

9 الصناعة والابتكار والهياكل الأساسية
يمكن استخدام البيانات المستمدة من أجهزة النظام العالمي لتحديد المواقع (جي بي أس) في مراقبة حركة المرور وتحسين وسائل النقل العام

10 الحد من التباين
يمكن للتحليلات المتعلقة بالانتقال من المقروء إلى المكتوب (في المحتوى الإذاعي المحلي) الكشف عن شواغل متصلة بالتميز، فضلاً عن قدرتها على دعم سياسات الاستجابة

11 مدن ومجتمعات محلية مستدامة
يمكن للاستشعار الساتلي عن بعد تتبع التعدي على الأراضي والمساحات العامة من مثل المتنزهات والغابات

12 الاستهلاك والإنتاج المسؤولين
يمكن أن تكشف أنماط البحث في الإنترنت أو معاملات التجارة الإلكترونية عن وتيرة الانتقال إلى المنتجات الموفرة للطاقة

13 العمل المناخي
يمكن بالجمع بين صور الأقمار الصناعية وشهادات العيان والبيانات المفتوحة تقديم المساعدة في تتبع إزالة الغابات

14 الحياة تحت الماء
يمكن لبيانات تتبع السفن البحرية أن تكشف عن أنشطة الصيد غير المشروعة وغير المنظمة وغير المبلغ عنها

15 الحياة في البر
يمكن لمراقبة وسائل التواصل الاجتماعي دعم إدارة الكوارث بمعلومات فورية عن موقع الضحايا، وأثار وقوة حرائق الغابات أو الضباب

16 السلام والعدل والمؤسسات القوية
ويمكن لتحليل المشاعر في وسائل التواصل الاجتماعي أن يكشف عن الرأي العام بشأن الحوكمة الفعالة أو تقديم الخدمات العامة أو حقوق الإنسان

17 عقد الشراكات لتحقيق الأهداف
يمكن للشراكات — التي تتيح الجمع بين الإحصاءات وبيانات الهواتف الخليوية وبيانات الإنترنت — أن تتيح فهماً فورياً أفضل للعالم شديد الترابط اليوم

الدور الذي تضطلع به الأمم المتحدة

وضع المبادئ والمعايير الموجهة للعمل الجمعي فيما يتصل بالاستخدام المأمون للبيانات الضخمة للتنمية والعمل الإنساني ضمن المجتمع العالمي

إتاحة أكثر انفتاحا وشفافية للبيانات

تجنب التعرض للخصوصيات أو الإساءة لحقوق الإنسان من خلال سوء استخدام البيانات في يتعلق بالأفراد والمجموعات

التقليل من التفاوت في إنتاج البيانات وسبل الحصول عليها واستخدامها

الفعاليات العالمية

القمة العالمية
للذكاء الاصطناعي
من أجل الصالح
العام

منتدى البيانات
العالمي للأمم
المتحدة

للبيانات الضخمة
للإحصاءات
الرسمية

مصادر البيانات الضخمة - تقرير بعنوان "ماذا تعنيه البيانات الضخمة للإحصاءات الرسمية"

