

Lecture 6: Understanding Data science and Big Data Motivation/Drivers for Adoption

Benjamin Simeon Harvey (Ben)
Faculty, EMSE Department
The George Washington University
E-mail: bsharve@gwu.edu
Web: TBD

Agenda

Chapter 2: Business Motivations and Drivers for Big Data Adoption

Marketplace Dynamics

Business Architecture

Business Process Management

Information and Communication Technology

- Data Analytics and Data Science
- Digitization
- Affordable Technology and Commodity Hardware
- Social Media
- Hyper-Connected Communities and Devices
- Cloud Computing
- Internet of Everything

Big Data Characteristics



Volume

- **The anticipated amount/volume of data that is processed** by Big Data solutions is substantial and ever-growing. High data volumes impose distinct data storage and processing demands, as well as additional data preparation, curation and management processes.

Velocity

- In Big Data environments, data can arrive at fast speeds, and enormous datasets can accumulate within very short periods of time. From an enterprise's point of view, the **velocity of data translates into the amount of time it takes for the data to be processed once it enters the enterprise's perimeter**.

Variety

- Data variety refers to **the multiple formats and types of data that need to be supported** by Big Data solutions. Data variety brings challenges for enterprises in terms of data integration, transformation, processing, and storage.

Veracity

- Veracity refers to the **quality or fidelity of data**. Data that enters Big Data environments needs to be assessed for quality, which can lead to data processing activities to resolve invalid data and remove noise.

Value

- Value is defined as the **usefulness of data for an enterprise**. The value characteristic is intuitively related to the veracity characteristic in that the higher the data fidelity, the more value it holds for the business.

Big Data Characteristics

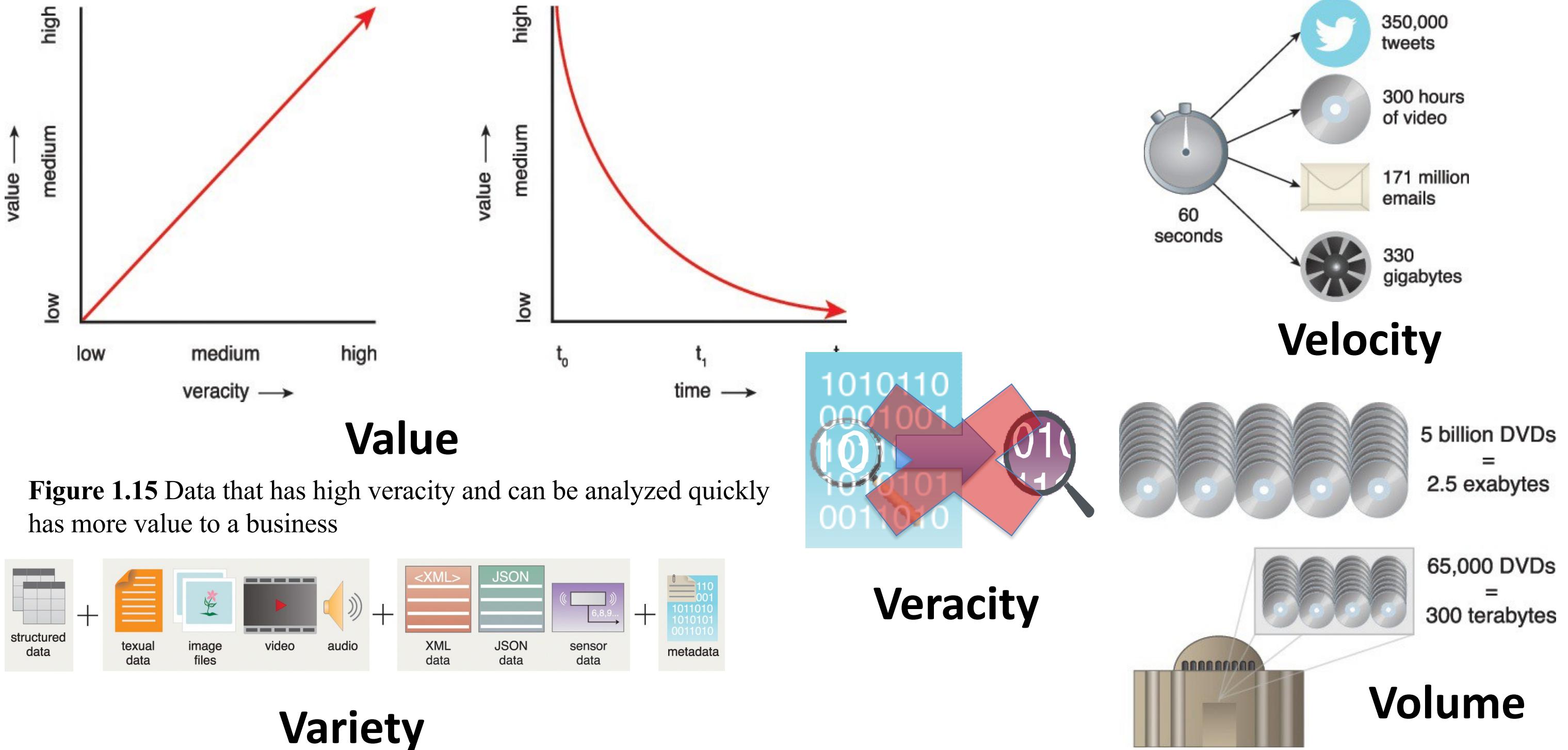


Figure 1.15 Data that has high veracity and can be analyzed quickly has more value to a business

Motivation and Drivers for Big Data Adoption

Here we will explore the business motivations and drivers behind the adoption of Big Data solutions and technologies:

- Marketplace Dynamics,
- An appreciation and formalism of Business Architecture (BA),
- The realization that a business' ability to deliver value is directly tied to Business Process Management (BPM),
- Innovation, Information, and Communications Technology (ICT), and
- The Internet of Everything

Motivation and Drivers for Big Data Adoption

Marketplace Dynamics:

- Businesses working to improve efficiency and effectiveness to stabilize their profitability by reducing costs.
- Companies conduct transformation projects to improve their corporate processes to achieve savings.
- Data – a discrete, objective facts about events.
- Information – data that makes a difference
- Knowledge – a fluid mix of framed experience, values, contextual information and expert insight that provides a framework for evaluating and incorporating new experiences and information.

Motivation and Drivers for Big Data Adoption

GW

- Latitude – Longitude
- Jacksonville FL
- Benjamin Harvey – 10668 Pine Rd Jax FL
- Ben will be at home at 6pm today

