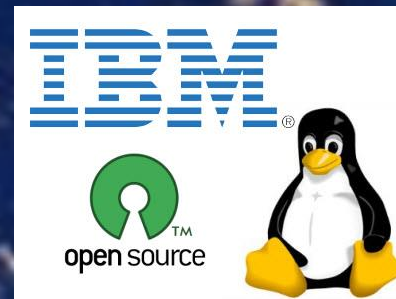


Modern Data Platform & Open Source Data Base



Bologna, 11 Aprile





75%

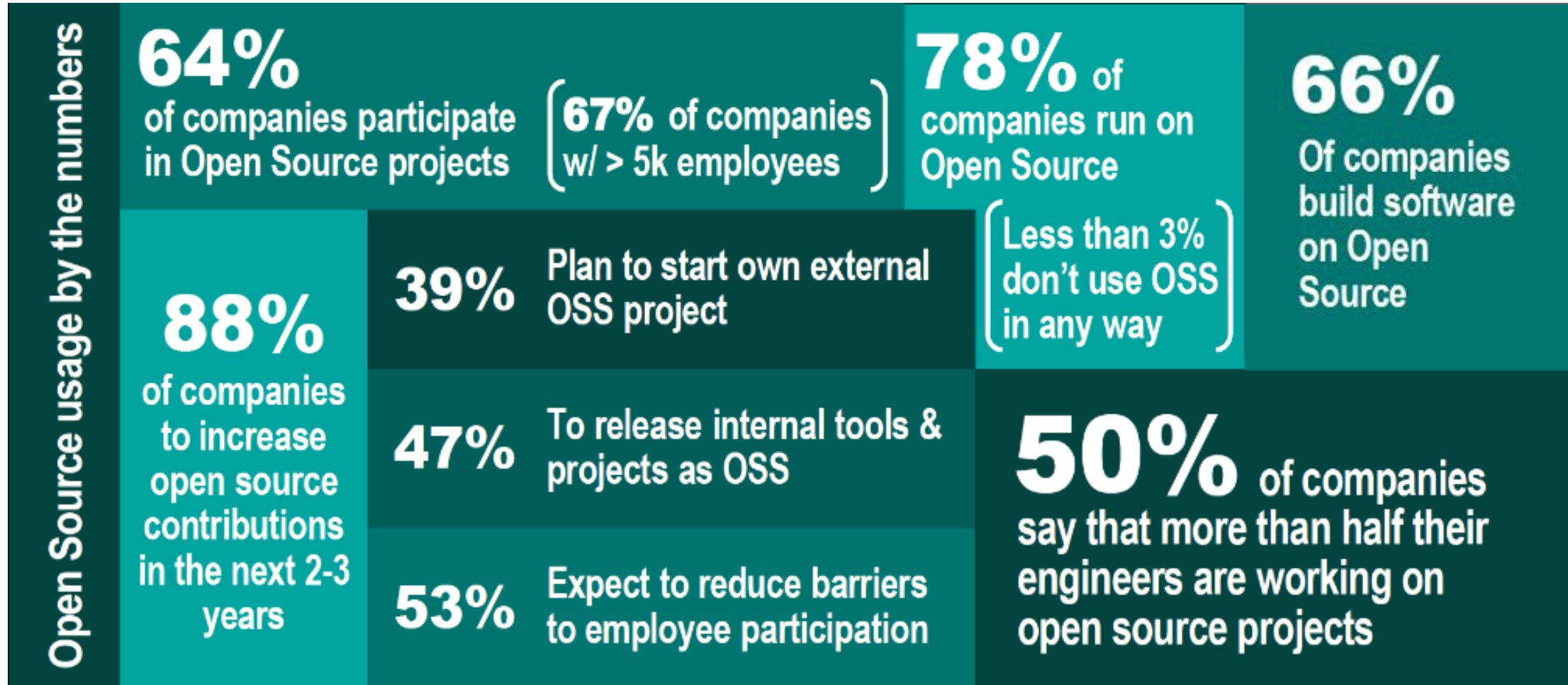
of application development supporting digital business will be built not bought by 2020

70%

of new applications will use Open Source and will run on Open Source Data Base by 2020

Open Source Software Guiding the Digital Transformation

“If 80 % to 90 % of your IT spend is on maintenance (such as proprietary software licensing and such), that leaves very little budget for innovation...”



“The secret for successful digital transformation are data”

Source: Forbes

25 gigabytes
of data per hour
is generated by a
connected car.

90% of cars will
be connected by 2020.

2.5
quintillion
bytes of data
generated daily
by connected
machines.



153 exabytes
of healthcare
data generated by
devices in 2013.

Increasing to 2,314
exabytes in 2020.

There
will be
**28 times
more
sensor-
enabled
devices
than
people**
by the
year 2020.



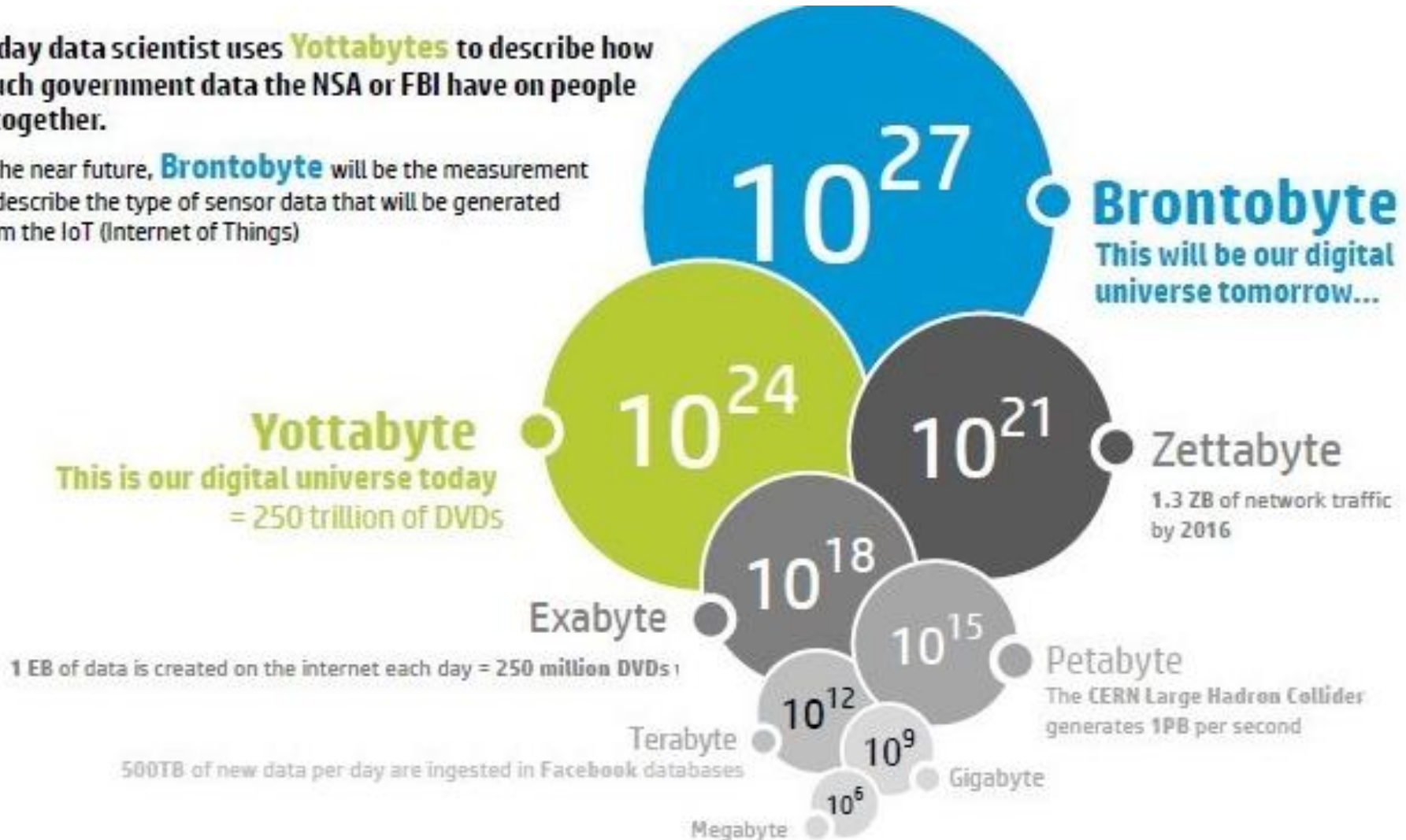
1.7 megabytes
of data per
second
generated by
every human
being on the
planet by 2020.

Data growth...Brontobyte Era is coming....

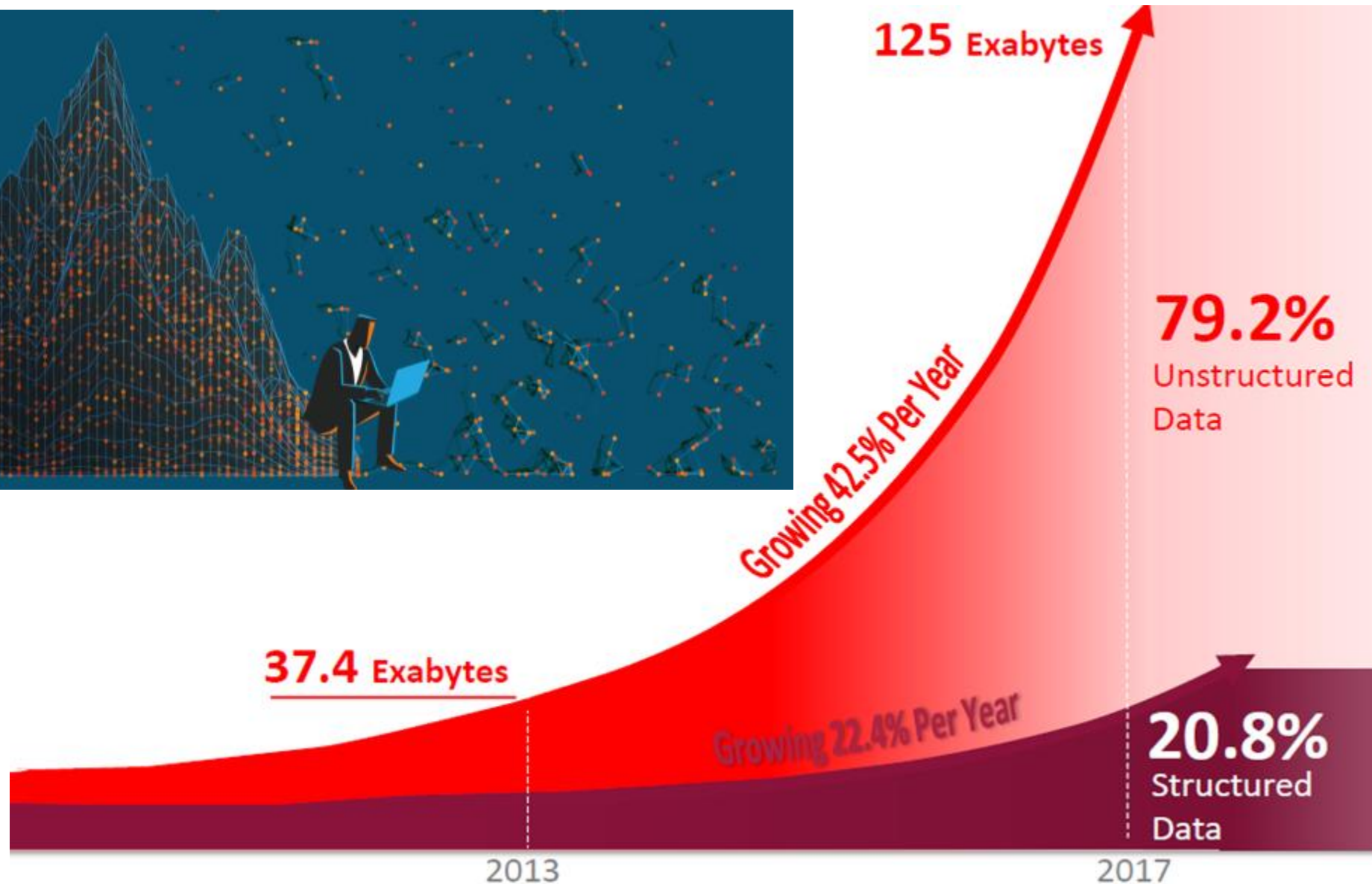
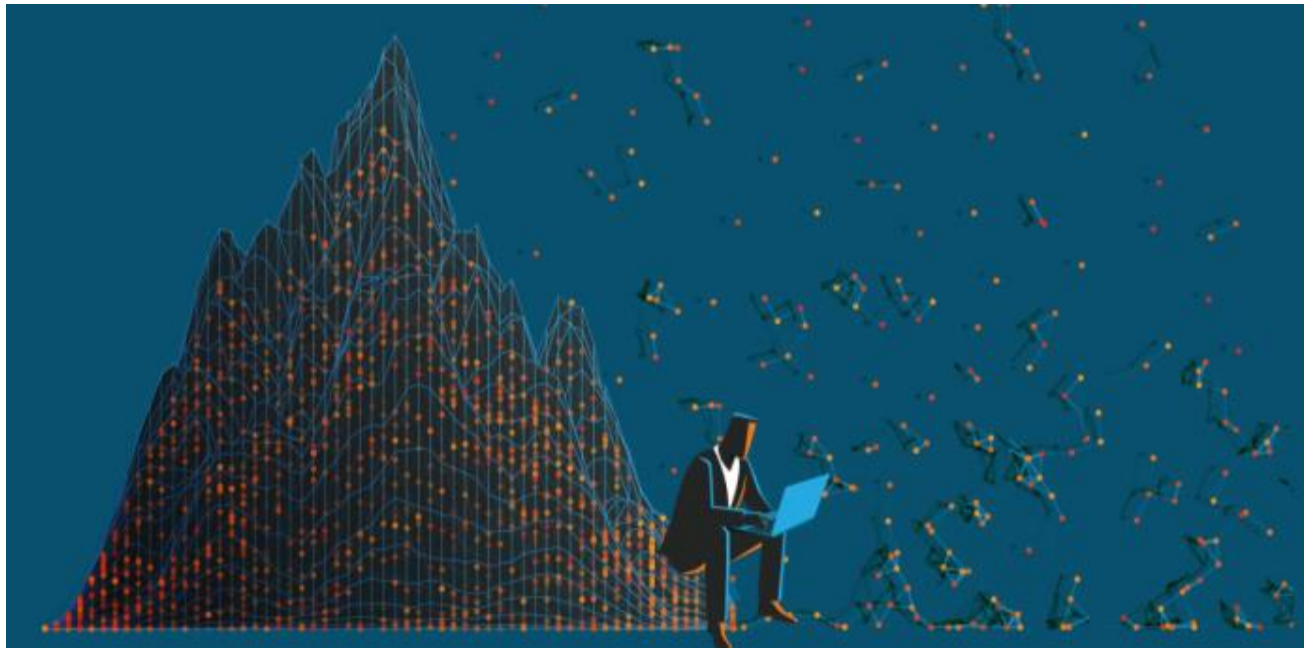
"There were five exabytes of information created between the dawn of civilization through 2003, but that much information is now created every two days, and the pace is increasing...People aren't ready for the technology revolution that's going to happen to them" Google CEO Eric Schmidt - August 2010

Today data scientist uses **Yottabytes** to describe how much government data the NSA or FBI have on people altogether.

