3rd Big Data Innovation Conference

7-8 June 2018
Frankfurt Germany

Improving Business Performance Through Big Data Benchmarking

DataBench

Todor Ivanov (todor@dbis.cs.uni-frankfurt.de)
Frankfurt Big Data Lab, Goethe University

IDC ANALYZE THE FUTURE

Atos

Institut "Jožef Stefan", Ljubljana, Slovenija

SINTEF

POLITECNICO MILANO 1863

Frankfurt Big Data Lab
GOETHE UNIVERSITÄT

Benchmarking

- The Term Benchmark:

A benchmark is a measured „best-in-class“ achievement recognised as the standard of excellence for that business process. (APQC 1993)

- ***Business Performance Benchmarking*** – comparison of performance measures for the purpose of determining how good one’s own company is compared to others.

- ***A software benchmark*** is a program used for comparison of software products/tools executing on a pre-configured hardware environment.

Performance Metrics

- ***Lord Kelvin defined KPIs as:*** "When you can **measure** what are speaking about and **measure** it **in numbers**, you know something about it, when you **cannot express it in numbers**, your **knowledge is of meager and unsatisfactory kind**; it may be the beginning of knowledge but you have scarcely, in your thoughts advanced to the stage of science."

  [Arora Amishi, Kaur Sukhbir. Performance assessment model for management educators based on KRA/KPI. In: International conference on technology and business management March, vol. 23; 2015.]

- ***Key Performance Indicators (KPIs)*** - tell you what to do to highly increase performance.

- ***Jim Gray*** describes the benchmarking as: "This *quantitative comparison* starts with the definition of a benchmark or workload. The benchmark is run on several different systems, and the **performance and price** of each system **is measured and recorded**. Performance is typically a **throughput metric (work/second)** and **price is typically a five-year cost-of-ownership metric**. Together, they give a ***price/performance ratio***."

  [Jim Gray (Ed.), The Benchmark Handbook for Database and Transaction Systems, 1992]

- TPC defines ***complex performance metrics*** for each standard benchmark:
  - Transaction per second (tps)
  - $/tps

Example Use Case

System A	System B
4 Nodes (servers)	30 Nodes (servers)
4 hours (execution time)	4 minutes (execution time)
5500 Euro (total cost)	50 000 Euro (total cost)

- Which system will have ***best price/performance*** for my application?

→ ***Need to use a benchmark***

- What is more important the ***minimal execution time or lower total cost***?

→ ***Depends on the business KPIs***

- The company decide to introduce Machine Learning model to improve product recommendations to customers. Different ML models have different Accuracy. How important is the accuracy? Should the company invest in improving the model accuracy?

→ ***Business decision***

Goals of DataBench

- *… There is a lack of objective, evidence-based methods measuring the correlation between Big Data Technology benchmarks and Business benchmarks as well as Big Data Technology capability to impact the competitiveness and market success of organizations.*

- Objective:

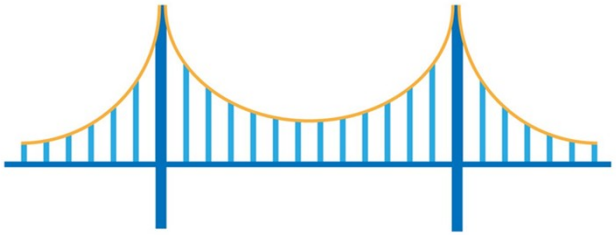
… The objective is to provide a model which correlates technical benchmarks to performance and business needs of different sectors and domains. …

