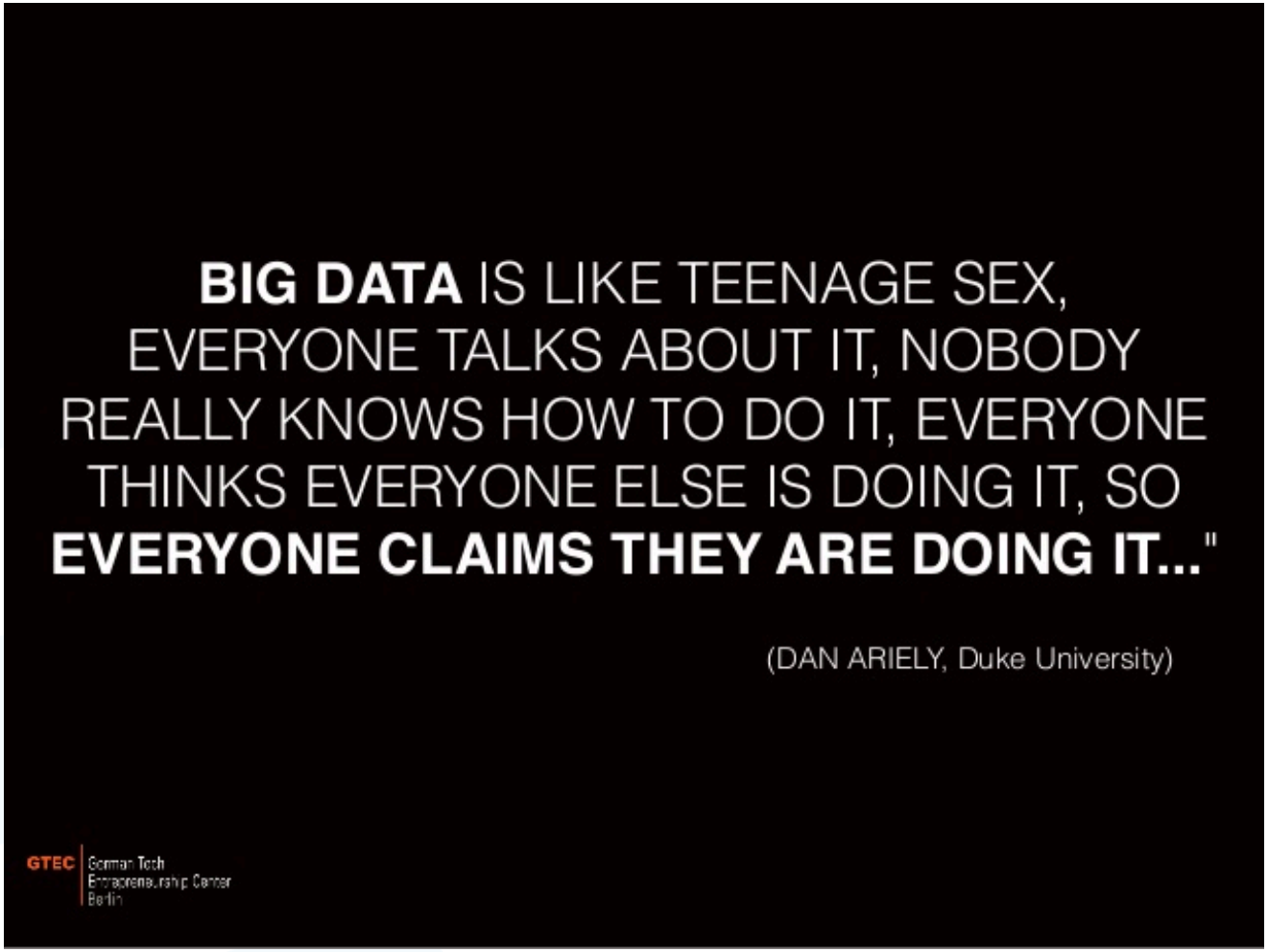
NZRS
an InternetNZ company

Big Data para operadores de red y DNS

GORE 18 – Madrid, Octubre 2016

Sebastian Castro – NZRS


BIG DATA IS LIKE TEENAGE SEX,
EVERYONE TALKS ABOUT IT, NOBODY
REALLY KNOWS HOW TO DO IT, EVERYONE
THINKS EVERYONE ELSE IS DOING IT, SO
EVERYONE CLAIMS THEY ARE DOING IT..."
(DAN ARIELY, Duke University)
GTEC
German Tech
Entrepreneurship Center
Berlin

¿Qué es Big Data?

