

Delivering a Flexible IT Infrastructure for Analytics

Click to edit Master subtitle style

with HDP on IBM Power Systems



Agenda

- Hortonworks and IBM Power Systems Partnership
- Customer Analytics Journey
- Open Community Innovation
- Leading Time to Insights

Hortonworks, IBM Collaborate to Offer Open Source Distribution on Power Systems

Latest Hortonworks Platform (HDP + HDF) to provide IBM customers with more choice in open source Hadoop distribution for big data processing

Las Vegas, NV (IBM Edge) - 19 Sep 2016: IBM (NYSE: IBM) and Hortonworks (NASDAQ: HDP) today announced the planned availability of Hortonworks Data Platform (HDP®) for IBM Power Systems enabling POWER8 clients to support a broad range of new applications while enriching existing ones with additional data sources.



Scott Gnau, CTO, Hortonworks
at Edge.

Youtube: <http://bit.ly/2dSOliW>



James Wade, Director of Application Hosting,
Florida Blue.

Youtube: <http://bit.ly/2dxVHIY>

INNOVATE

RENOVATE

About Hortonworks

The Leader in Connected Data Platforms

- ▣ Publicly traded on NASDAQ: HDP
- ▣ Hortonworks DataFlow for data in motion
- ▣ Hortonworks Data Platform for data at rest
- ▣ Powering new modern data applications

Partnering for Customer Success

- ▣ Leader in open-source community, focused on innovation to meet enterprise needs
- ▣ Unrivaled support subscriptions



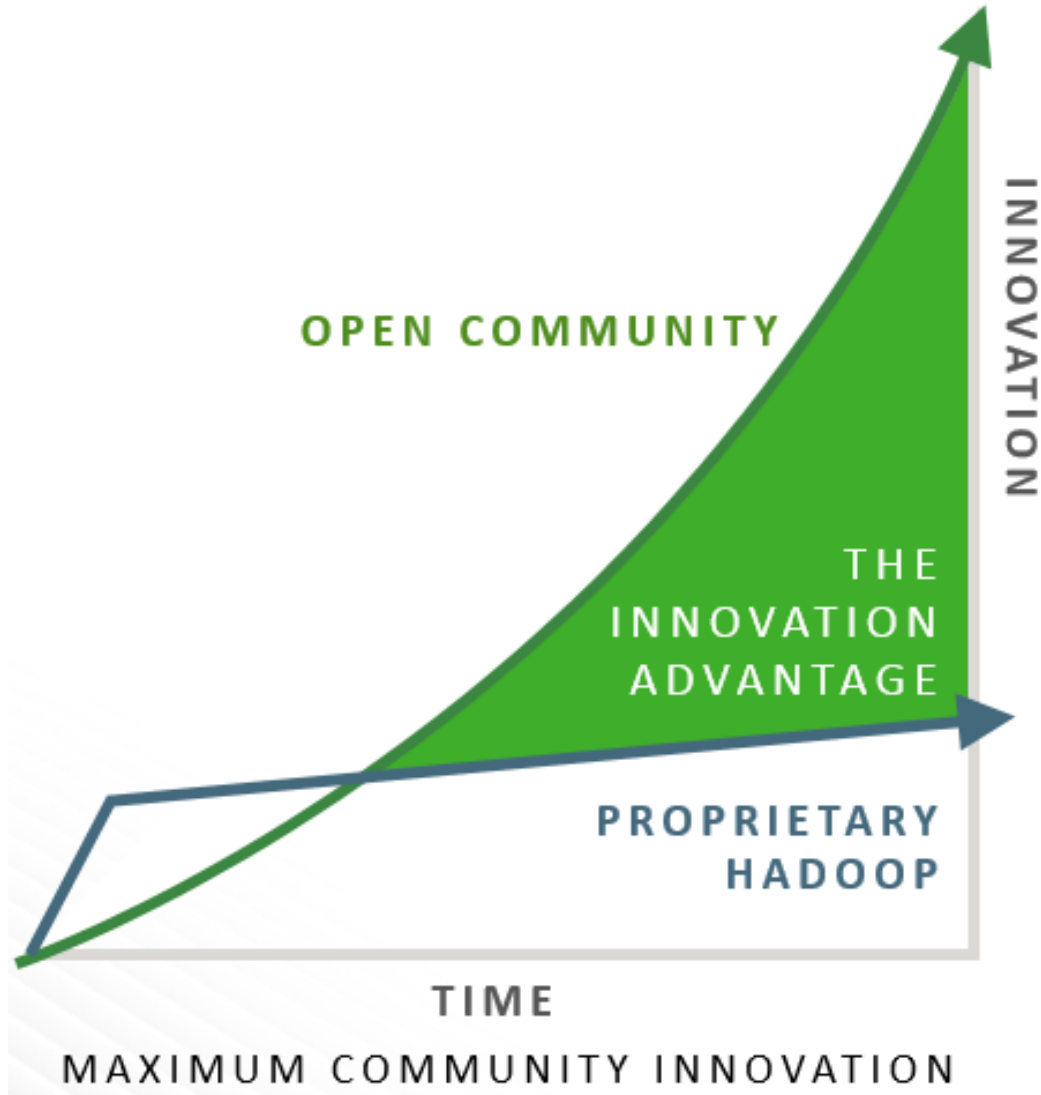
HORTONWORKS®

Original 24 Architects, Developers,
Operators of Hadoop from Yahoo!

1000
Employees
+

1500
Ecosystem
Partners

HPD is a 100% Open Source Connected Data Platform



Eliminates Risk

of vendor lock-in by delivering 100% Apache open source technology

Maximizes Community Innovation

with hundreds of developers across hundreds of companies

Integrates Seamlessly

through committed co-engineering partnerships with other leading technologies



We Employ the Committers

one third of all committers to the Apache® Hadoop™ project, and a majority in other important projects

Our Committers Innovate

and expand Open Enterprise Hadoop

We Influence the Hadoop Roadmap

by communicating important requirements to the community through our leaders

Apache Hadoop Committers

Hortonworks Nourishes the Community and Ecosystem

Hortonworks Community Connection



- Community Q/A Resources
- Articles & Code Repos!
- Community of (big data) developers