DataBench

Evidence Based Big Data Benchmarking to Improve Business Performance

Building a bridge between technical and business benchmarking

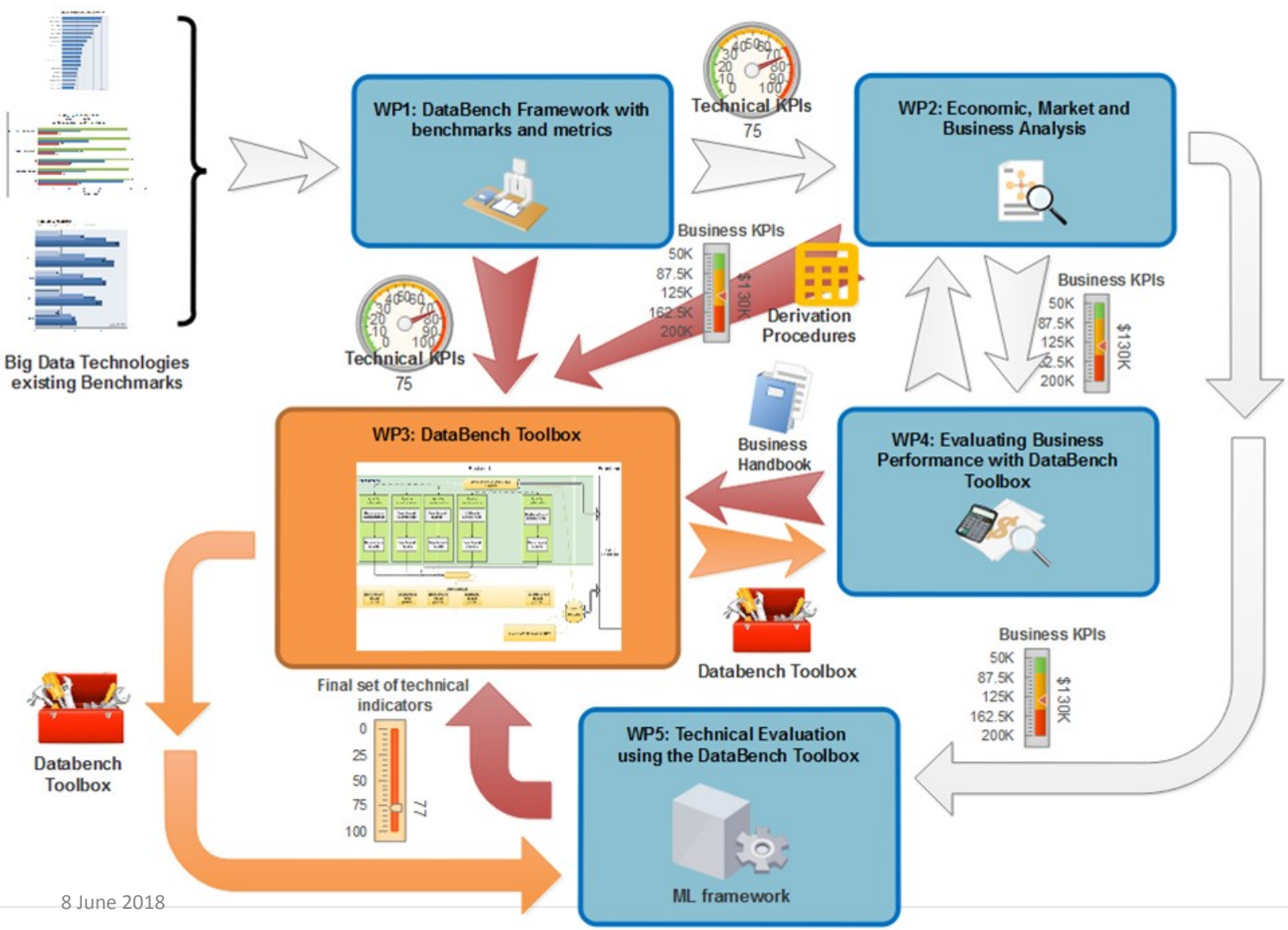
Mapping and assessing technical benchmarks



Evaluating business performance and benchmarks

Develop a Benchmarking Toolbox and Handbook

IDC ANALYZE THE FUTURE · POLITECNICO MILANO 1863 · SINTEF · Atos · IJS · Frankfurt Big Data Lab GOETHE UNIVERSITÄT