In the near future, **Brontobyte** will be the measurement to describe the type of sensor data that will be generated from the IoT (Internet of Things)



Data growth...driven by unstructured data



Structured and Unstructured Data

Structured Data



0.103	0.176	0.387	0.300	0.379
0.333	0.384	0.564	0.587	0.857
0.421	0.309	0.654	0.729	0.228
0.266	0.750	1.056	0.936	0.911
0.225	0.326	0.643	0.337	0.721
0.187	0.586	0.329	0.340	0.829
0.153	0.485	0.360	0.428	0.628

Unstructured Data



Quale oggi il più grande vantaggio derivato dall'uso dei dati? "Non pagare il prezzo della loro mancanza. Le decisioni migliori si prendono proprio grazie ai dati". Paolo Pelloni (Axians Sirecom) per Digitalic Marzo 2019

Enterprise Applications

- Enterprise Resource Planning systems
- Financial systems
- HR systems
- CRM systems

20% Structured Application Content

Enterprise Content Management

- Marketing
 - Outreach
 - Collateral

Document Management

- Emails
- Electronic docs and files
- Paper documents
- Forms

Finance

- Requisitions
- Purchase orders
- Invoices
- A/P & A/R supporting documentation

HR & Legal

- Staff & faculty onboarding
- Employee records
- Contracts

Digital Asset Management

- Rich media files such as images and videos

Records Management


- Governance
- Compliance
- Retention
- Disposition

80% Unstructured content

Structured and Unstructured Data & Big Data

	Internal	External
Unstructured	<ul style="list-style-type: none"> • Open Innovation/Collaboration Networks* • Employee Social Business Networks* • Supplier/Vendor Social Business Networks • Employee Feedback/Emails* • Employee Performance Reviews (Text) • Mobile Market Research Data* • Candidate Interviews (Text) • Product R&D* • Brand Social Media Properties* • Company Social Media Properties* • Consumer Call Center (audio) • Customer Service Centers (audio) • Consumer Survey/Panel Data* 	<ul style="list-style-type: none"> • Retail In-Store Video Tracking & Monitoring • Merchandising Photos/Photo Recognition • Social Referral* • Crowd Sourcing* • Web Scraping (Menus) • Social Media* • Viral Videos* • Consumer Product Reviews* • Comments, Chat Rooms, Blogs* • News Media* • Professional Product Reviews* • Quality and Safety Reports (text) • Consumer Apps
Structured	<ul style="list-style-type: none"> • Enterprise Resource Planning (ERP) Data • Shipment and Inventory Data • Invoice and P/O Data • Trade Promotion Management Data • Marketing Data (Pricing, Media, etc). • Consumer Data (CRM, Social Media) • Supply Chain Data • eCommerce Data • Accounts Payable/Receivable Data • General Ledger / Chart of Accounts Data • Financial Forecasts and Budget Data • Customer Deduction Data • Retail Store and Merchandising Data • Customer Master Data (Bill-to, Ship-to) • Product Master Data (BOM, COGS, SKUs) 	<ul style="list-style-type: none"> • eCoupon Data • Sensor Tags Data • GPS Data / Mobile Phone Data • Gamification Data • Virtual Shopping Data • Augmented Reality/Package Entertainment Data • Retailer POS data • Retailer Shopper Profile Data • Syndicated Data • Distributor Sell-Thru Data • Supplier Data • eTailer Data • Weather Data • Macro and Micro Economic Data (household incomes, commodity pricing)

* text, audio, photo, video
Bold = Newer & Emerging

 Indicates Generally More Velocity and Volume ("Big Data")

Big Data references “The Three V’s” – *Volume, Velocity and Variety.*

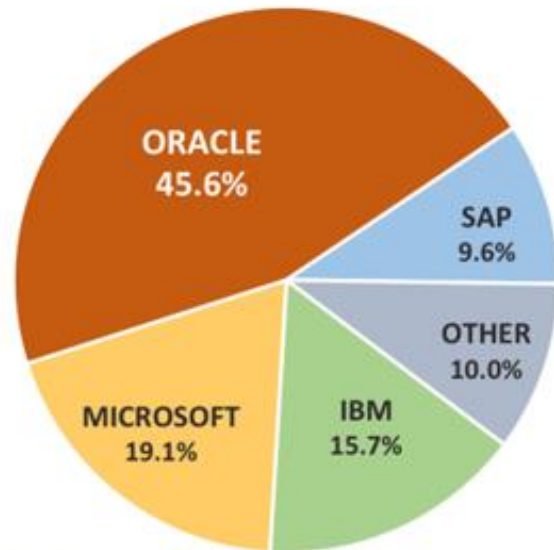
Typically, only one of these attributes needs to be present for a data asset to be considered Big Data however, often two or all three attributes co-exist.



...and so, what's going on in the market?

COMMERCIAL DATABASE MARKET SHARE

(Source: Gartner, Inc. 2016)



<https://seekingalpha.com>

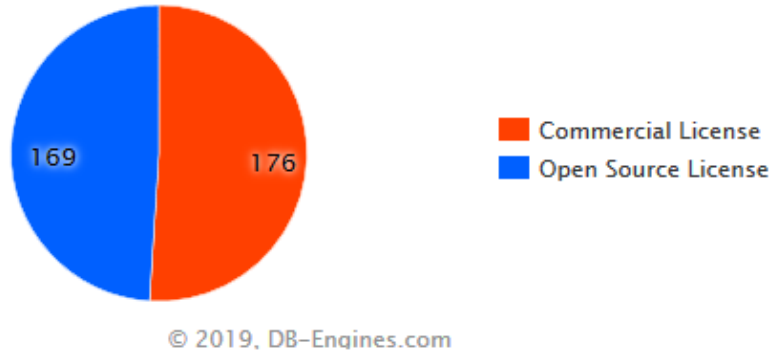
The *commercial database market* - 80% of which is an **oligopoly of Oracle, Microsoft and IBM** - has remained one of the most stable and sticky in all of tech for over two decades.

However, we think the velocity and magnitude of its decline is likely to surprise many investors...

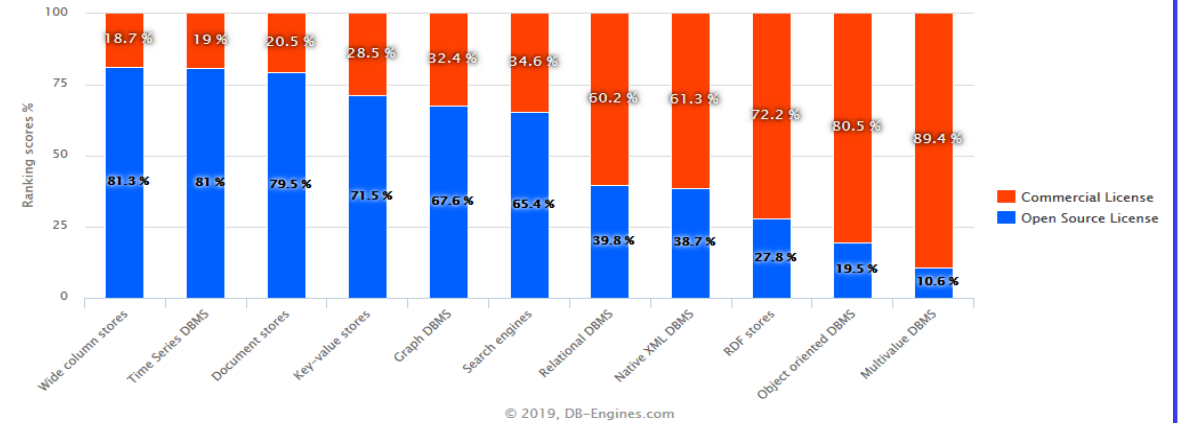


...popularity trend picture.....

Number of systems, April 2019

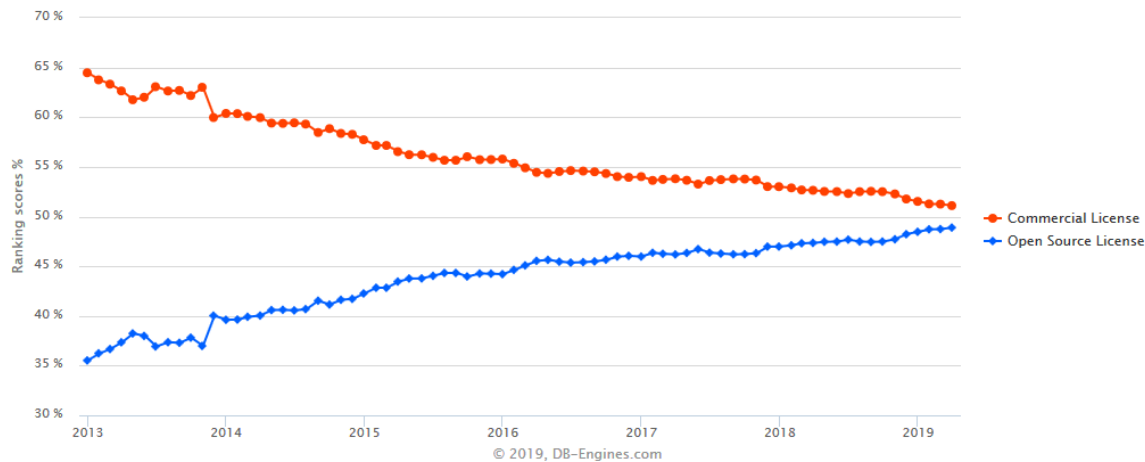


Popularity broken down by database model, April 2019



This chart demonstrates that the popularity of open source tools heavily depends on their database model.

Popularity trend



345 systems in ranking, April 2019

Rank	Rank			DBMS	Database Model	Score		
	Apr 2019	Mar 2019	Apr 2018			Apr 2019	Mar 2019	Apr 2018
1.	1.	1.	1.	Oracle +	Relational, Multi-model	1279.94	+0.80	-9.85
2.	2.	2.	2.	MySQL +	Relational, Multi-model	1215.14	+16.89	-11.26
3.	3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model	1059.96	+12.11	-35.55
4.	4.	4.	4.	PostgreSQL +	Relational, Multi-model	478.72	+8.91	+83.25
5.	5.	5.	5.	MongoDB +	Document	401.98	+0.64	+60.57
6.	6.	6.	6.	IBM Db2 +	Relational, Multi-model	176.05	-1.15	-12.89
7.	↑8.	↑9.	9.	Redis +	Key-value, Multi-model	146.38	+0.25	+16.27
8.	↑9.	8.	8.	Elasticsearch +	Search engine, Multi-model	146.00	+3.21	+14.64
9.	↓7.	↓7.	7.	Microsoft Access +	Relational	144.65	-1.55	+12.43
10.	10.	↑11.	11.	SQLite +	Relational	124.21	-0.66	+8.23
11.	11.	↓10.	10.	Cassandra +	Wide column	123.61	+0.81	+4.52
12.	12.	↑14.	14.	MariaDB +	Relational, Multi-model	85.23	+0.92	+20.67
13.	13.	13.	13.	Splunk	Search engine	79.09	-0.01	+18.03
14.	14.	↓12.	12.	Teradata +	Relational	78.00	+0.13	+1.67

Method of calculating the scores of the DB-Engines Ranking

The DB-Engines Ranking is a list of database management systems ranked by their current popularity.

We measure the popularity of a system by using the following parameters:

- **Number of mentions of the system on websites**, measured as number of results in search engines queries. At the moment, we use [Google](#) and [Bing](#) for this measurement. In order to count only relevant results, we are searching for <system name> together with the term database, e.g. "Oracle" and "database".
- **General interest in the system**. For this measurement, we use the frequency of searches in [Google Trends](#).
- **Frequency of technical discussions about the system**. We use the number of related questions and the number of interested users on the well-known IT-related Q&A sites [Stack Overflow](#) and [DBA Stack Exchange](#).
- **Number of job offers, in which the system is mentioned**. We use the number of offers on the leading job search engines [Indeed](#) and [Simply Hired](#).
- **Number of profiles in professional networks, in which the system is mentioned**. We use the internationally most popular professional network [LinkedIn](#).
- **Relevance in social networks**. We count the number of [Twitter](#) tweets, in which the system is mentioned.

Open Source applications and databases are changing the enterprise

Database Popularity

345 systems in ranking, April 2019

Rank			DBMS	Database Model	Score		
Apr 2019	Mar 2019	Apr 2018			Apr 2019	Mar 2019	Apr 2018
1.	1.	1.	Oracle	Relational, Multi-model	1279.94	+0.80	-9.85
2.	2.	2.	MySQL	Relational, Multi-model	1215.14	+16.89	-11.26
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1059.96	+12.11	-35.55
4.	4.	4.	PostgreSQL	Relational, Multi-model	478.72	+8.91	+83.25
5.	5.	5.	MongoDB	Document	401.98	+0.64	+60.57
6.	6.	6.	IBM Db2	Relational, Multi-model	176.05	-1.15	-12.89
7.	8.	9.	Redis	Key-value, Multi-model	146.38	+0.25	+16.27
8.	9.	8.	Elasticsearch	Search engine, Multi-model	146.00	+3.21	+14.64
9.	7.	7.	Microsoft Access	Relational	144.65	-1.55	+12.43
10.	10.	11.	SQLite	Relational	124.21	-0.66	+8.23
11.	11.	10.	Cassandra	Wide column	123.61	+0.81	+4.52
12.	12.	14.	MariaDB	Relational, Multi-model	85.23	+0.92	+20.67
13.	13.	13.	Splunk	Search engine	83.09	-0.01	+18.03
14.	14.	12.	Teradata	Relational	75.35	+0.13	+1.67

Top five Data Base



The top 5 commercial systems, April 2019

Rank	System	Score	Overall Rank
1.	Oracle	1280	1.
2.	Microsoft SQL Server	1060	3.
3.	IBM Db2	176	6.
4.	Microsoft Access	145	9.
5.	Splunk	83	13.

The top 5 open source systems, April 2019

Rank	System	Score	Overall Rank
1.	MySQL	1215	2.
2.	PostgreSQL	479	4.
3.	MongoDB	402	5.
4.	Redis	146	7.
5.	Elasticsearch	146	8.



PostgreSQL



Open source are changing the enterprise

“By 2020, more than 70% of new in-house applications will be developed on an OSDBMS, and 50% of existing commercial RDBMS instances will have been converted or will be in process...”