Hadoop & Big data ecosystem

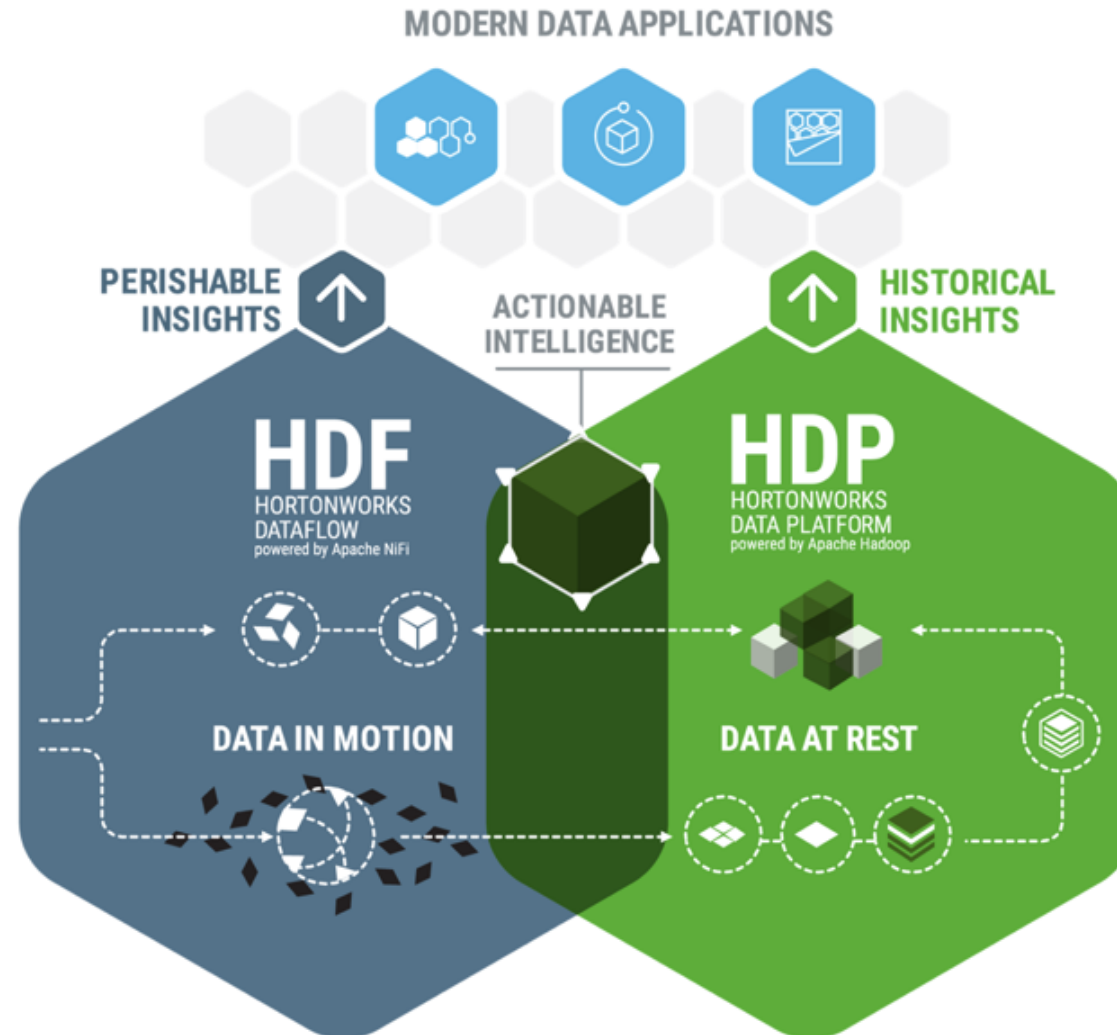


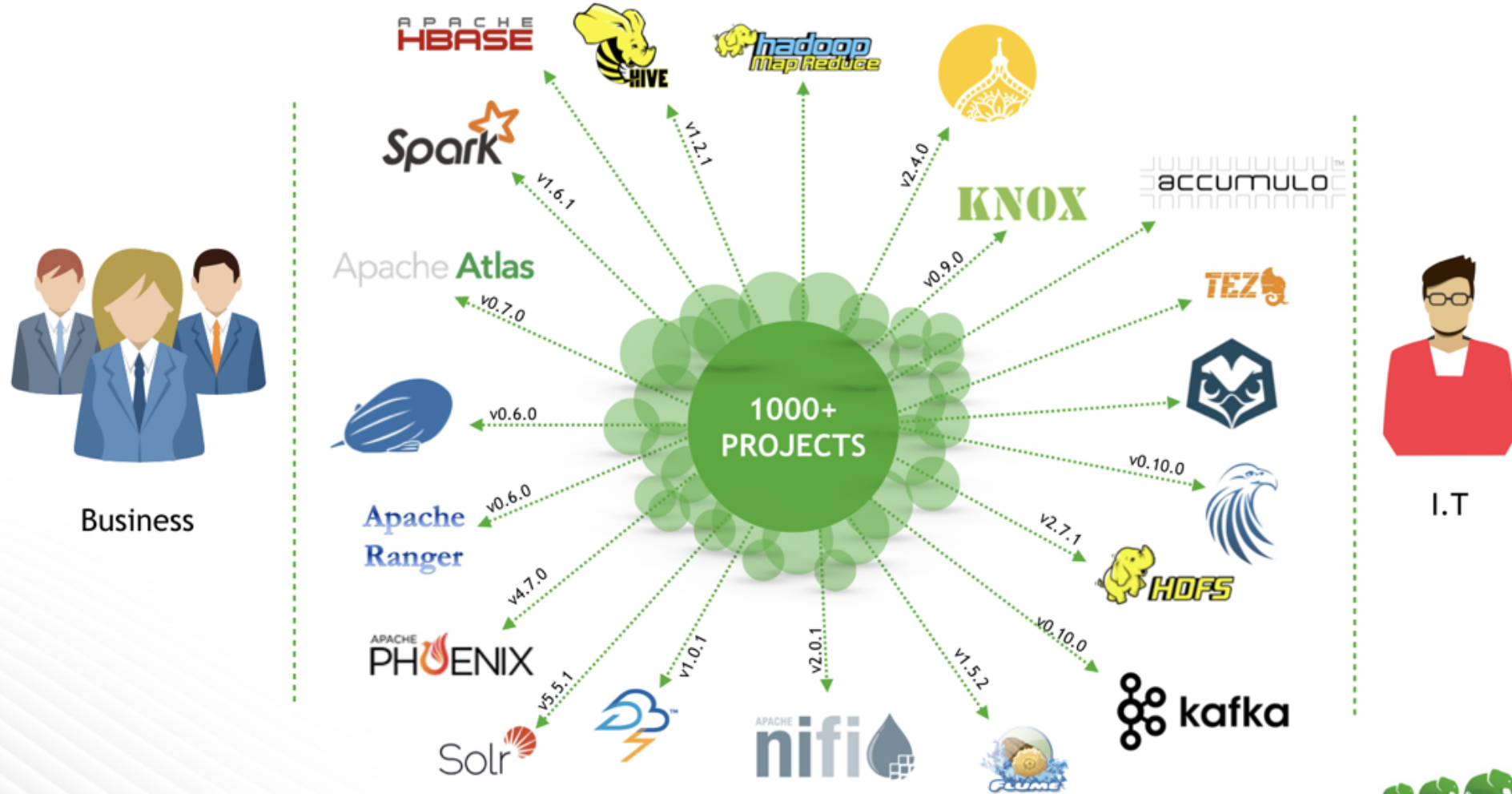
- Open Ecosystem of Big Data for vendors & end-users
- Advance Apache™ Hadoop®
- Enable more Big Data Apps