المصادر الناشئة
عن إدارة أحد
البرامج

المصادر التجارية
أو ذات الصلة
بالمعاملات

مصادر شبكات
أجهزة
الاستشعار

مصادر أجهزة
التتبع

مصادر البيانات
السلوكية

مصادر البيانات
المتعلقة بالآراء

الطرق القديمة لتخزين البيانات الضخمة

قواعد البيانات العلائقية Relational database

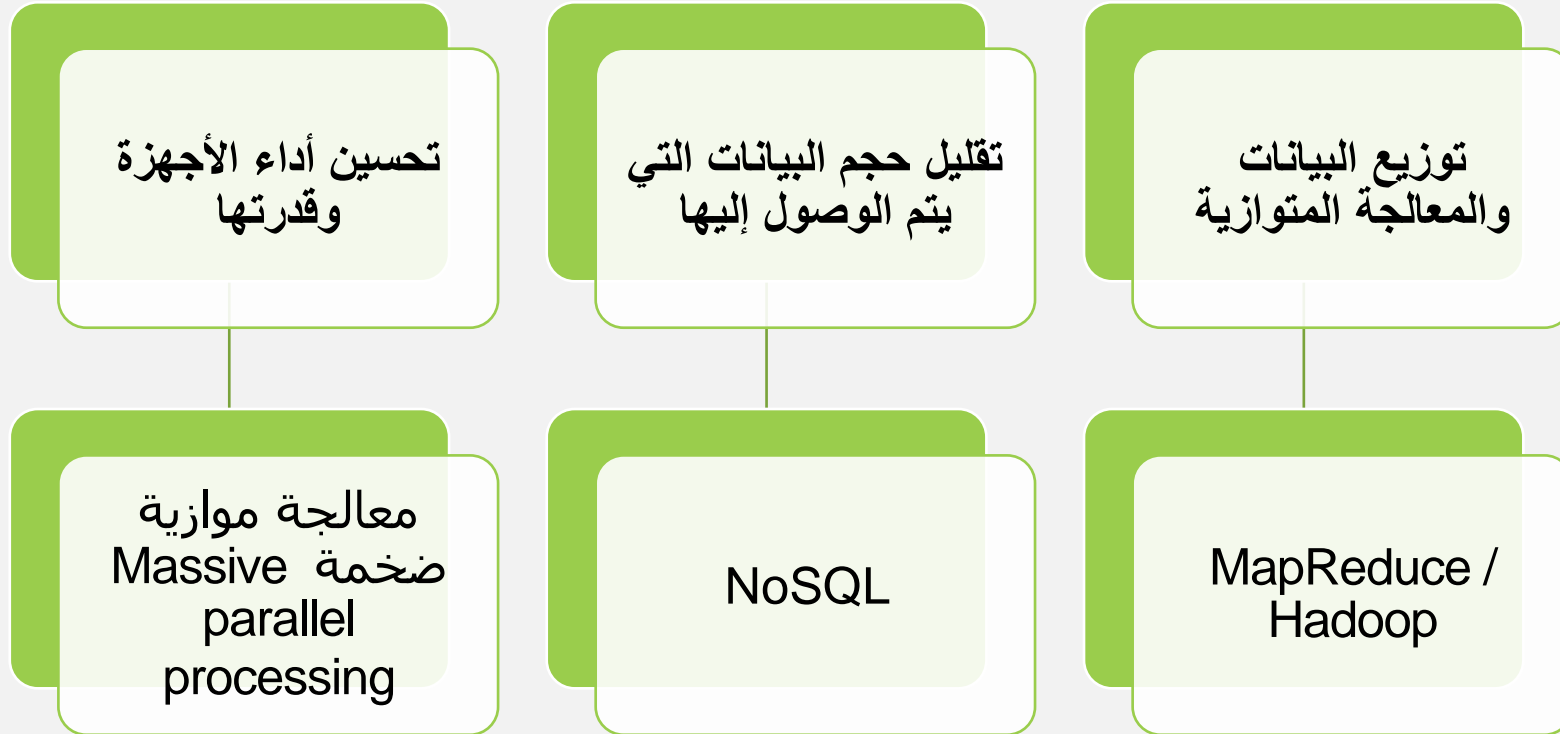
قواعد البيانات الكائنية Object-oriented database

