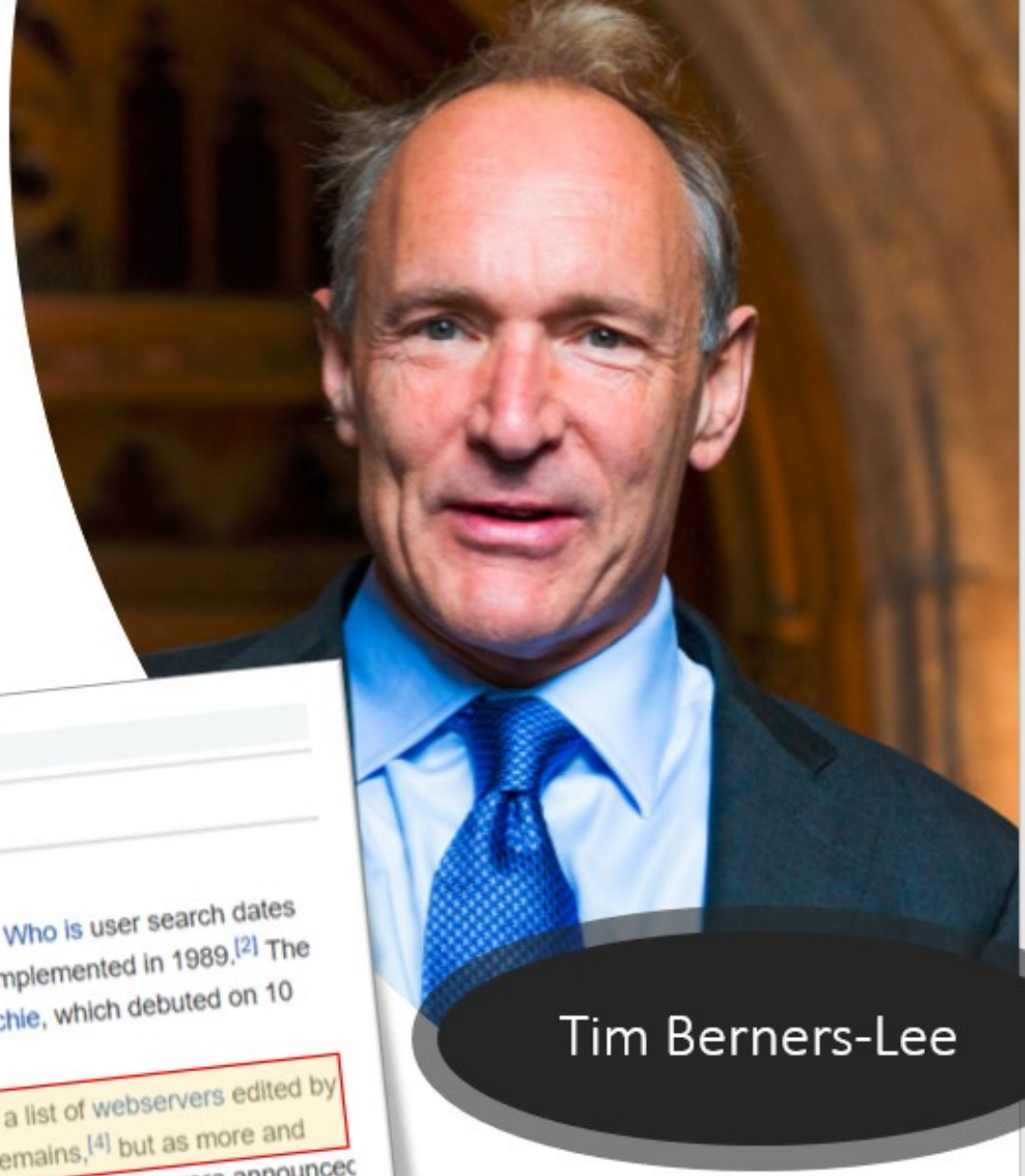
Kafka Streams
Real-time Stream Processing
by
LEARNING JOURNAL

LEARNING JOURNAL

How it started?

