

Strata+
Hadoop
WORLD

PRESENTED BY

O'REILLY®

cloudera®

strataconf.com
#StrataHadoop



Big Data Architecture: the Buy vs. Build Dilemma

Carme Artigas Co-founder & Partner
Synergic Partners

What makes Big Data Architectures different?

- Big data architecture starts with the data itself, taking a **bottom-up approach**.
- Big data introduces **new data sources** such as social media content and streaming data.
- The enterprise data warehouse (EDW) becomes a **source** for big data..
- The variety of big data and unstructured data requires a **new type of persistence**.
- **Analytics capabilities** need to be expanded to handle the variety, volume, and velocity of big data.
- Big data applications leverage reporting and visualization in new ways to integrate information and generate **new insights**.

OPTIONS

CHOICES

DECISIONS

DILEMMAS



KEEP
CALM
AND
PLAN
BIG DATA

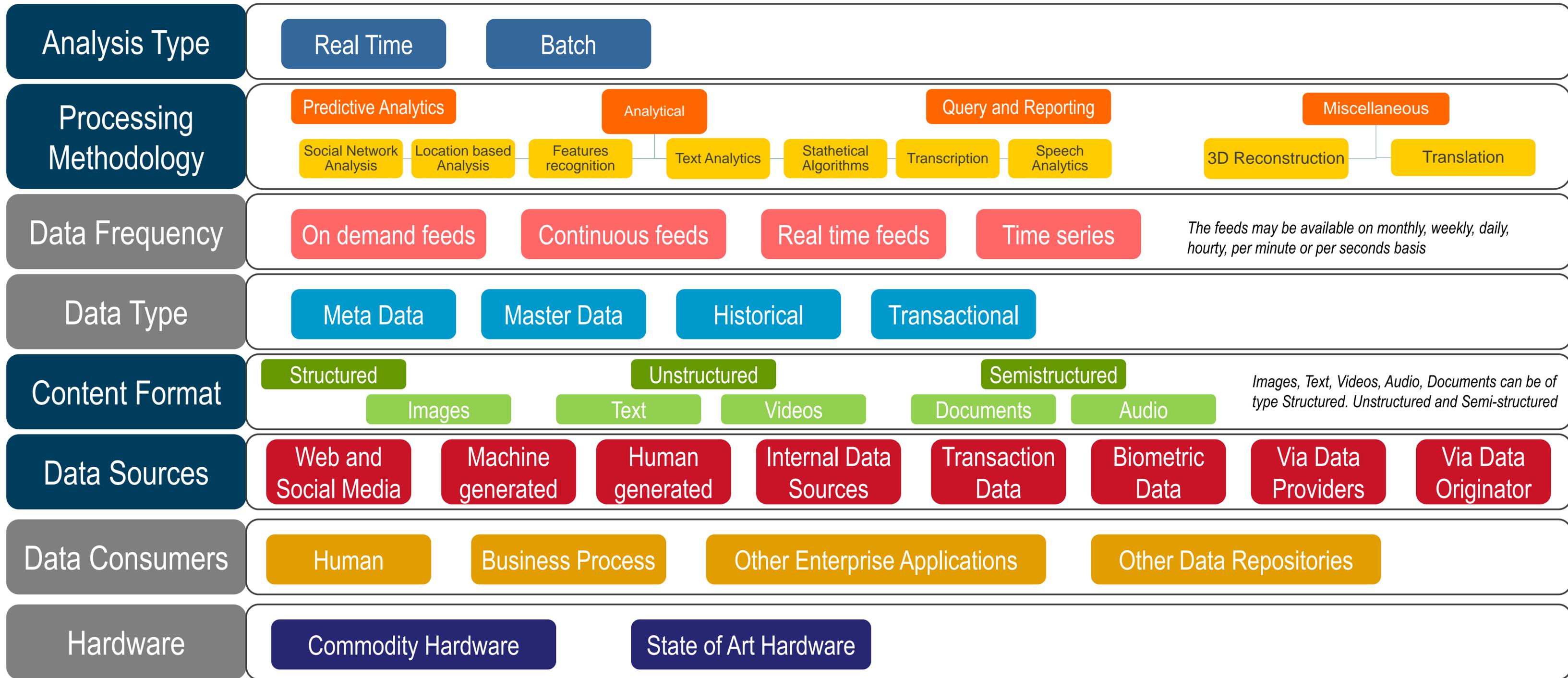
What are your data sources?