Gartner

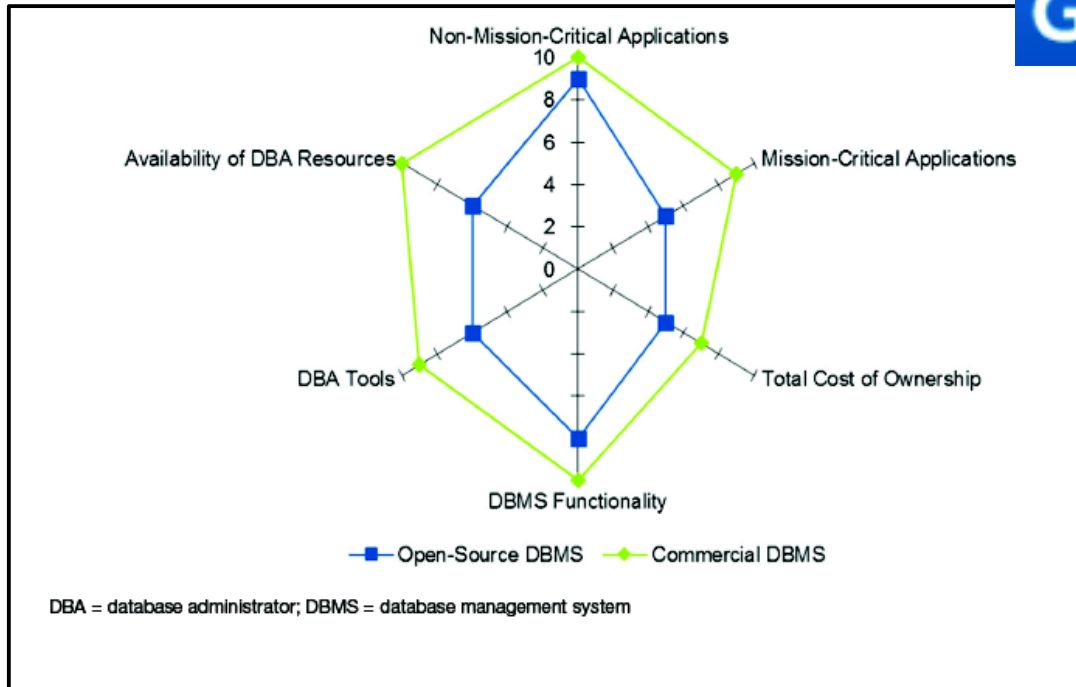


Figure 1. Relational Open-Source DBMS Maturity Evaluation, 2009

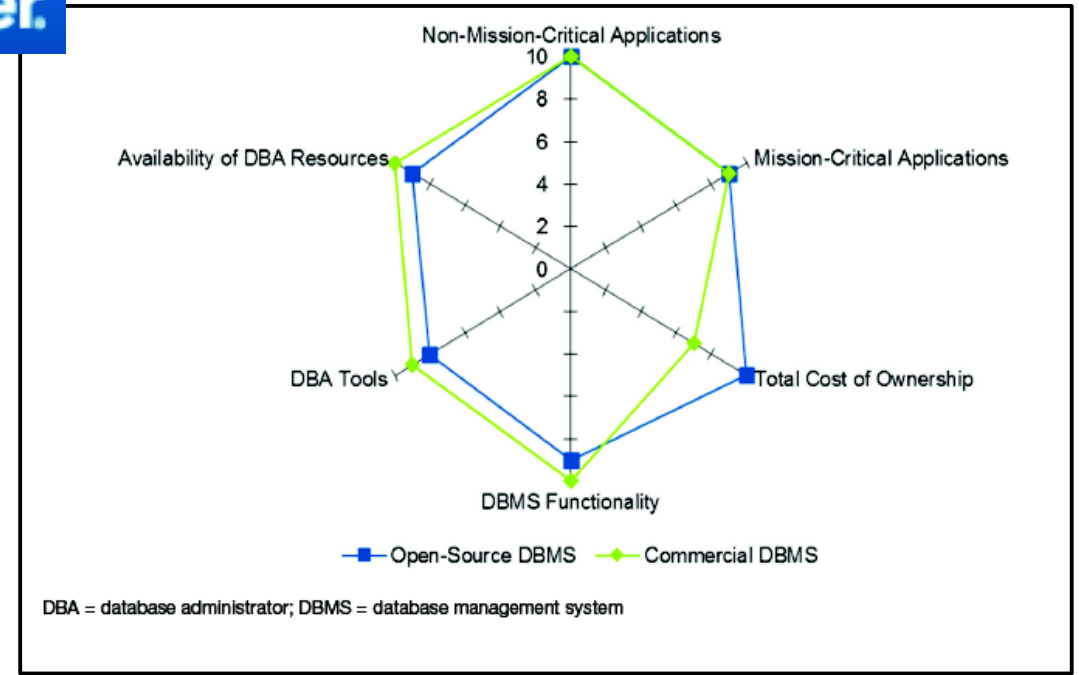
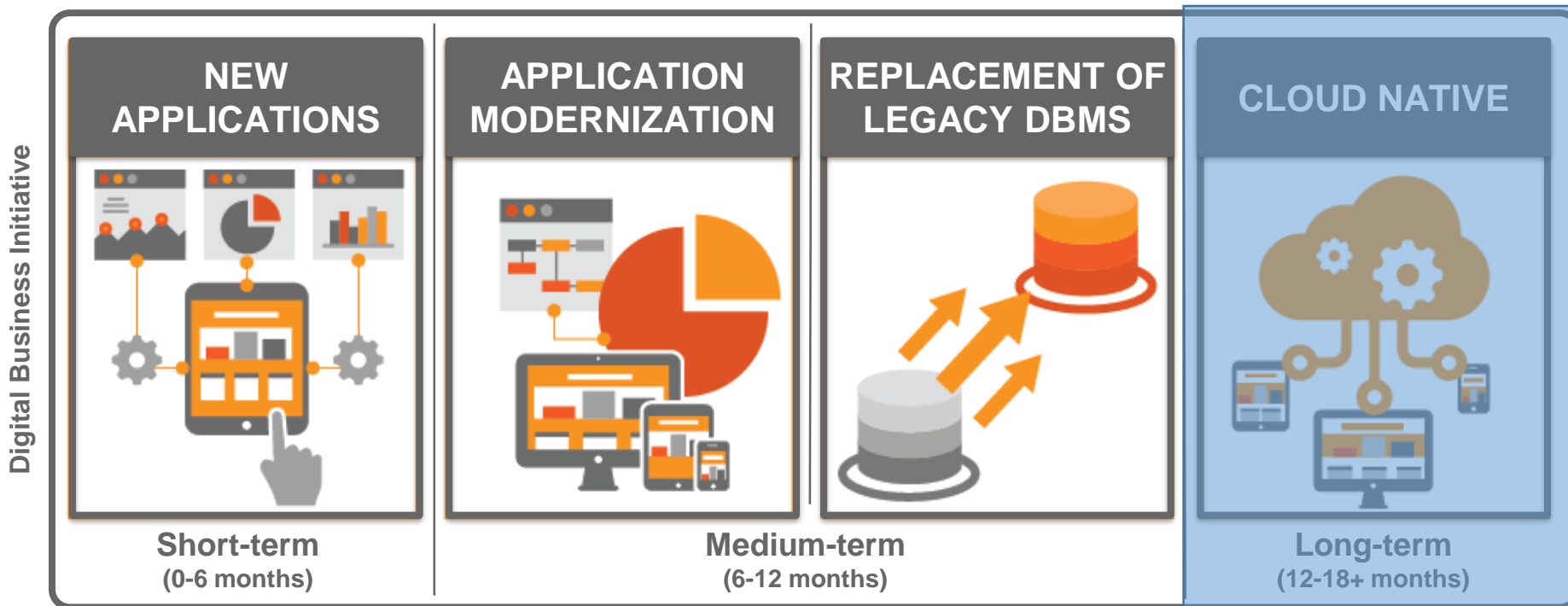


Figure 2. Relational Open-Source DBMS Maturity Evaluation, 2017

“The most demonstrable benefit of OSDBMSs, given their increasing suitability from a technology perspective, is the **TCO** of these products. When skills were at a minimum, management tools were few and the software was relatively immature, the TCO was not necessarily lower than those of commercial vendor offerings. That has changed to the point where we now believe that the cost of managing OSDBMSs and the availability of skills are now close to parity with those of the commercial DBMS offerings. We therefore believe there are clear savings in TCO for the OSDBMS. **With software costs skyrocketing, this has become a major focus of IT management and is a major impact of the OSDBMS.**”

Digital Business Initiatives of Clients Embracing Open Source



Which Open Source Data Base ?

The Dataflog Open Source Landscape 2.0



Created by: www.Dataflog.com

NEW Database paradigms



Relational Databases

Instant insight from real-time operational data

Simplified transactions and reporting

Reduced tuning and indexing



Distributed Hierarchical Databases

A scalable data architecture

A parallel and distributed programming model

Open source community innovation (Apache Hadoop)



NoSQL Databases

Analytics capability for *multiple data types*, often used in *mobile and social workloads*

Scalability and flexibility for different data store models

In Memory Databases

***Increased performance* by bringing data closer to compute**

The Modern Data Platform landscape

New data sources } Social Mobile Data Services Sensors Location

New Data Models
NoSQL

Document, Column, Key Value, Graph,



New Data Models
Hortonworks
(Hadoop/Spark)
Unstructured Data / Text

“New” Relational
Open Source SQL



SQL
MySQL
compatible



SQL
PostgreSQL
compatible

Conventional Data Platform

Systems of Record

RDBMS OLTP
ACID

Structured
Data

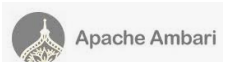
Systems of Insight

Data Warehouse
& Marts

Structured
Data



Transactions



SQL vs NoSQL (Not Only SQL)

SQL

Data uses Schemas

Relations!

Data is distributed across multiple tables

Horizontal scaling is difficult / impossible; Vertical scaling is possible

Limitations for lots of (thousands) read & write queries per second

NoSQL

Schema-less

No (or very few) Relations

Data is typically merged / nested in a few collections

Both horizontal and vertical scaling is possible

Great performance for mass (simple) read & write requests

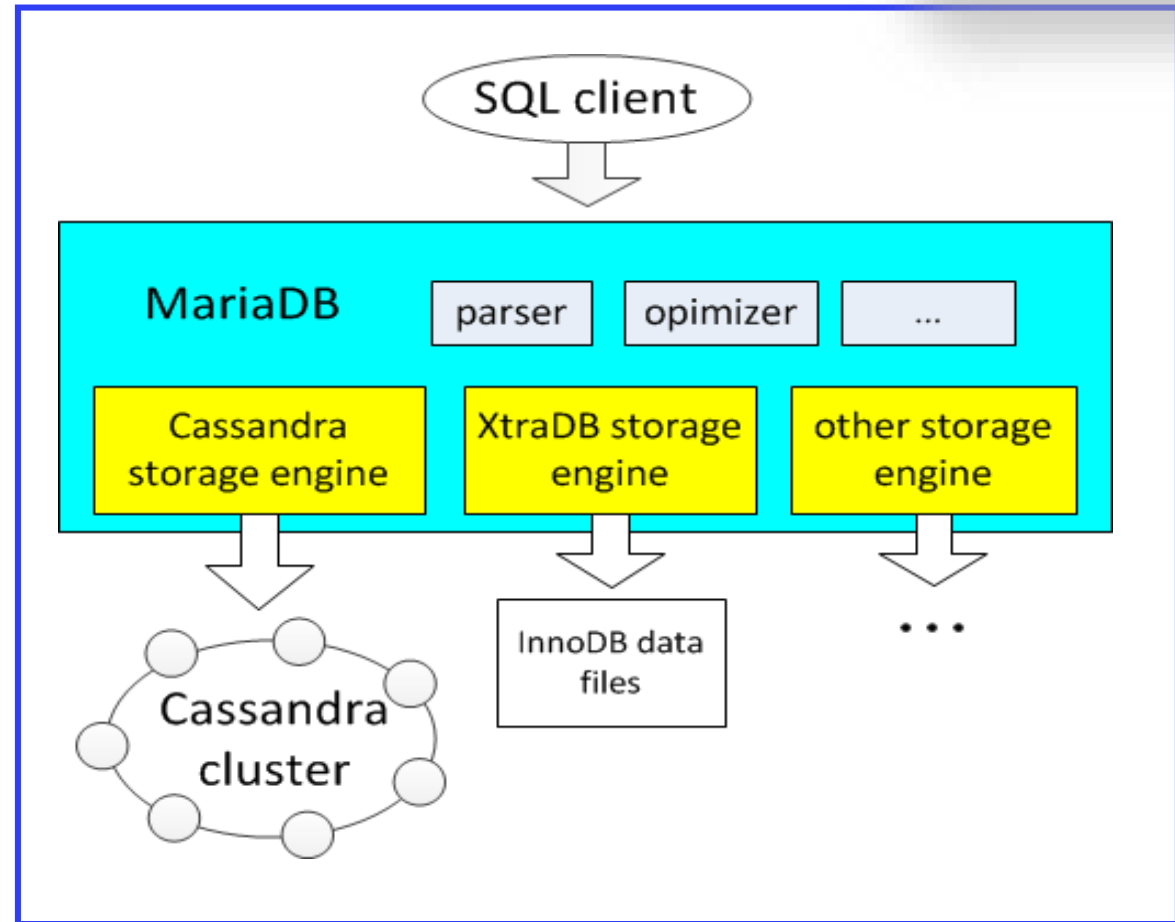
RDBMS + NoSQL : work together



Exemple:

Cassandra Storage Engine makes Cassandra's column family appear as a table in MariaDB that you can insert to, update, and select from.

You can write joins against this table, it is possible to join data that's stored in MariaDB with data that's stored in Cassandra.

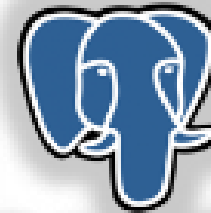


...so: why we are living the OSDB era ?

- **Migration to SaaS**, where most offerings use free open-source databases
- **Faster growth** in use cases such as social media, IoT, and **unstructured/semi-structured data** that are ill-suited to the SQL standard upon which the database oligopoly is based;
- The **increasing stability and functionality of a variety of free open-source** options, most of which are "Not Only SQL" (NoSQL) and are therefore much better suited to the aforementioned use cases;
- The **disproportionately large benefits that NoSQL** databases continue to realize as a result of Moore's Law-driven improvements in processors, memory, solid-state storage, and network throughput (*these improvements increase NoSQL's ability to handle both NoSQL and SQL use cases on-the-fly and progressively marginalize SQL-only databases just as SQL-only databases marginalized mainframe-based databases during the late '80s and '90s*).



Leading Open Source Technologies



PostgreSQL



SCYLLA.





mongoDB.

Humongous

MongoDB :

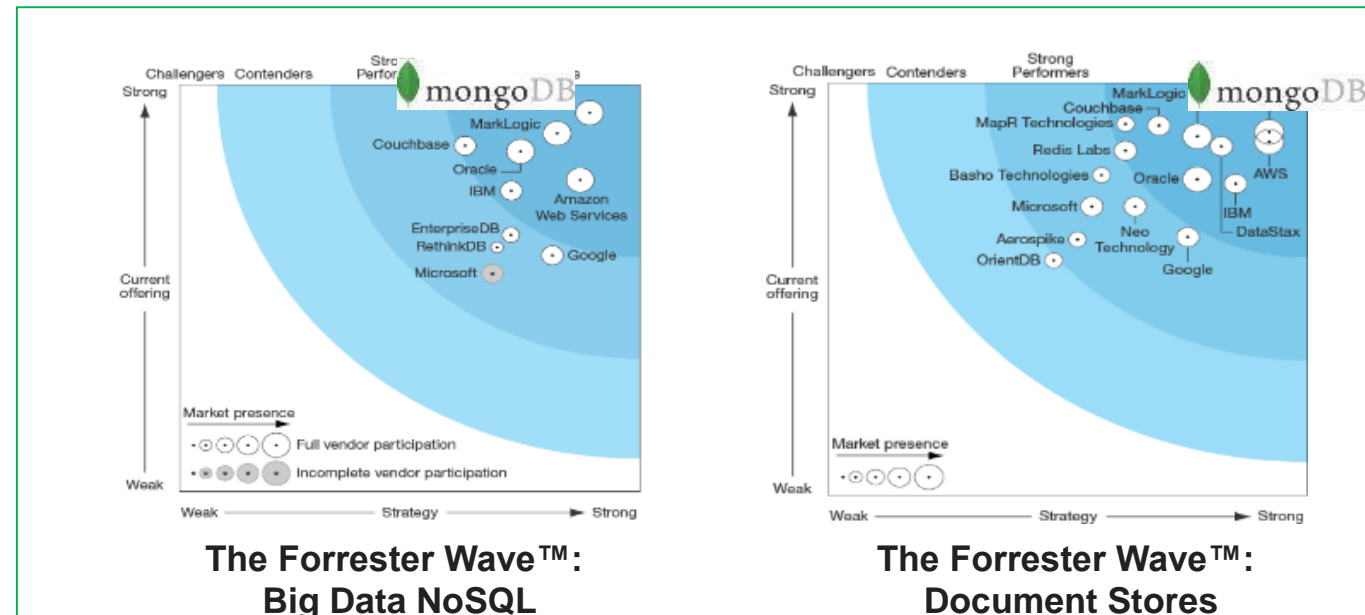
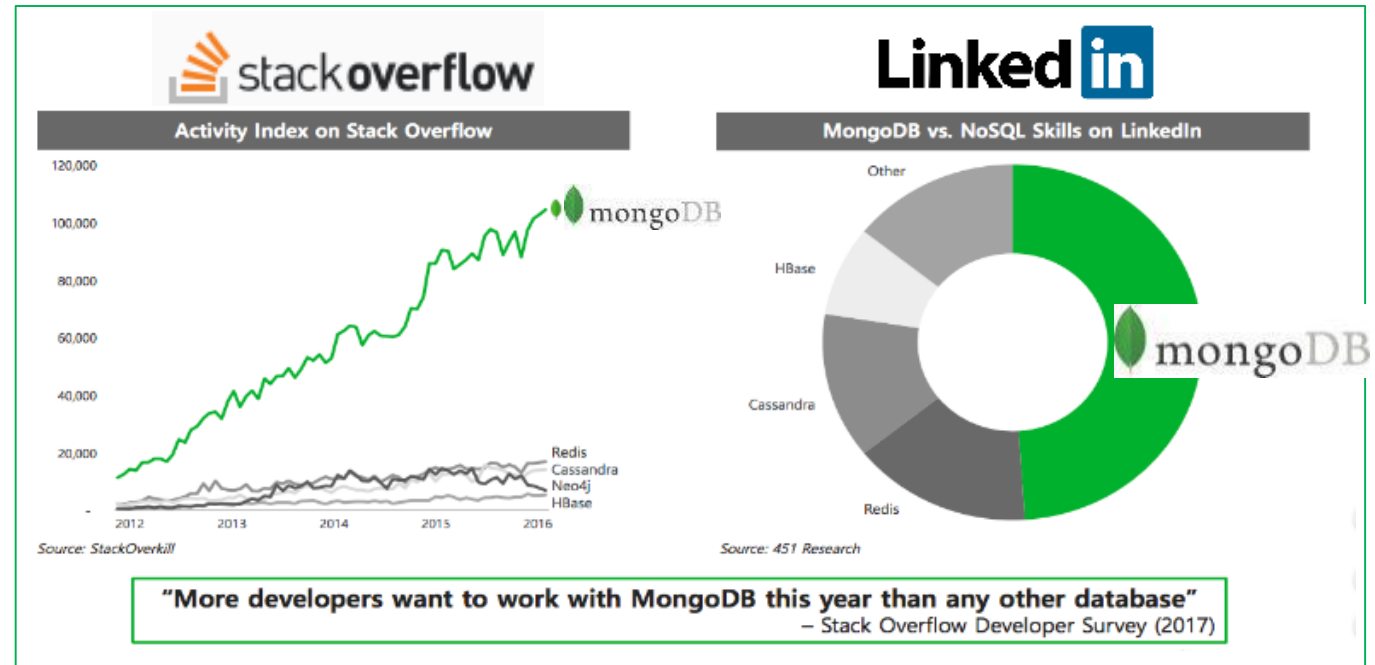
Most Developer Requested Database in the world