Big Data Technologies existing Benchmarks
WP1: DataBench Framework with benchmarks and metrics
Technical KPIs
75
WP2: Economic, Market and Business Analysis
Business KPIs
50K
87.5K
125K
162.5K
200K
$130K
Derivation Procedures
Business KPIs
50K
87.5K
125K
162.5K
200K
$130K
Technical KPIs
75
WP3: DataBench Toolbox
Business Handbook
WP4: Evaluating Business Performance with DataBench Toolbox
Databench Toolbox
Business KPIs
50K
87.5K
125K
162.5K
200K
$130K
Final set of technical indicators
0
25
50
75
100
77
Databench Toolbox
WP5: Technical Evaluation using the DataBench Toolbox
ML framework

DataBench ToolBox

DataBench

Goals & Objectives

Holistic benchmarking approach for big data

- The DataBench Toolbox will be a **component-based system** of both **vertical** (holistic/business/data type driven) **and horizontal** (technical area based) **big data benchmarks**. **following** the layered architecture provide by **the BDVA reference model**.

Not reinventing the wheel, but use wheels to build a new car

- It should be able to **work** or integrate **with existing benchmarking initiatives** and resources where possible.

Filling gaps

- The Toolbox will investigate **gaps of industrial significance** in the big data benchmarking field and contribute to overcome them.