How it started?



Tim Berners-Lee

https://en.wikipedia.org/wiki/Web_search_engine

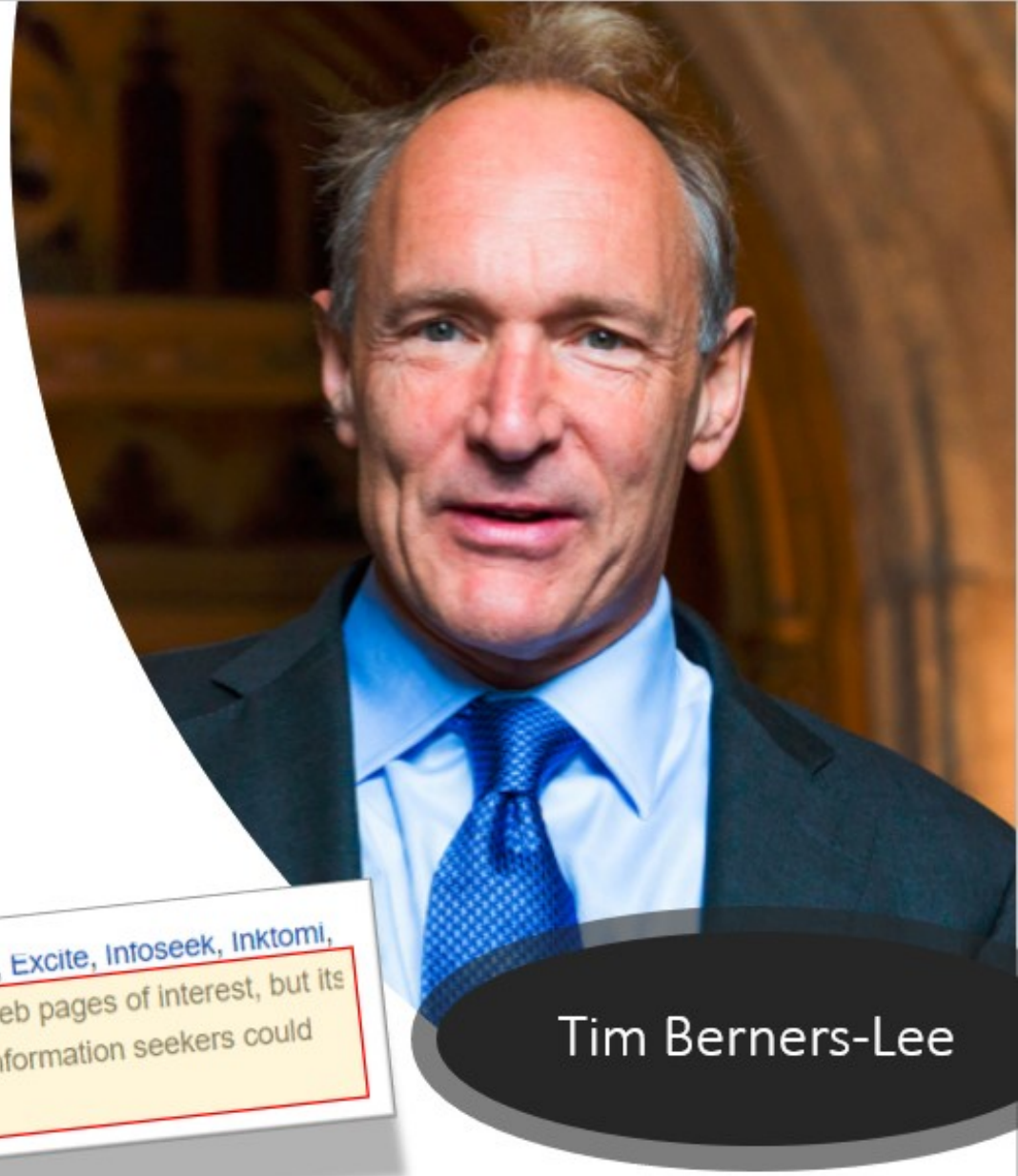
History [edit]

