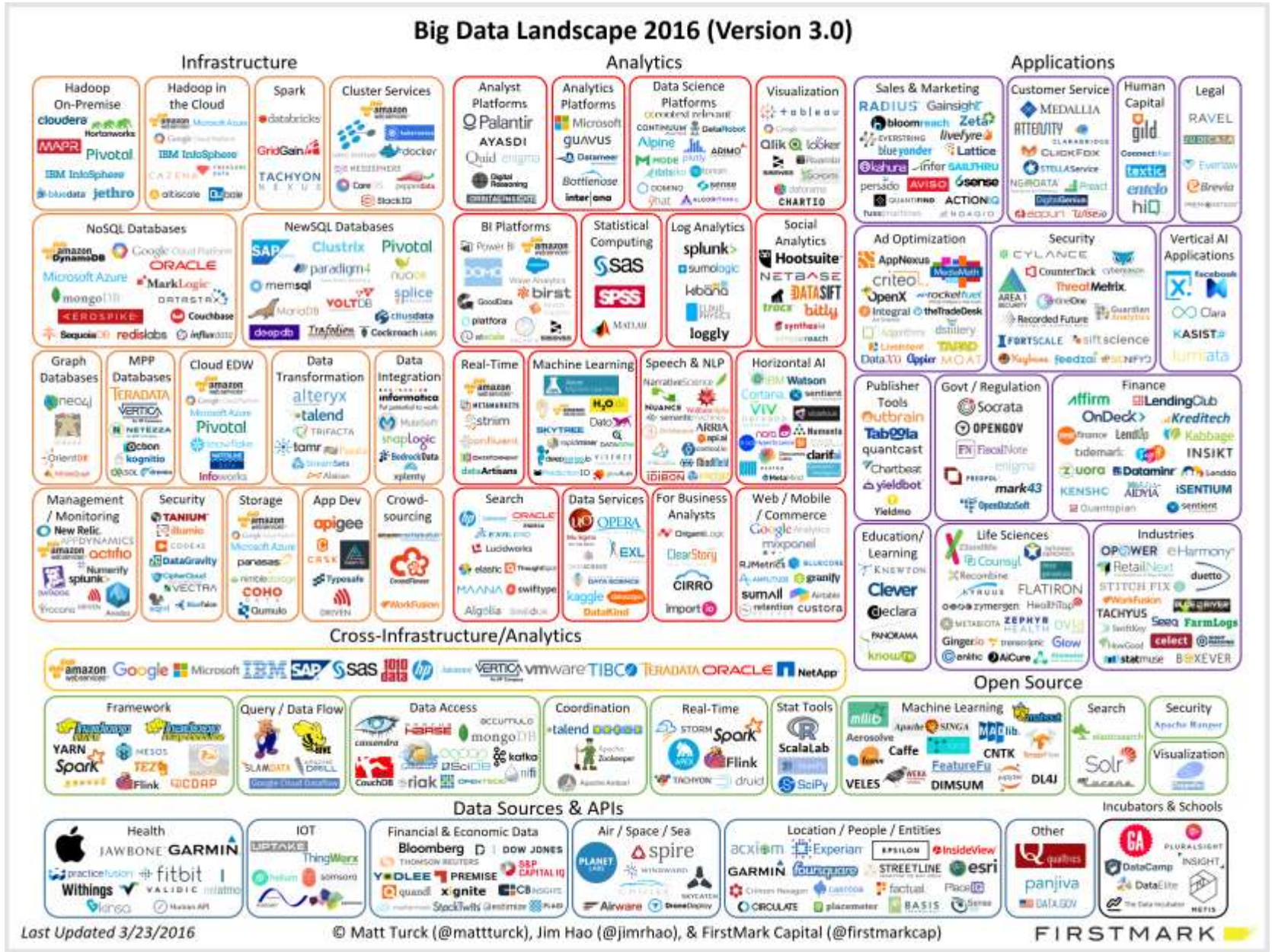
Big Data & The Cloud

- **Big Data**
  - Emergent Data sources generate a huge amount of information
  - New tools have the capability to handle high-volume data sources
- **“The Cloud”**
  - Speaks to WHERE data operations happen
  - Companies ‘rent’ hardware capacity, software & services
  - Operations happen in the third party’s data center(s)

