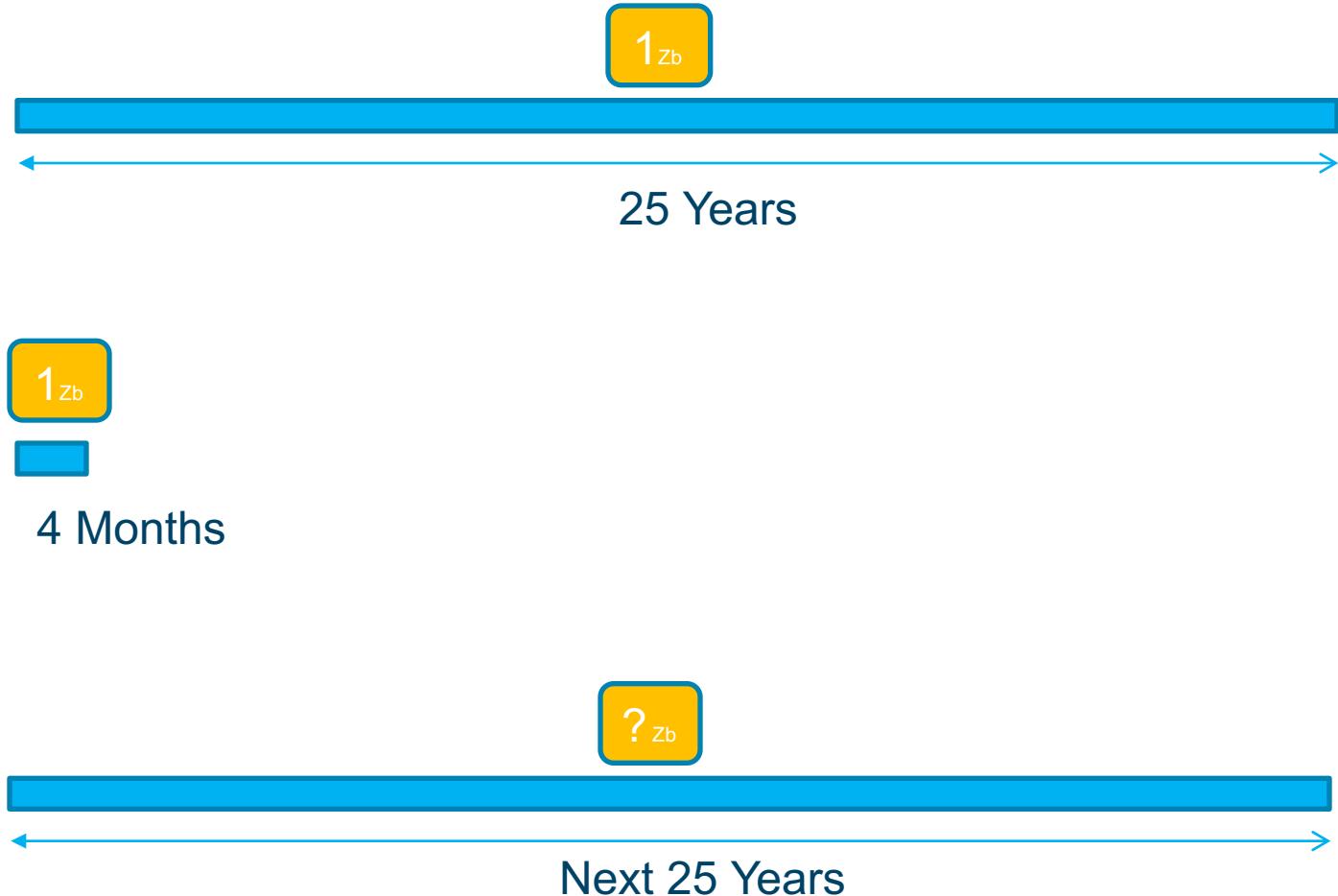


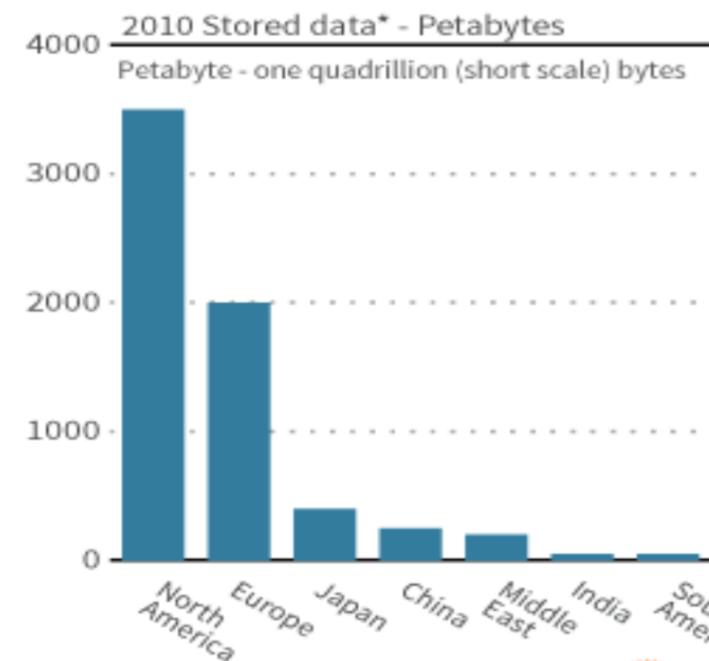
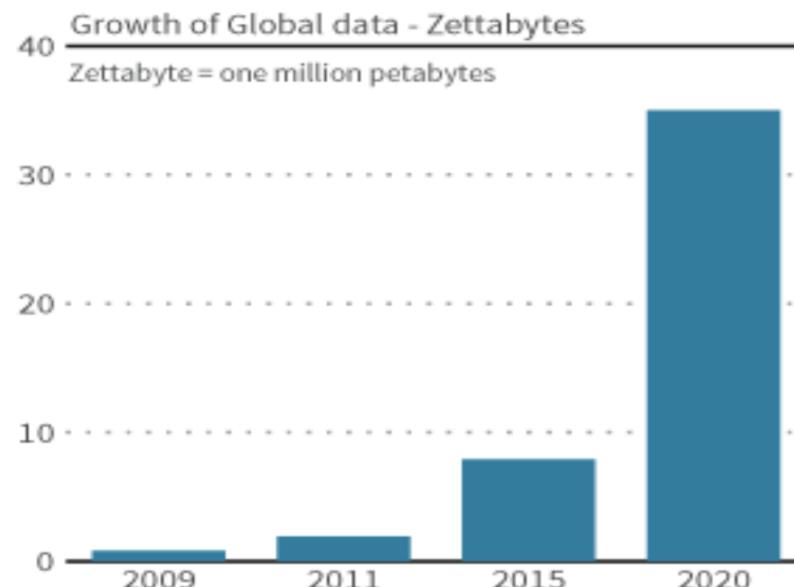
## Atlanta Data Science: Using R for distributed computing

- Goals:
  - Overview of distributed data analysis
  - Rstudio's tool for distributed R
  - Utilizing R as a scripting lang
  - Working with sparklyr