Further information: Timeline of web search engines

Internet search engines themselves predate the debut of the Web in December 1990. The Who is user search dates back to 1982[1] and the Knowbot Information Service multi-network user search was first implemented in 1989.[2] The first well documented search engine that searched content files, namely FTP files was Archie, which debuted on 10 September 1990.[3]

Prior to September 1993, the World Wide Web was entirely indexed by hand. There was a list of webservers edited by Tim Berners-Lee and hosted on the CERN webserver. One snapshot of the list in 1992 remains,[4] but as more and more web servers went online the central list could no longer keep up. On the NCSA site, new servers were announced under the title "What's New![5]

How it started?



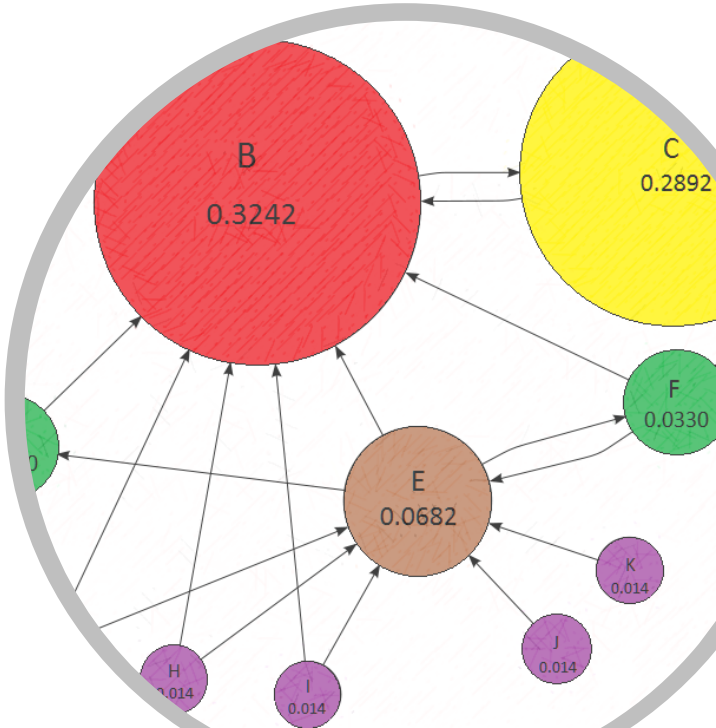
Tim Berners-Lee

Soon after, many search engines appeared and vied for popularity. These included Magellan, Excite, Infoseek, Inktomi, Northern Light, and AltaVista. Yahoo! was among the most popular ways for people to find web pages of interest, but its search function operated on its web directory, rather than its full-text copies of web pages. Information seekers could also browse the directory instead of doing a keyword-based search.

How it started?