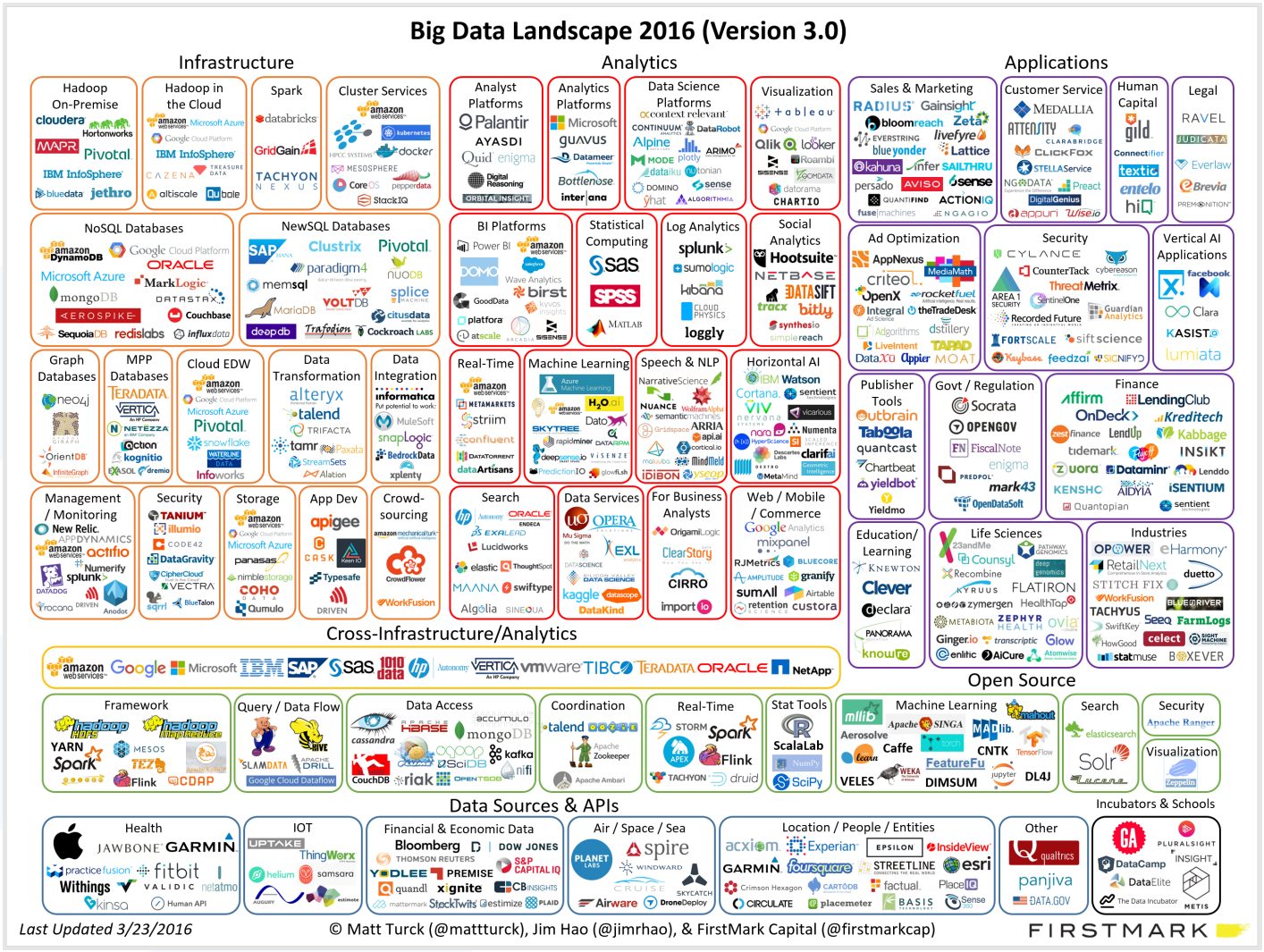
- ... like teenage sex ...
- Algo que no cabe en Excel
- Las tres V

  Volumen

  Velocidad

  Variedad

Big Data hoy


Big Data Landscape 2016 (Version 3.0)
Infrastructure
Analytics
Applications
Hadoop On-Premise
Hadoop in the Cloud
Spark
Cluster Services
Analyst Platforms
Analytics Platforms
Data Science Platforms
Visualization
Sales & Marketing
Customer Service
Human Capital
Legal
NoSQL Databases
NewSQL Databases
BI Platforms
Statistical Computing
Log Analytics
Social Analytics
Ad Optimization
Security
Vertical AI Applications
Graph Databases
MPP Databases
Cloud EDW
Data Transformation
Data Integration
Real-Time
Machine Learning
Speech & NLP
Horizontal AI
Publisher Tools
Govt / Regulation
Finance
Management / Monitoring
Security
Storage
App Dev
Crowd-sourcing
Search
Data Services
For Business Analysts
Web / Mobile / Commerce
Education/ Learning
Life Sciences
Industries
Cross-Infrastructure/Analytics
Open Source
Framework
Query / Data Flow
Data Access
Coordination
Real-Time
Stat Tools
Machine Learning
Search
Security
Visualization
Data Sources & APIs
Incubators & Schools
Health
IOT
Financial & Economic Data
Air / Space / Sea
Location / People / Entities
Other
Last Updated 3/23/2016
© Matt Turck (@mattturck), Jim Hao (@jimrhao), & FirstMark Capital (@firstmarkcap)
FIRSTMARK

¿Necesitas Big Data?