قاعدة البيانات الموزعة Distributed database

قاعدة بيانات NoSQL

قواعد بيانات الرسم البياني
Graph database

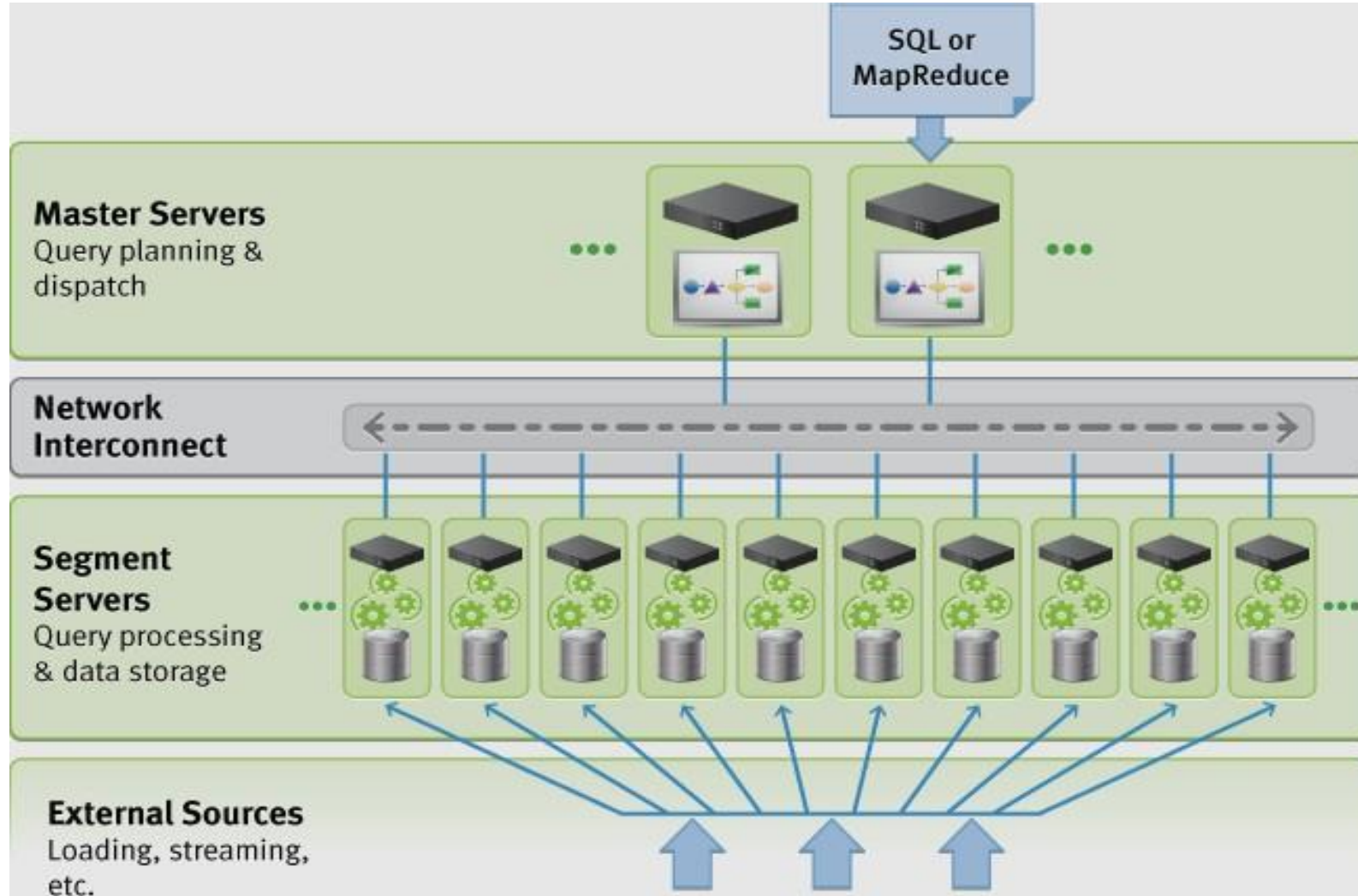
اي تكنولوجيا للبيانات الضخمة؟



تحسين أداء الأجهزة وسعتها

- استخدام وحدات المعالجة المركزية CPU الأسرع
- استخدام المزيد من CPU cores
- زيادة سعة القرص disk ونقل البيانات
- زيادة سرعة نقل البيانات عبر الشبكة

المعالجة المتوازية الضخمة



ماهي قاعدة No بيانات SQL؟

- تخزين البيانات بصيغة تختلف عن الجداول العلاقية
- لا يتم إصلاح مخطط البيانات data schema مسبقًا ولكن يمكن تكييفه بمرونة مع الشكل الفعلي للبيانات.
- المرونة التي تسمح للوصول إلى قابلية التوسع العالية
- تستخدم وحدة تخزين غير منظمة

