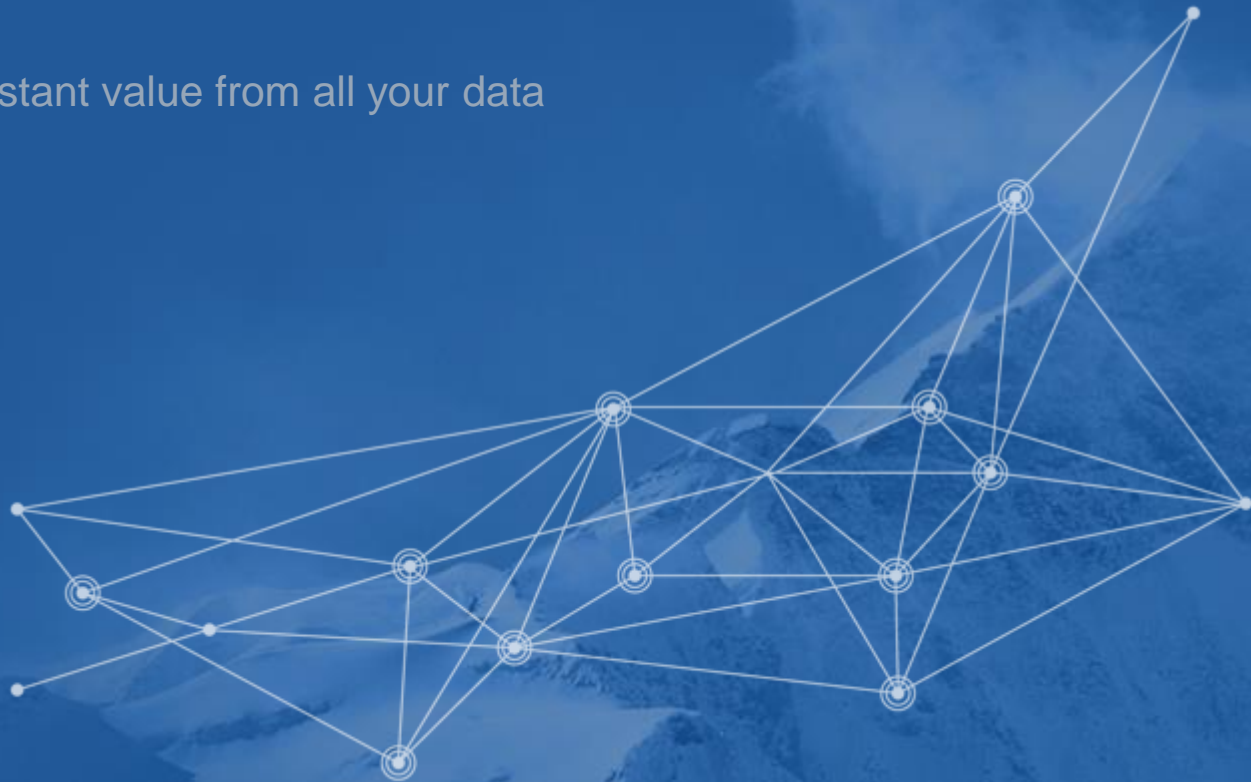


Talend Big Data

Delivering instant value from all your data



“I may say that this is the greatest factor: **the way in which the expedition is equipped.**”



Roald Amundsen
race to the south pole, 1911



The New Data Integration Economics

45x

savings. \$1,000/TB
for Hadoop vs
\$45,000/TB for
traditional

6x

faster ROI using
big data analytics
tools vs
traditional EDW

600x

active data.
Neustar moved
from storing 1% of
data for 60 days to
100% for one year

\$600B

revenue shift by
2020 to companies
that use big data
effectively

“Big data is what happened when the cost of keeping information became less than the cost of throwing it away.”