- ¿Tienes restricciones de tamaño en los datos que manejas?
- ¿Descartas data porque es demasiado grande?
- ¿Esperas que el volumen de datos que mantienes crezca en el futuro?
- ¿Esperas contestas preguntas de tu negocio usando datos?
- ¿Has considerado productos de valor agregado usando datos?

Big Data para operadores

- CenturyLink compró una compañía de análisis
- Posicionamiento
  Operación inteligente
  Provisionamiento
  Integración con otros servicios
- Herramientas
  Archivos de log
- Seguridad
  Identificación de amenazas e infecciones

Ejemplos

- OpenBMP + Kafka para análisis en tiempo real del estado de BGP

  https://www.nanog.org/sites/default/files/tuesday_general_evens_openbmp.pdf
- Network Machine Learning Research Group (nmlrg) en la IETF

  Explorar la habilidad para predecir o tomar decisiones en base a información en tiempo real de la red

Big Data para DNS

- Tráfico DNS
  - Servidores con autoridad
  - Servidores recursivos
- Actividad del registro
  - Operaciones: crear, renovar, cancelar
  - Consultas WHOIS
- Operaciones en general
  - Logs web, de correo, de firewall y otros dispositivos
- Otros
  - Verificaciones de delegación
  - Contenido Web
  - Registro de malware

Data Lake

- Un repositorio de grande escala
  Y capacidad de procesamiento asociada
- Operadores de DNS han hecho Big Data y Data Lakes hace años
  Conocen Day in the Life of the Internet (DITL)
  https://www.dns-oarc.net/oarc/data/ditl
- Almacenar primero, preguntar después
  En el futuro puedes contestar una pregunta con datos de hoy
  El valor oculto en los datos
  Y potencial combinación

Armando el puzzle

- Software base
  Cloudera, MapR, HortonWorks
- Almacenamiento
  HDFS, S3
- Procesamiento por lotes
  Pig, Hive, MapReduce
- Procesamiento en tiempo real
  Spark, Kafka, Apache Falcon
- Ejecutando consultas
  Apache Drill, Impala, Hue, Google's Big Query
- Herramientas de inteligencia de negocios
  Caravel, Tableau

Casos de uso para ccTLDs

- Inteligencia de negocios
  - Comportamiendo de los registradores/revendedores
  - Pronósticos de crecimiento
  - Efectividad de campañas
- Expandir el registro
  - Registrantes, dominios, contenido
- Detección de abuso
  - Usando Whois y DNS
- Detección de anomalías
  - Infecciones por malware
- Administración de la infraestructura
  - Crecimiendo del tráfico, cambios en patrones, indicadores de adopción de tecnología

Big Data en los TLDs

- SIDN y el proyecto “ENTRADA”
  Software libre
  http://entrada.sidnlabs.nl/
- Verisign
- Nominum para sus clientes
- Nominet y el proyecto “Turing”
  Servicio pagado
  Software no disponible
- OpenDNS
  3 plataformas diferentes para DNS, seguridad y análisis de los clientes

Ejemplos

- Big data security in .nl
  http://iepg.org/2015-11-01-ietf94/iepg-moura.pdf
- The Impact of a TTL Change at the TLD-level
  https://indico.dns-oarc.net/event/22/session/1/contribution/26/material/slides/0.pdf
- OpenDNS Tech Blog
  https://blog.opendns.com/2016/07/25/from-query-logs-to-visualization/
- AAAA Deep Dive: DNS Resolution Anomalies and Performance
  https://indico.dns-oarc.net/event/22/session/1/contribution/9/material/slides/0.pptx

Big Data en NZRS

- Hadoop basado en Cloudera
  22 servidores, medio PetaByte de almacenamiento
- Capturas DNS
  22TB de capturas DNS para .nz
  Más de 3 años
  Análisis de tendencias, usos, nuevos tipos de tráfico, fallas de configuración
- Escaneo de zona
  Adopción de IPv6, proveedores de servicios en la nube, Anycast, otras tendencias
- Escaneo web
  Clasificación por industria, clasificación de contenido, uso de CMS, etc

Ejemplos

- “Two years of .nz zone scans”
  http://blog.nzrs.net.nz/two-years-of-nz-zone-scans/
- “DNSSEC Validation at Spark NZ”
  http://blog.nzrs.net.nz/dnssec-validation-at-spark-nz/
- “The hunger for AAAA”
  Presentado en DNS-OARC 25
- ”Characterization of popular resolvers”
  http://blog.nzrs.net.nz/characterization-of-popular-resolvers-from-our-point-of-view-2/

Nuestros planes futuros

- Recolectar más fuentes
- Combinar, predecir, agregar valor
- Compartir datos a través de API

  Creación de oportunidades de negocios basados en datos
- Publicar estadísticas operacionales

  Contadores generales de DNS, WHOIS,

Sebastian sebastian@nzrs.net.nz
Gautier gautier.krings@realimpactanalytics.com

Contact:
www.nzrs.net.nz

NZRS
an InternetNZ company