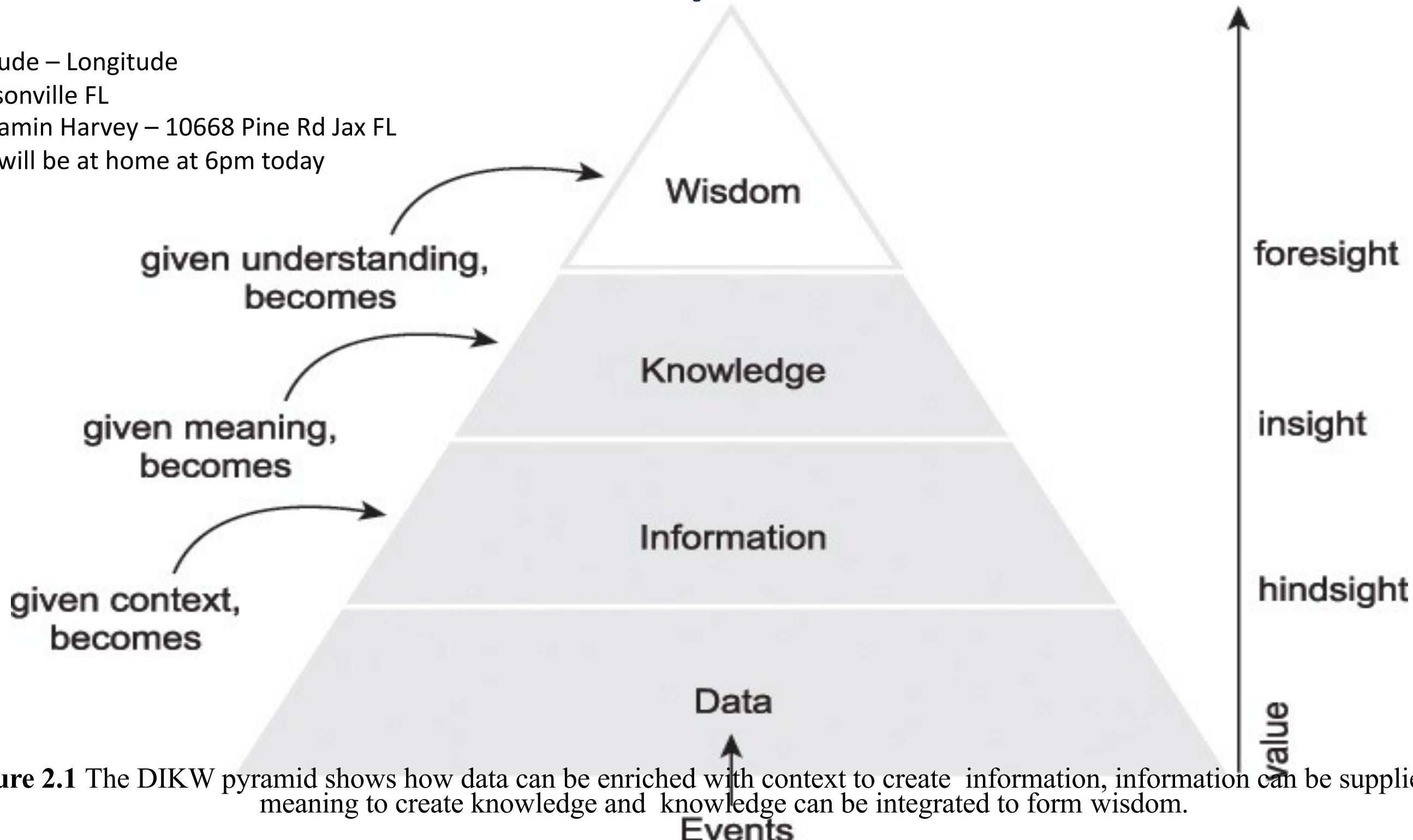


Figure 2.1 The DIKW pyramid shows how data can be enriched with context to create information, information can be supplied with meaning to create knowledge and knowledge can be integrated to form wisdom.

Erl, Thomas, Wajid Khattak, and Paul Buhler. Big data fundamentals: concepts, drivers & techniques. Prentice Hall Press, 2016.

Motivation and Drivers for Big Data Adoption

Business Architecture:

- Provides a means of blueprinting or concretely expressing the design and the processes of the business
- Helps and organization align its strategic vision, goals, objectives with its underlying execution.
- Layered system:
- Strategic Layer– C-Level Executives and Advisory Groups
- Managerial Layer – seeks to steer the organization in alignment with the strategy
- Operations Layer – where a business executes its core processes and delivers value to its customers

Motivation and Drivers for Big Data Adoption

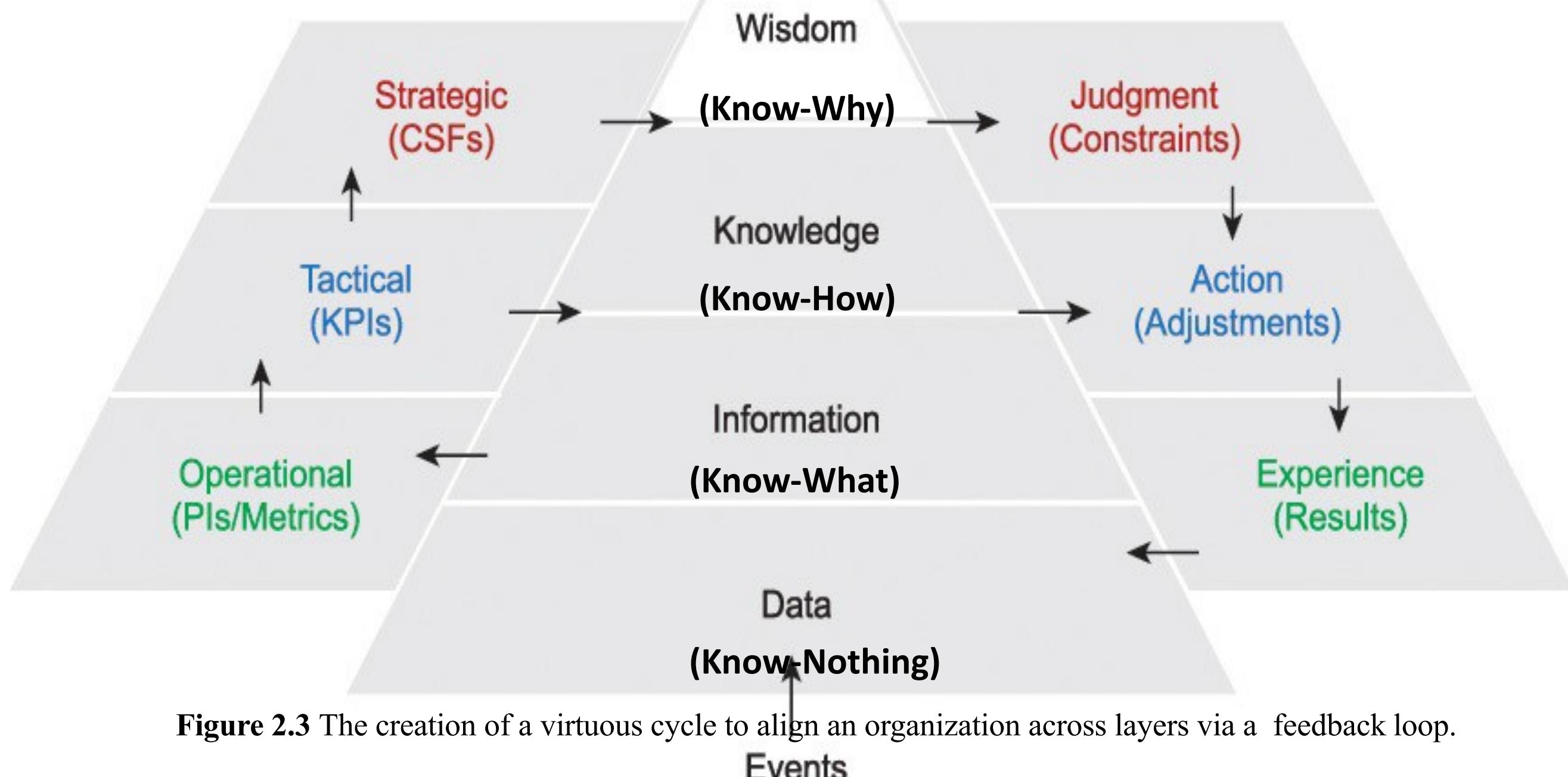


Figure 2.3 The creation of a virtuous cycle to align an organization across layers via a feedback loop.

Erl, Thomas, Wajid Khattak, and Paul Buhler. Big data fundamentals: concepts, drivers & techniques. Prentice Hall Press, 2016.

Motivation and Drivers for Big Data Adoption



Business Process Management:

- A business process is a description of how work is performed in an organization. It describes all work-related activities and their relationships, aligned with the organizational actors and resources responsible for conducting them.
- When the combination of Big Data analytic results and goal-driven behavior are used together, process execution can become adaptive to the marketplace and responsive to environmental conditions.