– Technology Historian George Dyson

Macro Trends Revolutionizing the Integration Market

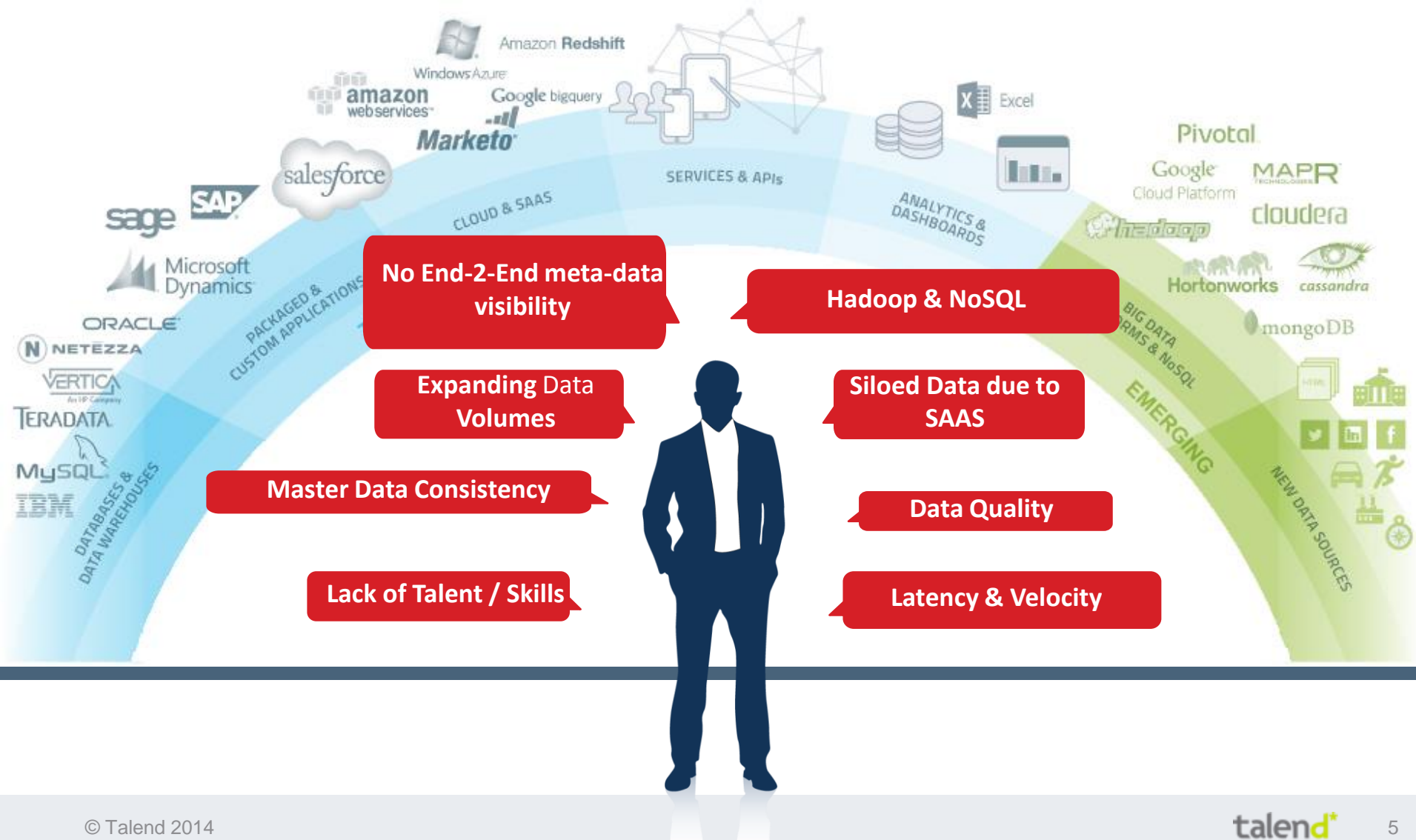
The amount of data will grow
50X from 2010 to 2020

64% of enterprises surveyed
indicate that they're
deploying or planning **Big
Data** projects