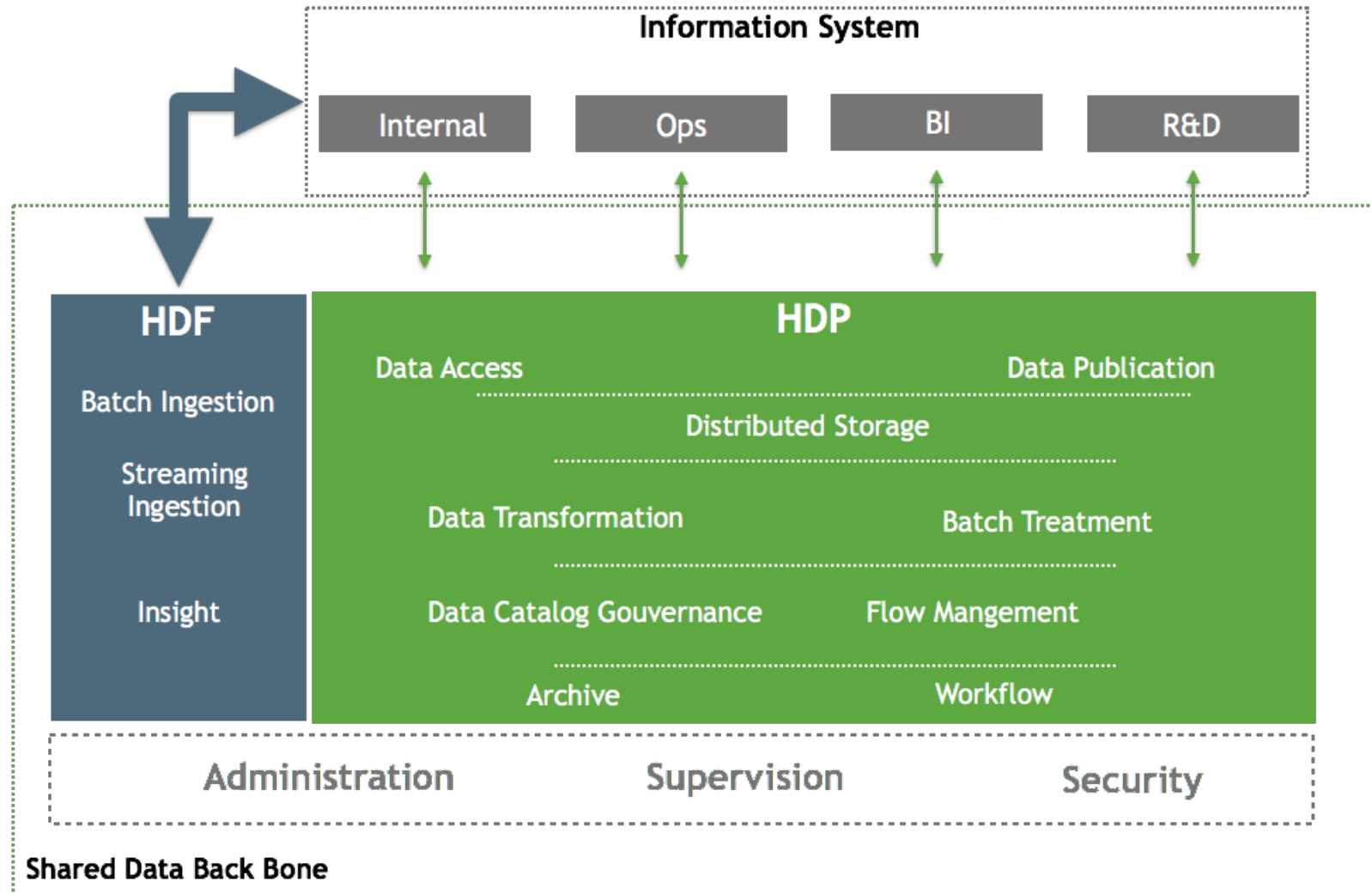
Hortonworks Partnerworks



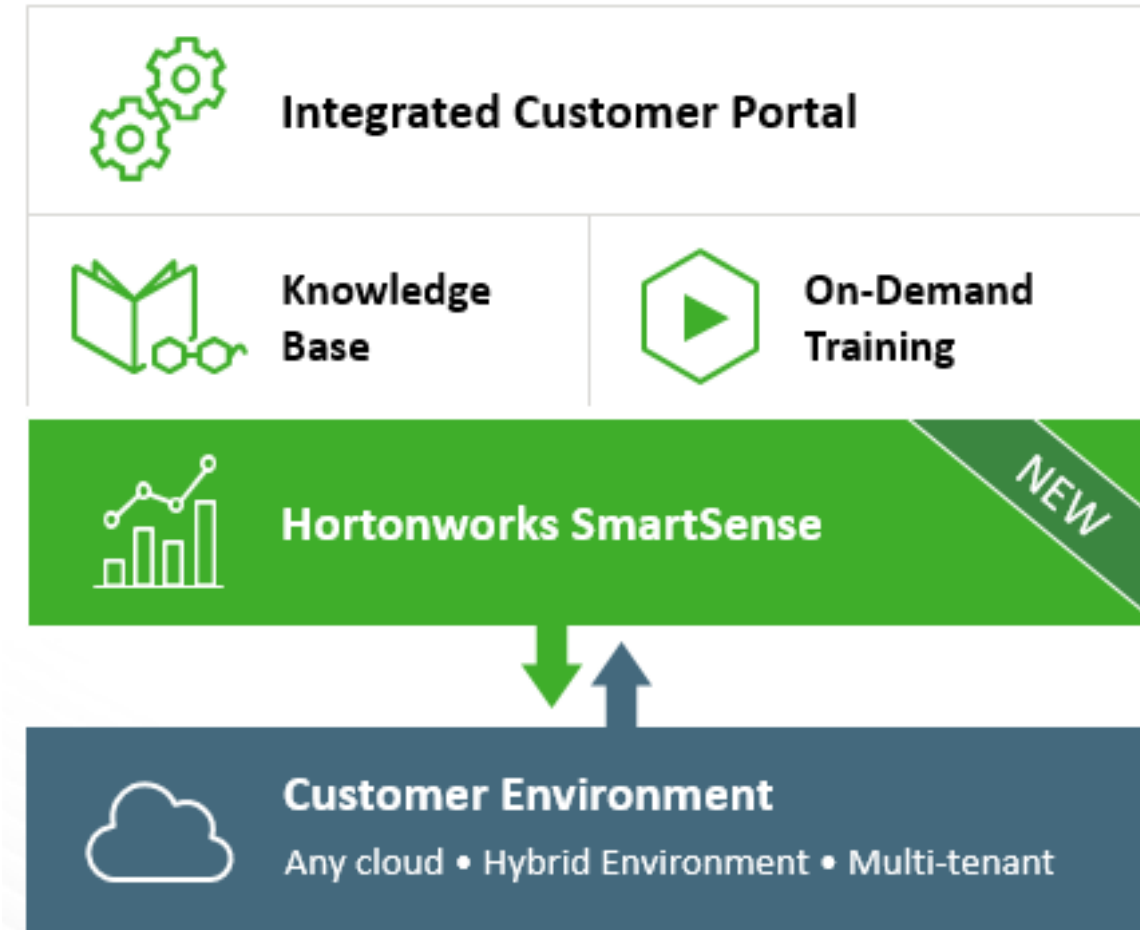
- World class partner program
- Network of partners providing best-in-class solutions







Hortonworks Delivers Proactive Support



Hortonworks SmartSense™
with machine learning and
predictive analytics on your cluster

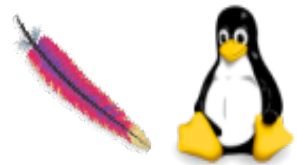
Integrated Customer Portal
with knowledge base and
on-demand training

- POWER

50k+ IBMers contributing to 150+ open organizations

Click to edit Master subtitle style

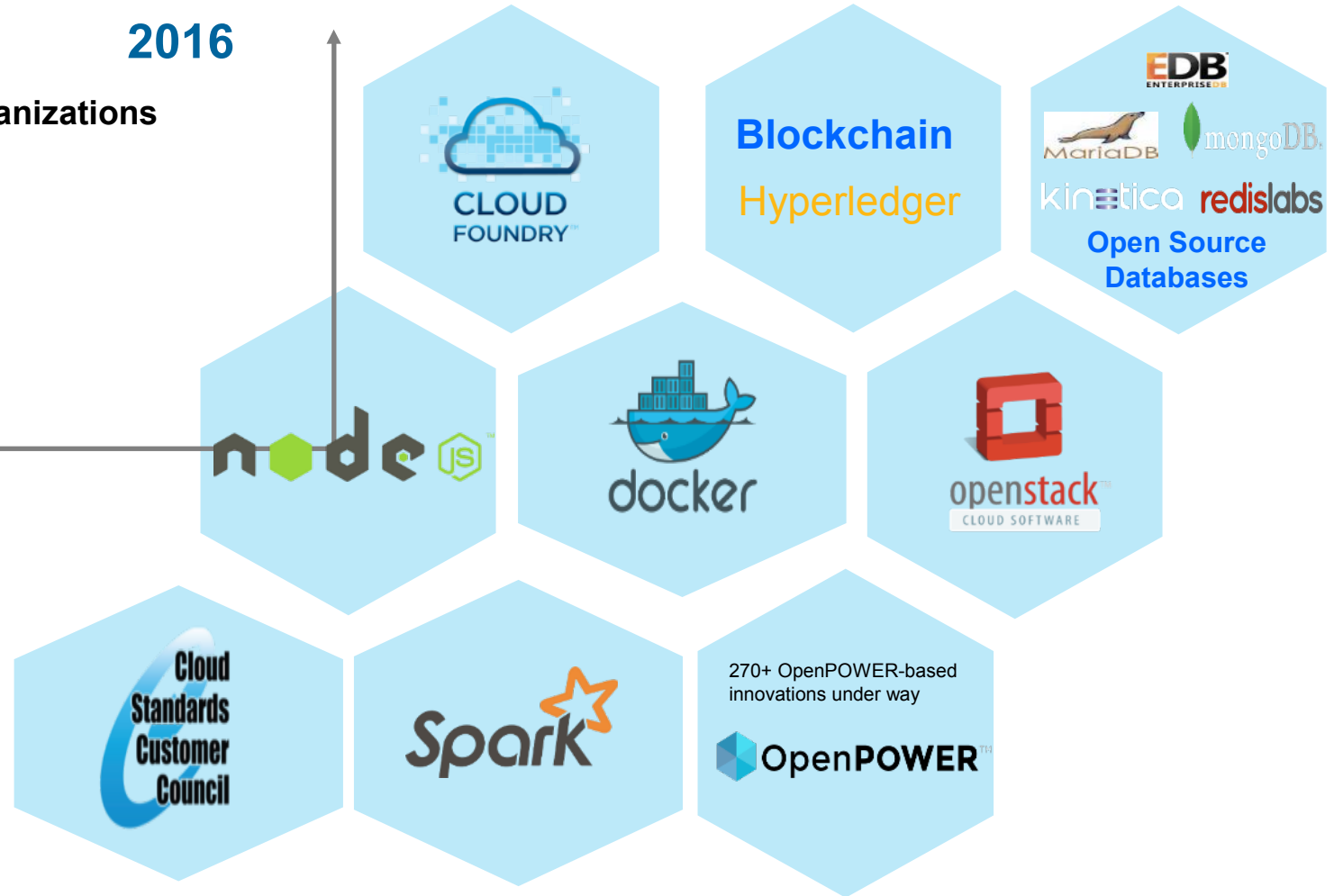
IBM is investing in the Linux ecosystems & open innovation



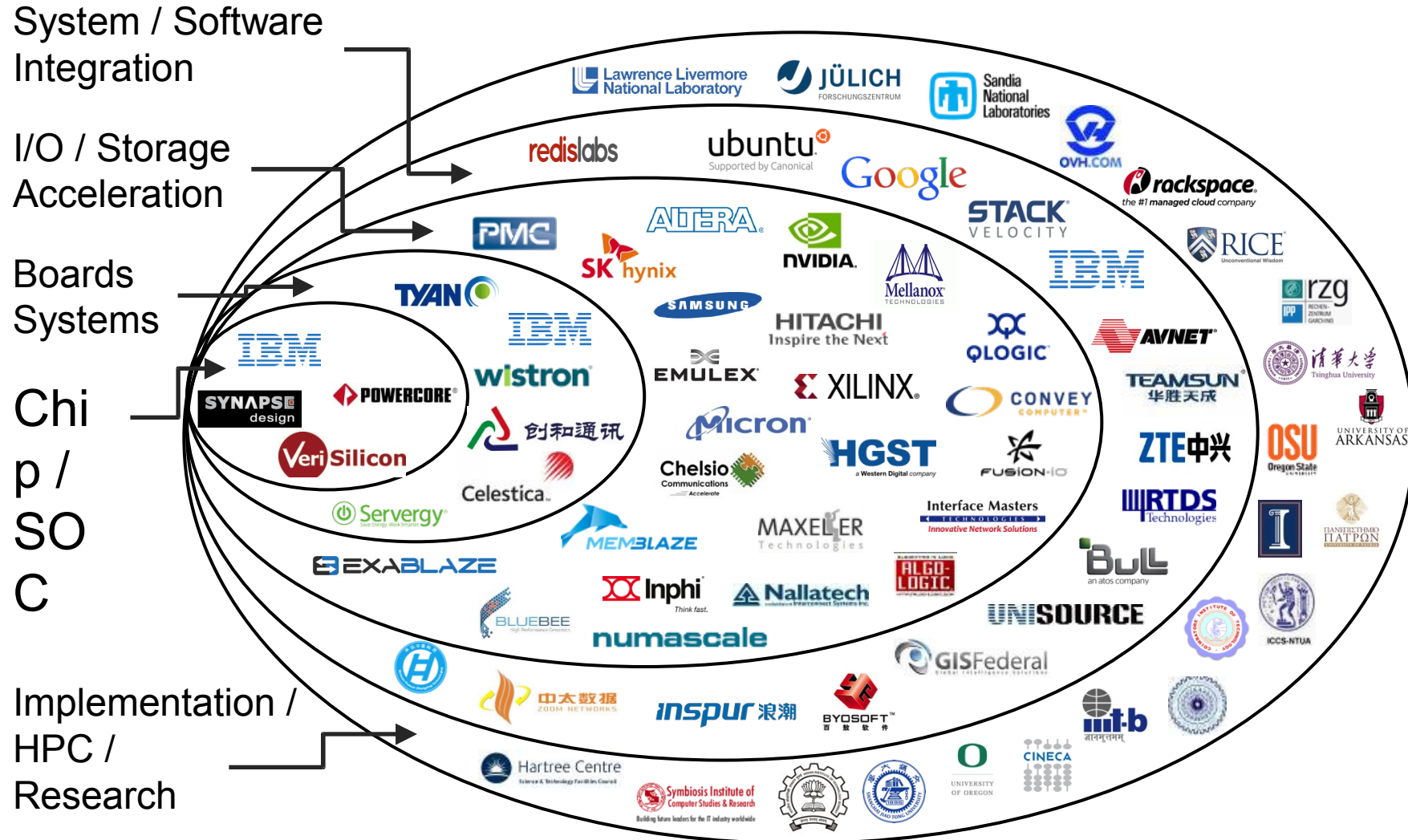
5 IBMers contributing to¹
Linux and Apache Projects

