

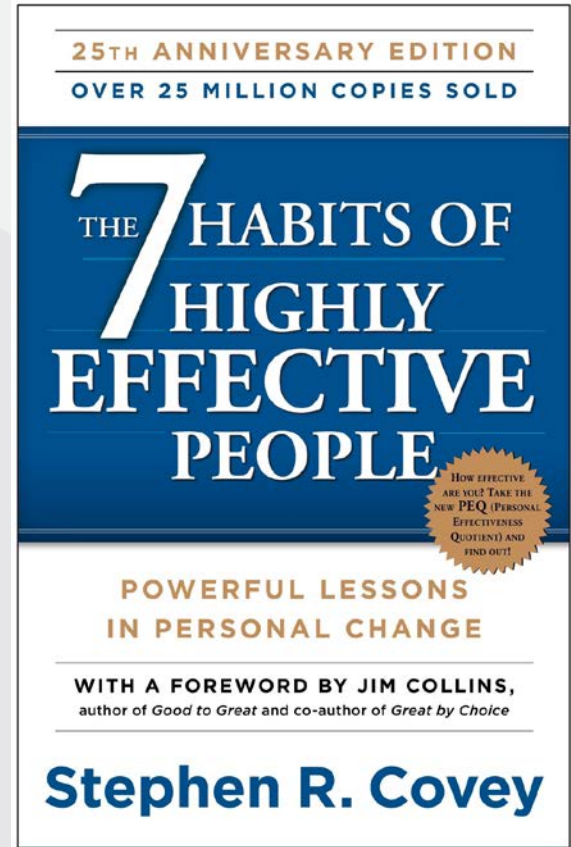
Big Data Analytics Platforms Architecture Requirements and Analysis Techniques

Spring 2019 NASPI
Sean Patrick Murphy

April 16th, 2019
San Diego, CA

Habit #2

“Start with
the end in
mind”



(1) Humans can learn from data

Fast (Human)
Access

Interactive
Visual
Interface

Translation
(Engineering
transformations)

Context
(Associated
metadata)