Company

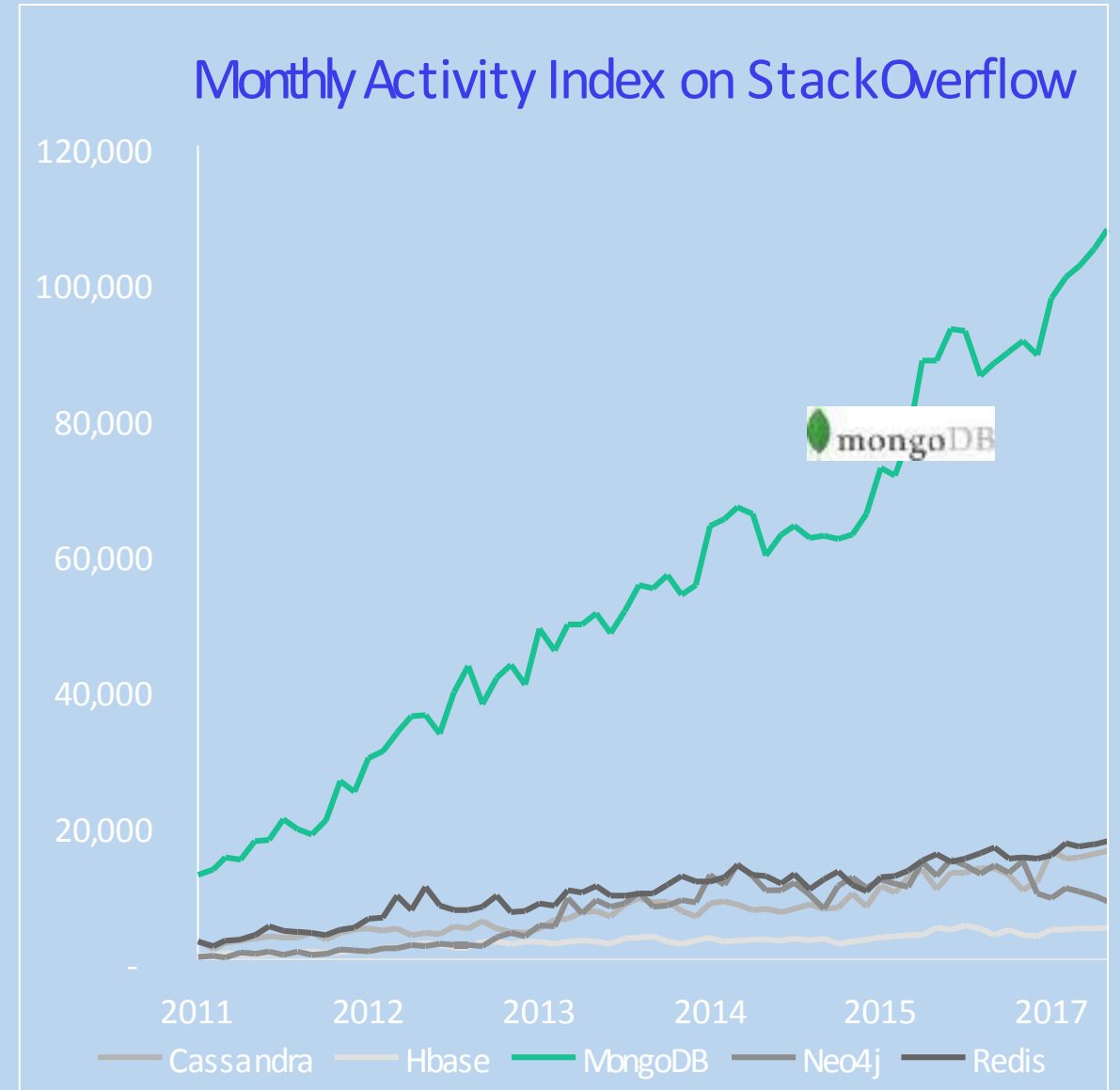
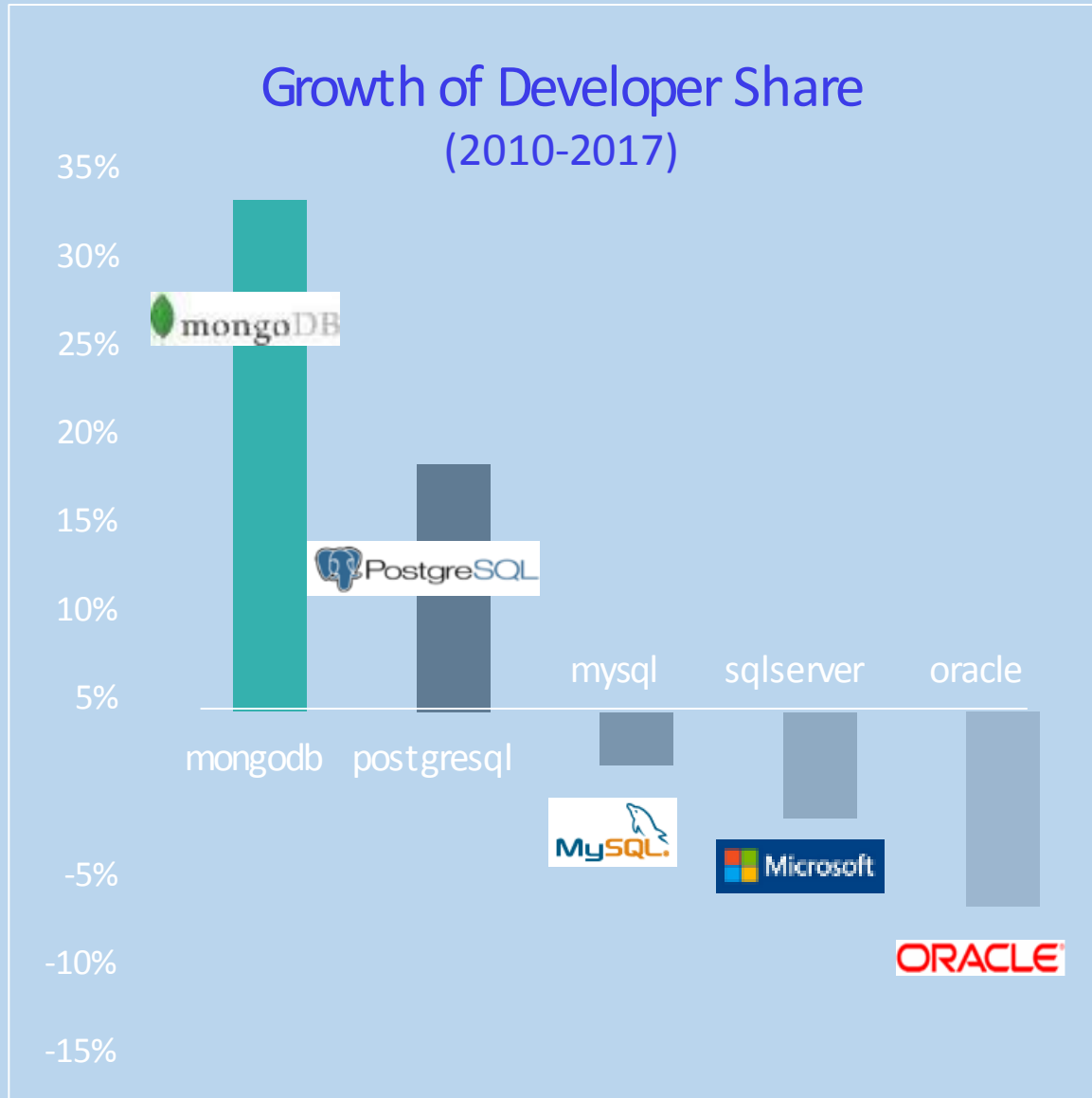
- Headquartered in New York, NY
- Founded in 2007
- Publicly listed in 2017
 - (Nasdaq Global Market: MDB)

Product

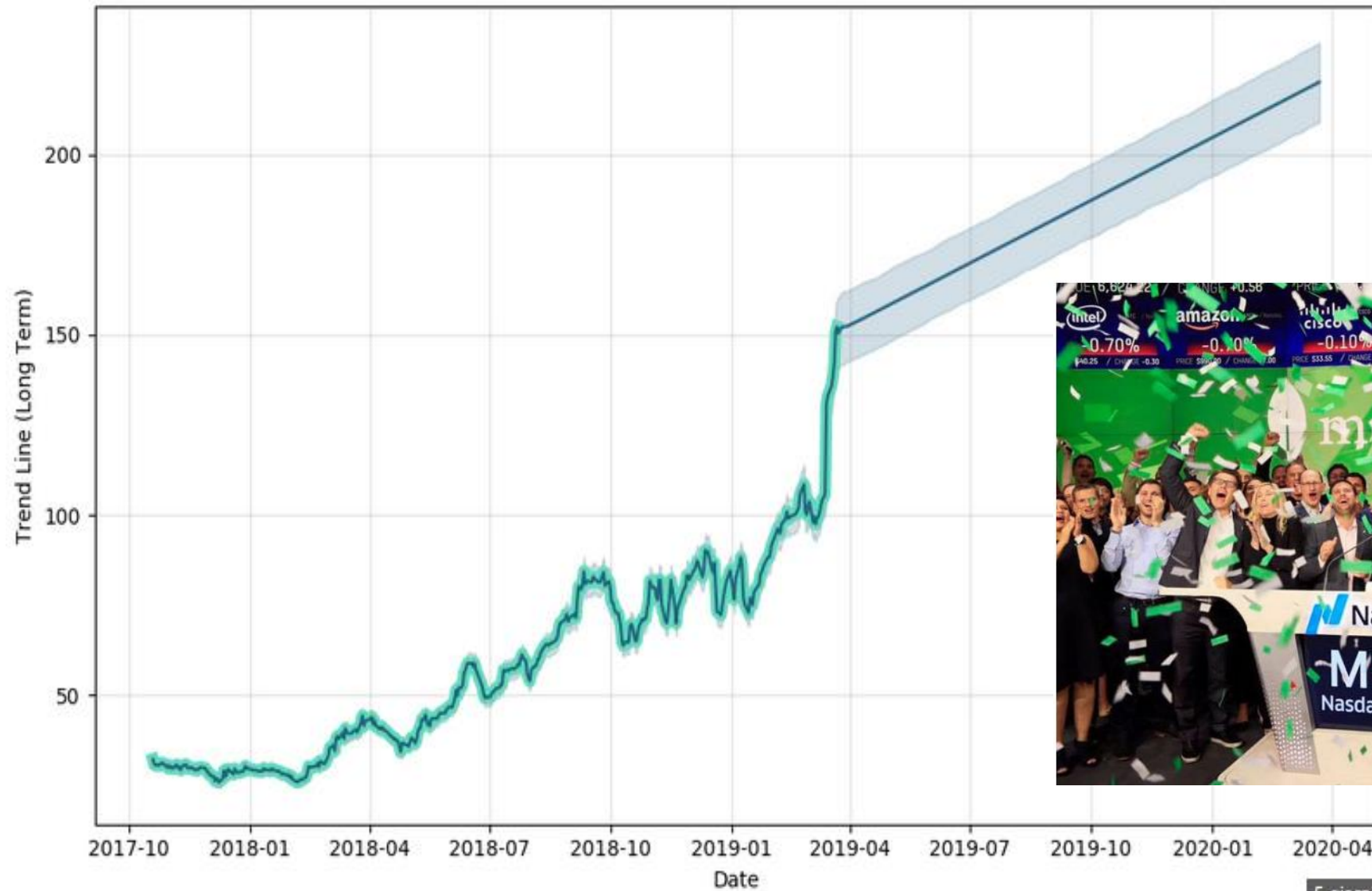
- #1 Fastest growing database in market
- Leader in The Forrester Wave™ for Big Data NoSQL and for Document stores
- 4,900+ customers in over 90 countries
 - 30mm+ downloads



Developers love Modern Databases and MongoDB for new Apps











MongoDB Inc. NASDAQ: MDB



5 giorni fa

MongoDB characteristics (1 di 2)

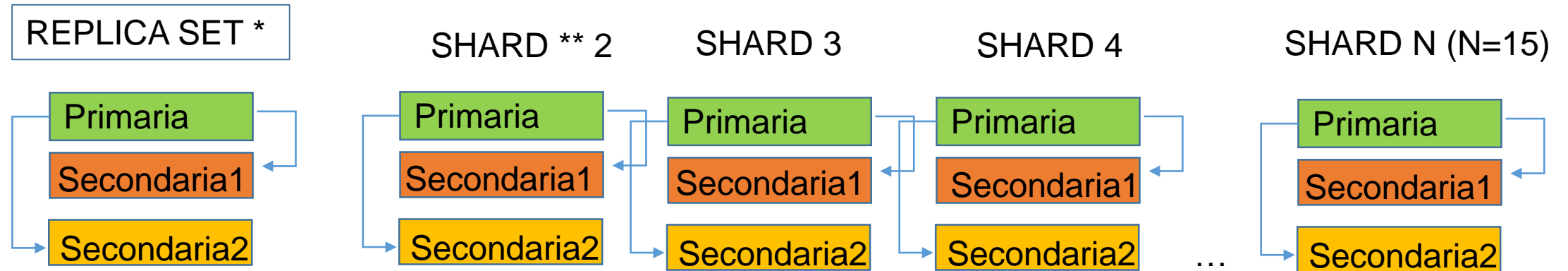
Name	MongoDB
Description	One of the most popular document stores
Primary database model	Document store
Secondary database models	Key-value store
Website	www.mongodb.com
Technical documentation	docs.mongodb.com/manual
Developer	MongoDB, Inc
Initial release	2009
Current release	4.0.2, August 2018
License 	Open Source 
Cloud-based 	no 
Implementation language	C++
Server operating systems	Linux OS X Solaris Windows
Data scheme	schema-free 
Server-side scripts 	JavaScript
Triggers	no
Partitioning methods 	Sharding
Replication methods 	Master-slave replication

RDBMS		MongoDB
Database	→	Database
Table	→	Collection
Row	→	Document
Index	→	Index
Join	→	Embedding & Linking

MongoDB characteristics (2 di 2)

Typical application scenarios	<ul style="list-style-type: none">• Internet of Things (Bosch, Silver Spring Networks)• Mobile (The Weather Channel, ADP, O2)• Single View (MetLife)• Real-Time Analytics (Buzzfeed, City of Chicago, Crittercism)• Personalization (Expedia, eHarmony, Gilt)• Catalogs (Under Armour, Otto)
Key customers	ADP, Adobe , Amadeus, AstraZeneca, Barclays , BBVA, Bond, Bosch , Cisco, CERN , City of Chicago, coinbase, Department of Veteran Affairs, Department of Works and Pensions, eBay , eHarmony, Electronic Arts, Epic Games, Expedia, Facebook's Parse , Forbes, Foursquare, Gap, Genentech, HSBC, MetLife, Morgan Stanley, NowTV, Pearson, RBS, Sage, Salesforce, SAP, Sega, Sprinklr, Telefonica, The Weather Channel, Ticketmaster, Under Armour, Verizon Wireless
Market metrics	<ul style="list-style-type: none">• 40 million downloads (growing at 30 thousand downloads per day).• 6,600+ customers.• Named a leader in the <i>Forrester Wave™: Big Data NoSQL, Q3 2016.</i>• Highest placed non-relational database in DB Engines rankings

Architettura di infrastruttura MongoDB



Sharding: meccanismo di scalabilità orizzontale tra i vari “nodi”

* Replica Set: consiste in due o più copie dei dati. Ogni replica può avere ruolo di Copia Primaria o Secondaria in qualunque momento

** Uno Shard è un Replica Set : i dati sono divisi in intervalli e distribuiti su molteplici shard. MongoDB include un meccanismo di balancing

MongoDB *running on IBM POWER Price-Performance Guarantee*



IBM Power Systems guarantees the Power S822LC for Big Data system built with POWER8 delivers at least a 2X price-performance advantage vs. x86 based servers when running a customer application/workload based on MongoDB.