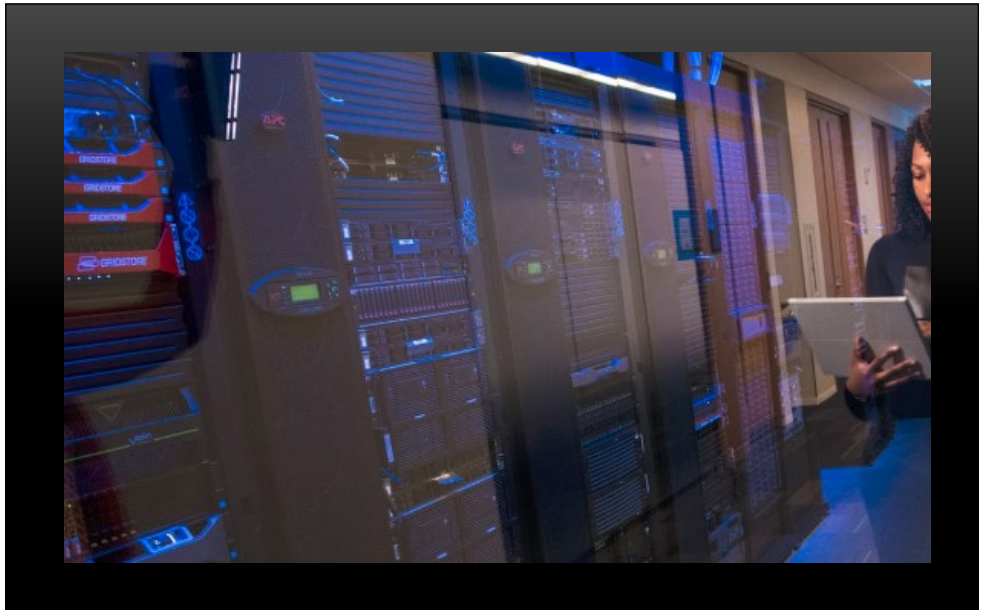
PageRank Algorithm

1. Domain Authority Links
2. Keywords & Metadata



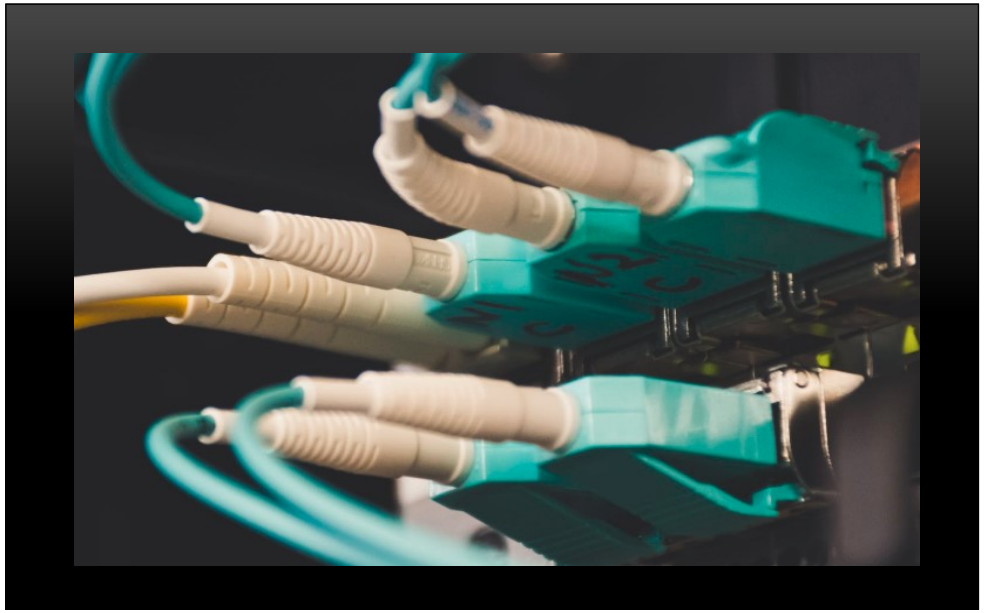
Bigdata Challenges

- Data Collection
- Data Storage
- Data Processing
- Data Access