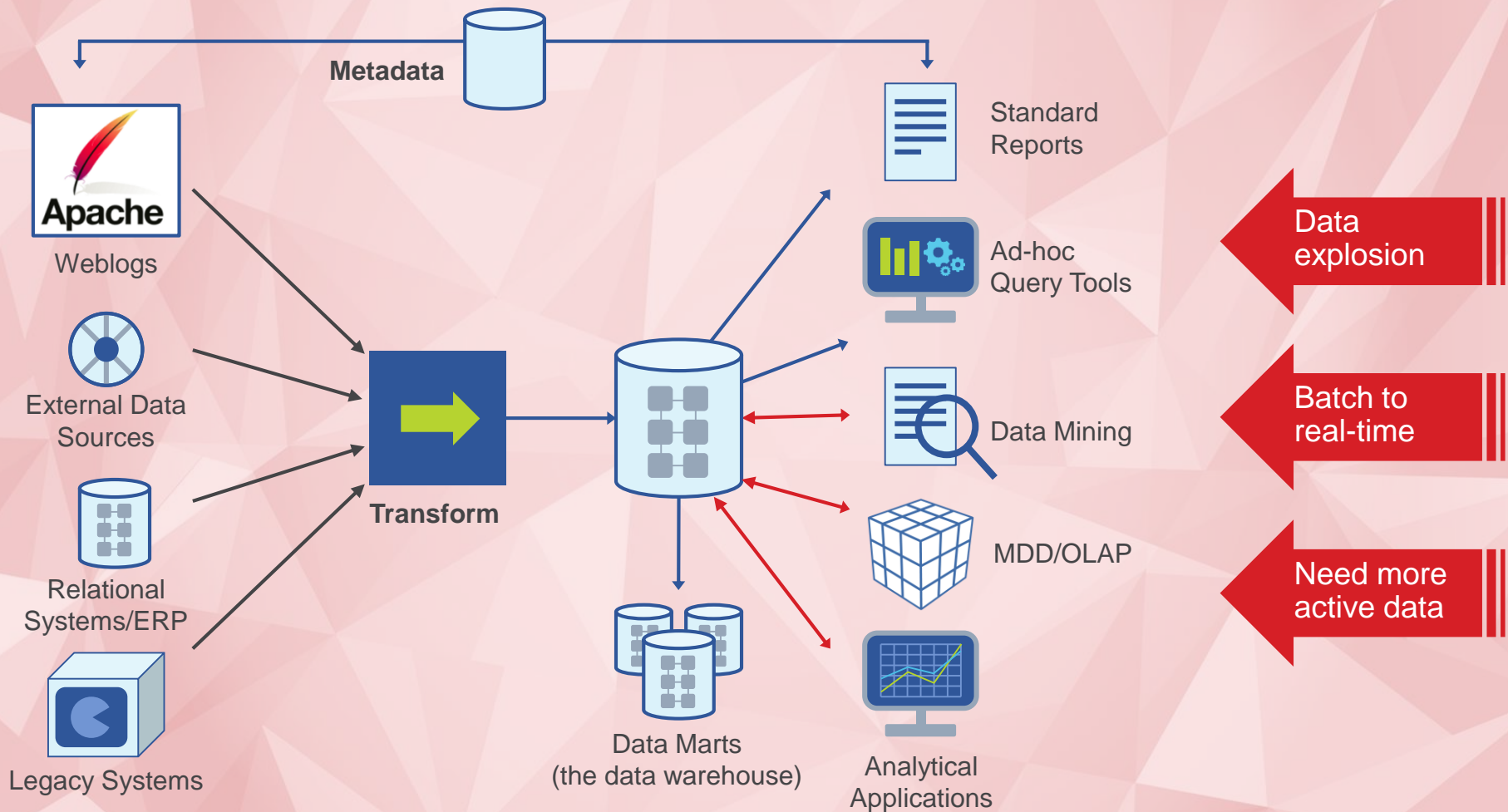
By 2020, 55% of CIOs will
source all their critical apps
in the **Cloud**