1999

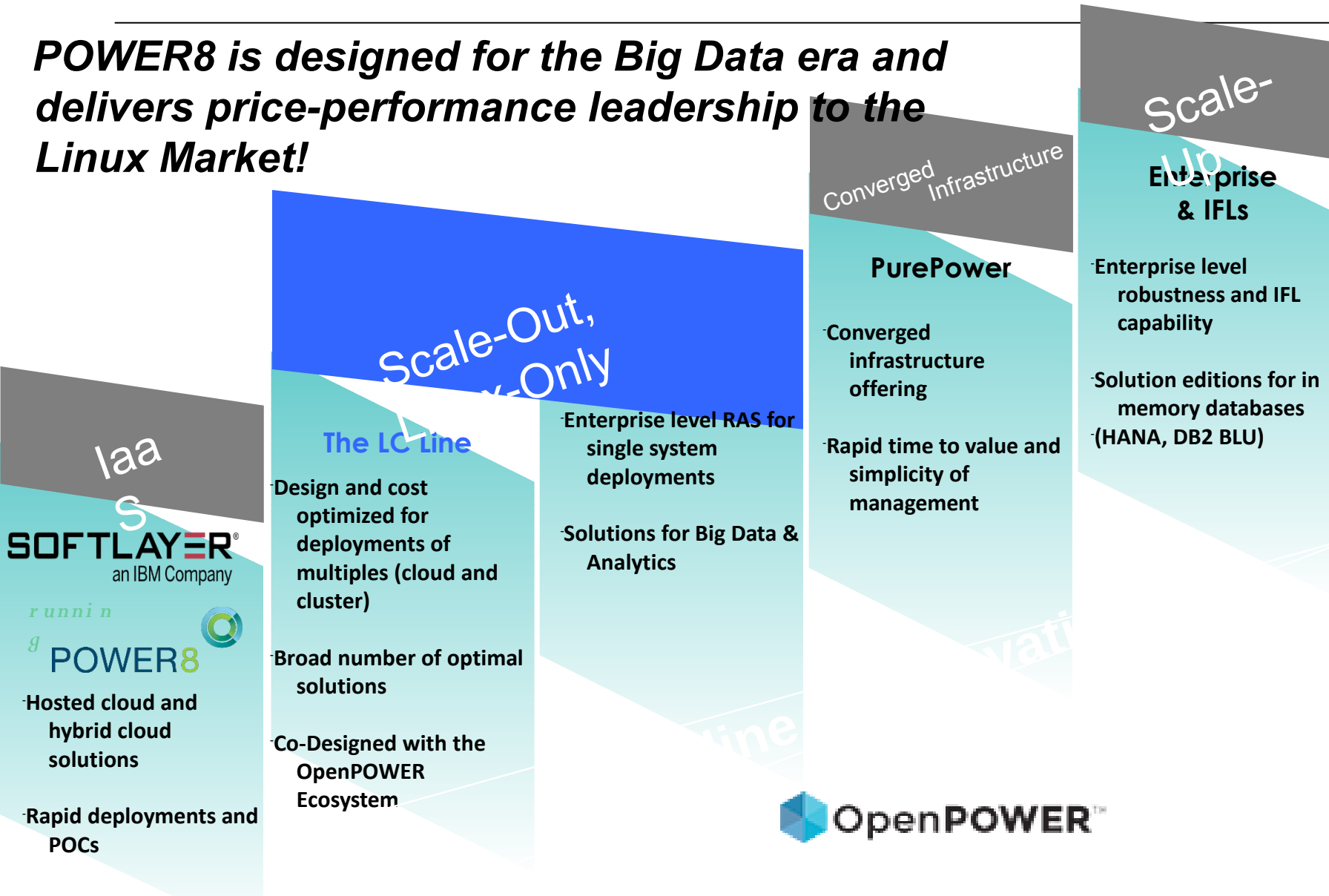
2016



1. Source:
<https://developer.ibm.com/start/>



POWER8 is designed for the Big Data era and delivers price-performance leadership to the Linux Market!



Broad Linux portfolio delivers all your Linux deployment needs **IBM Support**



Community / 3rd Party Support



Introducing the IBM Power Systems LC Line

OpenPOWER servers for cloud and cluster deployments that are different by design

High Performance Computing

Compute Intensive

Big Data

NEW

S822LC For Big Data



S812LC

- Storage rich single socket system for big data applications
- Memory Intensive workloads



- Ideal for storage-centric and high data through-put workloads
- Brings 2 POWER8 sockets for Big Data workloads
- Big data acceleration with work CAPI and GPUs

NEW

S822LC For High Performance Computing

Announce 9/8, GA 9/26



- Incorporates the new POWER8 processor with NVIDIA NVLink
- Delivers 2.8X the bandwidth to GPUs accelerators
- Up to 4 integrated NVIDIA "Pascal" GPUs

NEW

S821LC



- 2 POWER8 sockets in a 1U form factor
- Ideal for environments requiring dense computing

S822LC



- 2X memory bandwidth of Intel x86 systems
- Memory Intensive workloads

1

Open with Open
(Open Power)

Open distribution, leader of
Hadoop for open strategy

2

Enterprises Features
(Perf, RAS, Integration)

Brings +20 years experience of
Power Systems into Scale out

3

Accelerations / HPDA
(GPU, FPGA, CAPI)

Accelerations for high performance
Analytics and fastest data store

Value Proposition

Open Hadoop and Open Power –
embrace open solutions philosophy

Price performance leadership – fastest
time to innovation

Value Proposition

Performance – from 30 to 70% more
performance for MapReduce / SQL hadoop

Reliability and Availability from already
proven High End Power Technologies

Value Proposition

GPU more efficient for many (exponential and
non structured) data-intensive tasks than CPUs

GPU/FPGA accelerations to enable Machine
Learning and new Cognitive Breakthrough

SOLUTIONS

 OpenPOWER™

 SAMSUNG

 Google

 NVIDIA

 Mellanox

SOLUTIONS

 OpenCAPI™

 DELL

 Hewlett Packard
Enterprise

 AMD

SOLUTIONS

 torch

 Caffe

 TensorFlow™

 Chainer

1.7X Hadoop workload performance compared to Intel x86 (2)

(2) Performance results based on IBM internal testing.

 	IBM Power S822LC for BD (20-core, 256GB)	HP DL380 Gen9 (20-core, 256GB)
		
Server web price* -3-year warranty	\$18,598	\$23,194
QpH	288	169
Performance / Cost	9.14	4.31

1.70X @ 19%

More Performance

Lower HW costs and
maintenance

2.12X

Price-Performance

*Performance results are based on preliminary IBM Internal Testing of 10 TPC-DS based queries (simple, medium, and complex) with varying runtimes running against a 10TB database. The tests were run on 10 x IBM Power System S822LC for Big Data 20 cores / 40 threads, 2 X POWER8 2.92GHz, 256 GB memory, RHEL 7.2., HDP 2.5.3 compared to the published x86/Hortonworks results running on 10 x AWS d2.8xlarge EC2 nodes running HDP 2.5; details can be found at <https://hortonworks.com/blog/apache-hive-going-memory-computing/>. Conducted under laboratory condition, individual result can vary based on workload size, use of storage subsystems & other conditions. Data as of February 28, 2017

*Pricing is based on web prices for S822LC (<http://www-03.ibm.com/systems/power/hardware/s812lc/buy.html>) and HP DL380 Intel Xeon HP DL380; 20 cores / 40 threads, 2 X E5-2630 v4; , 256 GB (<http://h71016.www7.hp.com/dstoreHPE/MiddleFrame.asp?page=config&ProductLineId=431&FamilyId=3852&BaseId=45441&oi=E9CED&BEID=19701&SBLID=>)

Customer Story – Guidewell Health - Florida Blue

GUIDEWELL Florida Blue Companies

- Business Problem
 - Transformational journey resulting in rapid expansion of business models
 - Technology innovation required to keep up with the business expansion while improving client satisfaction, reducing costs and supporting the company's green IT initiatives
 - Existing x86 server sprawl not sustainable
- Solution with Hortonworks, IBM OpenPOWER servers and Sage Solutions Consulting
 - Embraces the open software and hardware model adopted by Florida Blue
 - Hortonworks supporting new fraud analytics initiative to reduce costs and client premiums
 - OpenPOWER to enable smaller datacenter footprint with stronger reliability