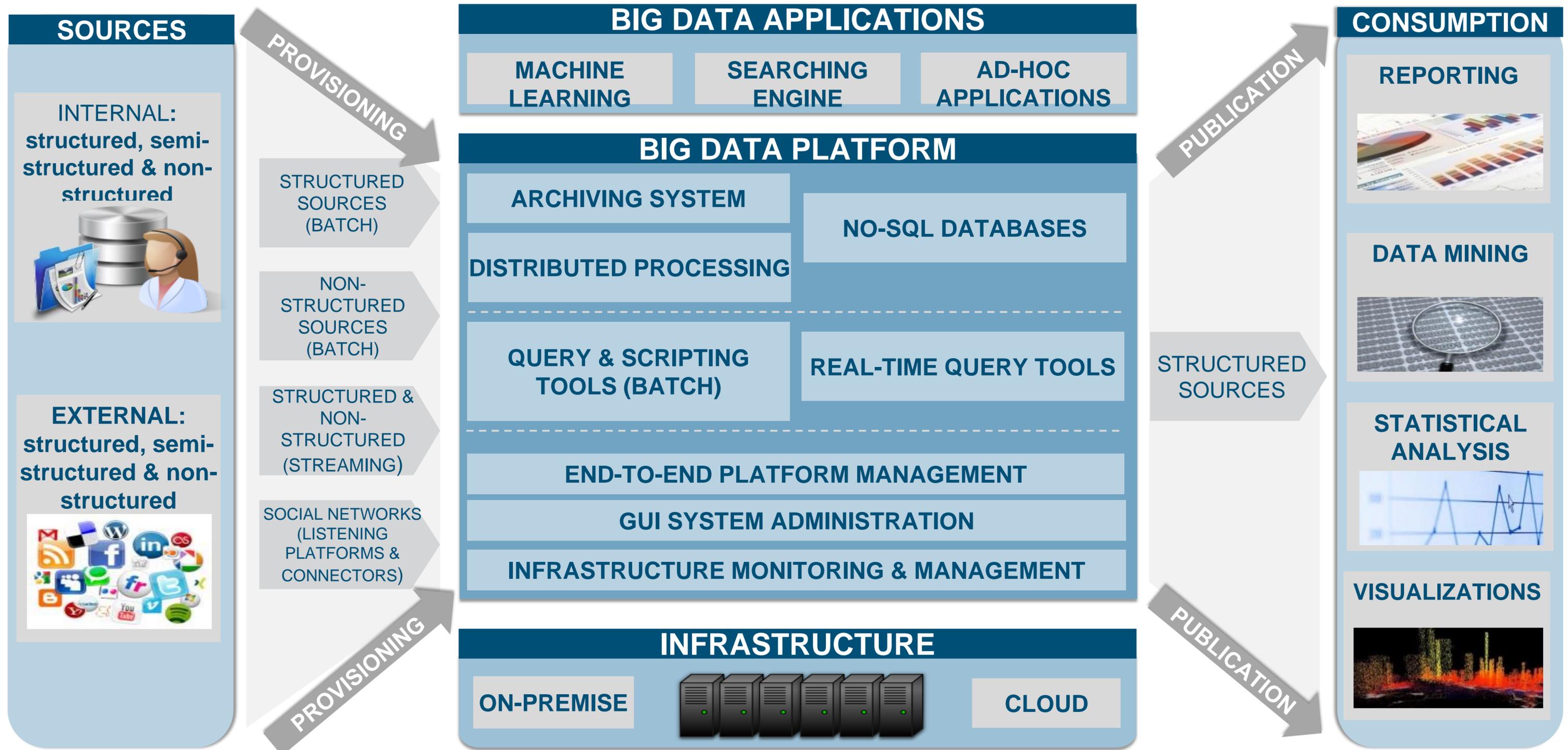
Types of data		Figure 1
Structured <ul style="list-style-type: none"> Relation & Legacy Databases Spreadsheets Flat Files with Proper Record Formats 	Semi-Structured <ul style="list-style-type: none"> XML EDI Documents 	UnStructured <ul style="list-style-type: none"> Web E-Mails Wikipedia Multimedia (Video, Audio) RSS Feeds Messages
	Within Corporate <ul style="list-style-type: none"> Static <ul style="list-style-type: none"> Internal Documents Sales Report Marketing Material Real Time <ul style="list-style-type: none"> Customer Center Call Logs Customer Center Representative Notes Trading Rooms 	Internet <ul style="list-style-type: none"> Static <ul style="list-style-type: none"> Analysts Reports Formal/Legal Filings (SEC, FDIC) Journals Real Time <ul style="list-style-type: none"> Breaking News Market Pricing Weather Events Corporate Actions Chat Rooms Social Media Platforms