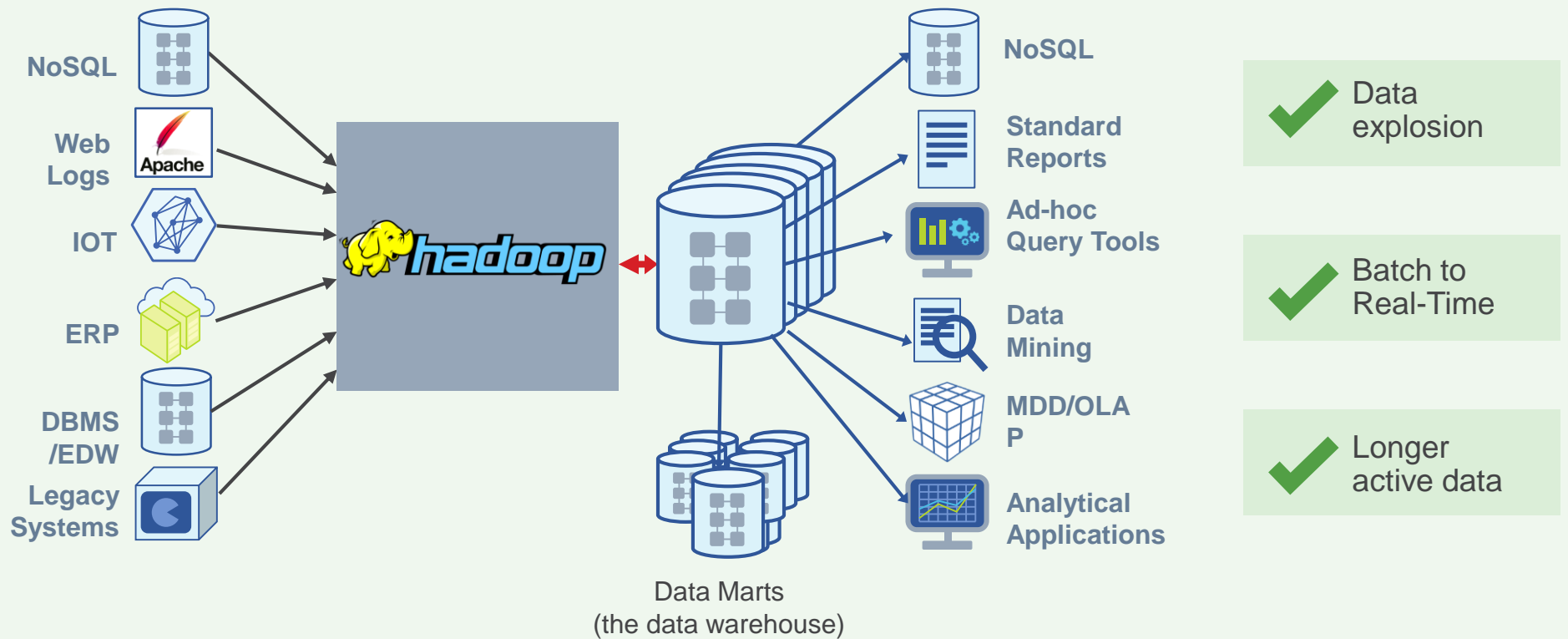
CIO: It's tough at the top



Existing Infrastructures Under Distress: Architecturally and Economically



Benefits of Hadoop and NoSQL



Different flavors of Big Data across industries

Manufacturing



- Product as a Service
- Innovation in R&D
- Preventive Maintenance

Retail



- Real time offers and personalization
- In store customer experience and clienteling
- Dynamic Pricing

Banking



- Multi Channel customer journeys
- Fraud, anti money Laundering
- Personalized offers

Insurance



- Frauds & Risk Mgmt
- Customer recommendations
- Pay per use and personalized services

Healthcare



- Adverse effects Mgmt
- Personalized Healthcare.
- Prevention and diagnoses
- Genomic computation

Transports/Travel



- Planning and management of events related to logistics
- Customer real-time service
- Energy saving
- Dynamic pricing

Public Sector



- Linked Data
- Frauds, crime, Public Safety
- Guided learning in Education
- Citizen relationship

Telecom



- Multi channel customer journeys
- Big Data Monetization (e.g. geo localization)
- Fraud and churn mgmt