LEARNING JOURNAL

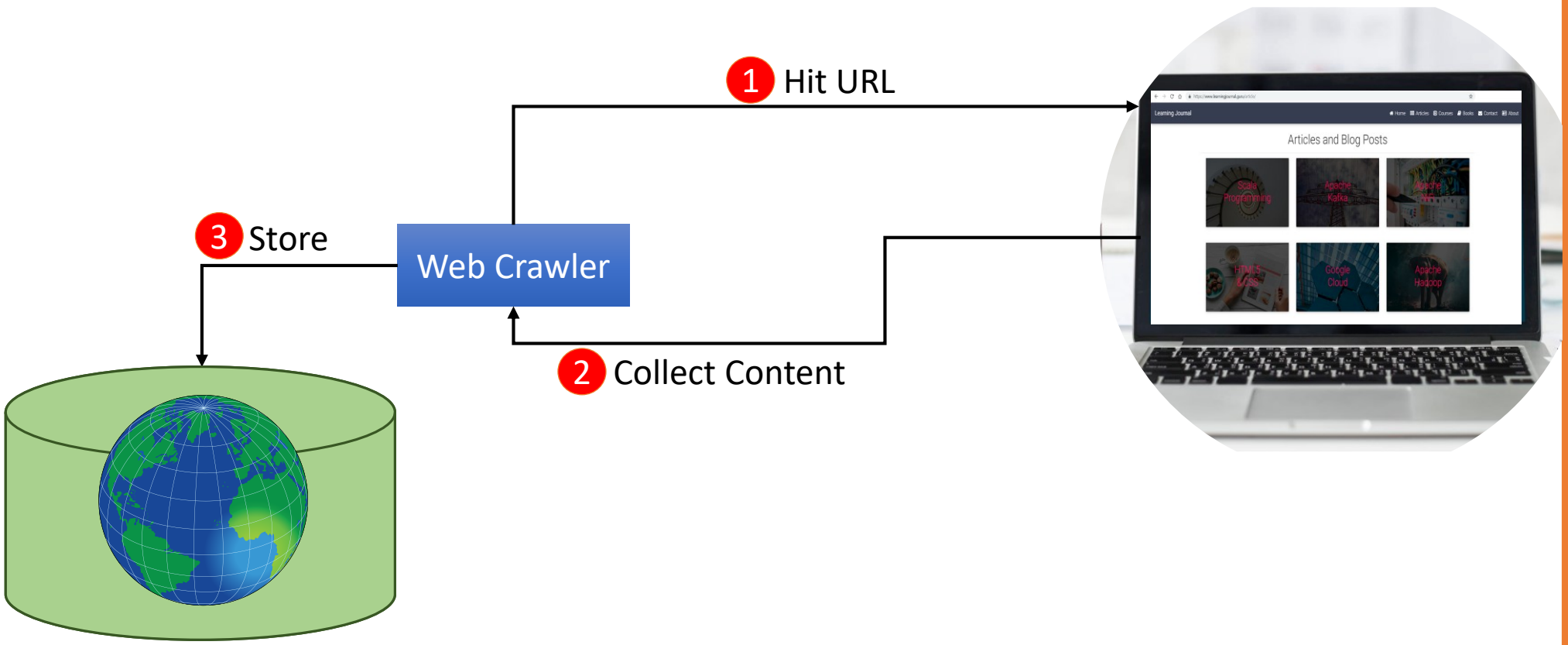
Bigdata Challenges



- Volume
- Variety
- Velocity

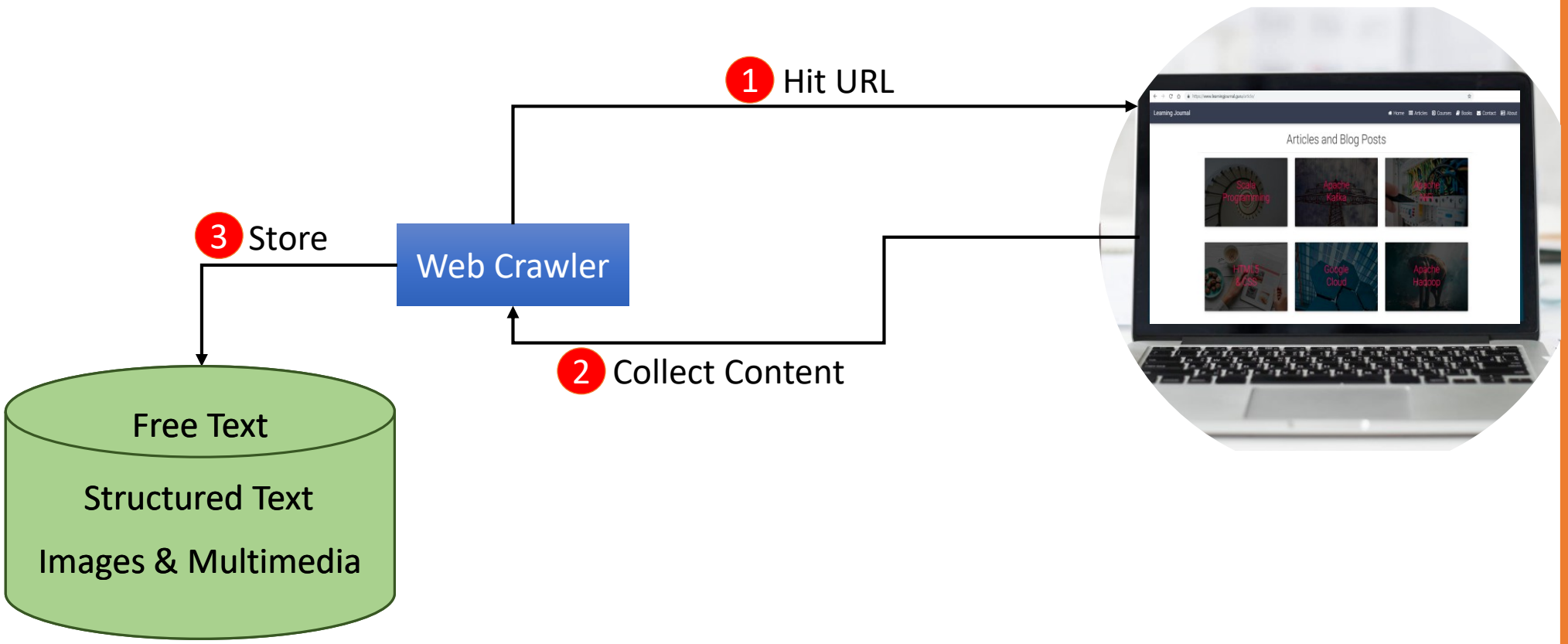
LEARNING JOURNAL

Volume



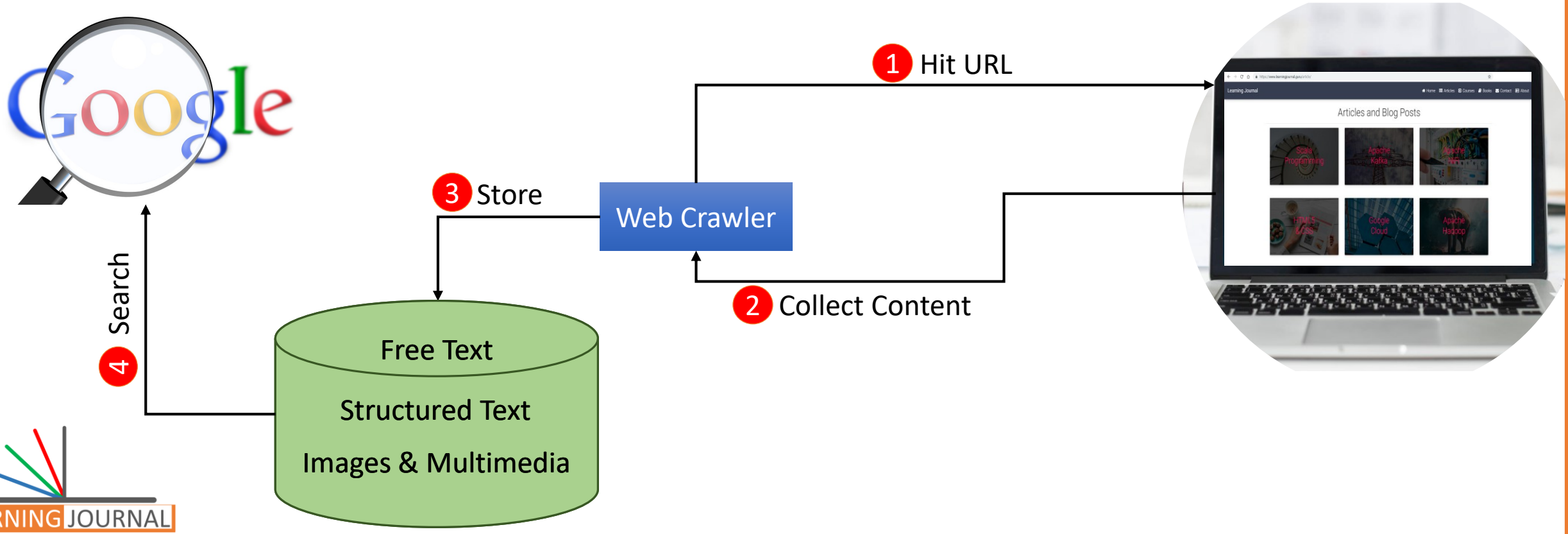
LEARNING JOURNAL

Variety



LEARNING JOURNAL

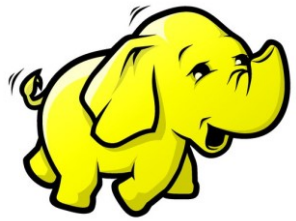
Velocity


Google
1 Hit URL
Articles and Blog Posts
3 Store
Web Crawler
2 Collect Content
4 Search
Free Text
Structured Text
Images & Multimedia
LEARNING JOURNAL