Choosing a Big Data Architecture

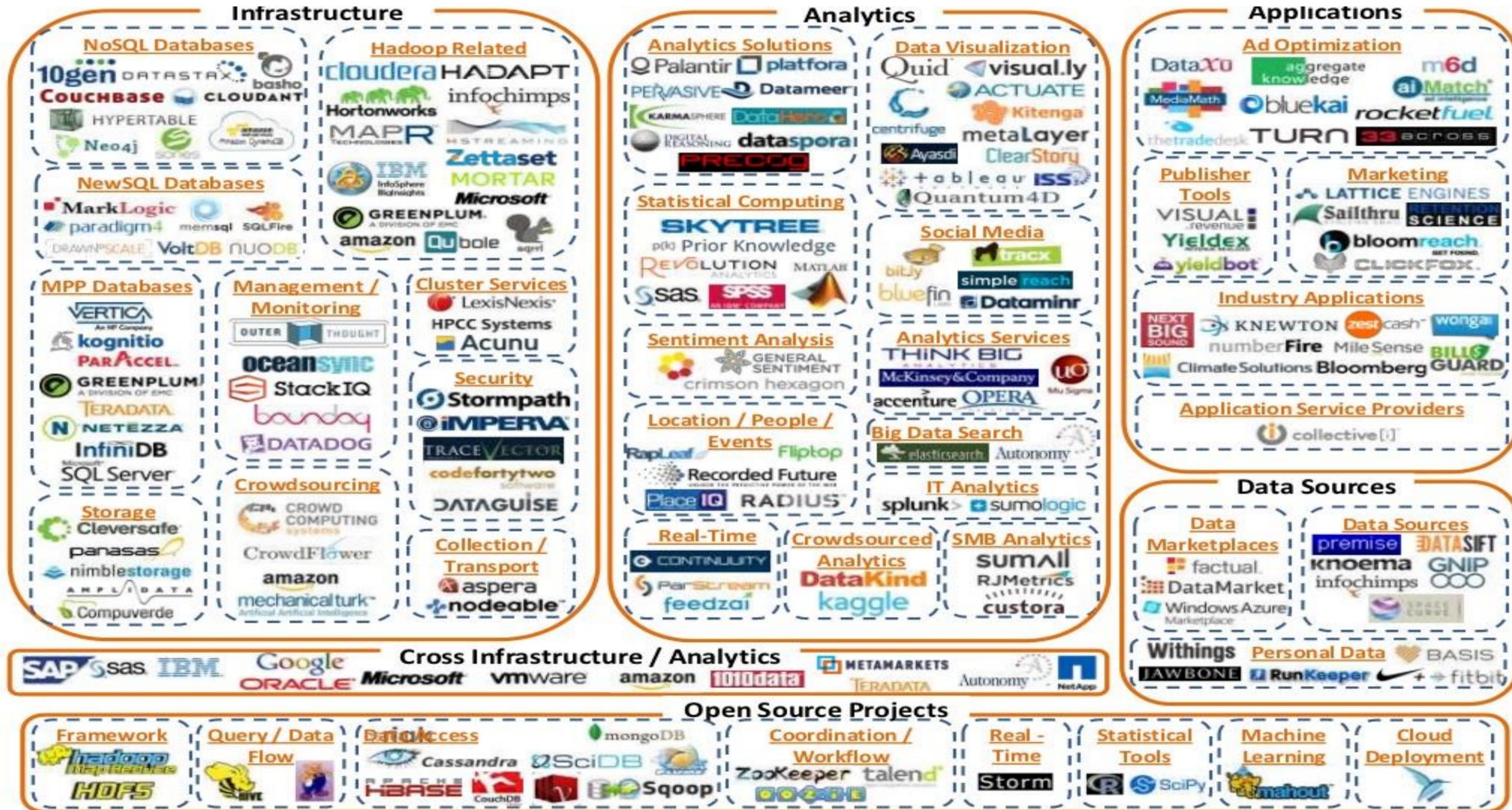
Key Categories for defining Big Data Classification Patterns