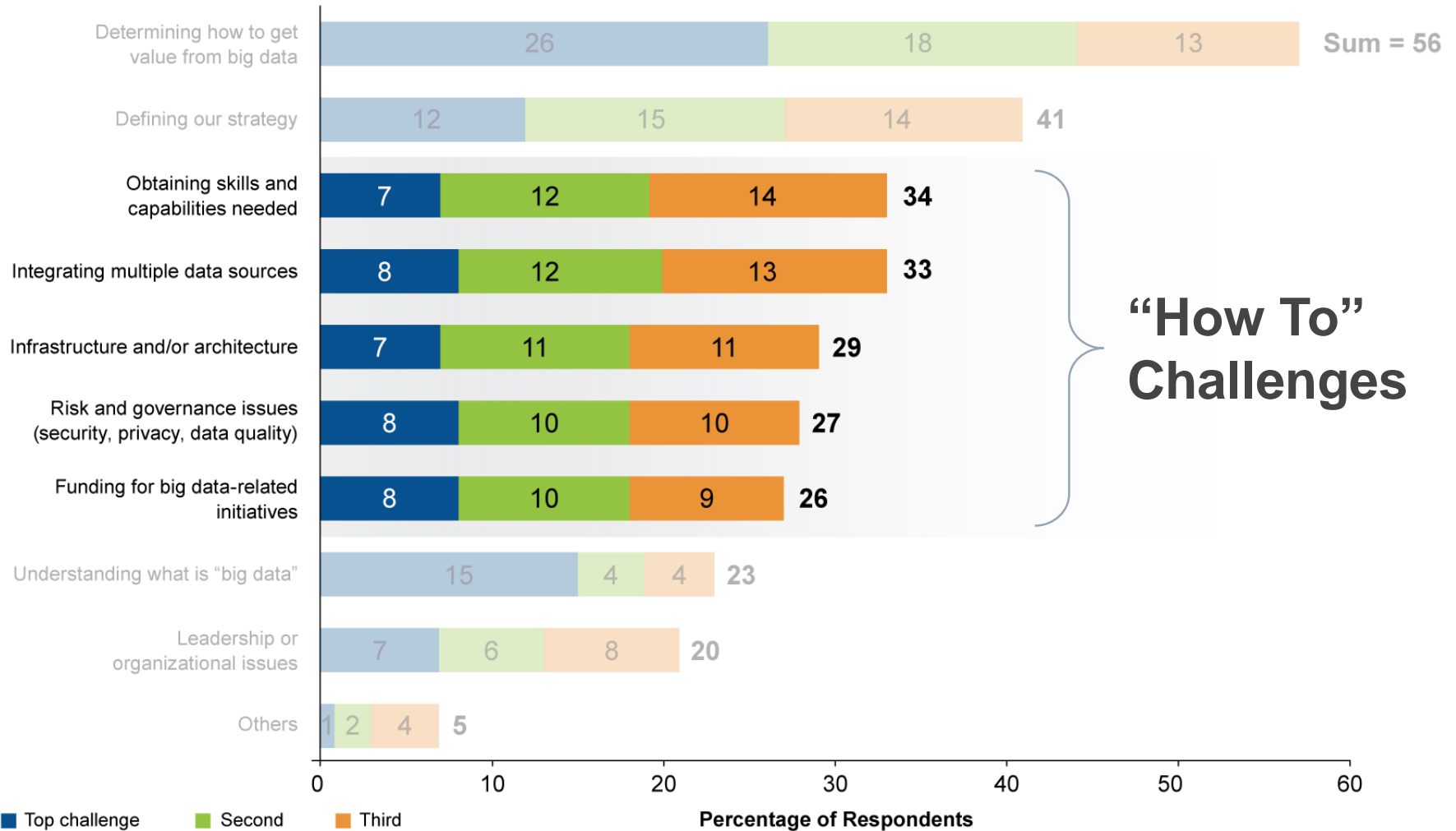
Consumer Product



- Sentiment analysis
- Consumer Relationship management
- Product as a service

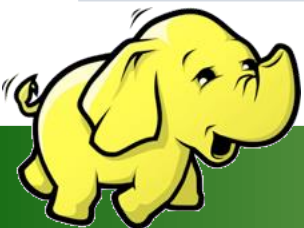
How is this related to your world ?

Top Big Data Challenges



Source: Gartner - Survey Analysis: Big Data Adoption in 2013 Shows Substance Behind the Hype - 12 September 2013 - G00255160

A Brief History of Hadoop and Talend



Apache Project
Established

Enterprise Hadoop distribution Vendors
Hortonworks, Cloudera, ...

HDP 2.0 release include
Hadoop2.0 and Yarn

2004

2006

2008

2010

2012

2014
Adopted
technology



1st release of Talend Open
Studio

April 2010 v4 include Hive
and HDFS support

Talend support YARN
/Hadoop2.0

2005

2006

2008

2010

2012

2014
Preferred
solution for
BigData integration

Talend is matching and supporting the Hadoop ecosystem

What is Talend for Big Data?

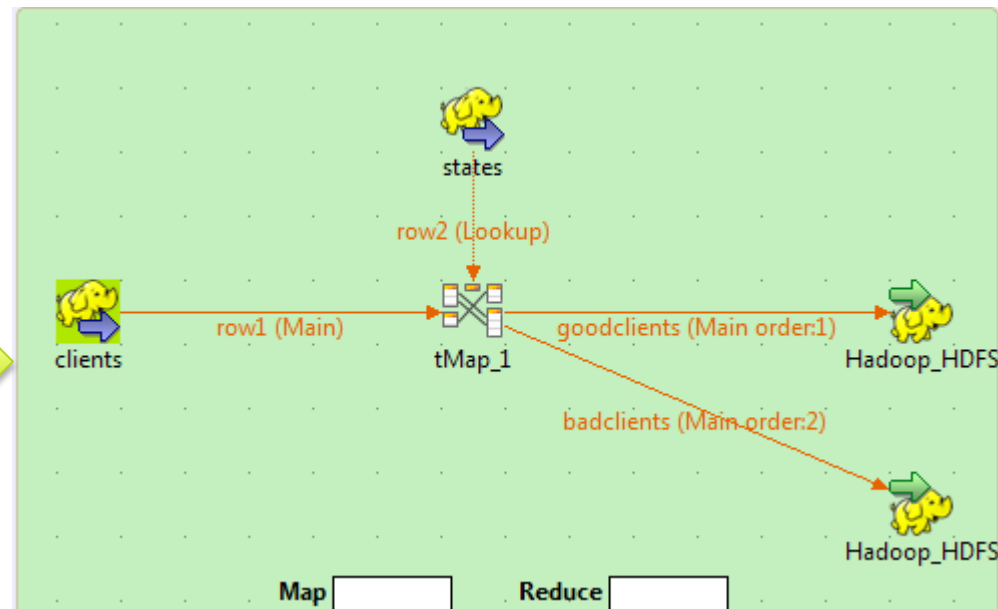
- The best way to get rid of manual/hand coding script.
- No need to learn : MapReduce, Pig, Hive, Spark, Flume, Kafka, Sqoop, Storm, etc....
- Leverage a nice, user-friendly Designer Studio to create your Big Data integration

```
chainMapper.setCid("tHDFSInput_1");
```