Bigdata Challenges

- Data Storage
- Data Processing
- Data Access

- Google File System
- Map Reduce
- Google Big Table

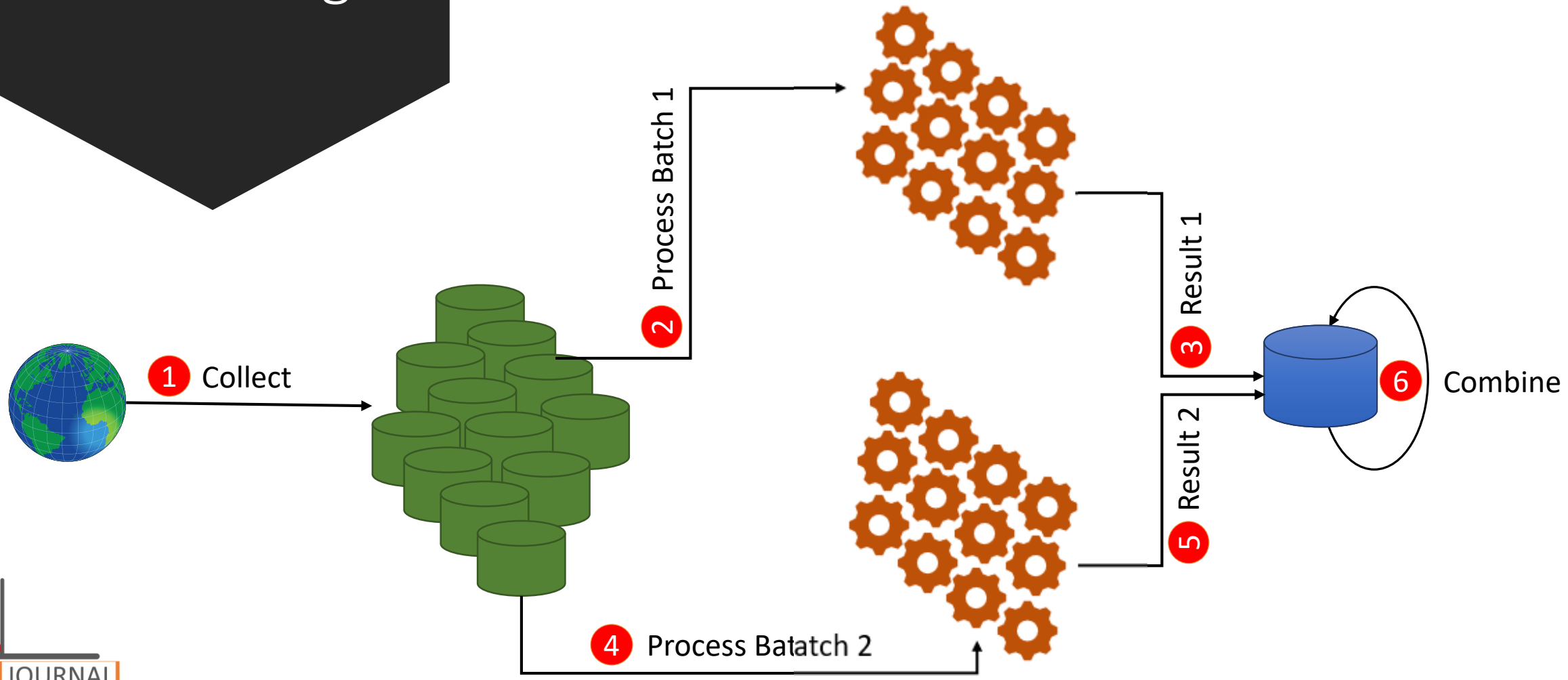


Hadoop

LEARNING JOURNAL

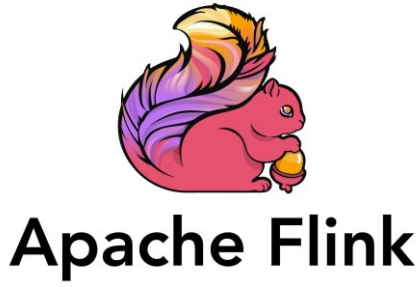
Batch Processing

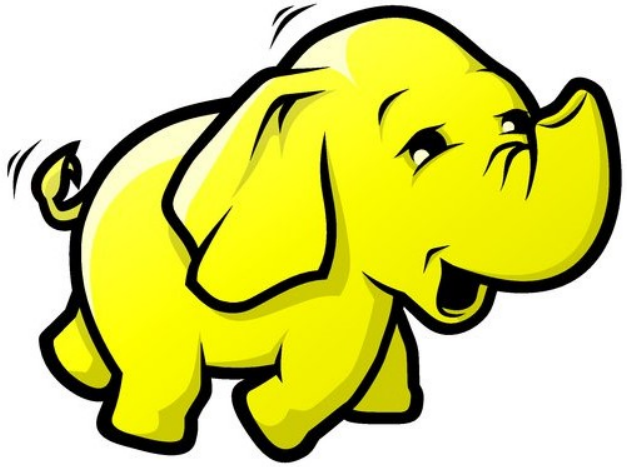