Read more: <http://bit.ly/PowerHDP>

Surrounding Solutions

IBM Data Science Experience

#1 Data Science Platform

- Community and social features to provide collaboration
- The best of open source and IBM value-add to create state-of-the-art data products
- Built-in learning to get started or learn more with advanced tutorials

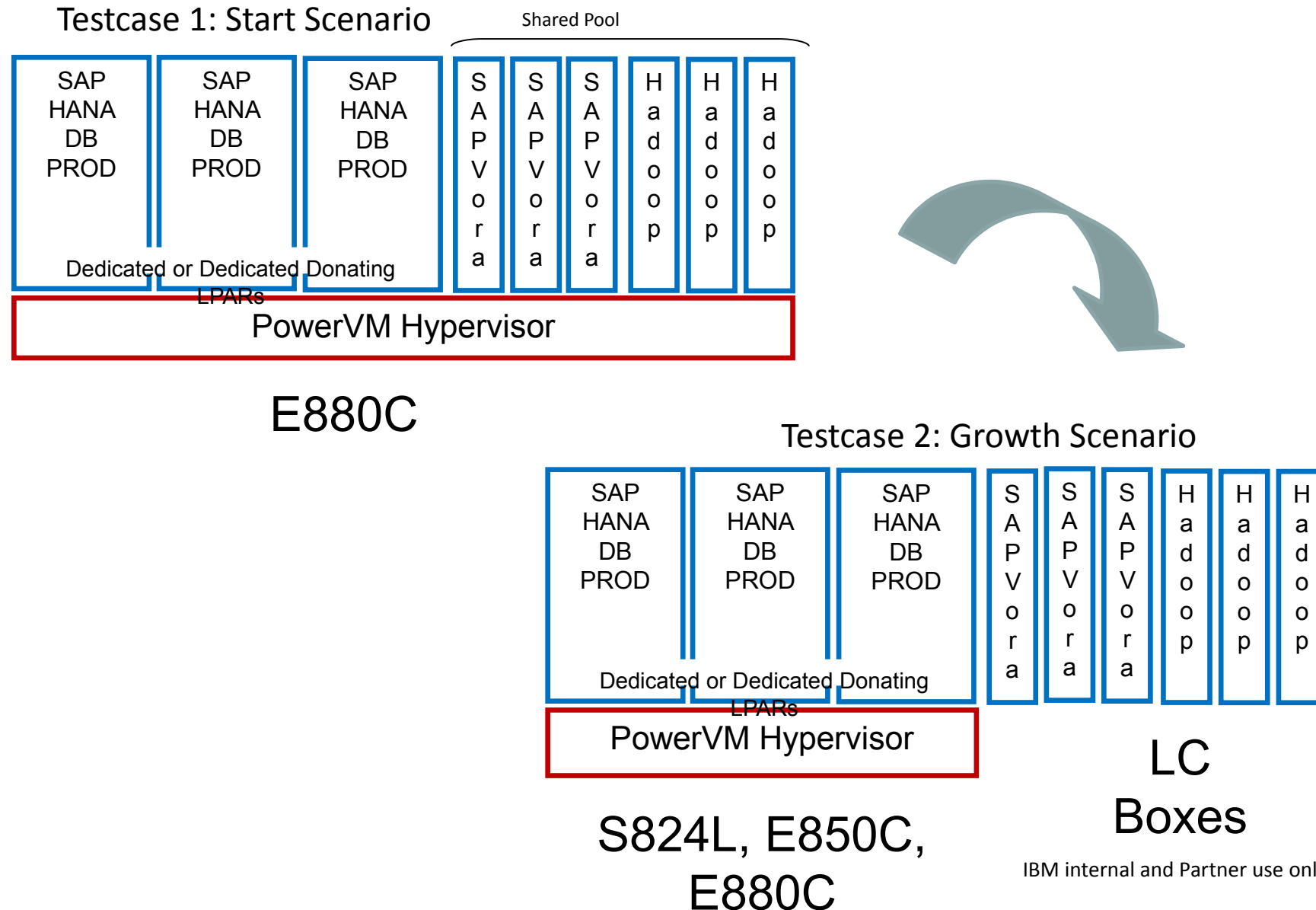
IBM Big SQL

#1 SQL Engine for Hadoop

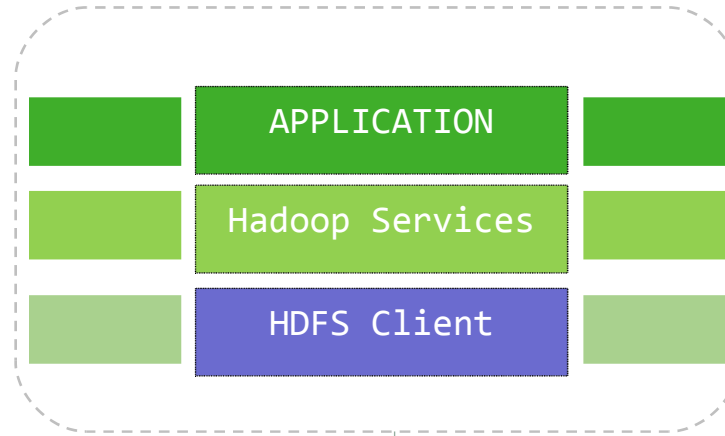
- Data virtualization layer
- Large data volume, extremely complex query support
- Supports low latency, high concurrency workloads



#1 Open Hadoop Distribution

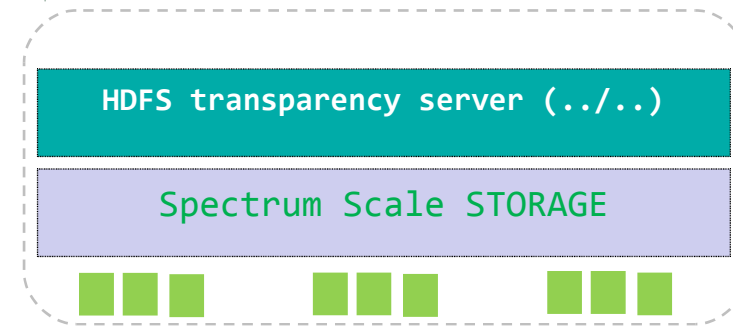
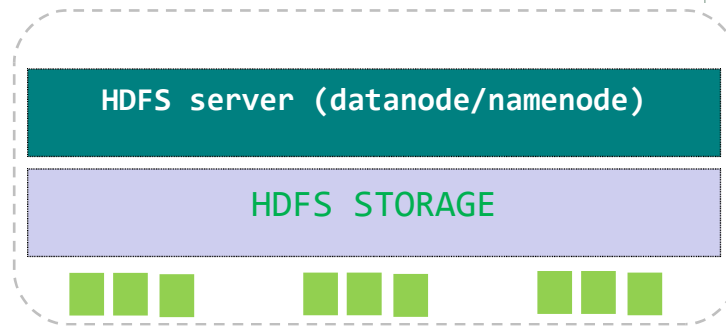


Seamless: Uses Standard HDFS client



Performance: Uses HDFS client cache
Own cache mechanism

RPC
Call



Improvement: replaces HDFS connector

Extreme Scalability

parallel architecture with no bottleneck
Every node serves as metadata or data

Better Data Management

Global namespace can span over multiple clusters
Datalake ready
Cloud Tiering for powerful ILM

Reduced Data Center Footprint

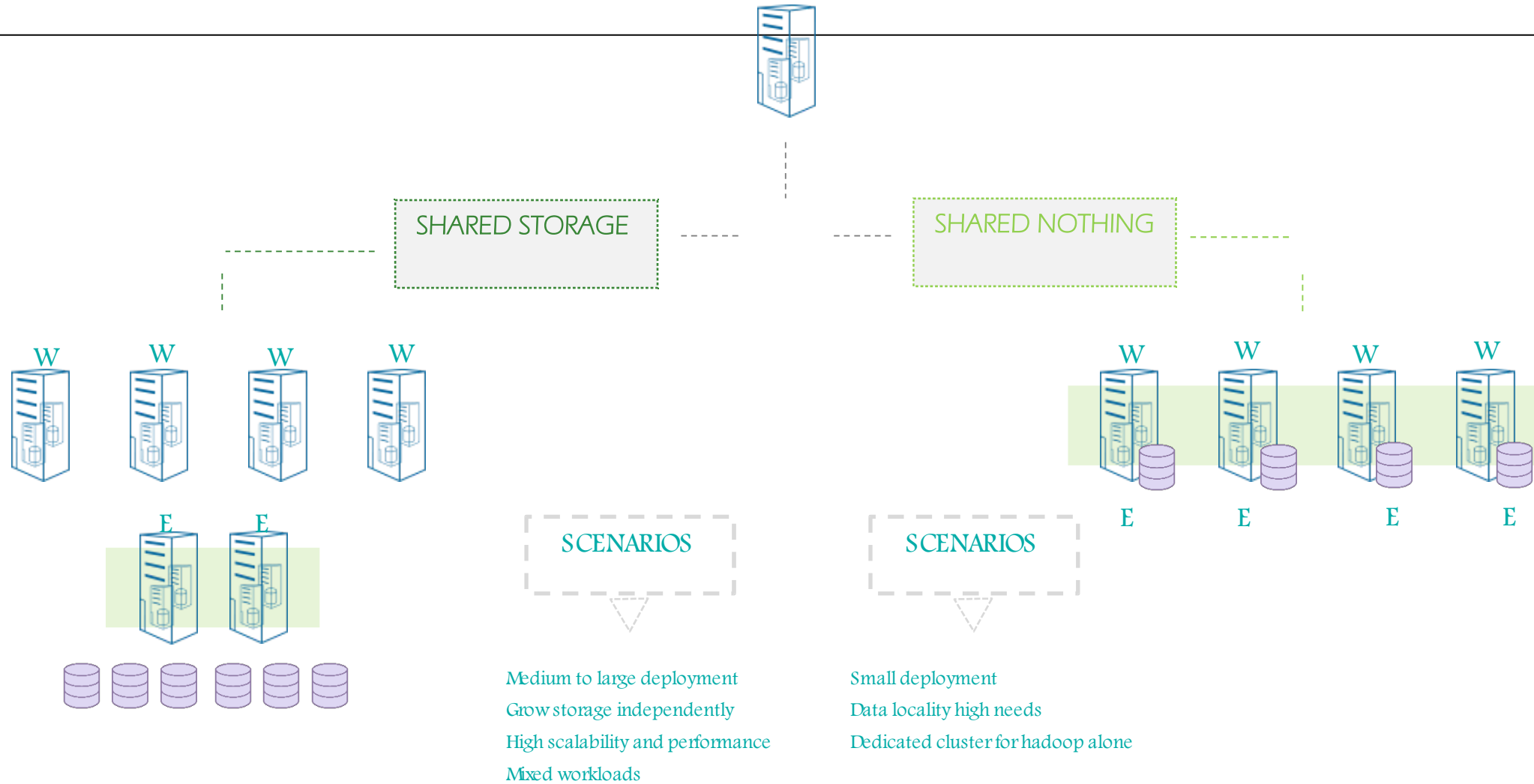
Multiple data access protocols (including POSIX)