Motivation and Drivers for Big Data Adoption

GW

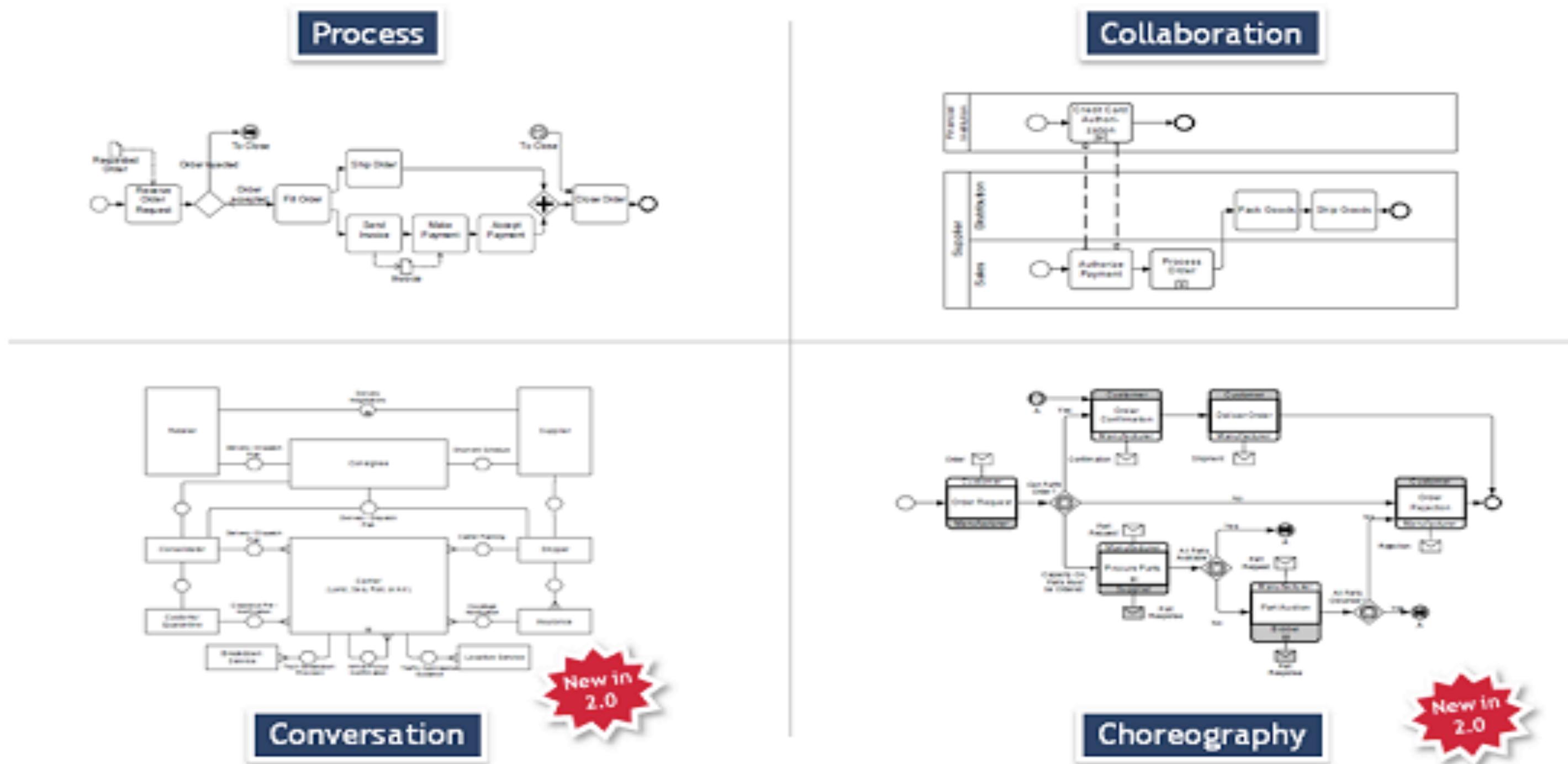


Figure: BPMN four model types

Source: <http://www.ariscommunity.com/users/roland-woldt/2011-01-28-learning-bpmn-2-which-models-are-available-bpmn>

Motivation and Drivers for Big Data Adoption

Information and Communications Technology

- ICT developments that have accelerated the pace of Big Data adoption include:
 - Data analytics and data science
 - Digitization
 - Affordable technology and commodity hardware
 - Social media
 - Hyper-connected communities and devices
 - Cloud computing

Information and Communications Technology

- Data analytics and data science
 - Enterprises are collecting, procuring, storing, curating, and processing increasing quantities of data to find new insights that can drive more efficient and effective operations.
 - Companies are looking for new ways to gain a competitive edge by using Data Science tools and technologies to extract meaningful information and insights.
- Digitization
 - Businesses have replaced physical mediums as the de facto communications and delivery mechanisms.
 - This allows for the opportunity to collect secondary data for analysis
- Affordable technology and commodity hardware
 - Technology capable of storing and processing large quantities of diverse data has become increasingly affordable.
 - Big data solutions also often leverage open-source software that executes on commodity hardware, further reducing costs.

Information and Communications Technology

GW

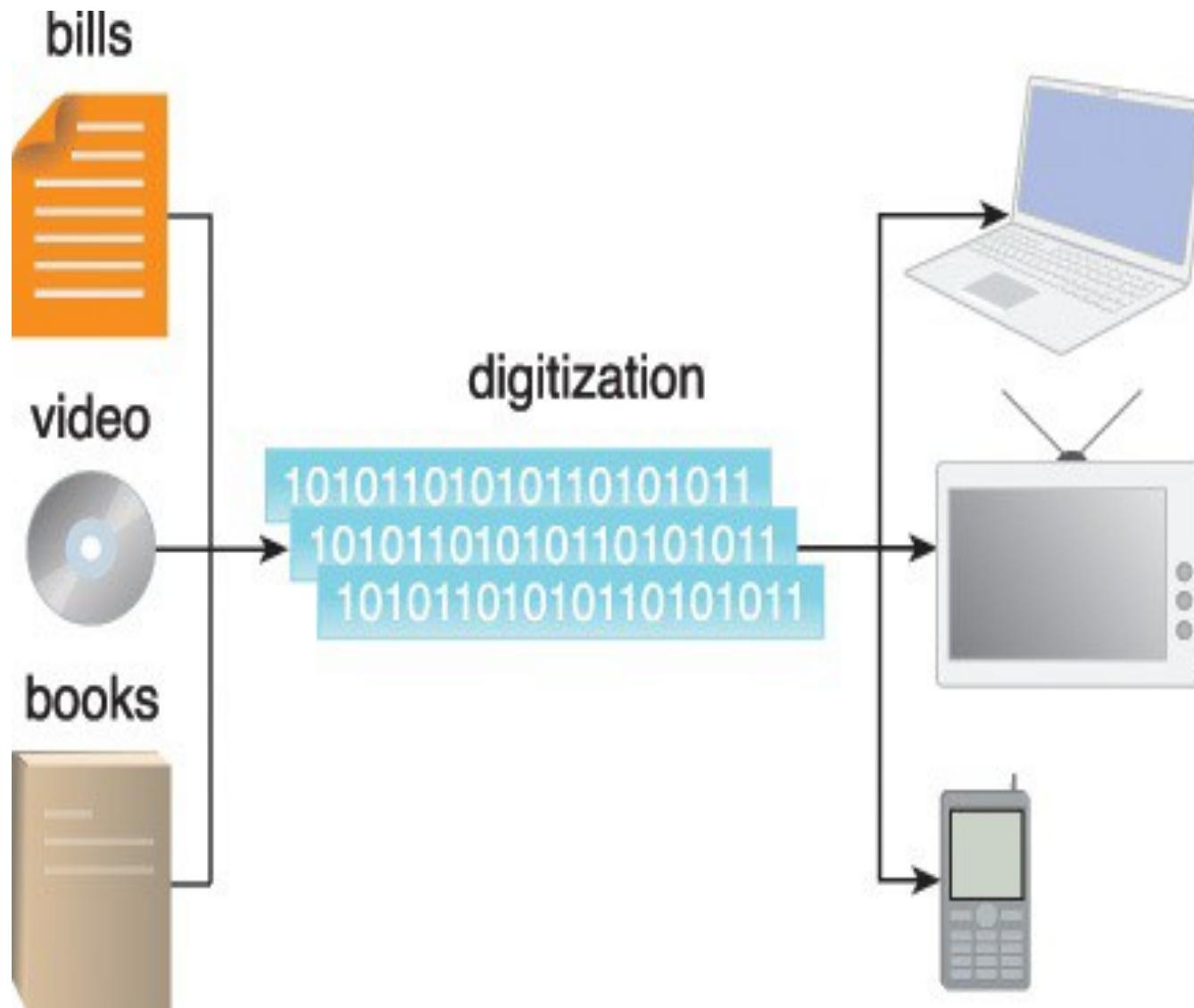


Figure 2.4 Examples of digitization including online banking, on-demand television and streaming video.