## Big data growth

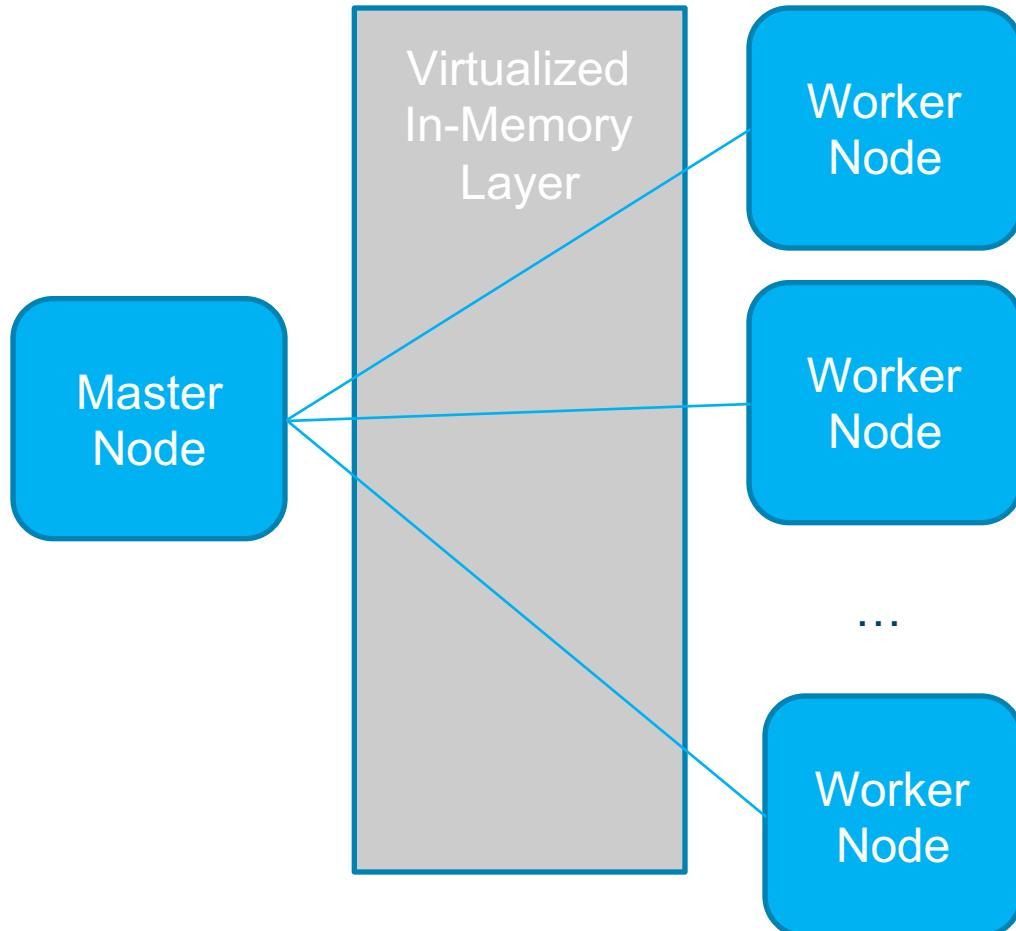
Big data market is estimated to grow 45% annually to reach \$25 billion by 2015