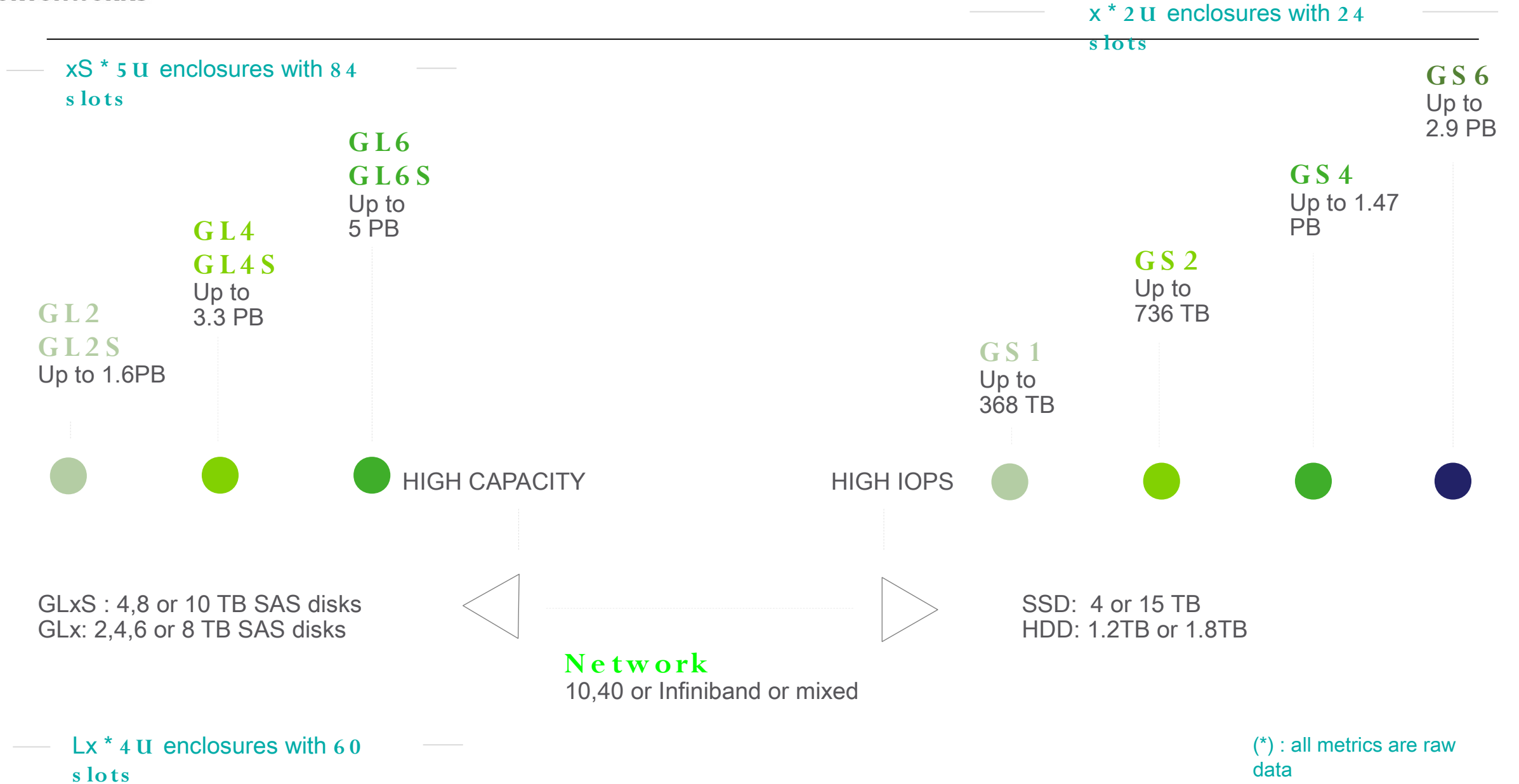
1.4 copies instead of 3

Enterprise Hardware

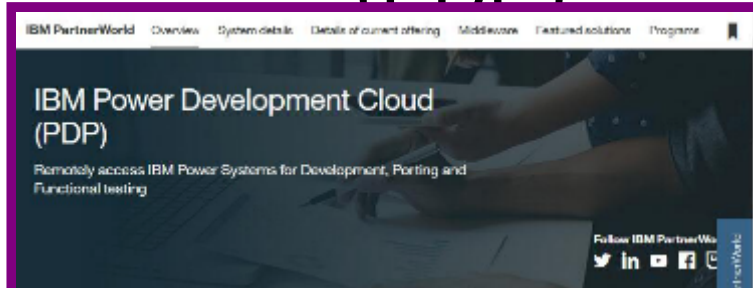
more performance and scalability
Less capacity requirements than competition
Regulatory compliance and encryption







Accessing POWER Hardware in IBM Power Dev Cloud (PDP)

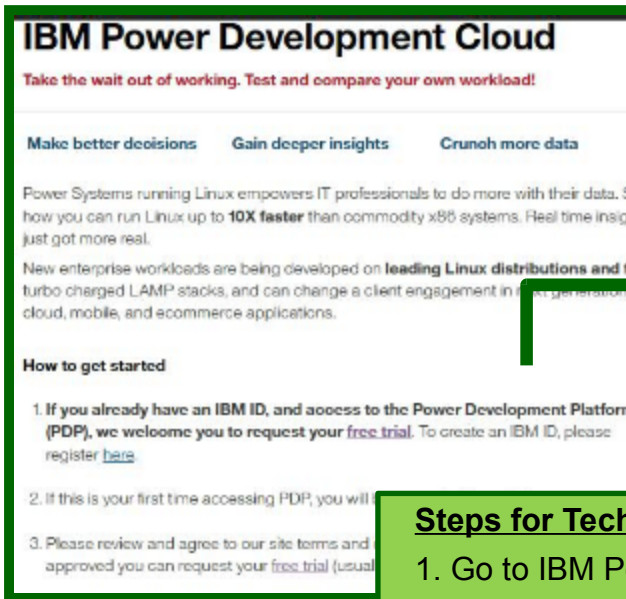


Steps for ISVs

1. Go to IBM Power Development Cloud (PDP) for PartnerWorld members: Link
2. Scroll down and click *How to* for resources on requesting/using hardware.
3. Click *Get Started* and follow instructions.