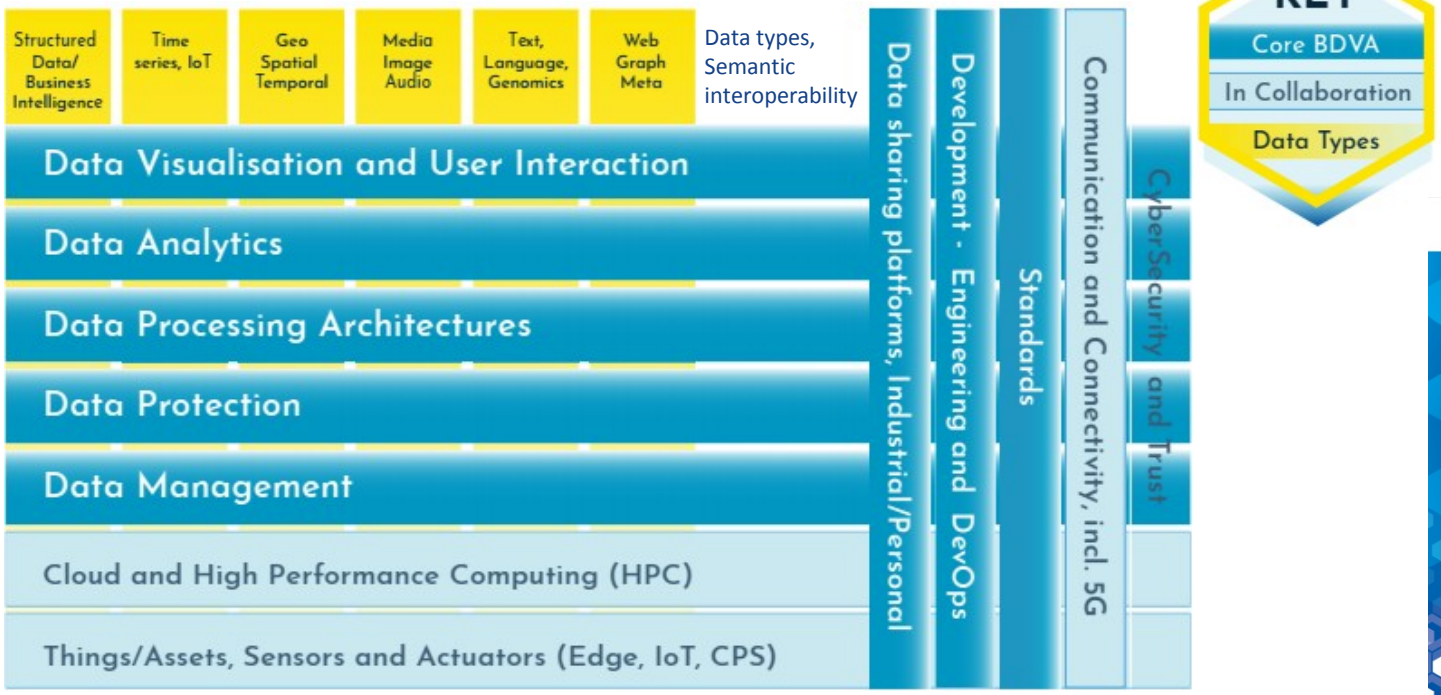
Homogenising metrics

- The Toolbox will implement ways to derive as much as possible **the DataBench technical metrics and business KPIs** from the metrics extracted from the integrated benchmarking.

Web user interface

- It will include a web-based visualization layer to **assist to the final users to specify their benchmarking requirements**, such as selected benchmark, data generators, workloads, metrics and the preferred data, volume and velocity, **as well as searching and monitoring** capabilities.

BDV – Big Data and Analytics/Machine Learning Reference Model



Source: http://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf

DataBench

BDVA Reference Model

Work-in-progress (WP1)

		Benchmarks	TPC-H	TPC-DS v1	Linear Road	Hadoop Workload Examples	GridMix	PigMix	MRBench	CALDA	HiBench	YCSB	SWIM	CloudRank-D	PUMA Benchmark Suite	CloudSuite	MRBS	AMP Lab Big Data Benchmark	BigBench	BigDataBench	LinkBench	BigFrame	PRIMEBALL	Semantic Publishing Benchmark (SPB)	Social Network Benchmark	StreamBench	TPCx-HS	SparkBench	TPCx-V	BigFUN	TPC-DS v2	TPCx-BB	Graphalytics	Yahoo Streaming Benchmark (YSB)	ShenZhen Transportation System (SZT)	DeepBench	DeepMark	TensorFlow Benchmarks	Fathom	AdBench	RIoTBench	Hobbit Benchmark	TPCx-HS v2	BigBench V2	Sanza	Penn machine learning benchmark (PM	OpenML benchmark suites	Senska	DAWNBench /MLPerf	IDEBench	ABench	MLPerf
			1999	2002	2004	2007		2008		2009	2010		2011	2012				2013						2014				2015								2016									2017							2018	
Verticals, incl. Data types		Domain/Sector/Business solutions KPIs ...																																																			
	30	Business	X	X													X		X			X			X				X		X	X		X					X			X	X										
	29	Transport			X																														X						X													
	28	Manufacturing																																														X						
	27	Energy																																							X													
	26	Bioinformatics															X																																					
	25	Health																																							X													
	24	Telecommunication																					X	X											X																			
	23	Finance																																																				
	22	Social Media															X				X	X	X		X					X											X													
	21	General Micro-benchmark									X			X	X					X						X	X	X		X			X						X	X		X	X			X	X	X		X			X
	20	Standardized Benchmark	X	X												X								X	X		X		X		X	X	X											X										
	19	MetaData																						X																														
	18	Graph, Network									X									X	X	X		X	X			X					X			X	X					X								X					
	17	Text, NLP, Web				X					X		X		X	X	X		X	X		X	X	X	X	X	X	X		X		X		X		X	X				X	X							X					
	16	Image, Audio																		X			X													X	X												X					
	15	Spatio Temp																		X															X							X													
	14	Time Series, IoT														X	X			X			X											X	X					X	X	X						X						
	13	Structured, BI	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X					X	X	X						X	X					
Analytics, Processing, Data Management, Infra	12	Visual Analytics																																								X				X								
	11	Industrial Analytics (Descriptive, Diagnostic, Predictive, Prescriptive)															X			X				X																	X					X								X
	9	Machine Learning, AI, Data Science									X			X			X		X	X				X				X				X				X	X				X	X			X	X								X
	8	Streaming/ Realtime Processing									X			X						X						X		X						X							X													
	7	Interactive Processing	X	X								X					X	X	X	X	X	X	X	X	X			X	X	X	X	X	X	X		X	X				X	X												
	6	Batch Processing	X	X		X	X	X	X	X	X		X	X	X		X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X			X	X					X												
	5	Data Privacy/Security																																																				
	4	Data Governance/Mgmt																		X																																		
	3	Data Storage	X	X		X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X								X	X										X	
	2	Communication & Connectivity				X																					X																											
	1	Cloud Services & HPC, Edge												X		X		X		X			X						X																									