No SQL

المزايا

- المرونة
- قابلية التوسع
- سرعة في استخراج البيانات الضخمة

العيوب

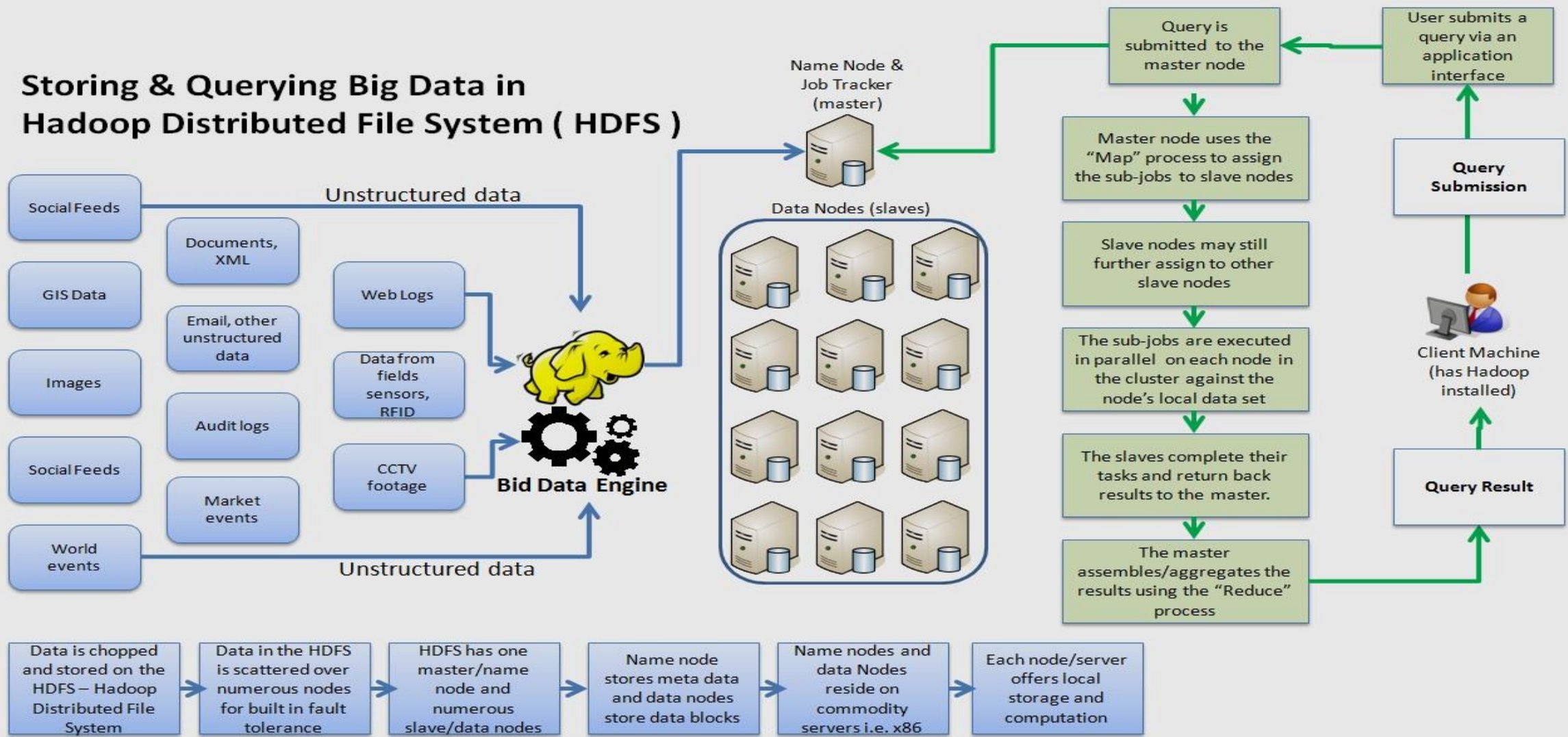
- عدم وجود تناسق قوي
- خصائص ACID غير صالحة
- صعوبة أداء الصلات join بين مجموعات البيانات datasets

ما هو؟

- برنامج او منصة برمجية مفتوح المصدر ، تديره مؤسسة Apache Software Foundation
- يهدف لتخزين ومعالجة البيانات الضخمة بشكل موزع لتسريع نتيجة المعالجة