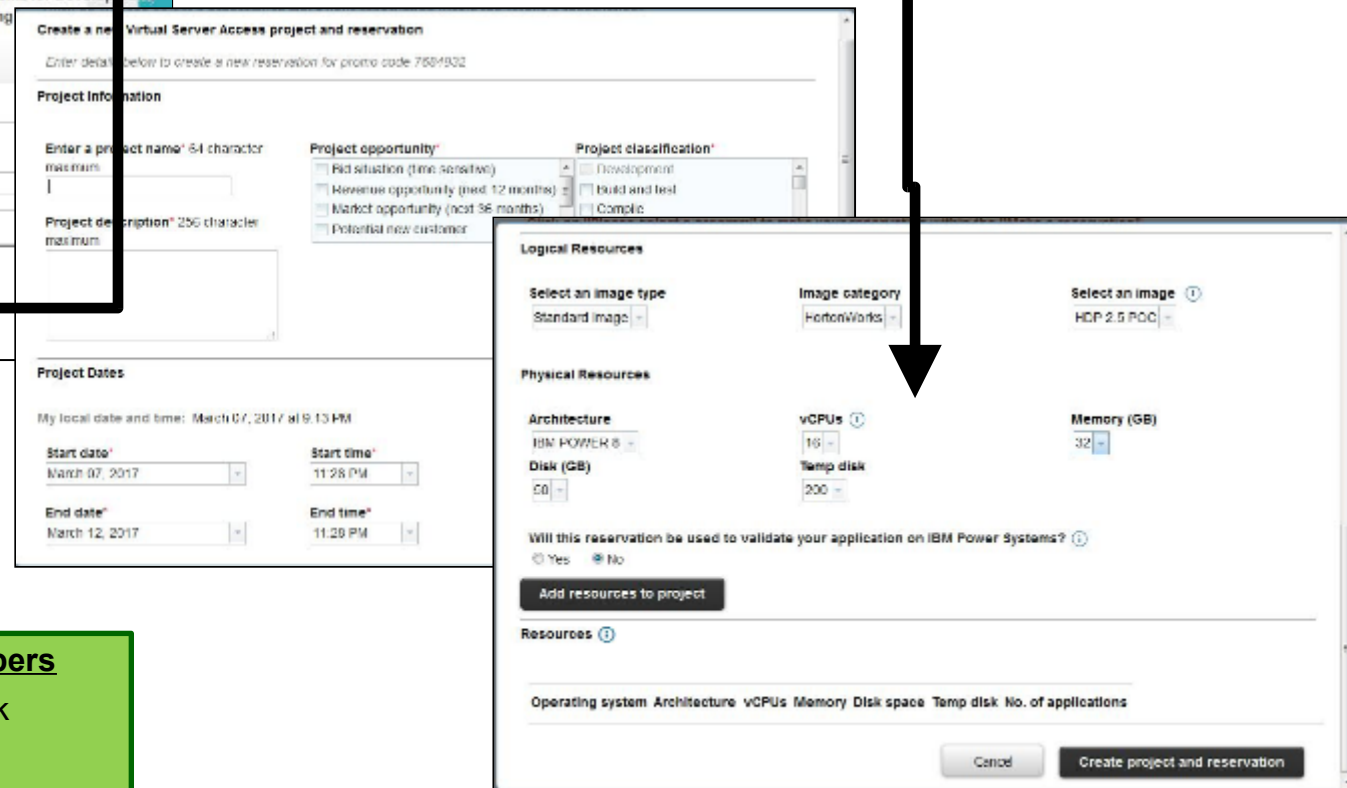
Note:

- When making the hardware request, use the **promo code** for your desired configuration size: **7657934** (small), **7684932** (medium), **7928463** (large), _____ (cluster)
- Select the max resources value allowed by the promo code.



Steps for Tech Sellers and Open Source Developers

1. Go to IBM Power Development Cloud (PDP): Link
2. Follow the instructions.



Logical Resources

Select an image type: Standard image
Image category: HortonWorks
Select an image: HCP 2.5 PDC

Physical Resources

Architecture: IBM POWER8
Disk (GB): 60
vCPUs: 16
Temp disk: 200
Memory (GB): 32

Will this reservation be used to validate your application on IBM Power Systems?
Yes No

Add resources to project

Resources

Operating system Architecture vCPUs Memory Disk space Temp disk No. of applications

Cancel Create project and reservation

Easy way for HDP to be introduced to existing large Power customers!

Digital Business Start for Power Enterprise Servers with Hortonworks Data Platform:

- Tailored for customers who already have IBM POWER8 enterprise class servers in their environment (E850C, E870C, E880C, E850, E870 or E880)
- Quickly deploy an HDP cluster through Ambari
- Great for education, light weight validation prior to production
- Free 30-day to 3-month trial offer
- Enable inactive cores and memory
- Small, medium, large config options



Quick Start Guide link: [coming soon](#)

- 📖 For POWER administrators
- 📖 Steps for install/config/verify

1	INTRODUCTION: "PURPOSE" OF GUIDE
2	CUSTOMER USE CASES
3	GETTING STARTED
4	RED HAT ENTERPRISE 7.2 LE INSTALLATION
5	POWERVC
6	PREPARE SERVER FOR HDP INSTALLATION
7	AMBARI SERVER INSTALLATION CONFIGURATION.....
8	HDP INSTALLATION CONFIGURATION
9	BASIC TEST FOR HDP CLUSTER VALIDATION

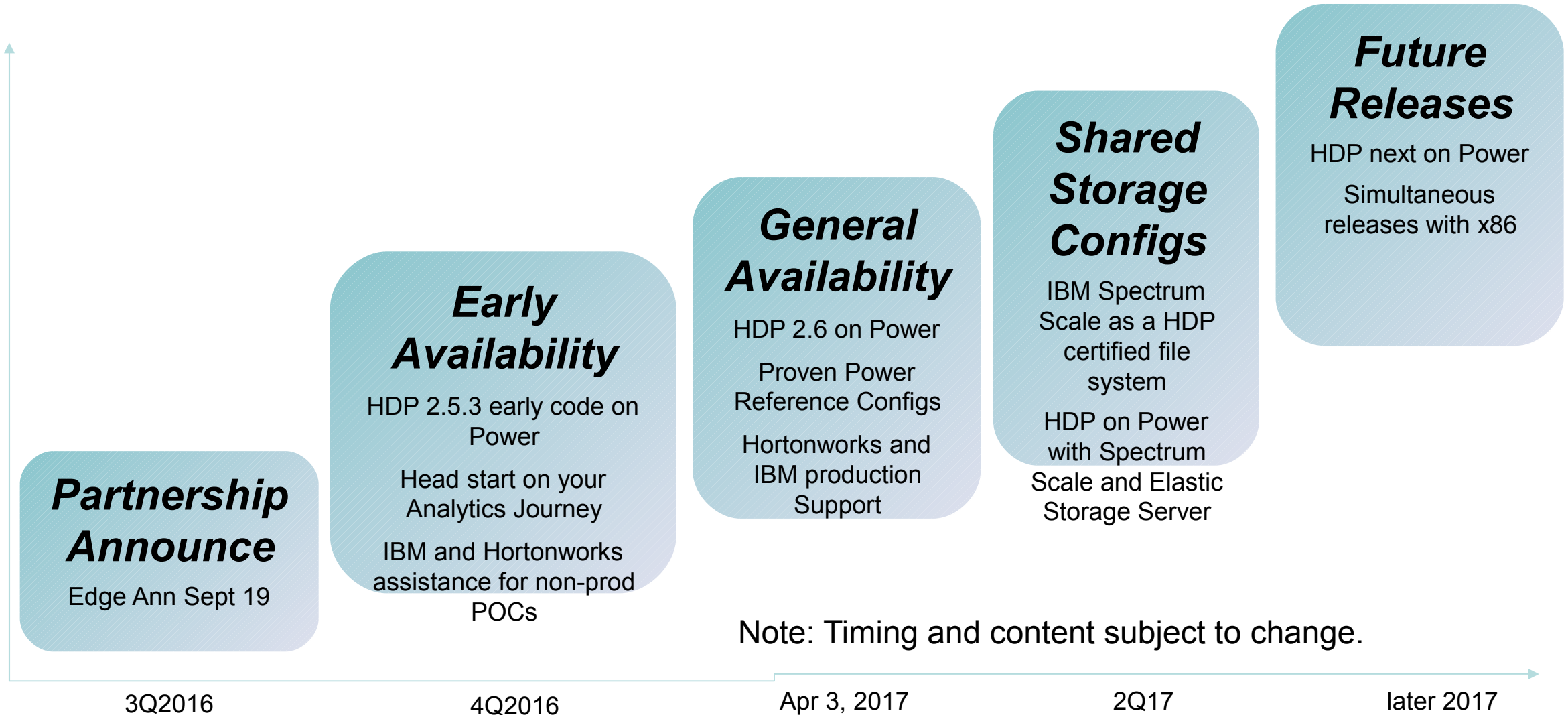
Medium Size Config Example

Digital start configs	Node cluster	VCPU (per node)	Memory (GB) (per node)	Storage FC-NPIV (per node)
HDP-Medium	1 HDP Master	4	32	100GB
	3 HDP worker	1	8	100GB

How to Get Started with HDP on OpenPOWER Systems

- Join the Hortonworks Community: <https://community.hortonworks.com/>
- Learn more about the benefits of Hortonworks: <http://hortonworks.com/training/>
- Learn more about the benefits of IBM Power Systems and OpenPOWER
- If you are interested in discussing a HDP on Power Systems option or proposal, talk to your Hortonworks or Power sales reps

Hortonworks (HDP) on Power Roadmap



POWER8: Designed for data to deliver breakthrough performance

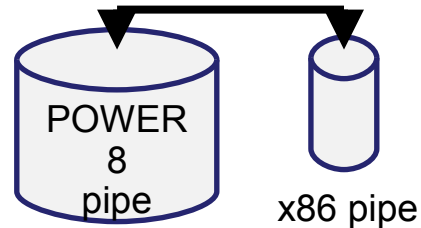
4X
Threads per core

POWER8
SMT8
Parallel Processing

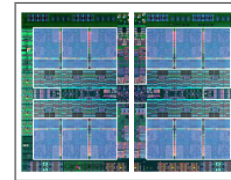
x86

Hyperthread

4X
Mem.
Bandwidth¹

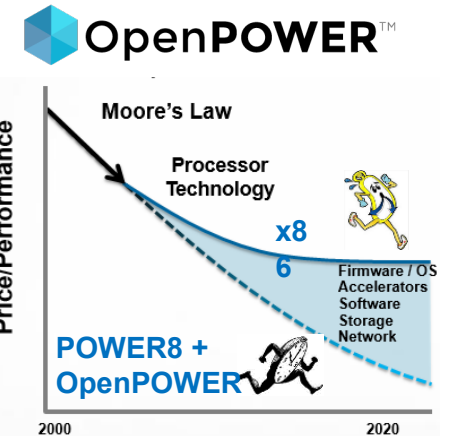


6X
More cache² @
Lower Latency



x86

POWER8



These design decisions result in best performance for data centric workloads like:

Spark, Hadoop, Database, NoSQL, Big Data Analytics, OLTP

SMT=Simultaneous Multi-Threading
OLTP = On-Line Transaction Processing

¹Up to 4X depending on specific x86 and POWER8 servers being compared
²Up to 6X more cache comparing Intel e7-8890 servers to 12 core POWER8 servers. See speaker notes for more details



Workload definition:

- **What is the peak ingestion rate required for the data?**
- **Please describe what type of data (structured, unstructured, semi-structured)?**

Capacity sizing.