Minutes to Hours



LEARNING JOURNAL

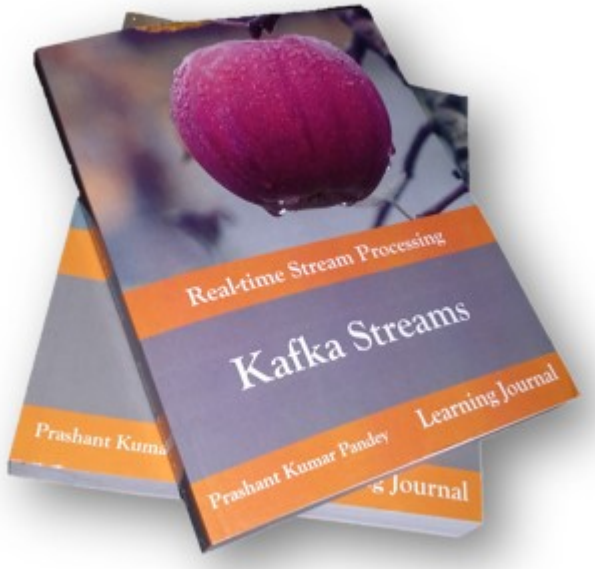
Real-time Processing





Real-time Data Processing

Video Based Course



eBook & Paperback

Available on Amazon, Google Books, iTunes, Nook, Kobo, Flipkart

LEARNING JOURNAL