2X price-performance means that the customer's documented throughput performance on the S822LC POWER8 divided by the price of the system will be at least 2 times higher than the customer's documented throughput performance on the x86 based system divided by the price of the comparable x86 system.

EX: If transactions per second on the S822LC are 20,000 and 10,000 on the x86 based system, while the price of the S822LC is \$10,000, and the price of the x86 based system is \$10,000, then the Throughput Performance Per Price would be exactly 2 times higher and the guaranty would be met."

- The IBM Power S822LC for Big Data server (22-core/2.89 GHz 128GB memory, 4 TB SATA Storage) must be purchased from IBM or an authorized IBM Business Partner prior to December 31, 2017. The guarantee period is valid for three (3) months from the date of purchase. The x86 based systems must be comparably configured branded servers from Cisco, Dell, HP, or Lenovo and the client is responsible for all MongoDB licenses.
- 2X throughput performance per price means that the customer's documented throughput performance on the S822LC POWER8 system based on either queries, operations or transactions per second divided by the price of the such system will be at least 2 times higher than the customer's same documented throughput performance on the x86 based system divided by the price of such comparable x86 system.
- Remediation: IBM will provide additional performance optimization and tuning services consistent with IBM Best Practices, at no charge. If unable to reach guaranteed level of price-performance, IBM will provide additional equally configured systems to those already purchased to reach the guaranteed level of price-performance.

Notes:

1. Client's POWER8 Machine and the x86 Machine must be running at similar utilization rates.
2. Client's POWER8 Machine's system performance cannot be constrained by I/O subsystem. Specifically, the I/O subsystem on the POWER8 Machines must achieve greater than or equal I/O bandwidth and operations per second than the x86 Machine.
3. Client's POWER8 Machine's physical memory must be the same or greater than the physical memory on the x86 Machine
4. Client is responsible for demonstrating comparable real-world representative workload between the POWER8 Machine and the x86 Machine through the use of the IBM provided tools and comparable tools on x86 systems.
5. 2.0x guarantee is available in countries with published list prices and is based on performance and list price for the x86 based server (Dell, Cisco, Lenovo, or HP) based on E5-2600 v4 or earlier and the IBM S822LC for Big Data.

IBM & MongoDB: a strong relationship since 2015....

Example POWER8 Configurations & Sizing Guidance

Entry Configuration

- IBM Power System S822LC for Big Data
- 10-core 2.92 GHZ
- 128 GB memory
- Two 960 GB SATA SSD drives
- 10k RPM 2 TB drives (for data)
- Linux (RHEL/Ubuntu LE)
- KVM (optional)



- For read-intensive workloads

Performance Configuration

- IBM Power System S822LC for Big Data
- 20-core 2.926 GHZ
- 256 GB memory
- Two 960 GB SATA SSD drives
- SSD drives (for data)
- Linux (RHEL/Ubuntu LE)
- KVM (optional)



- For read-write balanced workloads

Cloud Configuration

- IBM Power System S822LC for Big Data
- 20-core 2.926 GHZ
- 512 GB memory
- Two 960 GB SATA SSD drives
- FC i/o adapters (connect to external storage)
- Linux (RHEL/Ubuntu LE)
- KVM (optional)



- For write-intensive workloads

Note: These are sample configs provided as a starting point. They should be customized to meet specific client needs.

Installing, Tuning & Sizing on POWER8

- *Getting Started Guide on POWER8:* [Link](#)
- *Tuning Guide on POWER8:* [Link](#)
- *Sizing Guide on POWER8:* [Link](#)

High Availability & Redundancy

- MongoDB supports replica sets and auto replica failover.
- Review MongoDB's recommended replication architecture options.
- Multiple POWER8 servers will be required to implement replica sets.
- For details, read *MongoDB Getting Started Guide for Administrators: Replications:* [Link](#)

Sharding

- MongoDB supports sharded clusters.
- Review MongoDB's recommended sharding options.
- Multiple POWER8 servers or VMs will be required to implement sharding.
- For details, read *MongoDB Getting Started Guide for Administrators: Sharded Clusters:* [Link](#)



PostgreSQL



PGDay.IT 2019 - Bologna, Italy

Bologna 2019: Workshop 16 maggio e PGDay.IT 17 maggio



PostgreSQL : *DBMS of the Year 2017*

PostgreSQL is the DBMS of the Year 2017

by Paul Andlinger, Matthias Gelbmann, 2 January 2018

Tags: DBMS of the year, Elasticsearch, MariaDB, PostgreSQL

PostgreSQL is the database management system that gained more popularity in our [DB-Engines Ranking](#) within the last year than any of the other 341 monitored systems.

*We thus declare **PostgreSQL** as the **DBMS of the Year 2017**.*

For determining the DBMS of the year, we subtracted the popularity scores of January 2017 from the latest scores of January 2018. We use the difference of these numbers, rather than a percentage, because that would favor systems with a tiny popularity at the beginning of the year. The result is a list of DBMSs sorted by how much they managed to increase their popularity in 2017, or in other words, how many additional people started to communicate about it in one of the ways we measure in our [methodology](#), for example job offers, professional profile entries and citations on the web.

DBMS of the Year: PostgreSQL

While in our last year's popularity ranking [PostgreSQL](#) already ran in on place 3, 2017 was an even better year for PostgreSQL. With a total gain of 55.81 scoring points (+17%) and improving its score in each of the single monthly rankings of 2017, it outperformed all other systems in 2017.

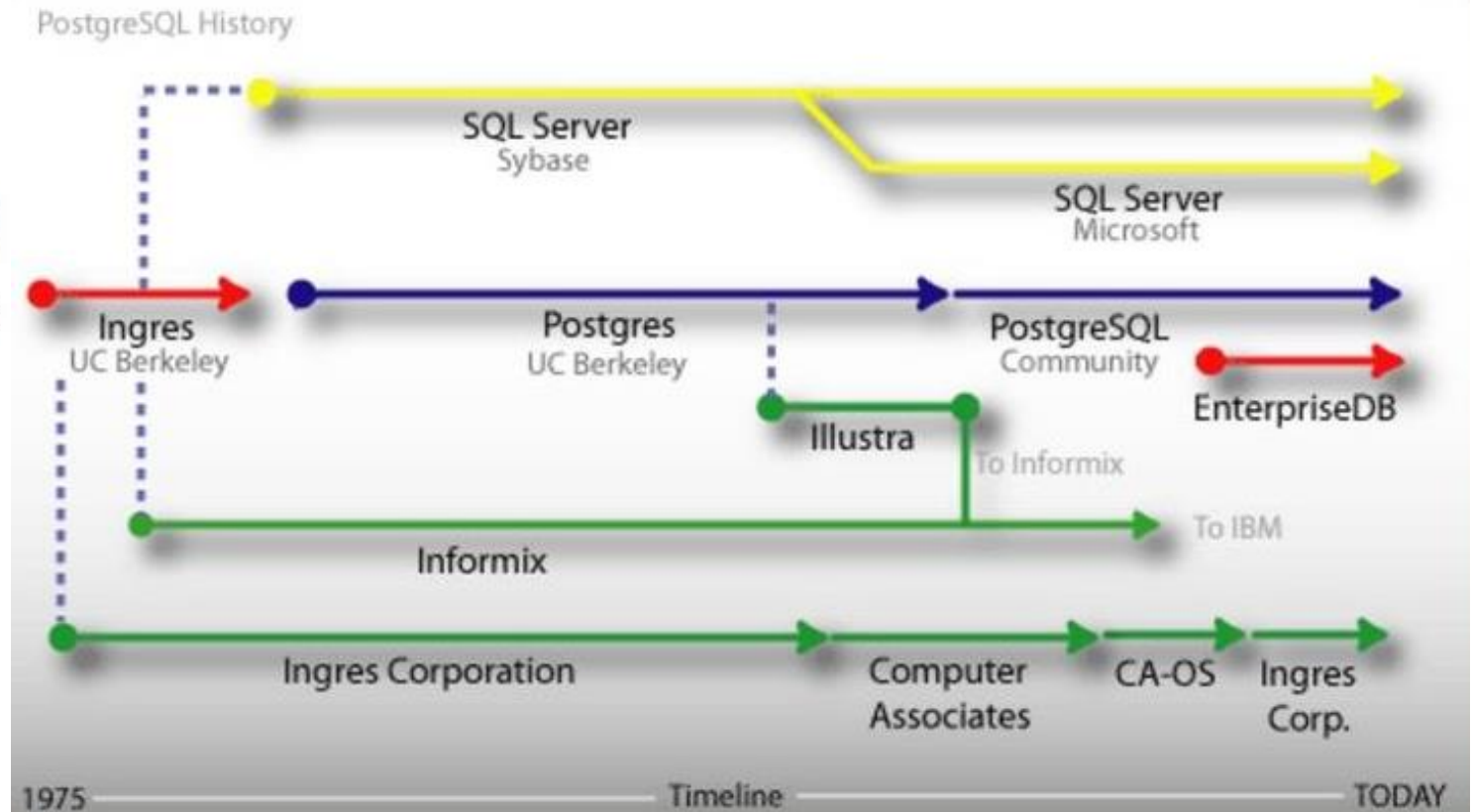


Postgres: una lunga storia....












Brief history

- 1977-1985: Ingres (Stonebreaker)
- 1986-1994: Postgres (Stonebreaker)
- 1995: Postgres + SQL = PostgreSQL
- 1996: Open-source project, run by the "global development group"
- Ever since, one major release per year





Current Release is 11.2 (Feb. 2019)



PostgreSQL characteristics (1 di 2)

Name	PostgreSQL
Description	Widely used open source RDBMS 
Primary database model	Relational DBMS 
Secondary database models	Document store Key-value store
DB-Engines Ranking  Trend Chart 	Score 406.43 Rank #4 Overall #4 Relational DBMS 
Website	www.postgresql.org
Technical documentation	www.postgresql.org/docs/manuals
Developer	PostgreSQL Global Development Group 
Initial release	1989 
Current release	10.5, August 2018
License 	Open Source 
Cloud-based 	no
DBaaS offerings (sponsored links) 	Google Cloud SQL : A fully-managed database service for the Google Cloud Platform
Implementation language	C

PostgreSQL characteristics (2 di 2)

Server operating systems	FreeBSD HP-UX Linux NetBSD OpenBSD OS X Solaris Unix Windows
XML support 	yes 
Secondary indexes	yes
SQL 	yes 
APIs and other access methods	native C library streaming API for large objects ADO.NET JDBC ODBC

PostgreSQL on Power Systems Price-Performance Guarantee

IBM Power Systems guarantees the S822LC for Big Data system built with POWER8 delivers at least a 1.8X price-performance advantage versus x86 based servers when running a virtualized customer application/workload based on EnterpriseDB Postgres 9.5.

1.8X price-performance means that the customer's documented throughput performance on the S822LC POWER8 divided by the sum of the price of the system and associated EnterpriseDB licenses will be at least 1.8 times that of the customer's documented throughput performance on the x86 based system divided by the sum of the price of the comparable x86 system and associated EnterpriseDB licenses

EX: If transactions per second on the S822LC are 18,000 and 10,000 on the x86 based system, while the price of the S822LC and associated EnterpriseDB licenses is \$10,000, and the price of the x86 based system and associated EnterpriseDB licenses is \$10,000, then the Throughput Performance Per Price would be exactly 1.8 times advantaged and the guaranty would be met."

The IBM Power S822LC for Big Data server (20-core/2.92 GHz 256GB memory, 4 TB SATA Storage) must be purchased from IBM or an authorized IBM Business Partner prior to March 31, 2017. The guarantee period is valid for three (3) months from the date of purchase. The x86 based systems must be comparably configured branded servers from Cisco, Dell, or HP and the client is responsible for all EnterpriseDB licenses.

1.8 X price-performance means that the customer's documented throughput performance on the S822LC POWER8 divided by the sum of the price of the system and associated EnterpriseDB licenses will be at least 1.8 times that of the customer's documented throughput performance on the x86 based system divided by the sum of the price of the comparable x86 system and associated EnterpriseDB licenses

Remediation: IBM will provide additional performance optimization and tuning services consistent with IBM Best Practices, at no charge. If unable to reach guaranteed level of price-performance, IBM will provide additional equally configured systems to those already purchased to reach the guaranteed level of price-performance.

Notes:

1. Client's POWER8 Machine and the x86 Machine must be running at similar utilization rates. Eligible Machine and the Compared Machine must be partitioned with at least 4 equal sized partitions.
2. Client's POWER8 Machine's system performance cannot be constrained by I/O subsystem. Specifically, the I/O subsystem on the POWER8 Machines must achieve greater than or equal I/O bandwidth and operations per second than the x86 Machine.
3. Client's POWER8 Machine's physical memory must be the same or greater than the physical memory on the x86 Machine
4. Client is responsible for demonstrating comparable real-world representative workload between the POWER8 Machine and the x86 Machine through the use of the IBM provided tools and comparable tools on x86 systems.
5. 1.8x guarantee is based on list price for the x86 based server (Dell, Cisco, or HP) and list price for the IBM S822LC for Big Data.
6. EnterpriseDB Postgres Advanced Server 9.5 license are priced at \$1750 per core - EDB 9.5 <http://www.enterprisedb.com/products-services-training/subscriptions-power>

IBM Power benefits for PostgreSQL ?



Performance

Power Hardware meets scale up requirements for PostgreSQL

Power's 3x cache vs. x86 retrieves data from memory with the lowest latency



Virtualization Flexibility

Power VM or KVM provide the ability to consolidate multiple database or applications while maintaining isolation

Power VC or Docker containers allow rapid deployment of new databases

Hardware resources can be shared and allocated/resized as needed

Availability

PostgreSQL is an RDBMS that scales up (although it does have replication support)

Highly available server technology is key to avoiding service interruption

Cost Savings ...great price performance....