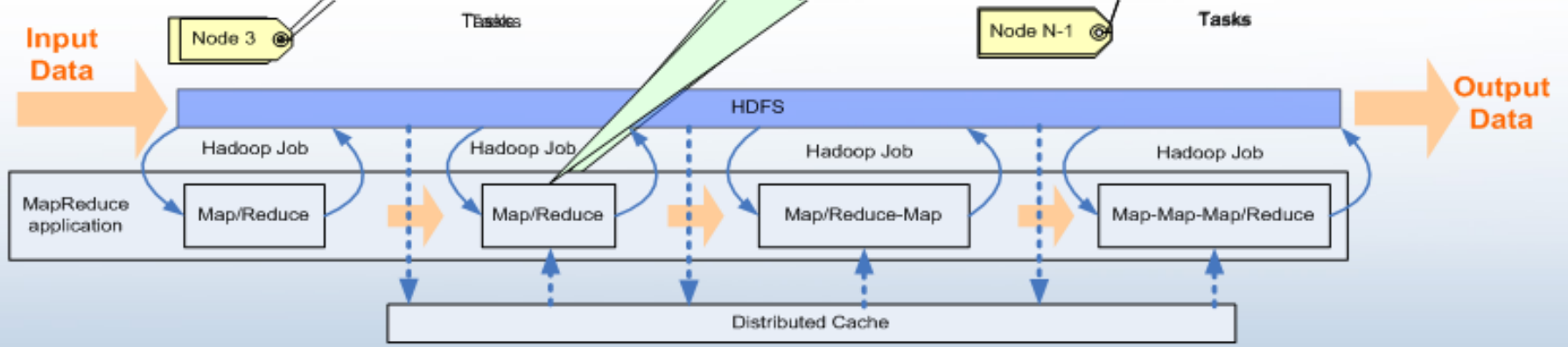
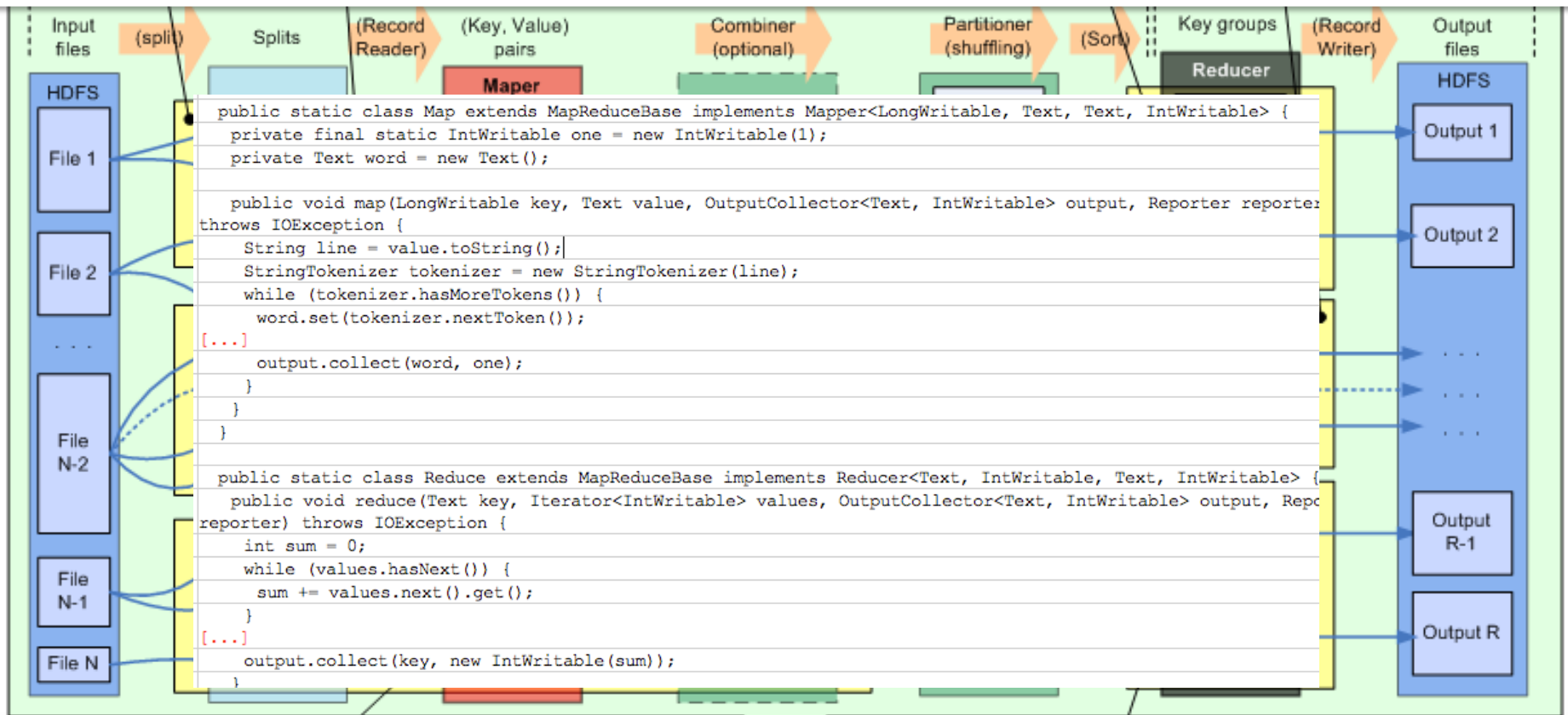
```
chainMapper.addMapper(job, tMap_1row1Mapper.class,  
    NullWritable.class, row1Struct.class,  
    TaggedJoinKey_tMap_1Struct.class,  
    TaggedMapOutput_tMap_1Struct.class, true,  
    new JobConf(false));
```

```
chainMapper.setCid("tHDFSInput_2");
```

