

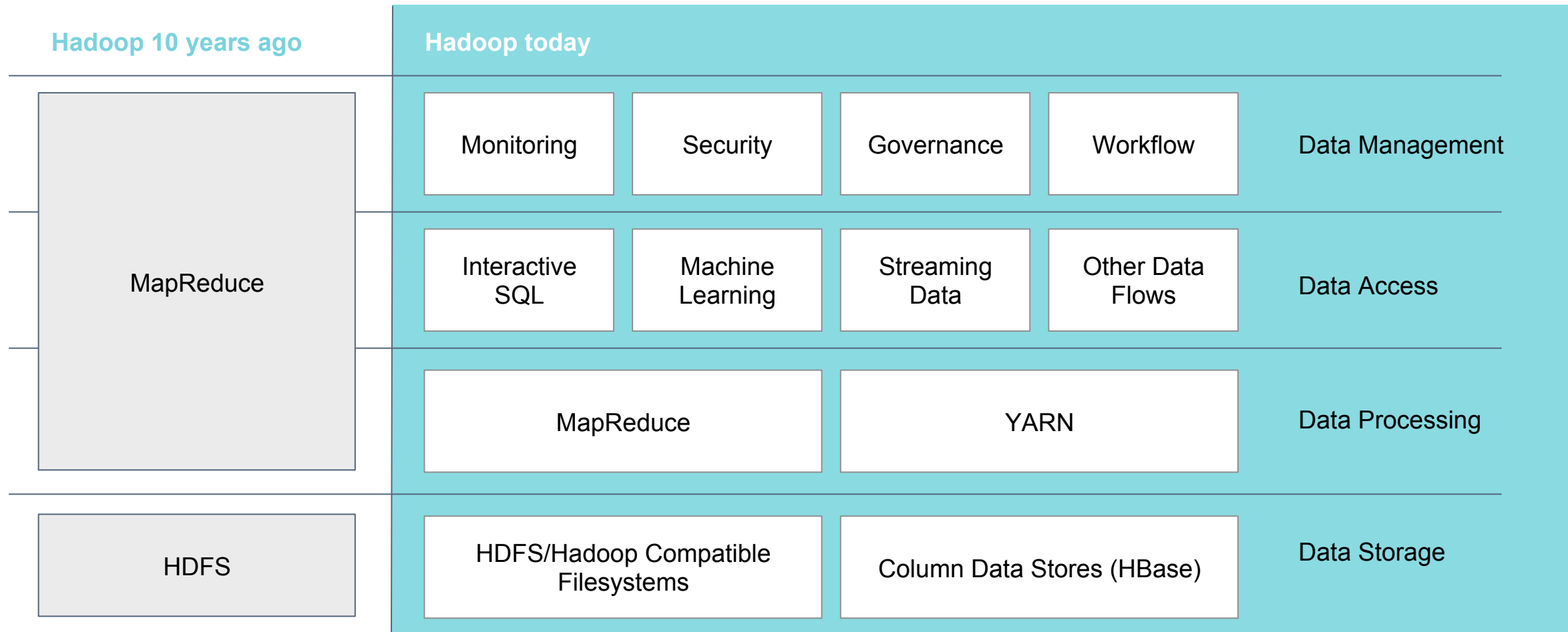


# Lessons from the Trenches: How Apache Hadoop is Being Used & The Challenges Its Users Face

John Mertic, Director, *ODPi*  
@jmertic - @ODPIorg

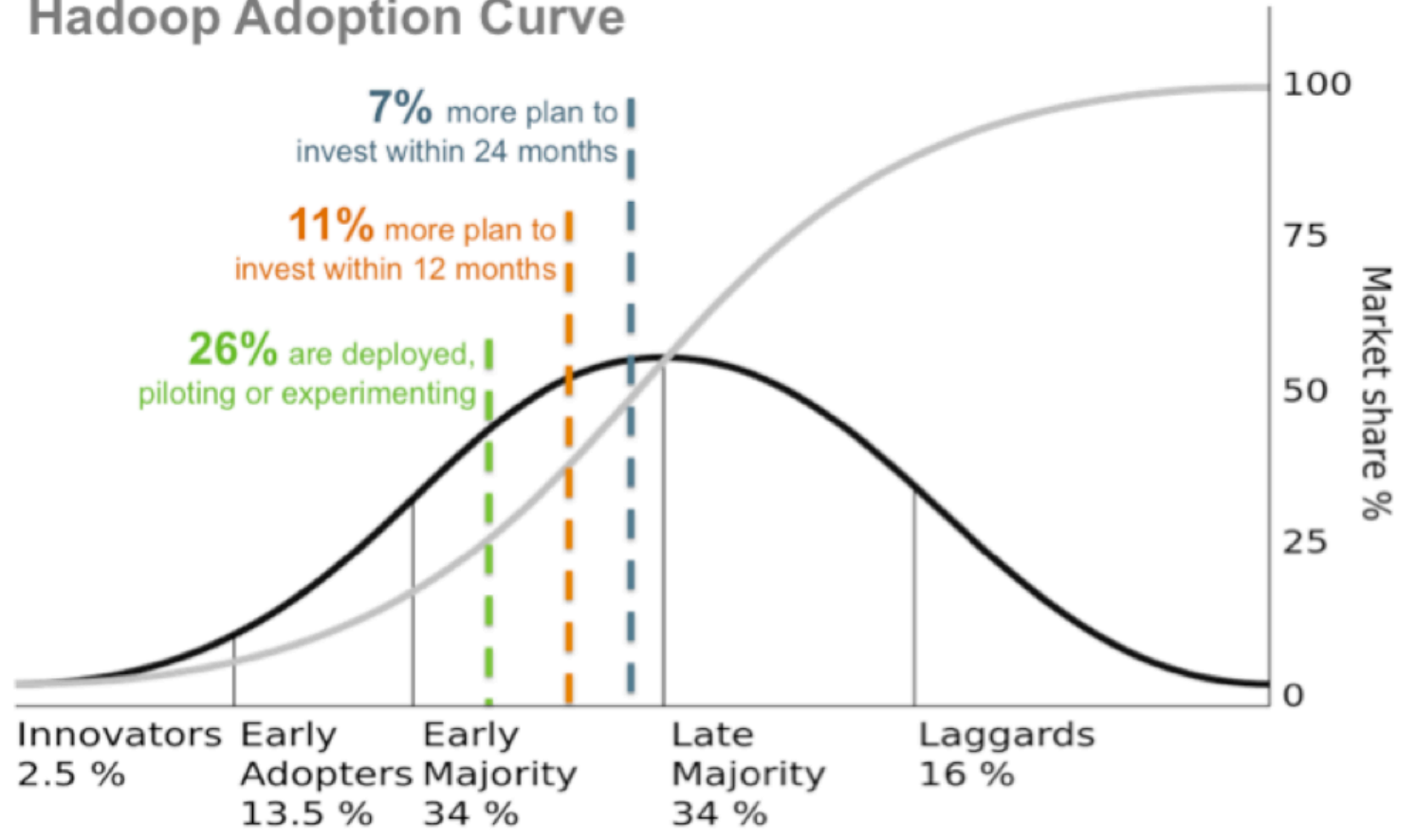
The open ecosystem of big data.

# How the Hadoop stack has grown



# Fast, unbridled growth has hurt adoption

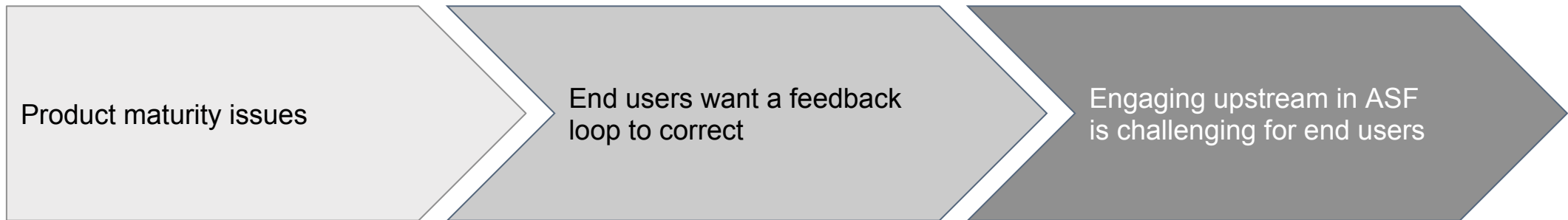
## Hadoop Adoption Curve



Source: <http://www.gartner.com/newsroom/id/3051717>



# Why is it so low?



# Let's look at some use-cases

# Linaro



## Improve supportability

- Hadoop projects have hardcoded x86 assumptions.
- Hard to connect to all relevant projects for upstream patches.