هادوب Hadoop

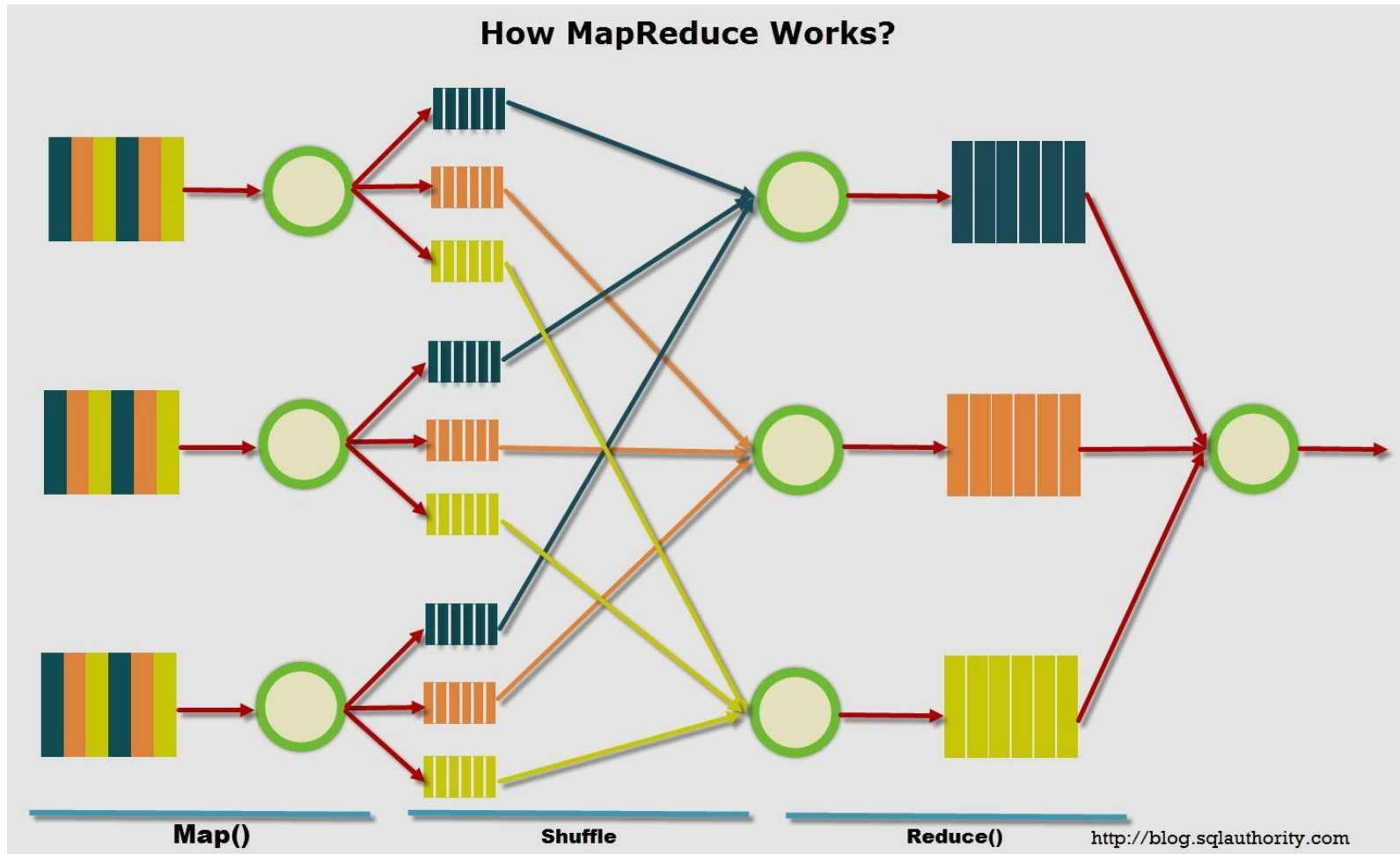
Storing & Querying Big Data in Hadoop Distributed File System (HDFS)



ماهو؟

- نموذج برمجة لمعالجة مجموعات البيانات الكبيرة
- Map التي تأخذ مجموعة من البيانات وتحولها إلى مجموعة أخرى من البيانات
- Reduce تأخذ المخرجات من Map كمدخلات وتجمع مجموعات البيانات هذه في مجموعة أصغر.
- يتم دائمًا إجراء Reduce بعد Map