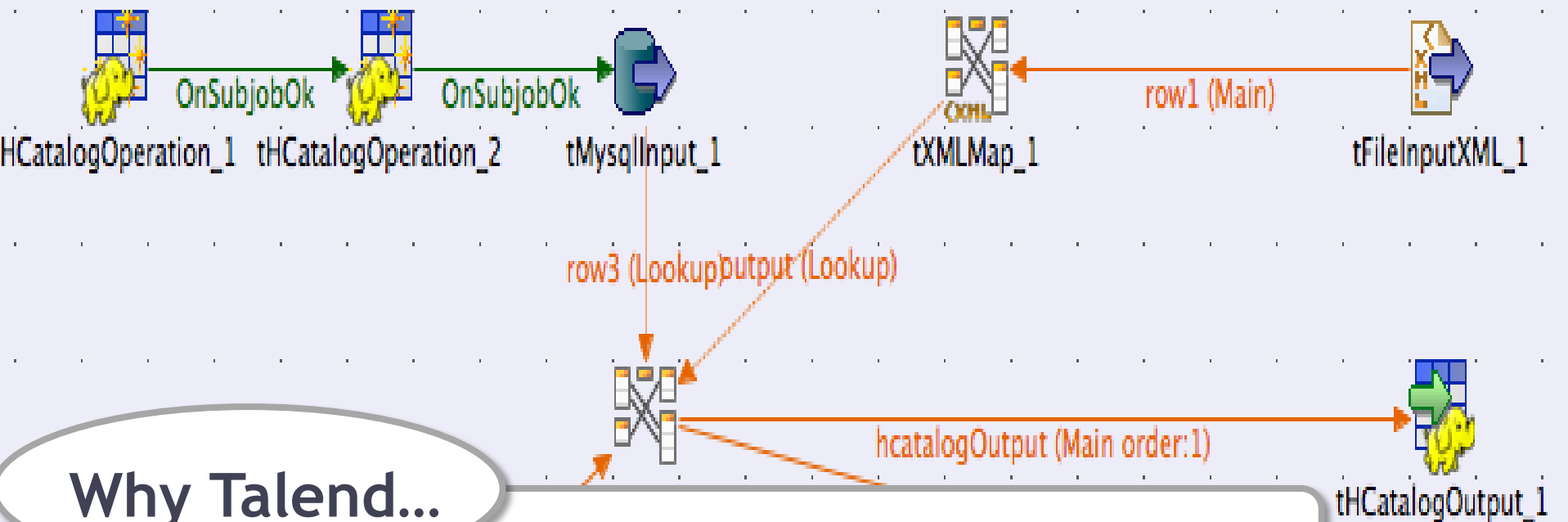
```
chainMapper.addMapper(job, tMap_1row2Mapper.class,  
    NullWritable.class, row2Struct.class,  
    TaggedJoinKey_tMap_1Struct.class,
```



Trying to get from this...

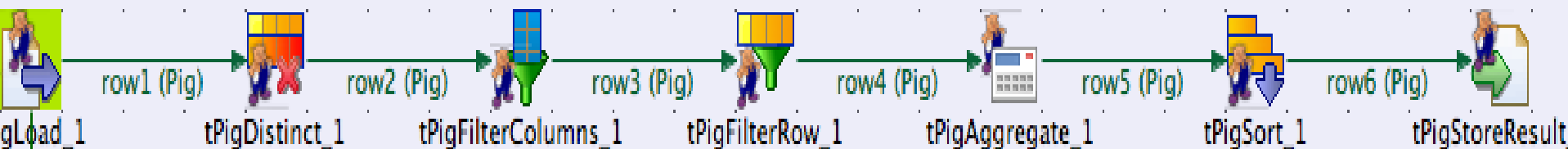


to this...