Synergic Big Data Reference Architecture



Market & Players



© Matt Turck (@mattturck) and ShivonZilis (@shivonz) Bloomberg Ventures

Non- relational

Relational

Analytic

Hadoop
Horton
Cloudera
MapR
Zettaset

Hadapt

Teradata
Aster

IBM InfoSphere
PureData

HP Vertica

Infobright
ParAccel
Calpont
VectorWise

Pivotal

SAP Hana
SAP Sybase IQ

Oracle
Times - Ten

Operational

InterSystems
Progress
Objectivity
Versant

Document

Lotus Notes

MarkLogic
McObject

Spark

Oracle

IBM DB2

SQLSrvr JustOneDB

MySQL

Ingres

PostgreSQL

Sybase ASE

EnterpriseDB

NoSQL

Key Value

Riak
Redis
Membrain
Voldemort
BerkeleyDB

Couchbase

Cassandra

CouchDB
MongoDB
RevenDB

Big Tables

HyperTable
HBase

“Data as a Service”

Cloudfant

App Engine
SimpleDB

Graph

Flock DB
InfiniteGraph
Neo4j
AllegroGraph

NewSQL

Amazon RDS
SQL Azure
Database.com
Xeround

FathomDB

SchoonerSQL
Tokutek
Continuent
Translattice

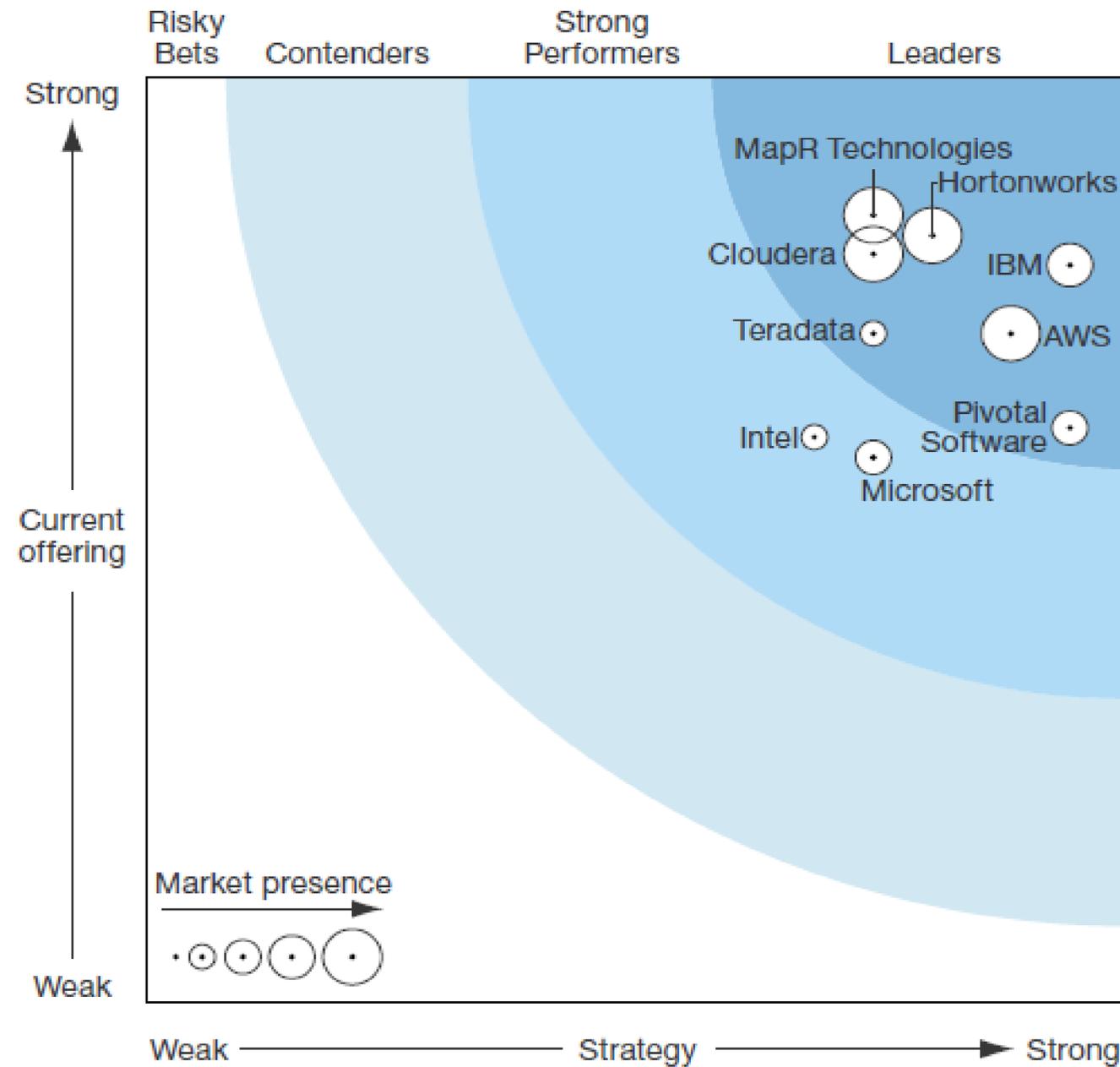
ScaleBase
CodeFutures
VoltDB

HandlerSpcket
Akiban
MySQL Cluster
Clustrix