Big Data

- < Use a hi-resolution version of the diagram on the next page>
- http://mattturck.com/wp-content/uploads/2016/03/Big-Data-Landscape-2016-v18-FINAL.png

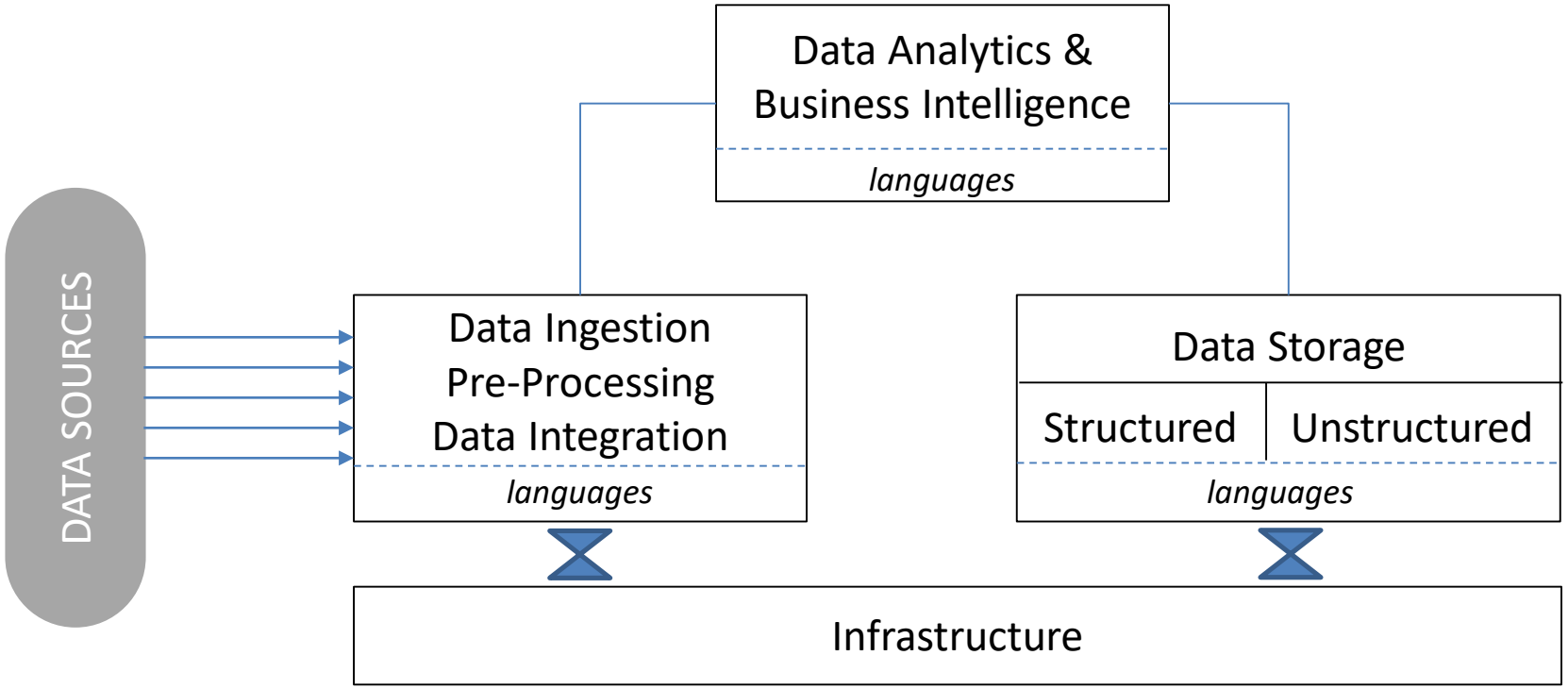
Big Data & The Cloud



Be Boulder.