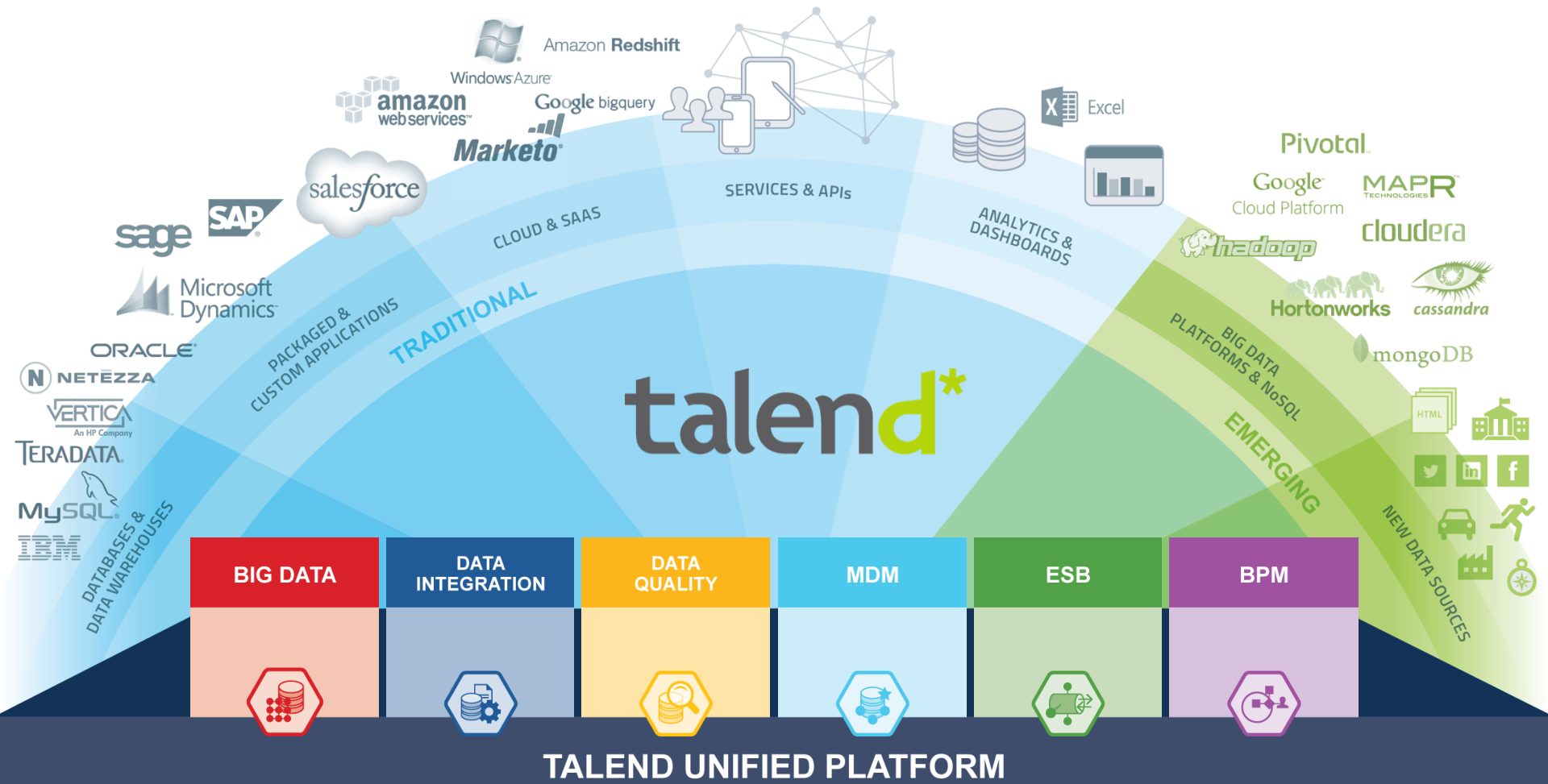


Why Talend...

Talend generates code that is executed within map reduce. This open approach removes the limitation of a proprietary “engine” to provide a truly unique and powerful set of tools for big data.



The Talend Platform



Talend Big Data Sandbox



Virtual Image installed with

- Four scenarios for you to try:
 - Clickstream data
 - Twitter sentiment
 - Apache weblogs
 - ETL offload