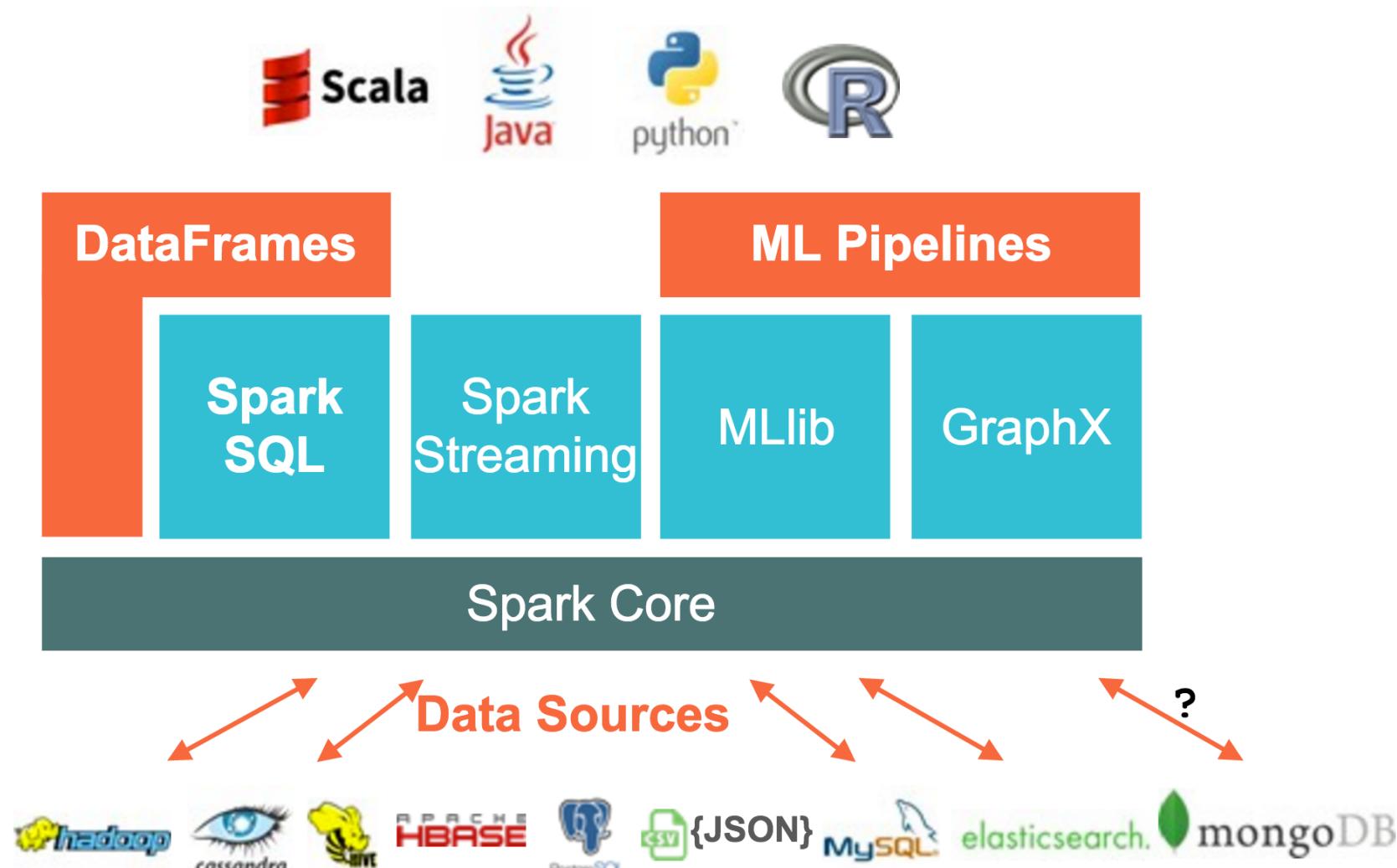
\*greater than

Sources: Nasscom -CRISIL GR&A analysis





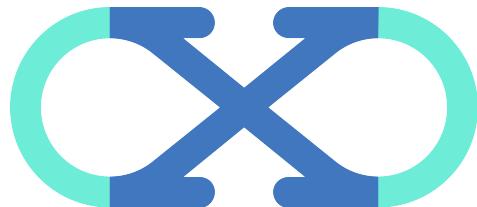
- Distribute the data across workers to increase time and efficiency for in-memory solution
- **10x-100x** efficiency increases from other methods that can handle large data



IBM is committed to training over 1 million data scientist over the next few years

## IBM DATA SCIENCE EXPERIENCE

ALL YOUR TOOLS IN ONE PLACE



[datascience.ibm.com](http://datascience.ibm.com)

The screenshot shows the homepage of Big Data University. At the top, there's a navigation bar with links for Courses, Events, Badges, Resources, Participate!, Blog, and About. It also shows a user profile for Deb Jenson and a Logout link. The main header reads "Analytics, Big Data, and Data Science Courses" with a sub-subtext "Your awesome career in Data Science and Data Engineering starts here." Below this is a "SIGN UP" button. To the right is a network graph visualization with nodes labeled Python, NoSQL, Data Science, Data Journalism, Analytics, Spark, Hadoop, Scala, R, Social Good, Big Data, and Machine Learning. A large section titled "RECOMMENDED FOR YOU" displays nine course cards arranged in a 3x3 grid. The courses are: Big Data Fundamentals (red), Data Science Fundamentals (orange), Introduction to OpenRefine (green), Introduction to R (orange), SQL Fundamentals (pink), Introduction to NoSQL (blue), Introduction to Data Analysis Using R (green), Hadoop Fundamentals I (light blue), and Spark Fundamentals I (orange). At the bottom left, there's a "RECOMMENDED EVENTS" section with a calendar icon showing "Jun 14" and a link to "IBM AT THE BIG DATA CONFERENCE & EXHIBITION". On the bottom right, there's a "WHAT'S NEW" section with a star icon and a link to "THIS WEEK IN DATA SCIENCE (MAY 24, 2016)".

**bigdatauniversity.com**

Overview dplyr and sparklyr