PGDay a Bologna: 16-17 Maggio 2019

Admiral Park Hotel - [Via Fontanella, 3, Zola Predosa - Bologna](#)



PGDay.IT
@PGDayIT

Account del PostgreSQL Day Italiano gestito da
[@ITPUGPRESS](#)

📍 Italia

[pgday.it](#)

📅 Joined September 2011



PGDay.IT

@PGDayIT

Follow



We are very proud to announce the Patronage of [#BolognaMetropolitana](#) [#PGDayIT](#) [#2019](#) 🔥🔥 [#Bologna](#) [#May_16th](#) and [#May_17th](#) 🔥🔥 [#ITPUG](#) [#PostgreSQL](#) [#OpenSource](#) 🐘🐘🐘 [2019.pgday.it/en/](#)
🐘🐘🐘



Conference – Bologna, May 17th, 2019

Con il Patrocinio della
LIBERTÀ
CITTÀ
METROPOLITANA
DI BOLOGNA

<https://2019.pgday.it/>

PGDay.IT 2019

Workshop Day – Bologna, May 16th, 2019





HORTONWORKS

IBM and Hortonworks: Our Journey

Others can simply not compete with this set of capabilities!

 **IBM Wins
2017 Global
Partner of the
Year!**

September
2016



IBM Power
Systems
partnership

February
2017



IBM Spectrum
Scale
Certification
Announced

April
2017



HDP Delivered
for Power
Systems

June
2017



HDP and HDF
Certified for
Spectrum Scale

Analytics
Partnership –
Hadoop & Data
Science

August
2017



BigSQL and
DSX on HDP

December
2017



HDF Delivered
for Power
Systems

July/Aug
2018

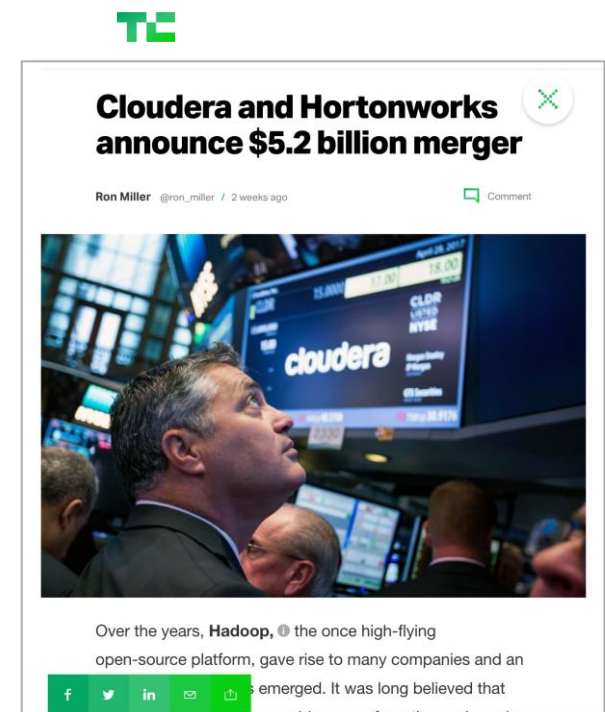
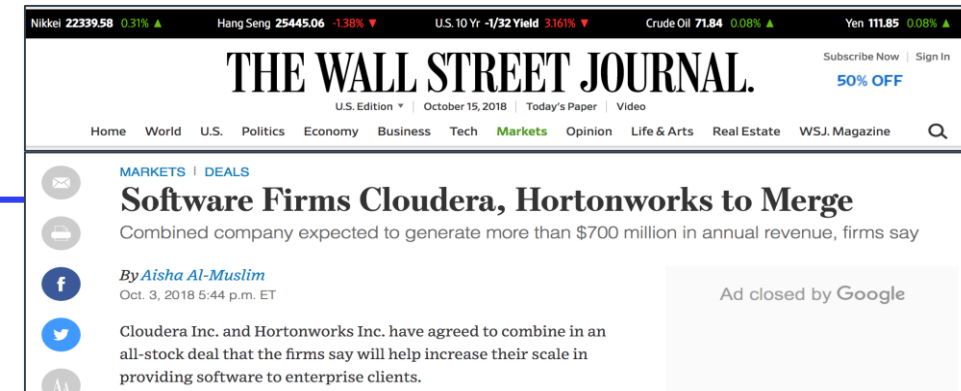


HCP, HDP 3 and
HDF 3.2 for
Power Systems

HDP and HDF
POWER9
Support

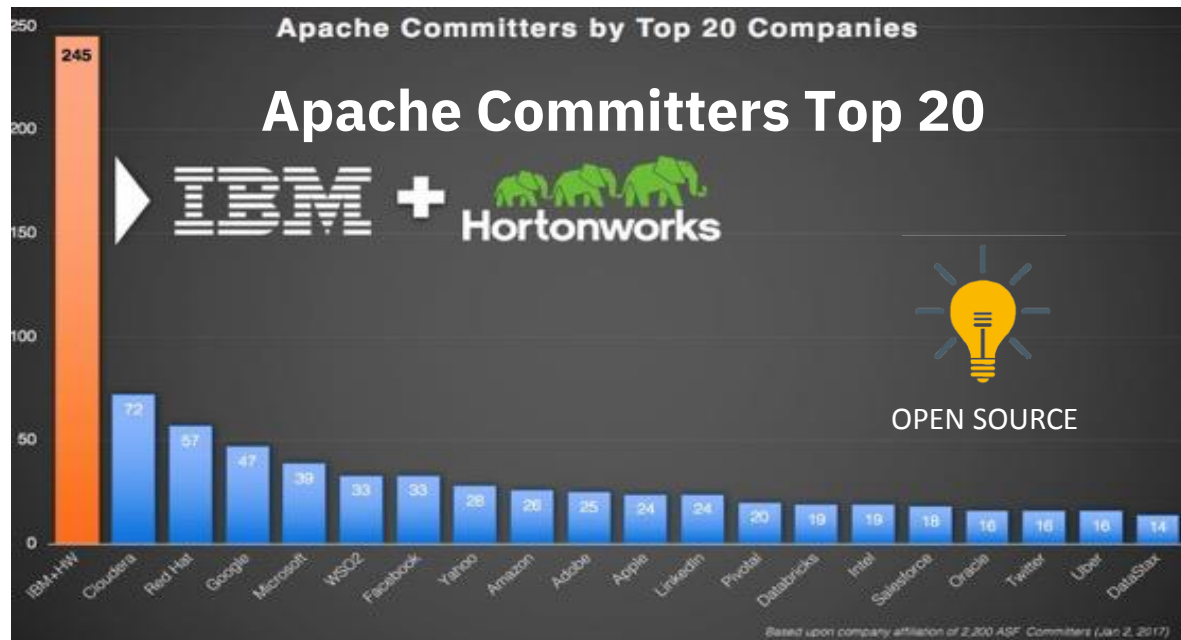
Hortonworks and Cloudera to Merge

- Announced 2018 Oct 3 – Deal closed on Jan 3, 2019
- Three year commitment to support current products (IBM Power and ESS support included)
- Post-merge: the combined company will create ‘Unity’ release
- Hortonworks Founder (Arun Murthy) becomes Chief Product Officer to lead product development at combined company
- **IBM relationship quoted by new CEO to be key partnership to be maintained post-merger**



IBM is the industry leader in analytics platforms

IBM Consensus Leader in Data Science & Business Analytics



Hortonworks is the leading contributor to Apache Hadoop

Together IBM and Hortonworks lead the open source community

IBM Power and Spectrum Scale lead in performance and flexibility

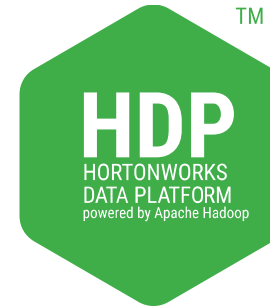
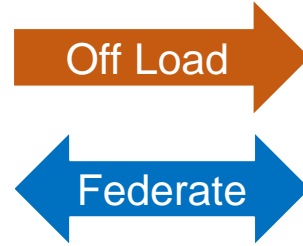


Hortonworks - Optimized Infrastructure with Power 9



Data At Rest

Relational Databases
(DB2, Oracle, Teradata, Netezza, SAP Hana, PostgreSQL, MySQL...)



Hortonworks Data Platform (HDP)

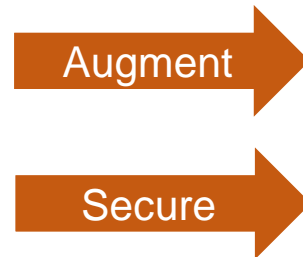


Advanced Analytics and Data Science
Explore, Discover, Analyze



Data in Motion

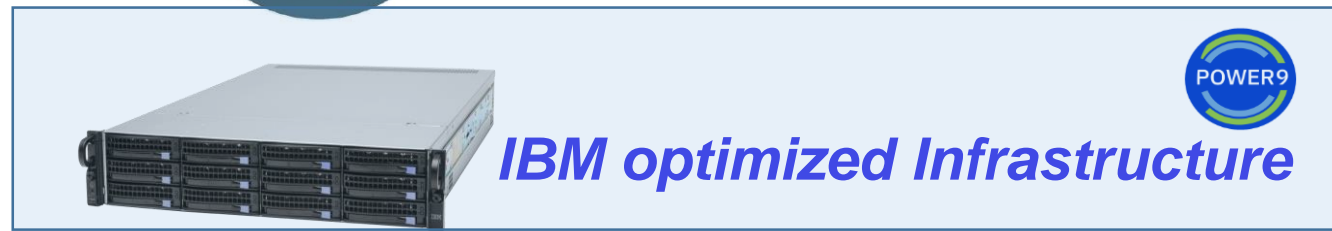
Live Streams



Hortonworks Data Flow (HDF)

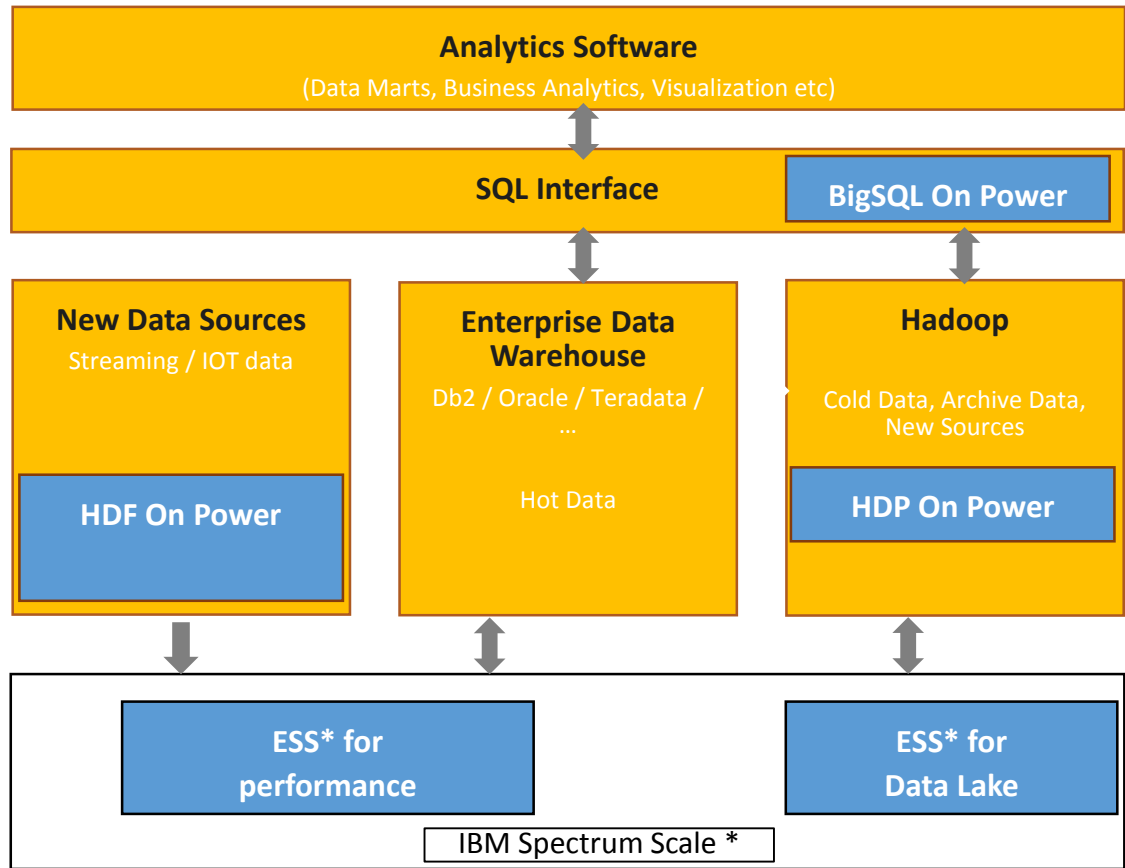


Hortonworks Cybersecurity Platform (HCP)



Enterprise Data Warehouse Modernization

Optimize the data warehouse by shifting right workload to Hadoop



*ESS= IBM Elastic Storage Server (hardware+software) → * IBM Spectrum Scale (software GPFS)

Archive Data away from EDW

- Move cold or rarely used data to Hadoop as active archive
- Get common namespace with Spectrum Scale between EDW and Hadoop for easier data movement
- Store more of data longer

Offload costly ETL process

- Free your EDW to perform high-value functions like analytics & operations, not ETL
- Use Hadoop for advanced ETL

Optimize the value of your EDW

- Use Hadoop to refine new data sources, such as web and machine data for new analytical context

Reduce migration effort & skillset gap

- Use existing investment in Oracle/Db2/Teradata...skills
- BigSQL allows you to migrate applications without major code rewrites and additional SQL development

Control cluster sprawl

- Grow storage independent of compute with ESS
- Significant savings in infrastructure costs with Power and ESS combo.

Hortonworks on IBM POWER9 Reference Architecture



	System Mgmt Node	Master Node	Edge Node	Worker Node		
Cluster Type	All	All	All	Balanced	Performance	Storage Dense
Server Model	1U LC921	1U LC921	1U LC921	2U LC922	2U LC922	2U LC922
# Servers (Min/Default/Max)	1 / 1 / 1	3 / 3 / Any	1 / 1 / Any	4 / 8 / Any	4 / 8 / Any	4 / 8 / Any
Sockets	2	2	2	2	2	2
Cores (total)	32	40	40	44	44	44
Memory	32GB	256GB	256GB	256GB	512GB	128GB
Storage - HDD (front)	2x 4TB HDD	4x 4TB HDD	4x 4TB HDD	12x 4TB HDD	8x 4TB HDD	12x 10TB HDD
Storage - SSD (front)					+ 4x 3.8TB SSD	
Storage - HDD (rear for OS)				2x 1.2TB HDD	2x 1.2TB HDD	2x 1.2TB HDD
Storage Controller	MicroSemi PM8069 (internal)	MicroSemi PM8069 (internal)	MicroSemi PM8069 (internal)	MicroSemi PM8069 (internal)	MicroSemi PM8069 (internal)	MicroSemi PM8069 (internal)
Network* - 1 GbE	Internal (4 ports OS)	Internal (4 ports OS)	Internal (4 ports OS)	Internal (4 ports OS)	Internal (4 ports OS)	Internal (4 ports OS)
Cables* - 1 GbE	3 (2 OS + 1 BMC)	3 (2 OS + 1 BMC)	3 (2 OS + 1 BMC)	3 (2 OS + 1 BMC)	3 (2 OS + 1 BMC)	3 (2 OS + 1 BMC)
Network** - 10 GbE	1x 2-port Intel (2 ports)	1x 2-port Intel (2 ports)	2x 2-port Intel (4 ports)	1x 2-port Intel (2 ports)	1x 2-port Intel (2 ports)	1x 2-port Intel (2 ports)
Cables** - 10 GbE	2 cables (DACs)	2 cables (DACs)	4 cables (DACs)	2 cables (DACs)	2 cables (DACs)	2 cables (DACs)
Operating System	RHEL 7.5 for P9	RHEL 7.5 for P9	RHEL 7.5 for P9	RHEL 7.5 for P9	RHEL 7.5 for P9	RHEL 7.5 for P9

* The 1 GbE network infrastructure hosts the following logical networks: campus, management, provisioning and service networks. See Section 7.4.1 for details.