Some of the benchmarks to integrate (I)

Micro-benchmarks: stress test very specific part of the Big Data technologies typically utilizing synthetic data.

Year	Name	Type
2010	HiBench	Big data benchmark suite for evaluating different big data frameworks. 19 workloads including synthetic micro-benchmarks and real-world applications from 6 categories which are **micro, machine learning, sql, graph, websearch and streaming.**
2015	SparkBench	System for benchmarking and simulating **Spark jobs**. Multiple workloads organized in 4 categories.
2010	Yahoo! Cloud System Benchmark (YSCB)	Evaluates performance of different **"key-value" and "cloud" serving systems**, which do not support the ACID properties. The YCSB++ , an extension, includes many additions such as multi-tester coordination for increased load and eventual consistency measurement.
2017	TPCx-IoT	Based on YCSB, but with significant changes. Workloads of data ingestion and concurrent queries simulating workloads on typical **IoT Gateway systems**. Dataset with data from sensors from electric power station(s).

Some of the benchmarks to integrate (II)

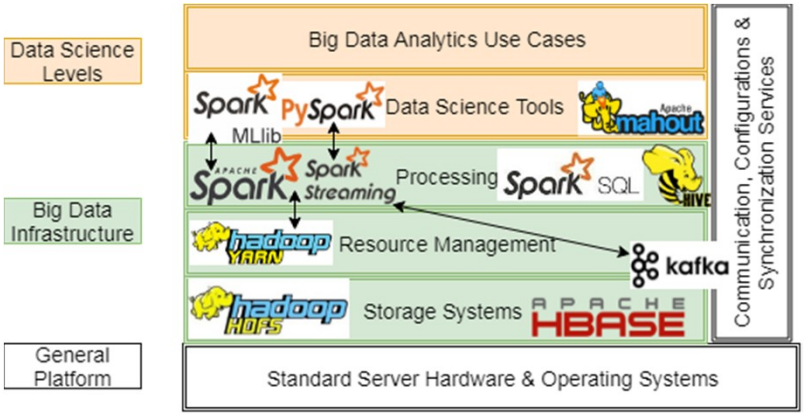
Application-level benchmarks: cover a wider and more complex set of functionalities involving multiple Big Data technologies and typically utilizing real scenario data.

Year	Name	Type
2015	Yahoo Streaming Benchmark (YSB)	The Yahoo Streaming Benchmark is a **streaming application benchmark** simulating an **advertisement analytics** pipeline.
2013	BigBench/TPCx-BB	BigBench is an end-to-end, technology agnostic, application-level benchmark that tests the analytical capabilities of a Big Data platform. It is based on a fictional product retailer business model.
2017	BigBench V2	Similar to BigBench, BigBench V2 is an end-to-end, technology agnostic, application-level benchmark that tests the analytical capabilities of a Big Data platform.
2018	ABench (Work-in-Progress)	New type of multi-purpose Big Data benchmark covering many big data scenarios and implementations. Extends other benchmarks such as BigBench.

ABench (Work-in-Progress)

- Growing number of new ***Big Data technologies*** and ***connectors*** in the Big Data Stacks

→ Challenges for Solution Architects, Data Engineers, Data Scientist, Developers, etc.