- **What is the raw data size (in TB) the environment is initially required ?**
- **What is the compression rate the customer assumes ?**
- **What is the data growth rate over what period of time?**

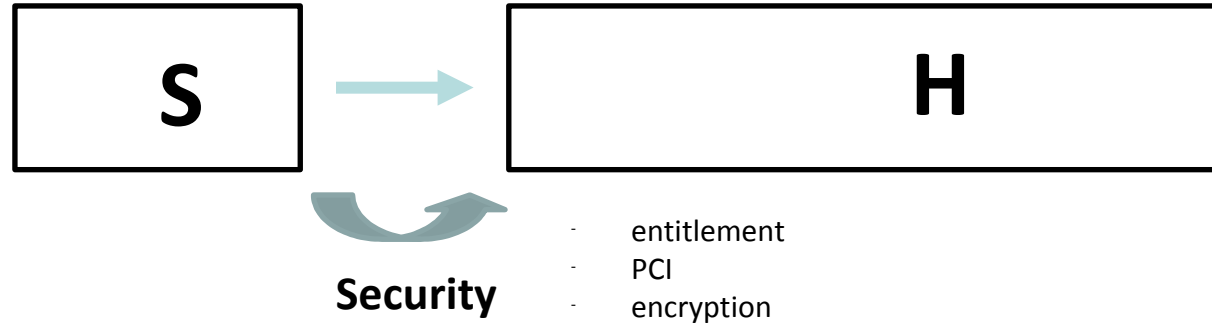
What are the HA / DR requirements for the environment ?

Rule of thumb is : $4 * \text{useable}$

Contact us for sizing and design for Hortonworks on power

1

Offload cold data from SAP HANA to Hadoop



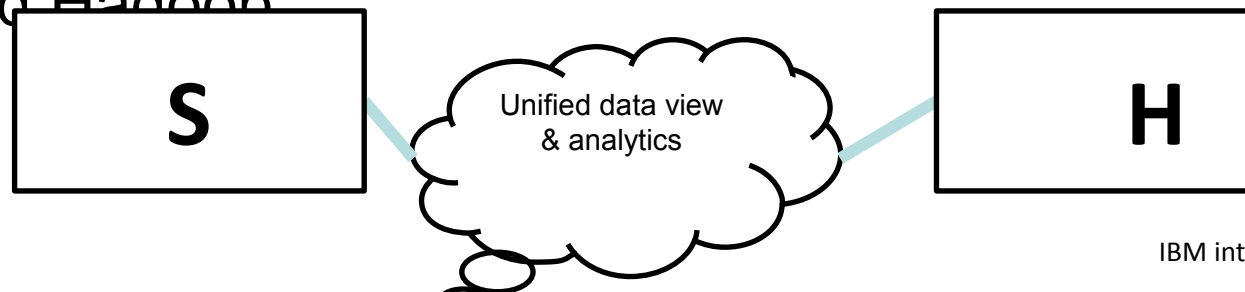
2

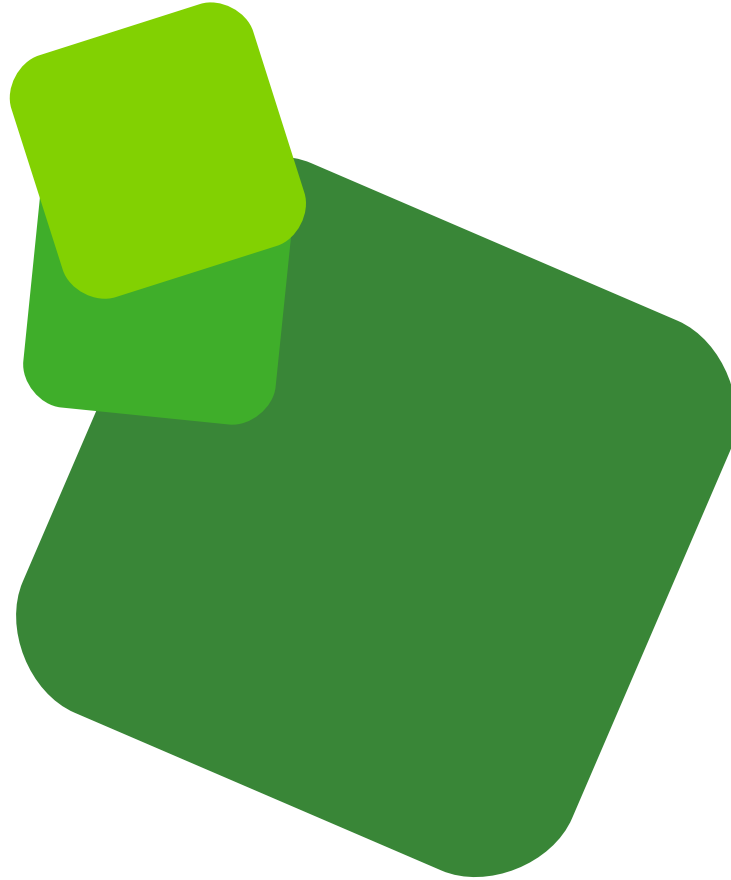
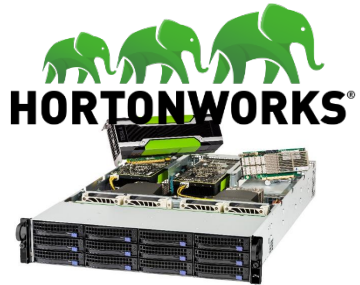
Data pre-analysis in Hadoop and Aggregation to Hana



3

Single Data Platform to span SAP HANA and Hadoop





Hortonworks Sizing



STORAGE DENSE

2-socket, 2U
Up to 20 cores (2.9-3.3Ghz)
512 GB Memory (16 DIMMs)
115GB/sec memory bandwidth
12 SFF/LFF (HDD/SSD) 96 TB 5 2
NVIDIA K80 GPU capable

Open Power

X86 Price

Performance

Enterprise



COMPUTE DENSE

2-socket, 1U
Up to 20 cores (2.9-3.3Ghz)
512 GB Memory (16 DIMMs)
115GB/sec memory bandwidth
4SFF/LFF (HDD/SSD) 32 TB NVIDIA
K80 GPU capable

Sys Mgt Node	Edge Node	Master Node(s)	Worker node(s)		
(min/max)1 / 1	(min/max)1 / 1	1 / 3 / any	3 / 8 / any		
			balanced	Performanc	Storage
CPU : 8 cores	CPU : 20 cores	CPU : 20 cores	CPU : 22 cores	CPU : 22 cores	CPU : 11 cores
Mem : 32 GB	Mem : 256 GB	Mem : 256 GB	Mem : 256 GB	Mem : 512 GB	Mem : 128 GB
Storage : 2 x 4TB HDD	Storage : 2 x 4TB HDD	Storage : 2 x 4TB HDD	Storage : 12 x 4TB	Storage : 8 x 6TB 4x3.8 ssd	Storage : 12 x 8TB
Network : 4 ports internal 2 ports	Network : 4 ports internal 2 ports	Network : 4 ports internal 2 ports	Network : 4 ports int 2 ports	Network : 4 ports int 2 ports	Network : 4 ports 2 ports
	Model : S821LC		Model : S822LC		

SWITCHES

- 1 GbE (1x or 2x): Lenovo G8052 Switch (48x 1GbE + 4x 10GbE ports)
- 10 GbE (2x typical, 1x allowed): Lenovo G8264 Switch (48x 10GbE + 4x 40GbE), or Mellanox SX1410 Switch (48x 10GbE + 12x 40GbE)

Challenge

RETAIL

The Retail industry is very competitive so customer loyalty programs are essential. Metro collected a significant amount of information about their clients but didn't keep any of the data in-house, sending it out to an external firm for analysis.

by retaining their store level data, they could provide more efficient and flexible in-store experiences for their clients application. They needed to

How IBM won

- Existing Power client (IBM i), would like to test Power
- truly a team effort by quickly pulling the resources together
- The client team also took advantage of the IBM Power Deal Closer Program to fund IBM Lab Services.
- strong relationship with Hortonworks, where we are working together to identify opportunities and close business

Details

9 x S822LC - CANADA

Challenge

INDUSTRY

The client needed to start with Hadoop to store data from their products (Industry 4.0) and also from Legacy systems. They did not know, how they will analyze the data in future, but they needed to start storing it. They choose Hortonworks as a partner to start with Hadoop. In parallel they asked their hardware business partner about possible hardware solutions.

We had the challenge that Power systems are not mainstream for Hadoop. The client is a loyal IBM i client and knows Power Systems

How IBM won

- We showed good benchmark values with a good price
- Hortonworks also recommended Power Systems as a very good alternative to Intel
- A good relationship of ebm-papst to the business partner and also to IBM (loyal client)
- The client did not think about Power as a solution for Hadoop. When discussing the project with the business partner and IBM, we recommended him also to think about Power Systems.

Details

3 x S822LC - GERMANY