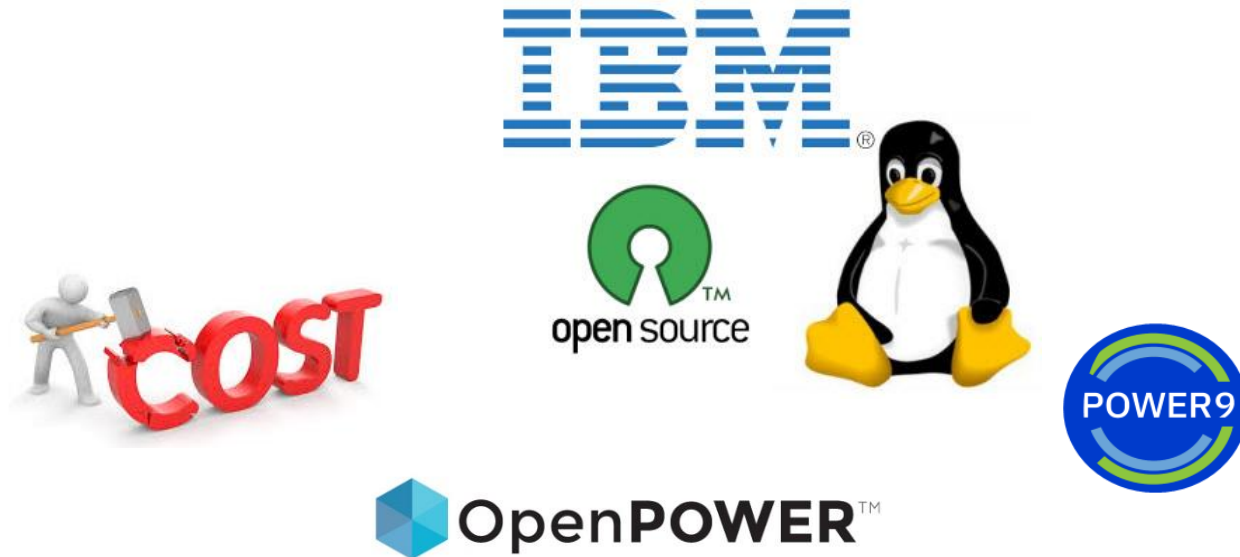
** The 10GbE network infrastructure hosts the data network.

Refer to the published Reference Architectures for [Local Storage](#) and [Shared Storage](#) configurations.

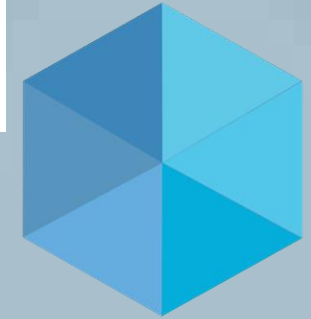
IBM Power Linux

Best platform for Open Source Data Base

(focus on: Scale out Servers)



Innovation from an ecosystem of partners across the stack and open to the core



OpenPOWER™ READY

OPEN POWER

350+ OpenPOWER members co-design around the core to accelerate cognitive and general workloads



OPEN SOURCE

Density, speed-up, compaction of the most ubiquitous open source engines



OPEN CAPI

Laying the groundwork for faster coherent open interfaces to attach to accelerators



OPEN FRAMEWORKS

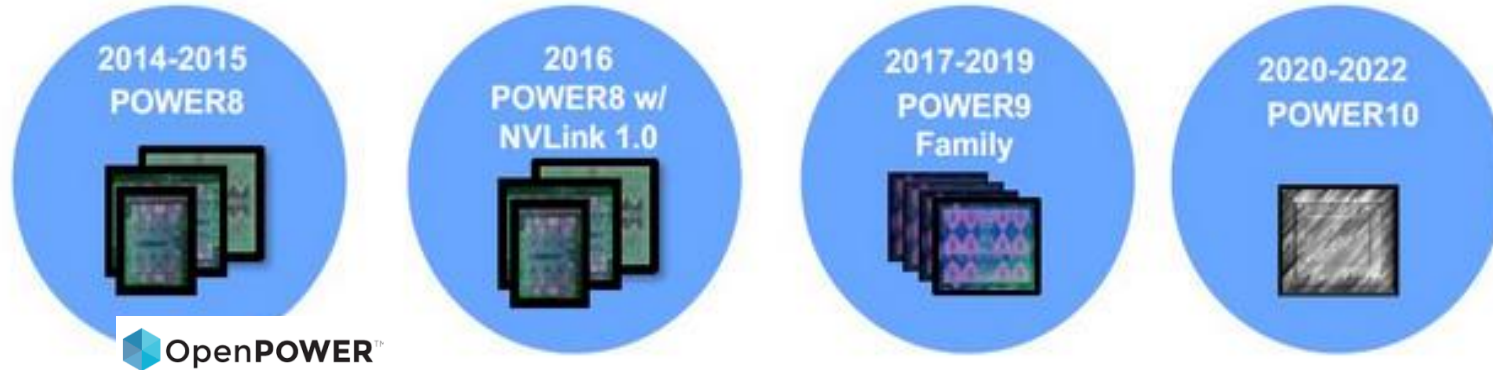
The industry's most ubiquitous Cognitive/AI frameworks - optimized and accelerated.



Infrastructure matters: *IBM Power System*



POWER Processor



Acceleration

PCIe ● OpenCAPI ● NVIDIA NVLink ● Persistent Memory

Solutions

Machine/Deep Learning ● PowerAI ● AI Vision

GPU DBs ● OpenSource DBs ● In-Memory Analytics

System SW

AIX ● i for Business ● Linux

PowerVM ● KVM ● PowerSC

PowerVC ● openstack. ● PowerHA

Why IBM POWER 9 is *ideal* for the MDP Era



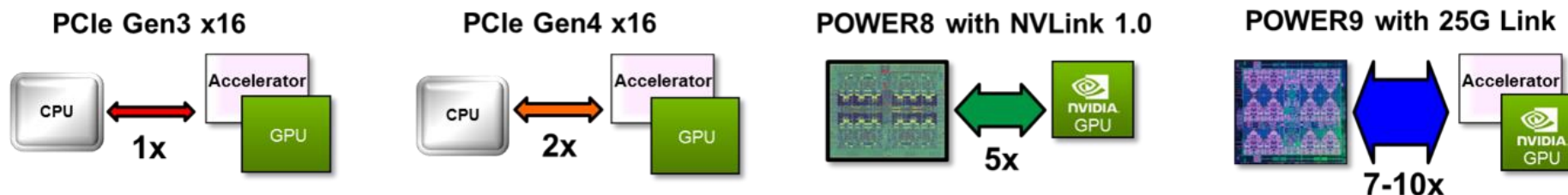
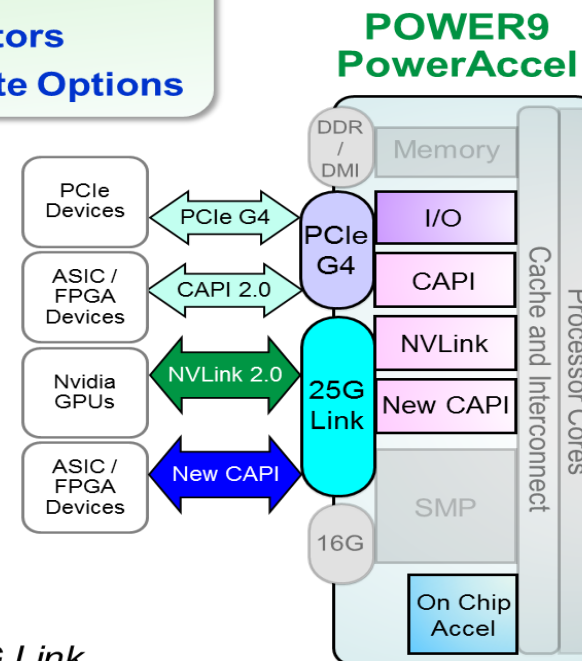
- Extreme Processor / Accelerator Bandwidth and Reduced Latency
- Coherent Memory and Virtual Addressing Capability for all Accelerators
- OpenPOWER Community Enablement – Robust Accelerated Compute Options

- **State of the Art I/O and Acceleration Attachment Signaling**

- PCIe Gen 4 x 48 lanes – 192 GB/s duplex bandwidth
- 25G Link x 48 lanes – 300 GB/s duplex bandwidth

- **Robust Accelerated Compute Options with OPEN standards**

- On-Chip Acceleration – Gzip x1, 842 Compression x2, AES/SHA x2
- CAPI 2.0 – 4x bandwidth of POWER8 using *PCIe Gen 4*
- NVLink 2.0 – Next generation of GPU/CPU bandwidth and integration
- New CAPI – High bandwidth, low latency and open interface using *25G Link*

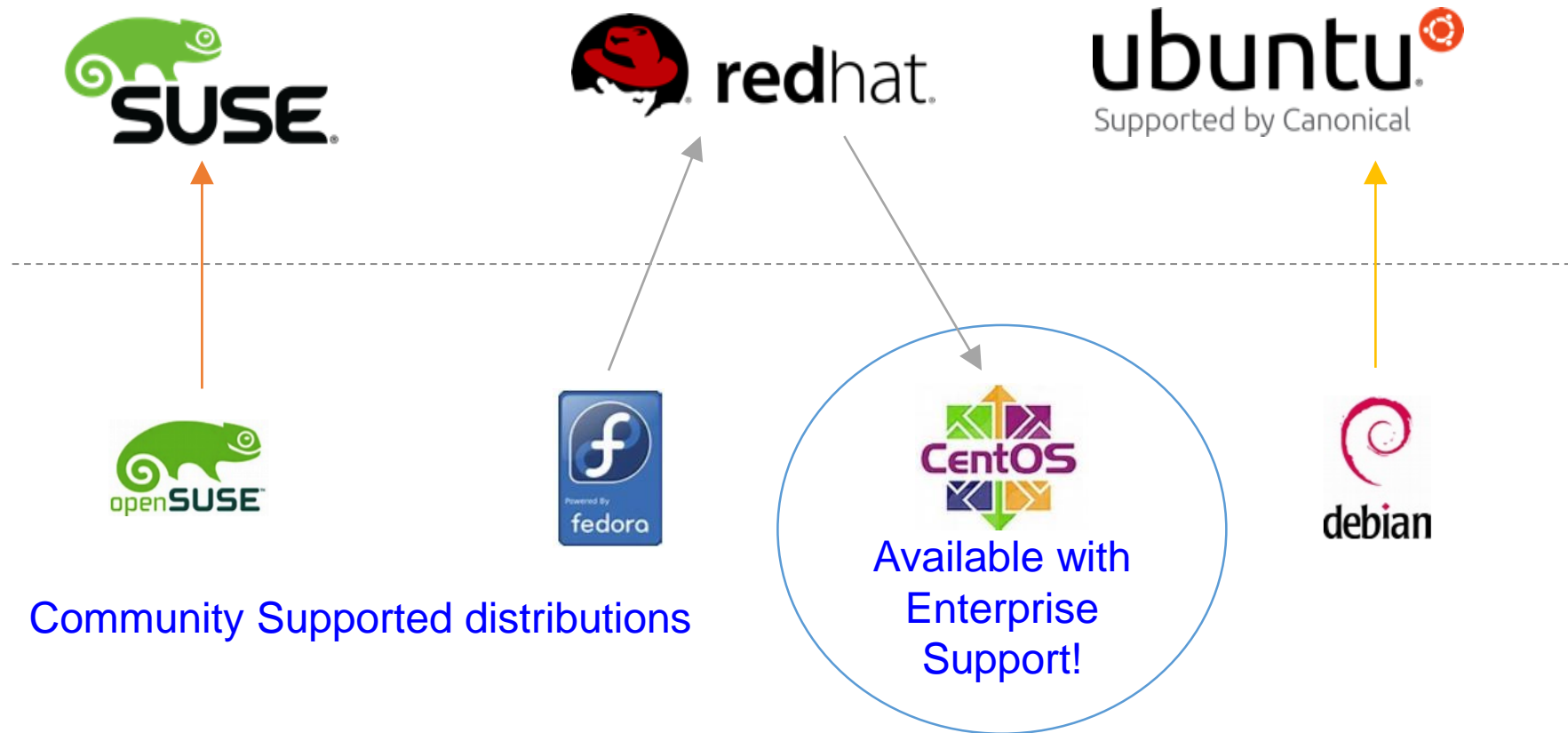


Increased Performance / Features / Acceleration Opportunity

Linux Distributions on IBM Power Systems



- › Built from the same source as x86
- › Delivered on same schedule as x86
- › Supported at the same time as x86



POWER9 - IBM Power Systems Virtualization



Fastest and most efficient virtualization system for Linux environment



IBM POWER 9



PowerVM



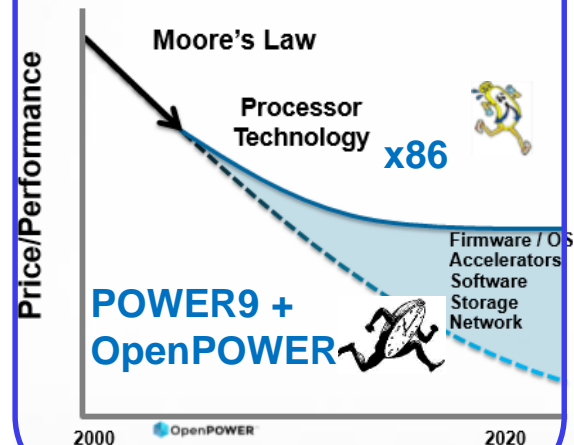
Key Advantages of IBM Power Systems versus x86 servers



CPU Performance

3x more cache per processor
2-4x Memory Bandwidth
4x Threads per core
Vs x86
Broadwell/Skylake

Hardware Accelerators



Flexibility



Availability

Less Downtime



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

Sources: Broadwell EP: <http://ark.intel.com/products/family/91287/Intel-Xeon-Processor-E5-v4-Family#@Server>
POWER8: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=BR&infotype=PM&appname=STGE_PO_PO_USEN&htmlfid=POB03046USEN
ITIC reliability survey : <http://itic-corp.com/category/itic-survey-results/>
VMware: <https://www.vmware.com/pdf/vsphere6/r60/vsphere-60-configuration-maximums.pdf>



A portfolio for the new Era

From Mission-Critical workloads to AI and Cloud Computing leadership

PowerVM and high RAS

L922



- Industry leading reliability and computing capability
- PowerVM ecosystem focus for outstanding utilization
- Focus on memory capacity with up to 4TB of RAM

Accelerated Compute

AC922



- Industry first and only in advanced IO with 2nd Generation CPU - GPU NVLink delivering ~5.6X higher data throughput
- Up to 4 integrated NVIDIA "Volta" GPUs air cooled (GTH) and up to 6 GPUs with water cooled (GTX) version
- Only platform with memory coherence for data capacity needed in the AI era
- ~4x better performance for model training⁽¹⁾

Big Data

LC922



Announce 5/8 GA 5/25

- Highest storage capacity (120 TB) in the Power portfolio, KVM support, leveraging P9 compute for a compact design
- Advanced IO with PCIe 4.0/CAPI 2.0
- Up to 44 cores and 2TB RAM
- Superior performance while having a list price which is ~30% less⁽⁴⁾

LC921



Announce 5/8 GA 5/25

- Density focused with 2 POWER9 sockets in a 1U form factor
- Up to 40 cores and 2TB RAM- ideal for environments requiring dense computing

IBM Power System and The Modern Data Center

IBM Power9 LC922

“The Big Data Crusher”



Data lakes and data intensive applications

- Combining the right balance of compute capabilities delivered with the P9 processor and up to 120TB of storage capacity
- Scaling efficiency with advanced IO (PCIe 4.0/CAPI 2.0)
- 2X the number of active users
- 2X the price-performance



IBM Power9 AC922

“The Best Server for Enterprise AI”



ML/DL Model Training

- Only server with 2nd Gen NVLink between CPU-GPU in the industry delivering leadership acceleration
- Co-optimized infrastructure and software with PowerAI
- ~4x the performance over x86 for AI workloads

IBM Power Systems LC922



Designed to crush Big Data Workloads in the New Era

The LC922 enhances the LC product line's open heritage and cost optimization with the new P9 processor which delivers the compute and scaling capability needed in today's AI Era.