Drizzle
GenieDB
ScalArc
NimbusDB

Source: Infochimps

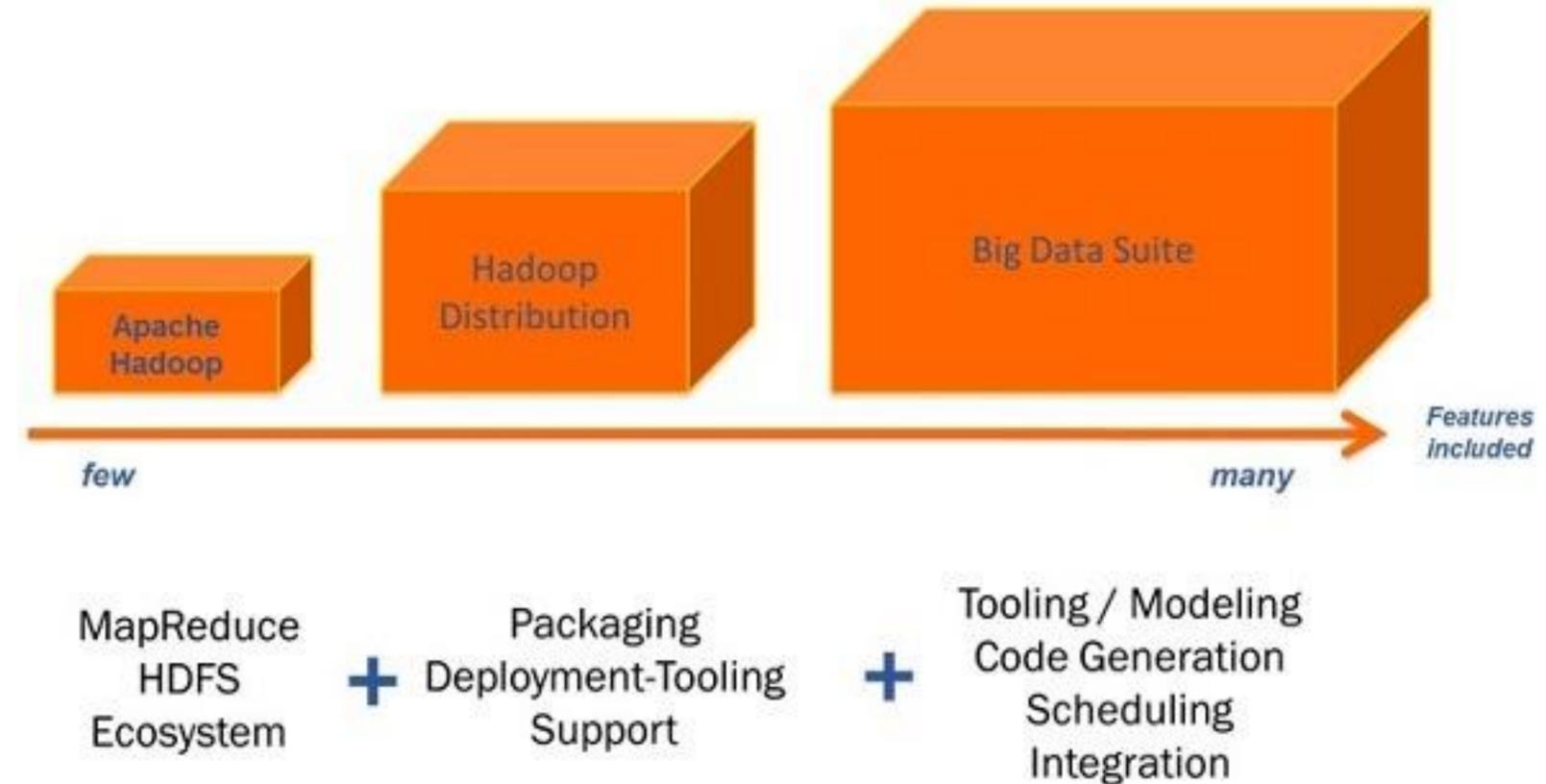
Buy vs Build: Hadoop

Figure 2 Forrester Wave™: Big Data Hadoop Solutions, Q1 '14



Build

Buy

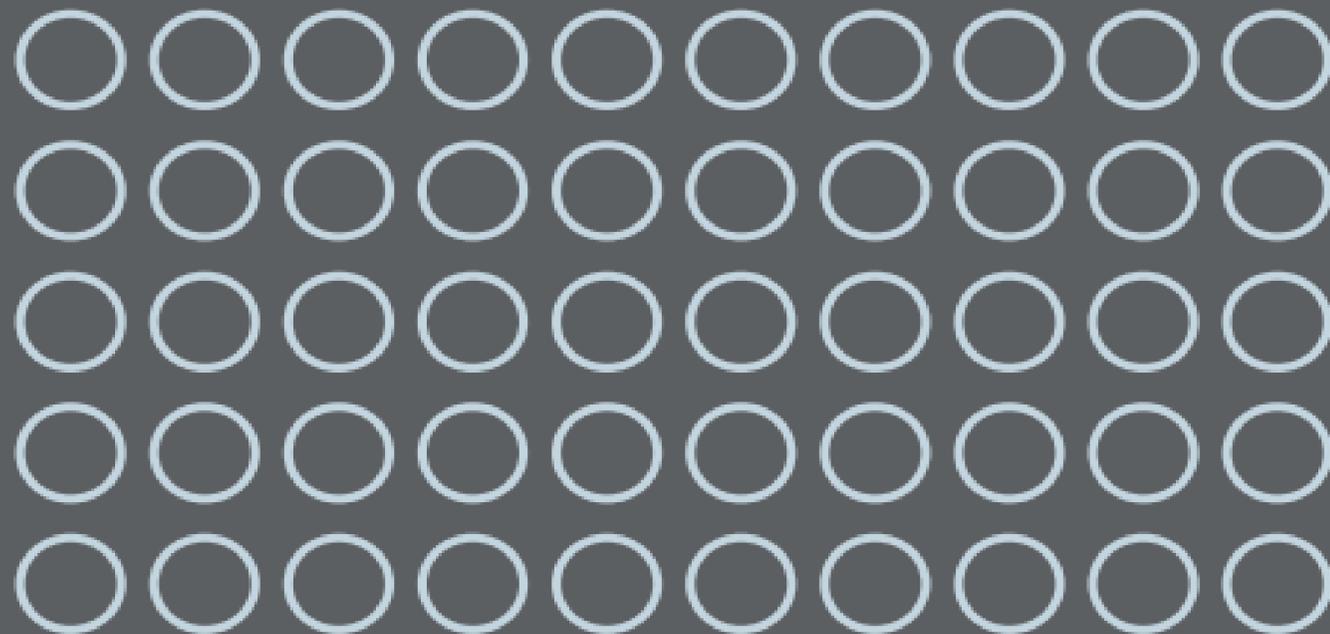


DIY- Hadoop can be difficult

HADOOP DISTRIBUTIONS IN USE

Regarding your organization's Hadoop deployment is your organization using a free/open source distribution or is your organization a paying subscription customer of one or more of the commercial Hadoop vendors?