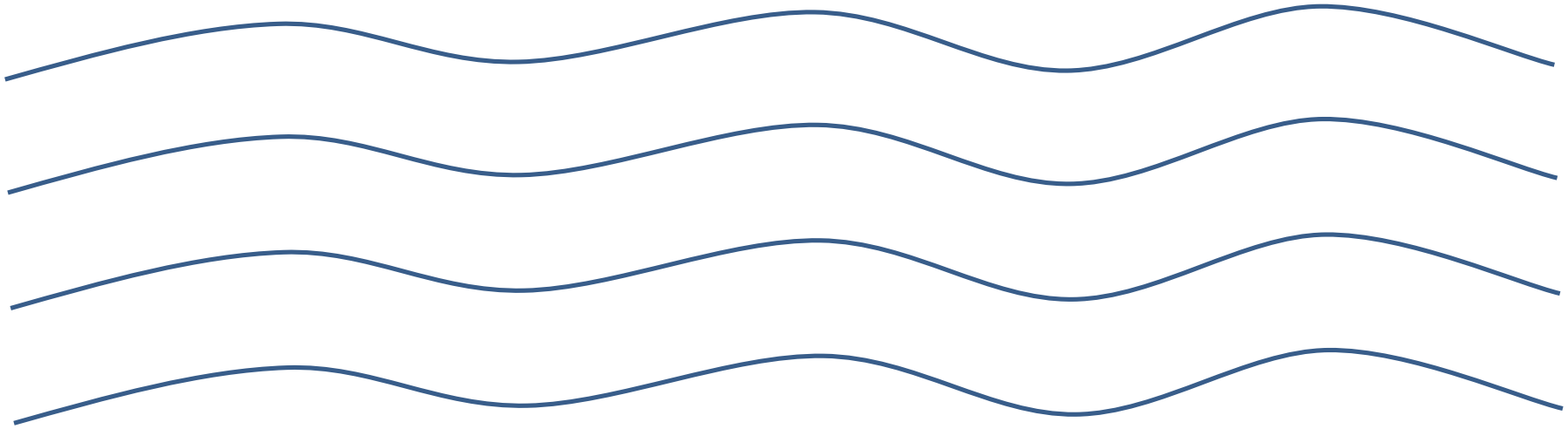
University of Colorado Boulder

Big Data



Big Data - Streaming Data

- < Streaming – ideally we have an image or animation that shows the fish analogy >



Big Data - Streaming Data

- < clip the “Spark” “Data Transformation” “Data Integration” and “Real-Time” boxes from the landscape image and line up left to right”
- < add in the following language icons >

Be Boulder.
University of Colorado Boulder

Big Data - Storage

- < clip the Hadoop On-Premise, Hadoop in the Cloud, NoSQL Databases, NewSQL databases, Graph databases, MPP Databases, and Cloud EDW boxes from the landscape arrange >

Big Data - Analytics

- < show analytics section from landscape>

Be Boulder.