ABench: Big Data Architecture Stack Benchmark.
Todor Ivanov, Rekha Singhal:
Companion of the 2018 ACM/SPEC International Conference on Performance Engineering, ICPE 2018, Berlin, Germany, April 09-13, 2018.

- Missing benchmarks for ***each technology***, ***connector*** or a ***combination of them***
- Consequence → ***Increasing complexity in the Big Data Architecture Stacks***
- Our approach → ***ABench: Big Data Architecture Stack Benchmark***

ABench Features

- Benchmark Framework
  - Data generators or plugins for custom data generators
  - Include data generator or public data sets to simulate workload that stresses the architecture

- Reuse of existing benchmarks
  - Case study using BigBench (on-going Streaming and Machine Learning extensions)

- Open source implementation and extendable design

- Easy to setup and extend

- Supporting and combining all four types of benchmarks in ABench

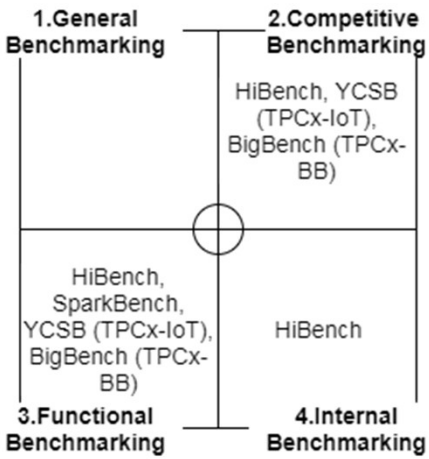
Benchmarks Types (adapted from Andersen and Pettersen)

1. **Generic Benchmarking:** checks whether an implementation fulfills given business requirements and specifications (*Is the defined business specification implemented accurately?).*

2. **Competitive Benchmarking:** is a *performance comparison* between the best tools on the platform layer that offer similar functionality (*e.g., throughput of MapReduce vs. Spark vs. Flink*).

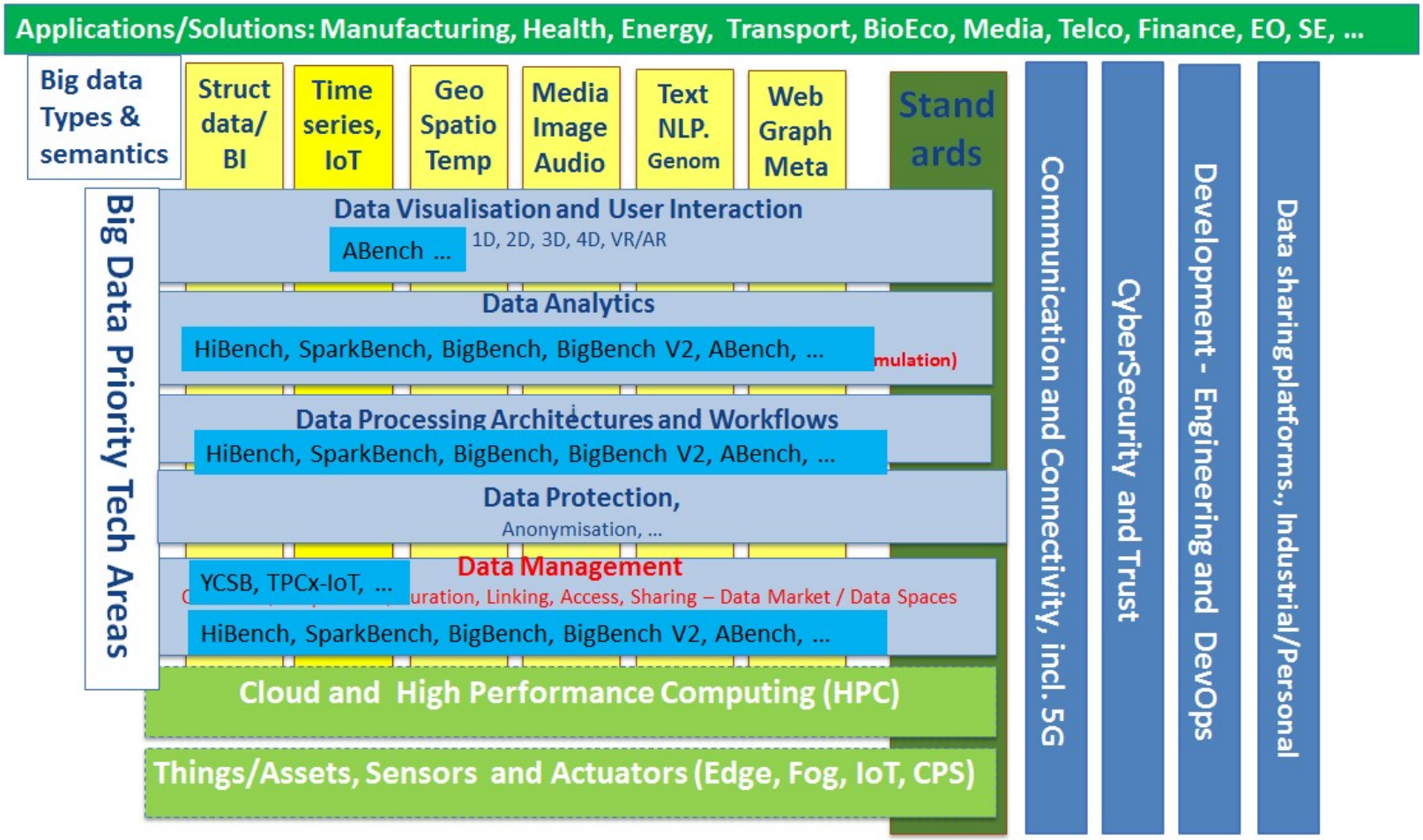
3. **Functional Benchmarking** is a *functional comparison* of the features of the tool against technologies from the same area. (*e.g., Spark Streaming vs. Spark Structured Streaming vs. Flink Streaming*).