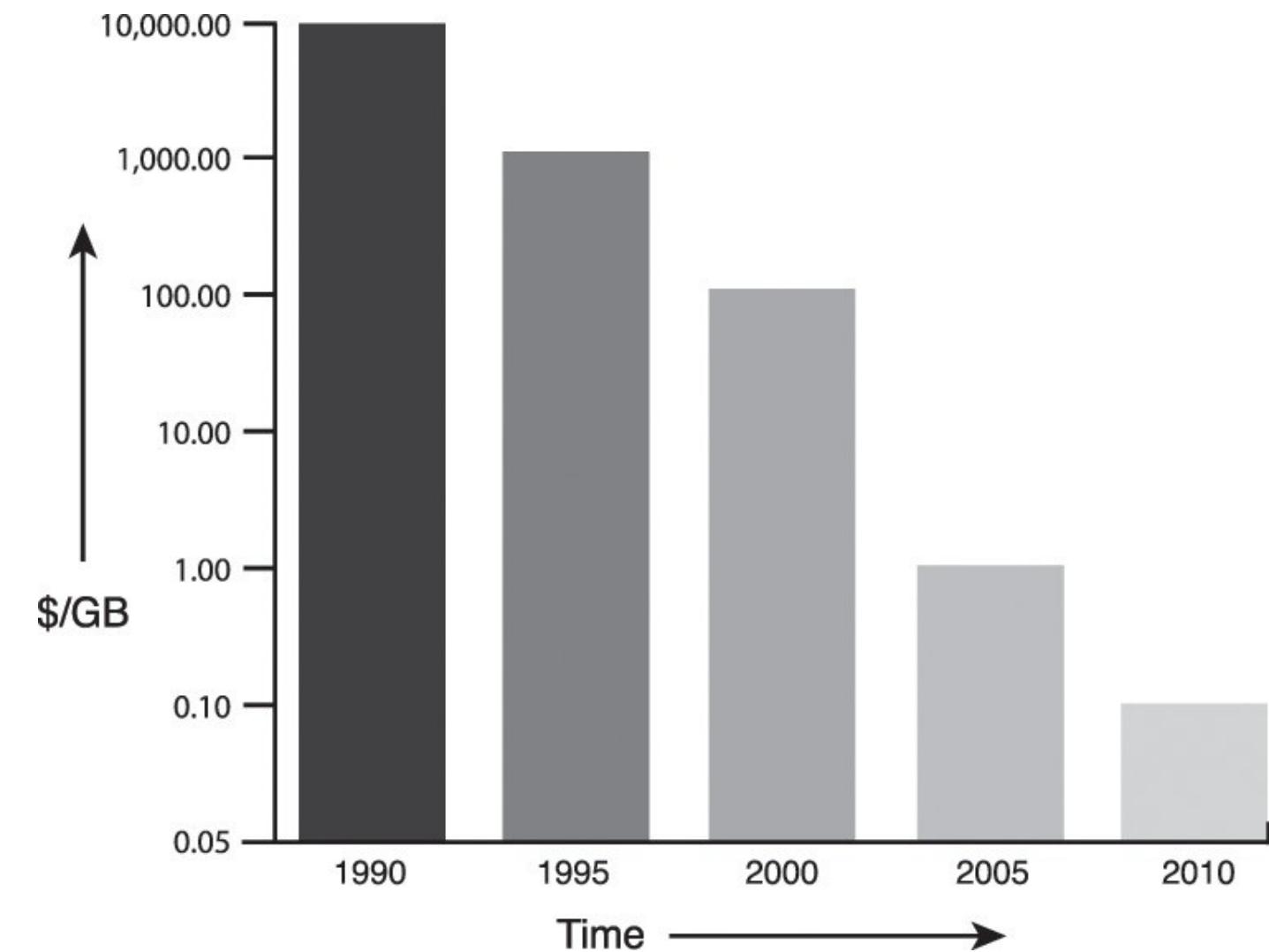


Figure 2.5 Data storage prices have dropped dramatically from more than \$10,000 to less than \$0.10 per GB over the decades.

Information and Communications Technology



- Social media
 - The emergence of social media has empowered customers to provide feedback in near real-time via open and public mediums.
 - The shift has forced businesses to utilize CRM's to analyze customer feedback on their service and product offerings
- Hyper-connected communities and devices
 - The broadening coverage of the Internet and the proliferation of cellular, Wi-Fi and sensor networks has enabled more people and their devices to be continuously active in virtual communities.
 - IoT – a vast collection of smart Internet-connected devices
 - IoE – combines the services provided by smart connected devices of the IoT into meaningful business processes that possess the ability to provide unique differentiating value propositions.
- Cloud computing
 - Cloud computing advancements have led to the creating of environments that are capable of providing highly scalable, parallel, distributed on-demand IT resources.

Information and Communications Technology

GW

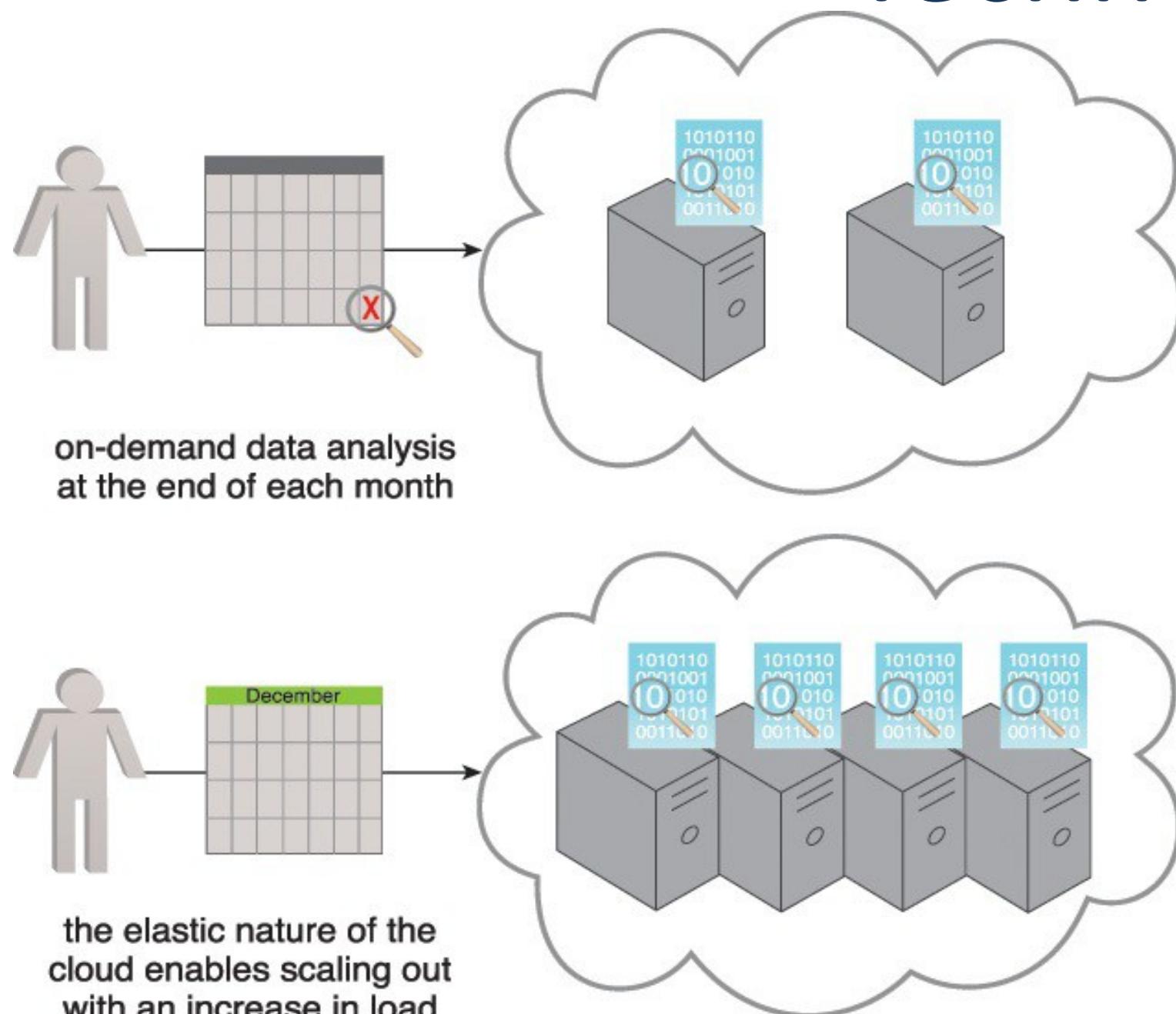


Figure 2.7 The cloud can be used to complete on-demand data analysis at the end of each month or enable the scaling out of systems with an increase in load.