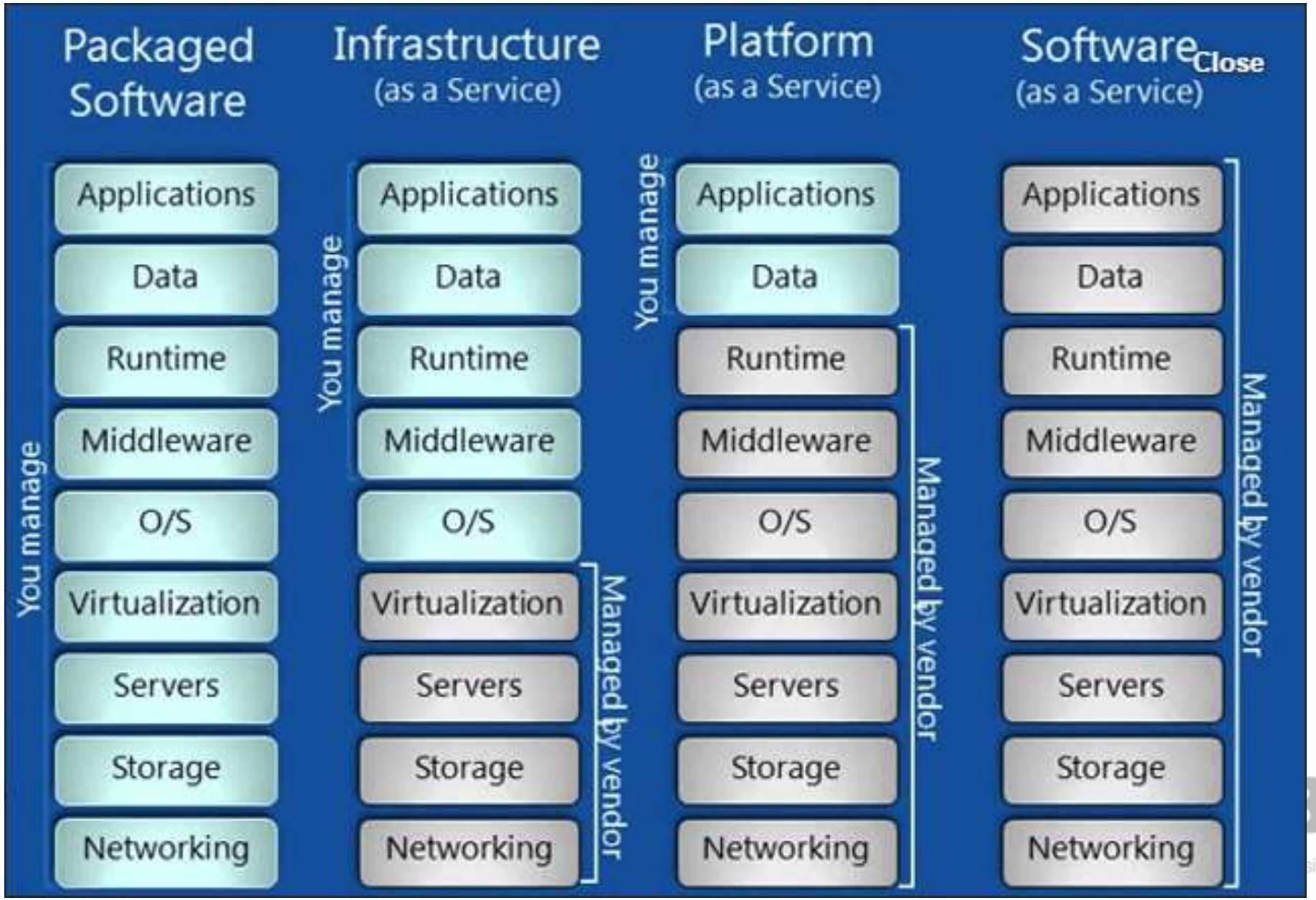
University of Colorado Boulder

Big Data - Infrastructure

The Cloud

- **Software as a Service (Saas)**
  - Software hosted on machines provided by third party
  - Applications accessed remotely via client and/or the Web
  - Targeted at application end users
- **Platform as a Service (PaaS)**
  - Development environment hosted by third party
  - Targeted at developers
- **Infrastructure as a Service (IaaS)**
  - Raw building blocks of data environment provided by third party
  - Processing capacity, storage, connectivity, security, etc.

The Cloud



The Cloud

- **Why the Cloud is Attractive**
  - Inexpensive & easy to set up
  - Can scale quickly and easily
  - Less distraction & overhead
- **Why not go to the Cloud?**
  - Control & security
  - Inertia & existing resource investment
  - Highly unique needs
  - Location of source data
  - Scale

Recap

- **Big Data**
  - Ingest & process data
  - Store structured & unstructured data
  - Enable analytics on data
  - Provide infrastructure
- **Cloud Computing**
  - SaaS, PaaS, IaaS
  - Pro/cons
  - How it can work with Big Data