مابریڈیوس Hadoop MapReduce



فرص جديدة للأجهزة الإحصائية

بيانات الهواتف
المحمولة

رصد الأرض

النظام الآلي
لتحديد هوية
السفن

الماسحات
الضوئية

تقنيات الحفاظ
على الخصوصية

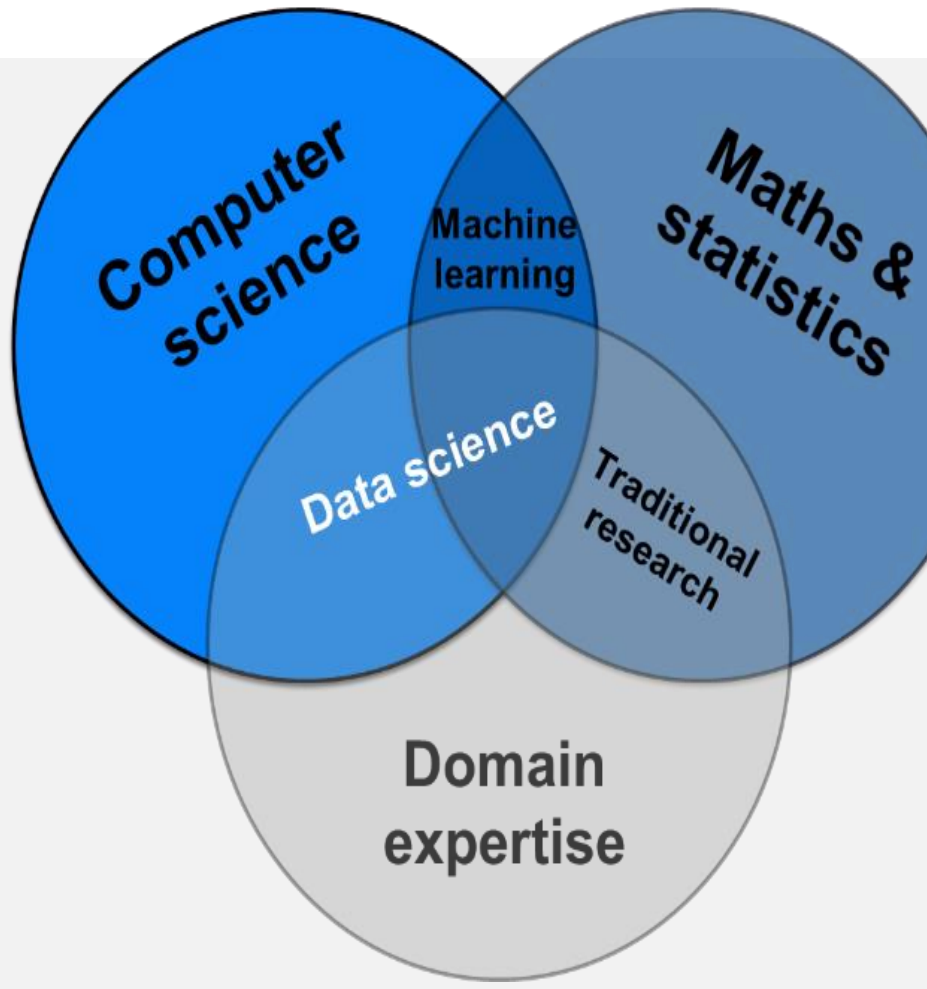
بناء القدرات

البيانات الضخمة
وأهداف التنمية
المستدامة

بناء القدرات

قياس سبل
الوصول إلى
المناطق الريفية

مهارات جديدة ايضا في الاجهزة الإحصائية



	Data acquisition	Data processing	Data analysis	Data visualization
Core competencies	Ethics and privacy	Ethics and privacy	Ethics and privacy	Ethics and privacy
	Data management	Data management	Mathematics	Statistics
	Machine Learning	Mathematics	Statistics	Programming
	Programming	Programming	Programming	Data visualization
Generic skills		Machine Learning	Machine Learning	
	product understanding	curiosity	curiosity	product understanding
	critical thinking	business acumen	adaptability	business acumen
	business acumen	critical thinking	critical thinking	storytelling
	curiosity	communication	communication	communication
	team player	team player	team player	team player
	agile project management	agile project management	agile project management	agile project management

بي إذن مسبق.

كيف واين

بيانات مراقبة الأرض	بيانات الماسح الضوئي	نظم التعرف التلقائي (AIS)	بيانات الهاتف المحمول	الإحصاءات الرسمية
X			X	التعداد السكاني
X			X	الاجتماعية
X			X	الديموغرافية
	X			الأسعار
X		X		البيئة
X	X			الطاقة
X		X	X	النقل
X			X	الهجرة
X	X	X	X	أهداف التنمية المستدامة



ازدهار البلدان كرامة الإنسان



شكرا لكم