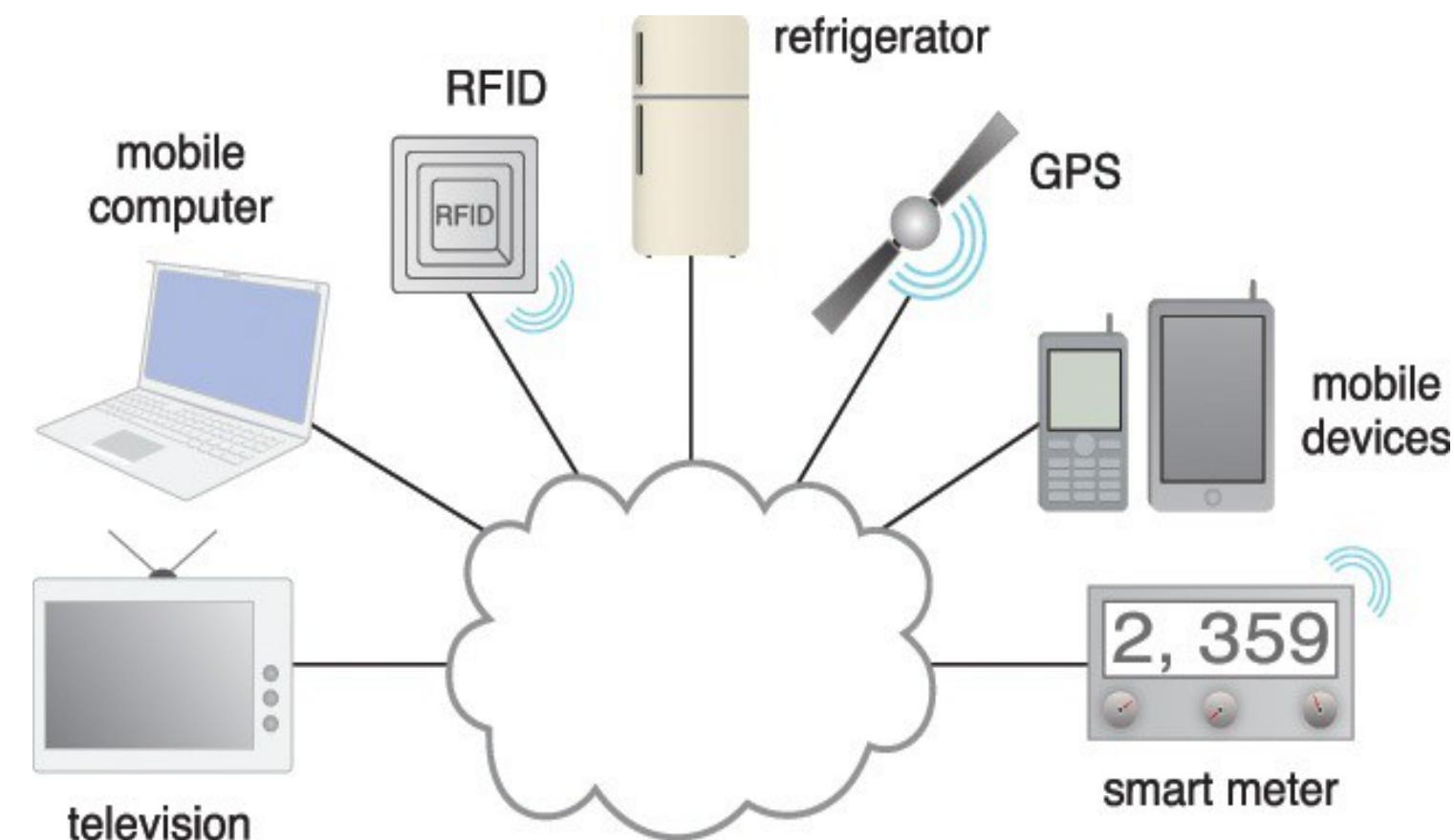
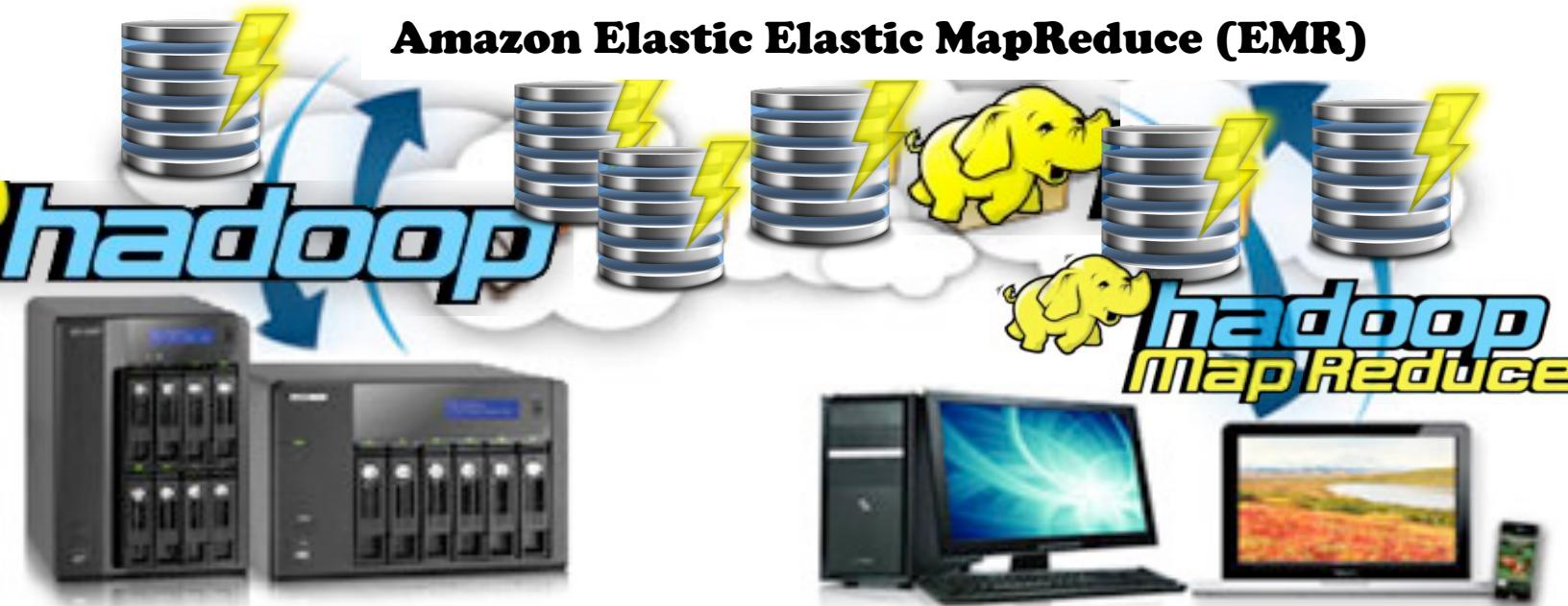


Figure 2.6 Hyper-connected communities and devices include television, mobile computing, RFIDs, refrigerators, GPS devices, mobile devices and smart meters.