*“We submit a lot of patches [to Hadoop] as we found things hardcoded for x86 and need to add ARM support. Patches then go to both ODPI and Apache Software Foundation. That’s the whole point of being in ODPI: to simplify testing. Slow is smooth, and smooth is fast.”*

# GE



## Lack of consistency causes vendor lockin - very anti-open source

- HDFS lacks standard ways to interact with third party storage solutions (for example S3).
- User authentication/authorization is not standard across the stack - must implement their own solutions for managing.
- Forced to implement one-off strategies for Hadoop deployment per scenario.

*“Lack of standardization causes results in a “ripple effect” of other technical compromises to be considered and made”*

# Capgemini



## Create downstream value and an enterprise experience from a stable, consistent upstream

- Scaling up security and governance to the needs of a large organization doesn't work.
- Intra-cluster collaboration impossible with any degree of security.

*“This stuff is irrelevant, not that it’s not important, but it’s dial tone. [We want to] move the battleground out of Hadoop, not worry about version 2.5.something of whatever. We should be talking about everything that goes on top instead.”*





# Eastern Pools



## Providing answers to big data questions in an accessible way to SMB

- Tie into standard LOB applications ( accounting, scheduling, CRM ) in a turn-key way
- But, deliver less rigid results and enable them to drive unique differentiation
- Hadoop is inaccessible as-is to them.



*“How can we ensure that we are servicing our customer in personable, yet cost effective way, building on 60+ years of relationships?”*

# Downstream challenges solved through standardization

## Challenges



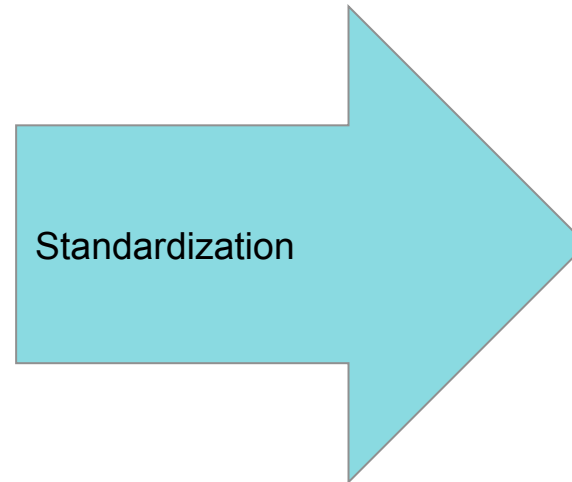
Limited flexibility in tool options, interoperability



Inconsistent/lack of support of stable tools



Ecosystem incompatibility across product lines



## Outcomes

Ability to match right tools to the task

Lowered costs of support - more time innovate

Broader offering of tools to a wider customer base

# Working on this as a community

EMC<sup>2</sup>

 DATATORRENT

 SQUID SOLUTIONS



NEC

 wanDISCO<sup>®</sup>

 altiscale

IBM<sup>®</sup>

 PLDT

 syncsort

 AsialInfo | 亚信数据

 4C Decision<sup>®</sup>

 Capgemini  
CONSULTING. TECHNOLOGY. OUTSOURCING

 REDOOP

splunk<sup>™</sup>

TOSHIBA  
Leading Innovation >>>

 Linaro

 CASK

 HORTONWORKS<sup>®</sup>  
POWERING THE FUTURE OF DATA<sup>™</sup>

Infosys<sup>®</sup>

 DriveScale

 XAVIENT  
Information Systems  
Great Ideas... Simple Solutions

 UNIFI

 Data<sup>INC.</sup>

 ampool

Pivotal<sup>™</sup>

 SAS<sup>®</sup>  
THE POWER TO KNOW.

vmware<sup>®</sup>

 Zettaset  
The Leader in Big Data Security

Large NA  
Automotive  
Vendor

 Xillab [싸이랩]  
eXperience, Idea and Insight

 ARENADATA





The open ecosystem of big data.