4. **Internal Benchmarking:** *comparing different implementations* of a functionality (*e.g., Spark Scala vs. Java vs. R vs. PySpark*)



Category	Year	Name	Type	Domain	Data Type
Micro-benchmarks	2010	HiBench	Micro-benchmark Suite	Micro-benchmarks, Machine Learning, SQL, Websearch, Graph, Streaming Benchmarks	Structured, Text, Web Graph
	2015	SparkBench	Micro-benchmark Suite	Machine Learning, Graph Computation, SQL, Streaming Application	Structured, Text, Web Graph
	2010	YCSB	Micro-benchmark	cloud OLTP operations	Structured
	2017	TPCx-IoT	Micro-benchmark	workloads on typical IoT Gateway systems	Structured, IoT
Application Benchmarks	2015	Yahoo Streaming Benchmark	Application Streaming Benchmark	advertisement analytics pipeline	Structured, Time Series
	2013	BigBench/TPCx-BB	Application End-to-end Benchmark	a fictional product retailer platform	Structured, Text, JSON logs
	2017	BigBench V2	Application End-to-end Benchmark	a fictional product retailer platform	Structured, Text, JSON logs
	2018	ABench (Work-in-Progress)	Big Data Architecture Stack Benchmark	set of different workloads	Structured, Text, JSON logs

Big Data Value Reference Model



DataBench Summary

- A **framework for big data benchmarking** for all Big Data practitioners.
- Will provide **methodology** and **tools.**

- Added value:
  - An **umbrella to access to multiple benchmarks**
  - Homogenized **technical metrics**
  - Derived **business KPIs**
  - **Connecting** the different **benchmark communities**

- Open initiative: Public-private partnership **(**PPP) projects, industrial partners (BDVA and beyond) and benchmarking initiatives are **welcomed to work with us, either to use our framework or to add new benchmarks**.

Contacts

info@databench.eu

@DataBench_eu

DataBench

DataBench Project