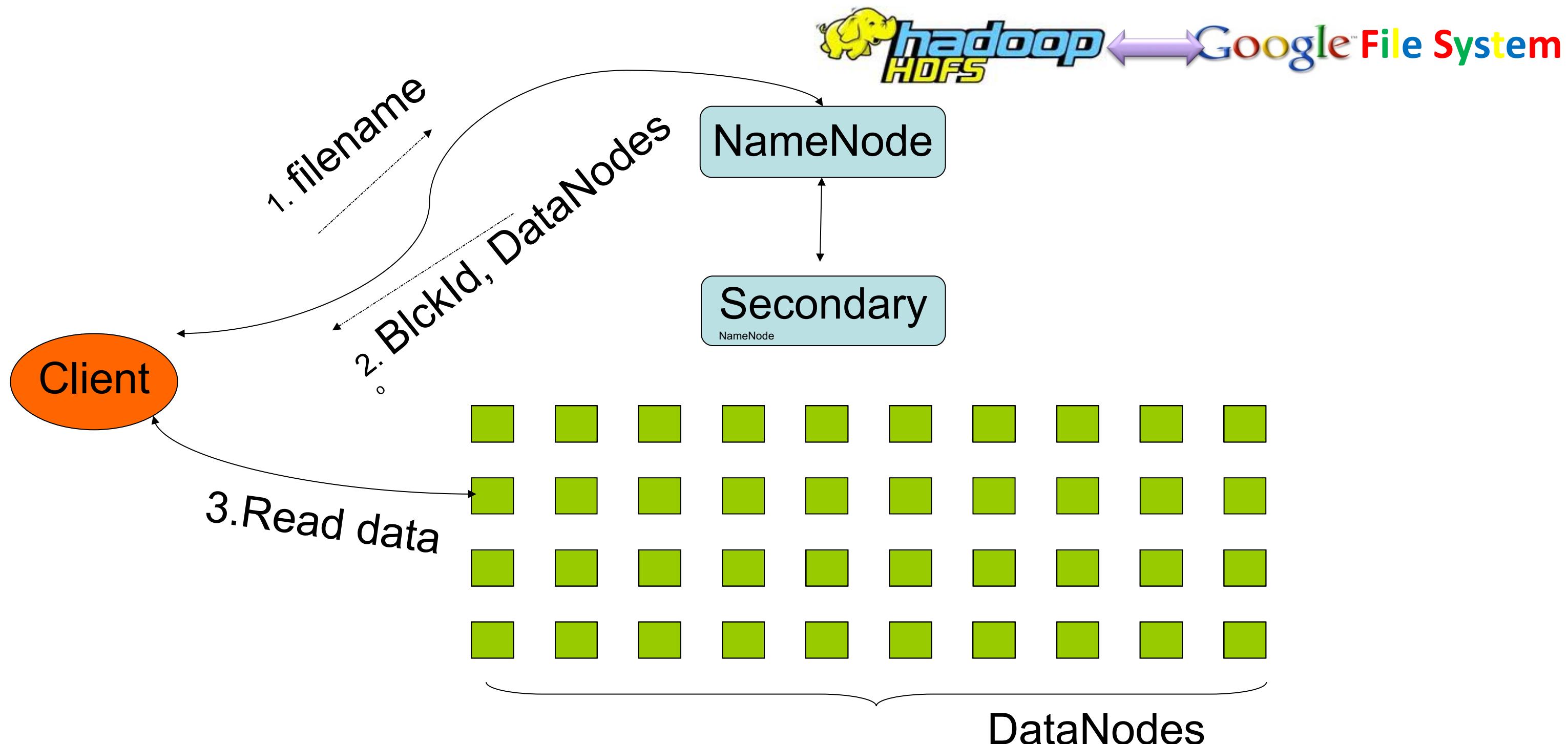
APACHE
HBASE



Cloud Infrastructures



HDFS: The System Structure

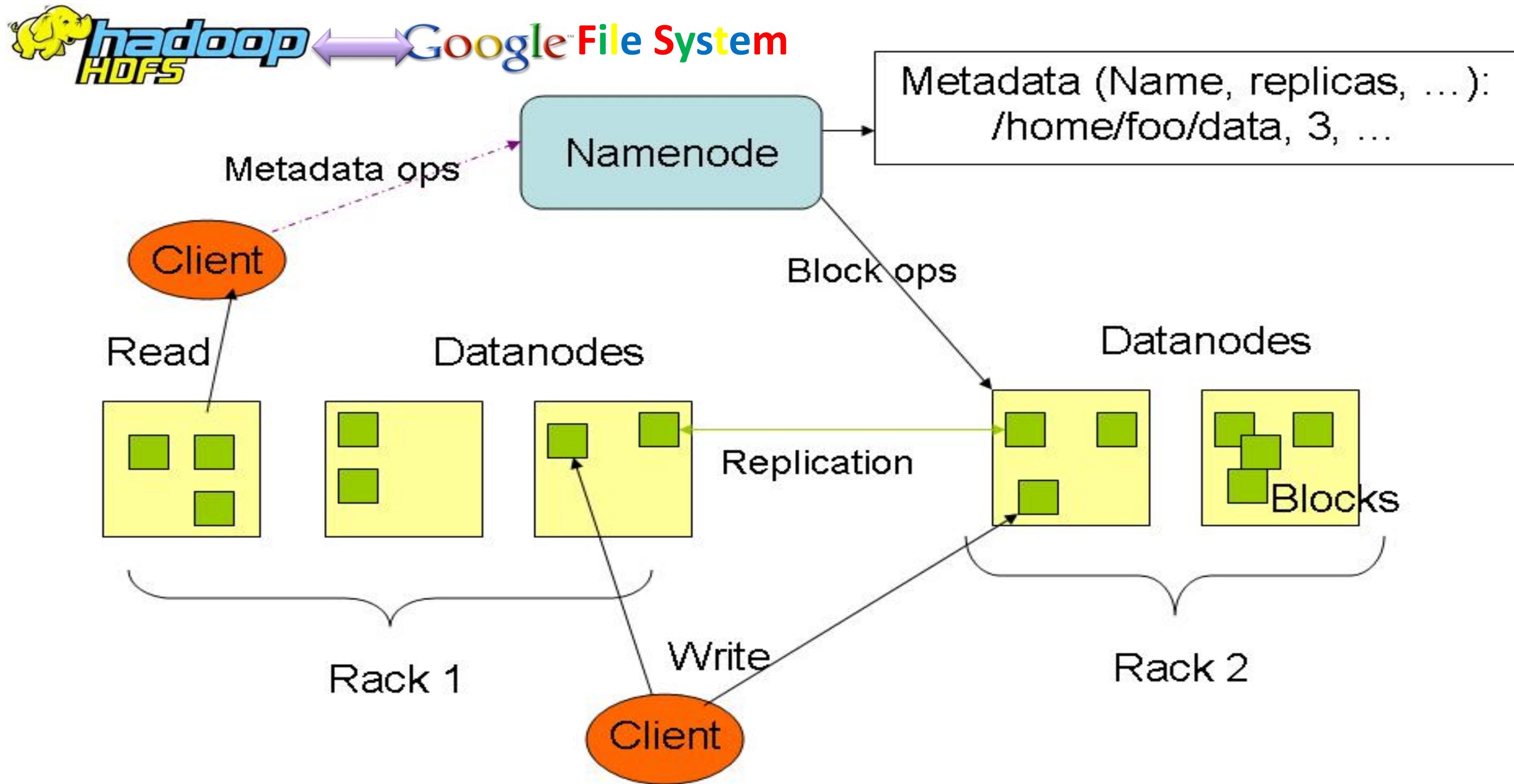


NameNode : Maps a file to a file-id and list of MapNodes

DataNode : Maps a block-id to a physical location on disk

Source: hadoop.apache.org

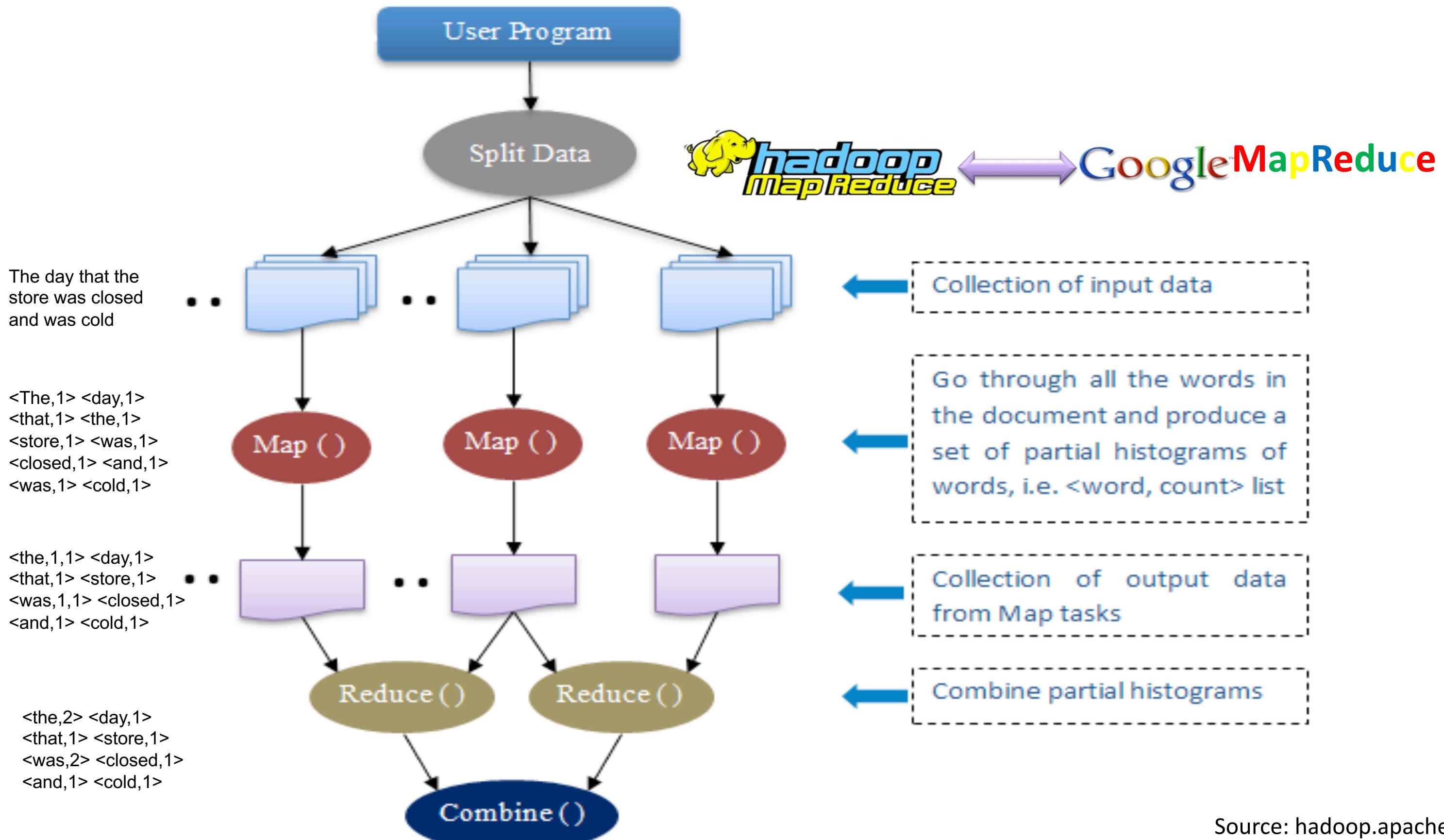
HDFS Architecture



Source: hadoop.apache.org

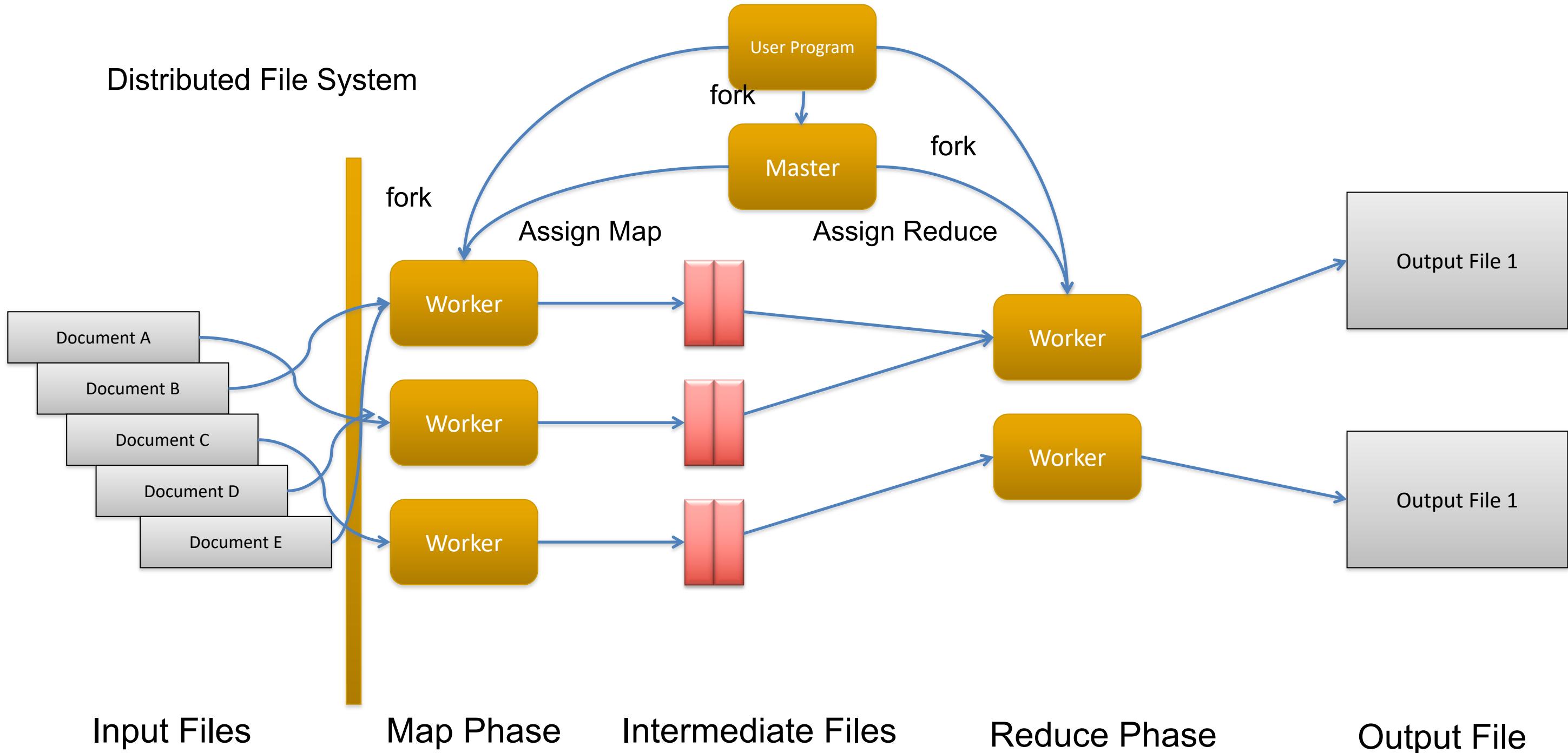
MapReduce WordCount

GW



Source: hadoop.apache.org

MapReduce Execution Overview



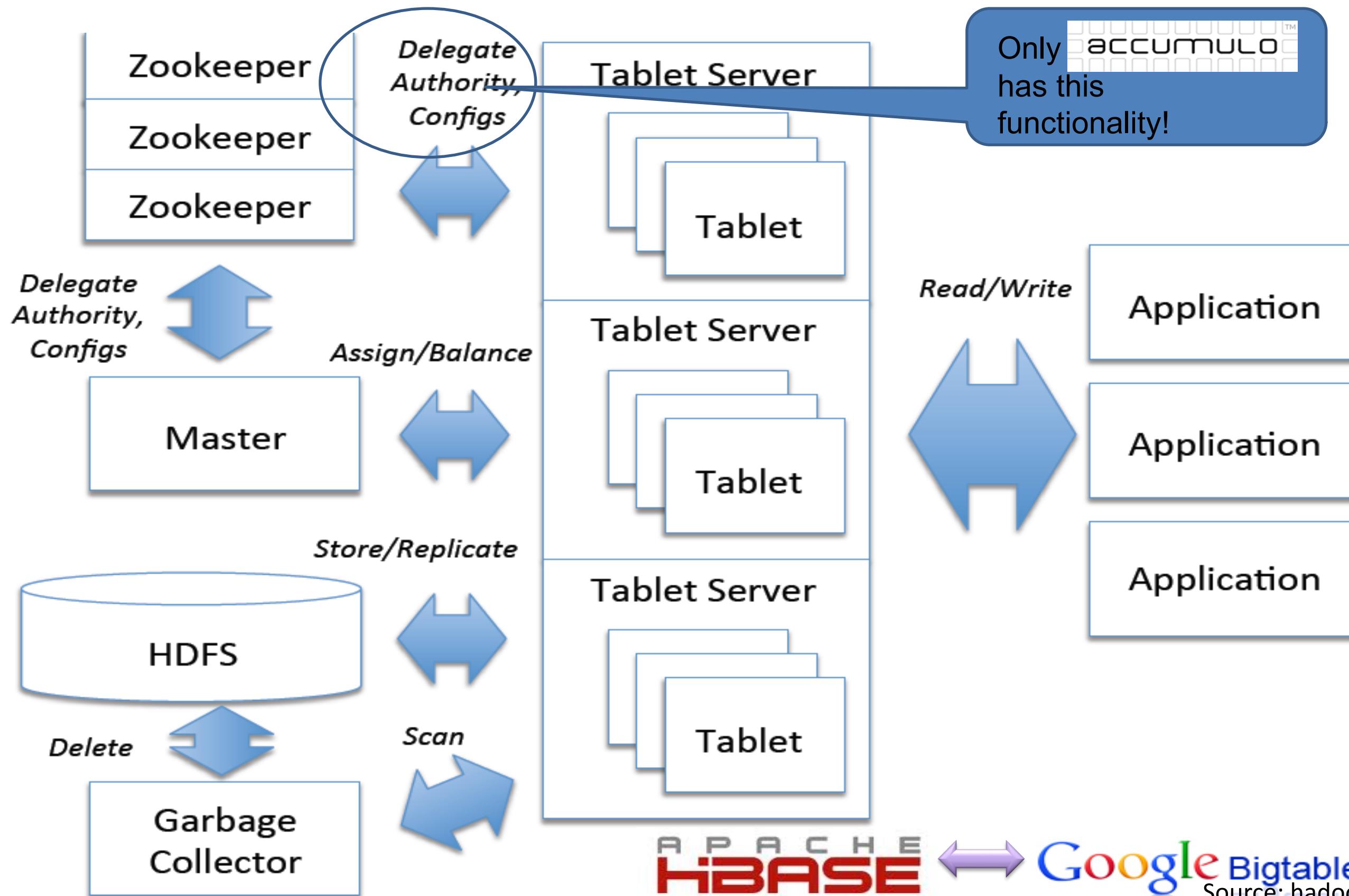
**hadoop
MapReduce**

Google™ MapReduce

Source: hadoop.apache.org

Hbase/Accumulo Architecture

GW



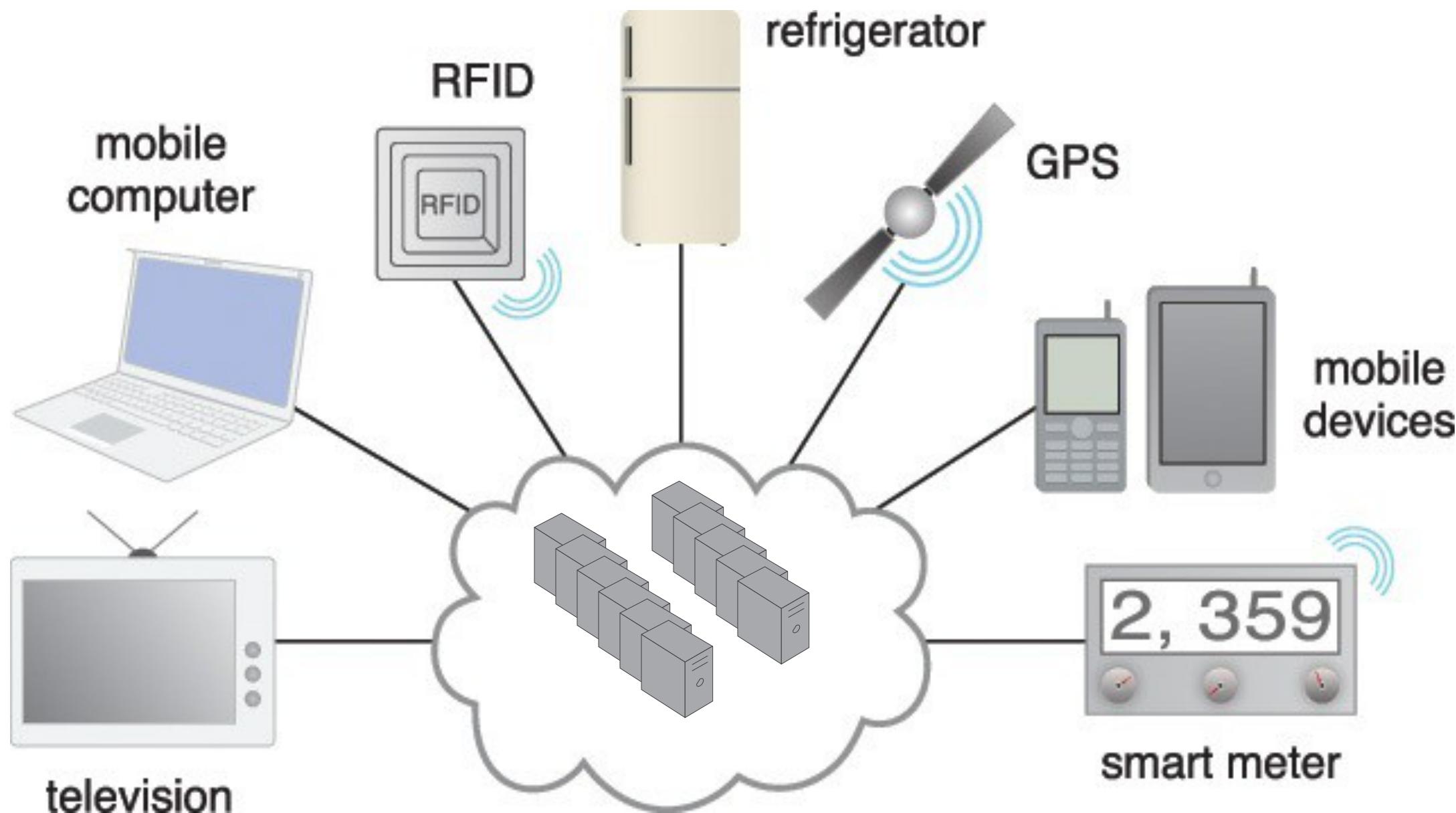
APACHE
HBASE

↔ Google Bigtable

Source: hadoop.apache.org

Internet of Everything

- Hyper-connected communities and devices running on affordable technology and commodity hardware stream digitized data that is **subject to analytic processes hosted in elastic cloud computing environments**.



Internet of Everything

- The results of the analysis can provide insights as to how much value is generated by the current process and whether or not the process should proactively seek opportunities to further optimize itself.
- (1) Business processes in combination with analysis of (2) streaming data and customer context, being able to (3) adapt the execution of these process to align with the customer's goals will become a key corporate differentiator in the future

Case Study: Ensure to Insure

Introduction to the Case Study

ETI new transformation and innovation corporate priorities:

- Considering transformation, business process management disciplines will be adopted to document, analyze, and improve the processing of claims
- Risk assessment and fraud detection will be enhanced with the application of innovative Big Data technologies that will produce analytic results that can drive data-drive decision-making.
- Redefining CSFs and KPIs has helped ETI link and align the strategic, tactical, and operating levels of the business
- Organization role responsible for innovation management.

Do you think the new transformation and innovations priorities will help? Justify your answer.

What's missing?