- LC922 delivers 2x price-performance over x86 alternatives for leading data rich applications (3)
- Enable ~2x more active users on a single server with LC922 improving infrastructure flexibility for data scientist (2)
- Industry leadership in IO with the industry's first general compute system containing PCIe 4.0 and CAPI 2.0
- Optimized design to deliver superior performance at a ~30% lower list price than a equivalent HP DL380 (4)



System Details

- 2 sockets, 2U
- Up to 44 cores (2.6-2.91Ghz)
- 2TB Max Memory (16 DIMMs)
- 12 SFF/LFF (HDD/SSD) (4x NVMe enabled)
 - Optional additional rear 2x SFF SAS/SATA for OS
- Max 120 TB storage (HDD)
- Max 45.6TB (SSD) capacity
- 6 PCIe Gen4 slots
 - 2 PCIe G4 x16 FHFL slots, CAPI2.0 enabled
 - 3 PCIe G4 x8 FHFL slots, physically x16
 - 1 PCIe G4 x8 LP slot
- Default 3 year 9x5 warranty, 100% CRU
- OS Support – Linux LE
 - Ubuntu 18.04 LTS
 - RHEL 7.5 for Power9 (i.e. Pegas 1.1)
 - Multiple VM support will be available Mid June 2018

IBM Power Systems LC921

Drives the most compute in the smallest rack space

Industry leading compute in a dense 1U form factor with 2 P9 processors to meet the challenges of data growth in the AI Era



System Details

- Ideal for dense compute, virtualization and container deployments requiring a balanced system design of through-put and density
 - Ideal scaling capabilities leveraging PCIe 4.0 with InfiniBand CX-5 for dense computing deployments
 - Maximize data center efficiency and reduce infrastructure cost
- 2 socket, 1U
 - Up to 40 cores (2.13-2.2Ghz)
 - 2TB Max Memory (16 DIMMs)
 - 4 LFF/SFF Bays (4x NVMe enabled)
 - Max 40 TB storage (HDD)
 - Max 15.2TB (SSD) capacity
 - 4 PCIe Gen4 slots
 - 2 PCIe G4 x16 FHFL Slot, CAPI 2.0 enabled
 - 1 PCIe G4 x8 LP slot
 - 1 PCIe G4 x8 LP slot, CAPI 2.0 enabled (internal)
 - Default 3 yr 9x5 Warranty 100% CRU
 - OS Support – Linux LE
 - Ubuntu 18.04 LTS
 - RHEL 7.5 for Power9 (i.e. Pegas 1.1)
 - Multiple VM support will be available Mid June 2018

Open Source Myth :

"x86 is the best platform for Open Source Databases"

Open Source DBs deliver **1.8-2X+** greater value on
POWER9 vs. x86.....and

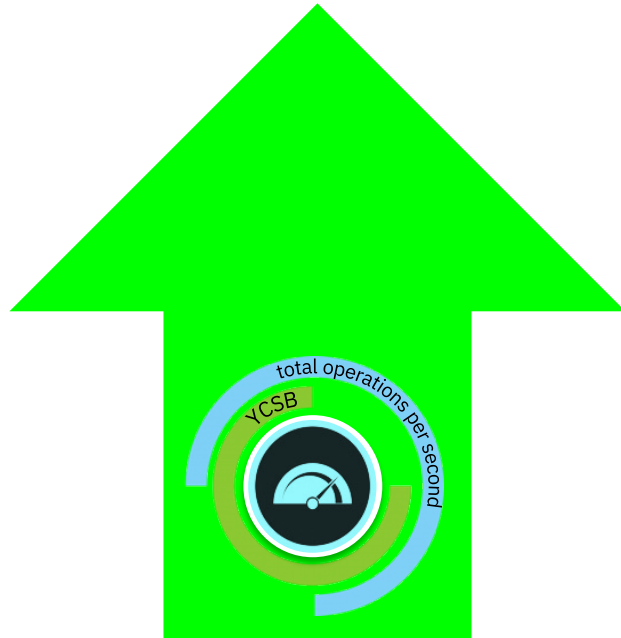


Power LC922 Server: Improved Price-Performance for Clients

Better Performance and Lower Cost running **YCSB** (Yahoo Cloud Serving Benchmark) with **MongoDB** than tested Intel Xeon SP servers

Power LC922
472,927 Ops/sec

Intel Xeon SP Gold
6150 server:
\$30,587



28% LOWER
Price^{2,3,4}

Intel Xeon SP Gold
6150 server:
322,738 Ops/sec

Power LC922
\$21,878



Power L922 Delivers
2X
Price-performance

1. Based on IBM internal testing of MongoDB 3.6.2 using YCSB workload, Results valid as of 4/11/18 and conducted under laboratory condition with speculative execution controls to mitigate user-to-kernel and user-to-user side-channel attacks on both systems, individual results can vary based on workload size, use of storage subsystems & other conditions.
2. IBM Power LC922 (2x22-core/2.6 GHz/256 GB memory) using 2 x internal HDD, 10 GbE two-port, 1 x 16gbps FCA running 4VM's of Mongo 3.6 and RHEL 7.5 LE for Power9Competitive stack: 2-socket Intel Xeon SP (Skylake) Gold 6150 (2x18-core/2.7 GHz/256 GB memory) using 2 x 300GB SATA 15K rpm HDD, 10 GbE two-port, 1 x 16gbps FCA , running 3 VM's Mongo 3.6 and RHEL
3. Pricing is based on Power LC922 <http://www-03.ibm.com/systems/power/hardware/linux-lc.html> and publicly available x86 pricing.

Reduce DBaaS operating costs with Power L922 Server running EnterpriseDB PostgreSQL Advance Server 10

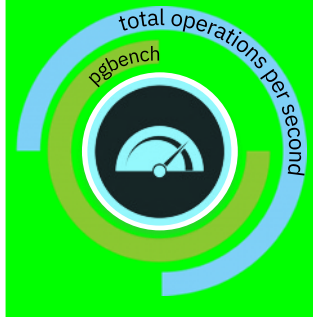
2.4X price-performance leadership over tested Intel Xeon SP Gold 6148 servers

Power L922
853,709 tps

Intel Xeon SP Gold
6148 server:
\$243,166



40% MORE
Performance¹



40% LOWER
Solution
Price^{2,3,4}

Intel Xeon SP Gold
6150 server:
611,118 tps

Power L922
\$144,557

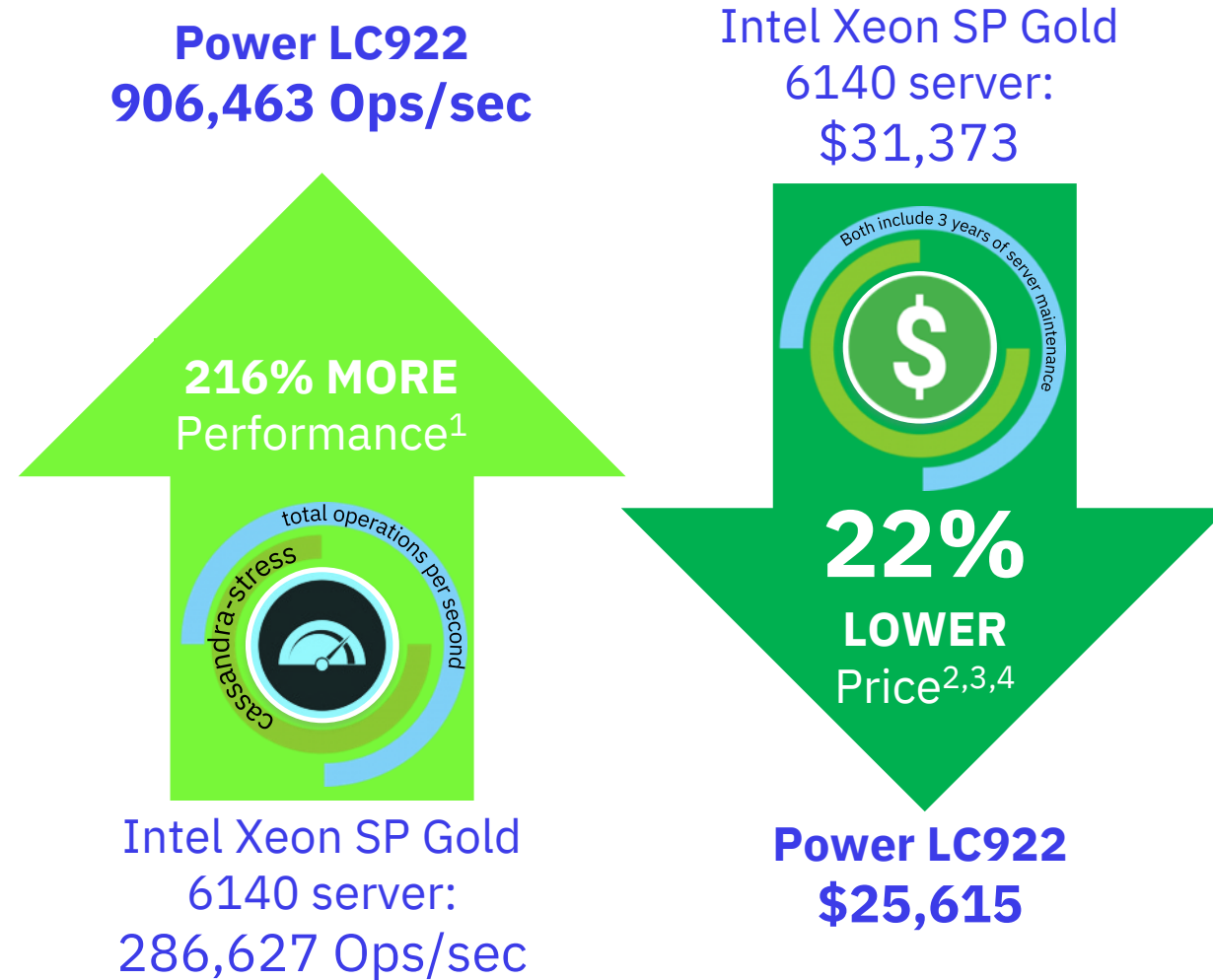


Power L922 Delivers
2.4X
Price-performance

1. Based on IBM internal testing of multiple VM images running pgbench benchmark at scale factor of 300, 20 GB buffer size. Results valid as of 4/19/18. and conducted under laboratory condition with speculative execution controls to mitigate user-to-kernel and user-to-user side-channel attacks on both systems, individual result can vary based on workload size, use of storage subsystems & other conditions.
2. IBM Power L922 (2x10-core/2.9 GHz/256 GB memory) 2 x 300GB SATA 7.2K rpm LFF HDD, 10 Gb two-port, 1 x 16gbps FCA, EDB Postgres Advanced Server 10, RHEL 7.5 with PowerVM (4 partitions@5-cores each),
3. Competitive stack: 2-socket Intel Xeon Skylake Gold 6148 (2x20-core/2.4 GHz/256 GB memory), 2 x 300GB HDD, 1 Gb two-port, 1 x 16gbps FCA , EDB Postgres Advanced Server 10, RHEL 7.5, KVM (4 VMs@10-cores each)
4. Pricing is based on Power L922 pricing, EDB <https://webcms.enterprisedb.com/products/subscriptions> , and publicly available x86 pricing.

Power LC922 running cassandra-stress delivers superior performance¹ with **ScyllaDB** vs Cassandra on tested x86 systems and at a lower price

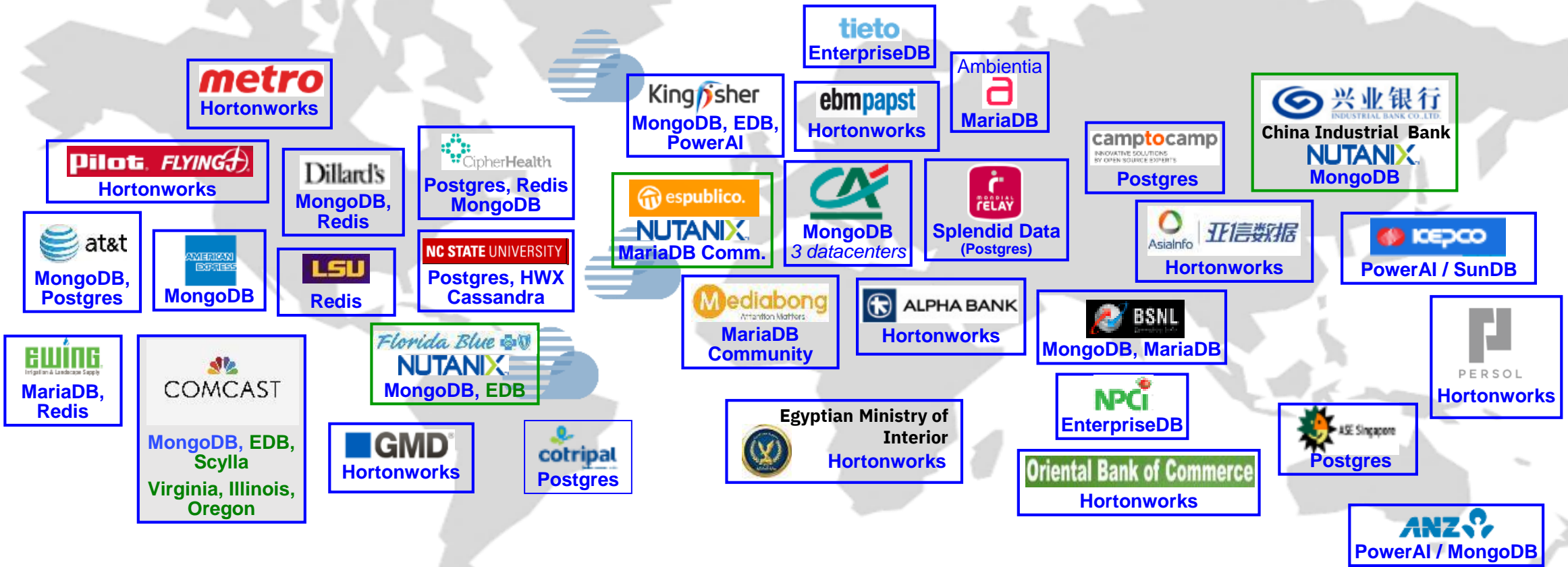
3.9X price-performance versus Intel Xeon SP Gold 6140 based servers



**Power LC922 Delivers
3.9X
Price-performance**

1. Based on IBM internal testing of cassandra-stress workload using Gaussian 9M,4.5M,10K model with 80% reads/20% writes , Results valid as of 5/21/18 and conducted under laboratory condition with speculative execution controls to mitigate user-to-kernel and user-to-user side-channel attacks on both systems, individual results can vary based on workload size, use of storage subsystems & other conditions.
2. IBM Power LC922 (2x22-core/2.6 GHz/256 GB memory) using 2 x internal HDD, 40 GbE, 1 x 1.6TB NVMe adapter running Scylla Enterprise 2018.1.0 on RHEL 7.5 LE for Power9
3. Competitive stack: 2-socket Intel Xeon SP (Skylake) Gold 6140 (2x18-core/2.3 GHz/256 GB memory) using 2 x internal HDD, 40 GbE, 1 x 1.6TB NVMe adapter running Open Source Cassandra 3.11.2 on RHEL 7.5
4. Pricing is based on Power LC922 <http://www-03.ibm.com/systems/power/hardware/linux-lc.html> and publicly available x86 pricing.

Momentum for Open Source DBs & Cognitive



Grazie per l'attenzione