We are using open source Hadoop (i.e. Roll-your-own Apache Hadoop)

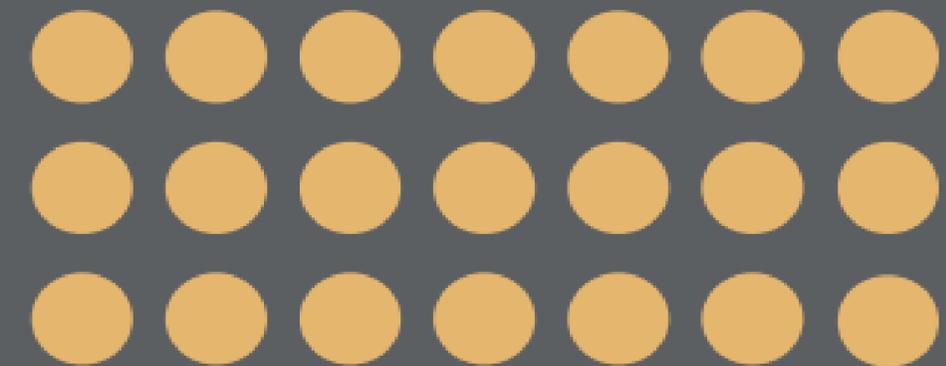


51%



We are a paying subscription customer of one or more a commercial vendors

25%



We are using a free Hadoop distribution from a vendor

24%

SOURCE: WIKIBON 2014

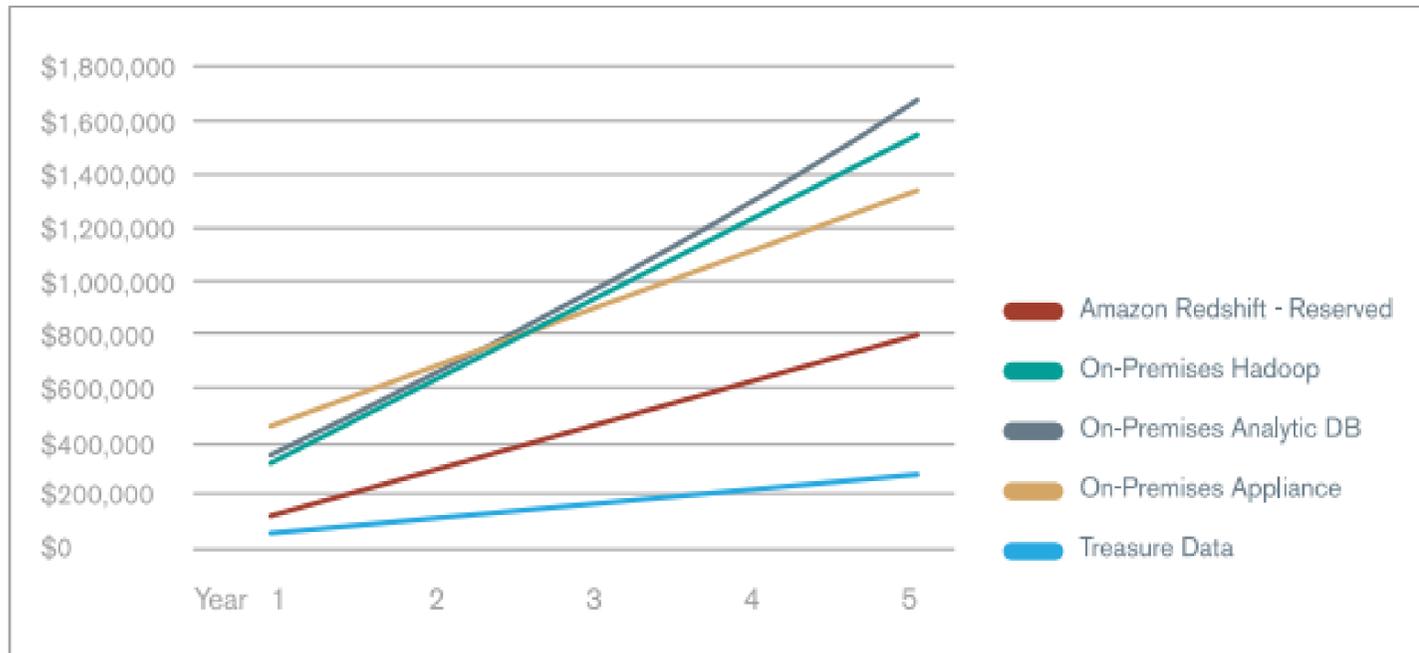
Buy vs. Build : Data Scientist Tools



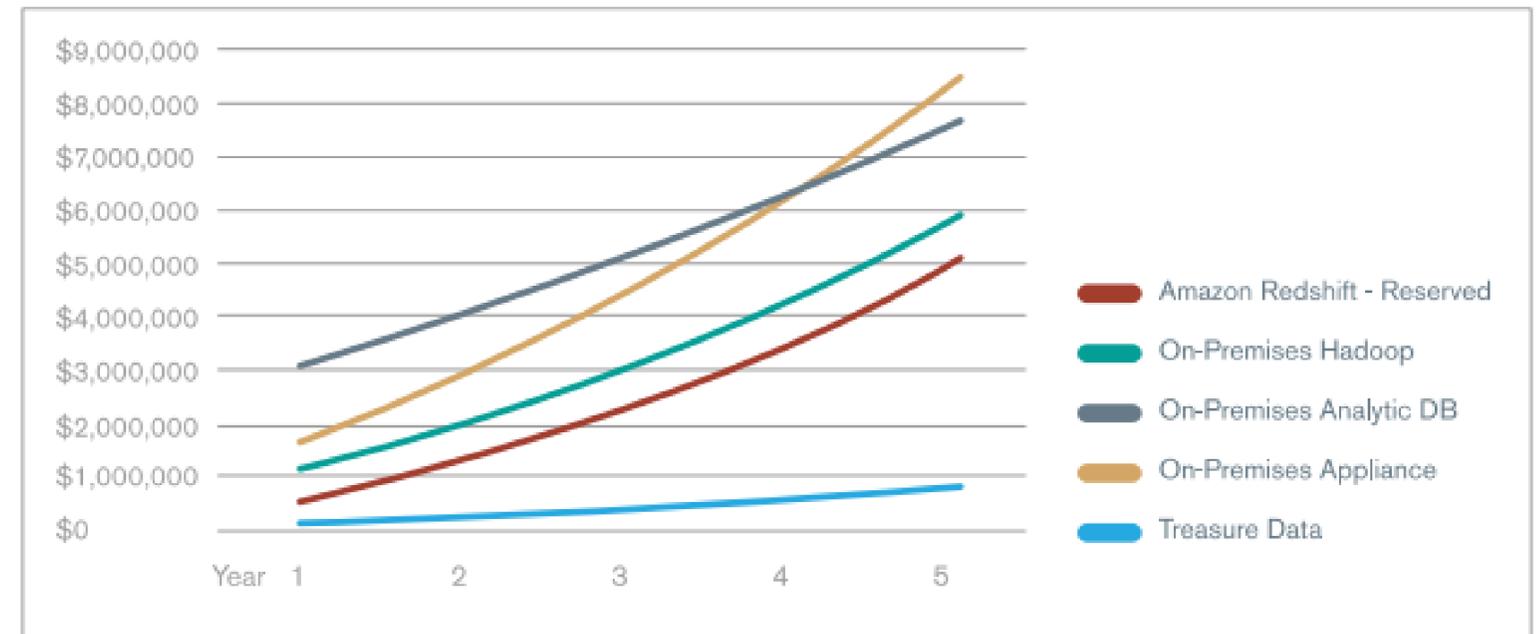
Cloud vs On-premise = Capex vs Opex



5 TERABYTES



500 TERABYTES



Source: Bootstrap resesarch

Buy vs. Build: Step-by-step

- 1 Start with the Business Case in mind
- 2 Identify Data Sources
- 3 Define your Big Data Reference Architecture & pattern
- 4 Understand the competitive landscape
- 5 Leverage your internal skills
- 6 Take some preliminary decisions: CAPEX vs OPEX, Cloud vs On-premise, legal issues, etc.
- 7 Technical criteria: decide, evaluate & validate
- 8 Calculate TCO
- 9 Define a flexible roadmap: new technologies will appear
- 10 Start small & Scale-up!

Strata+ Hadoop

WORLD

PRESENTED BY

O'REILLY®

cloudera®

strataconf.com
#StrataHadoop

Thank you!
cartigas@synergicpartners.com



@synergicpartner



/Synergic-Partners



/synergicpartners

synergicpartners.com