(2) Machines can learn from data

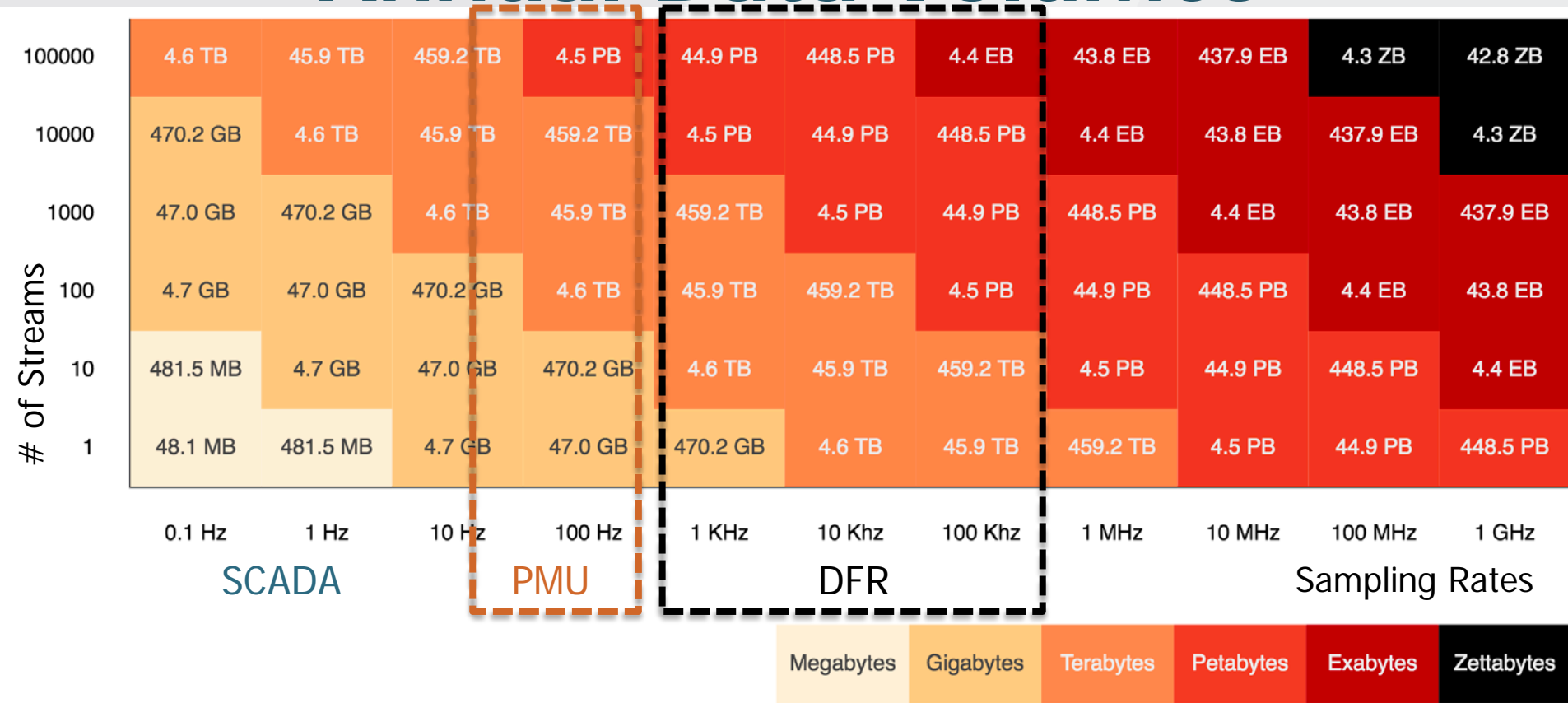
Fast (Machine)
Access

Unlimited
Compute
Resources

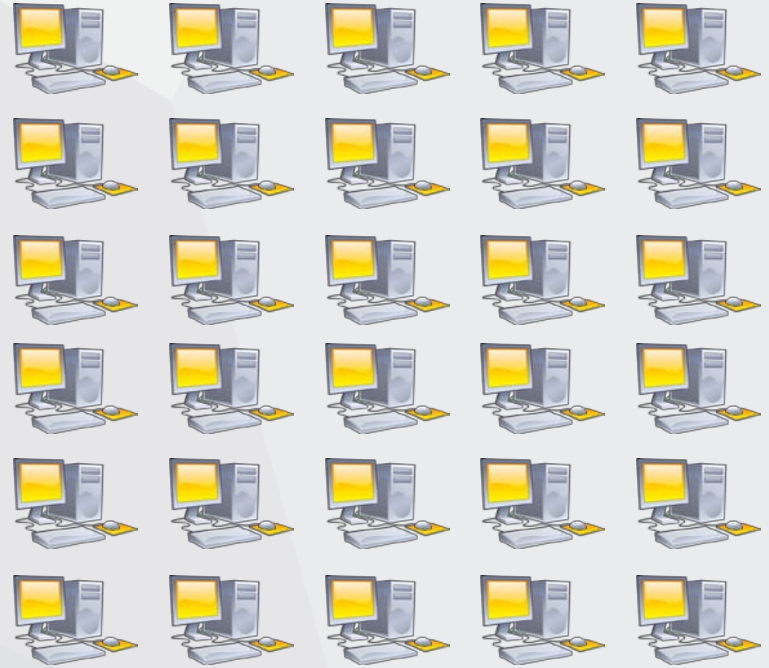
Integration w
Open Source
Libraries

Automated
Data Quality
Assessment

Annual Data Volumes



Solving Large Compute Problems



Machine Learning

Old Paradigm - Software Engineering

- Humans write the code
- Limited by ability to describe exactly what must be done without error

New Paradigm - Machine Learning

- Data teaches algorithms to perform function or task
- Limited by the amount of data and algorithms
- Algorithms need *****ALL***** available data
- Capable of tackling high dimensional problems



Deep Blue
beats Gary
Kasparov 1997



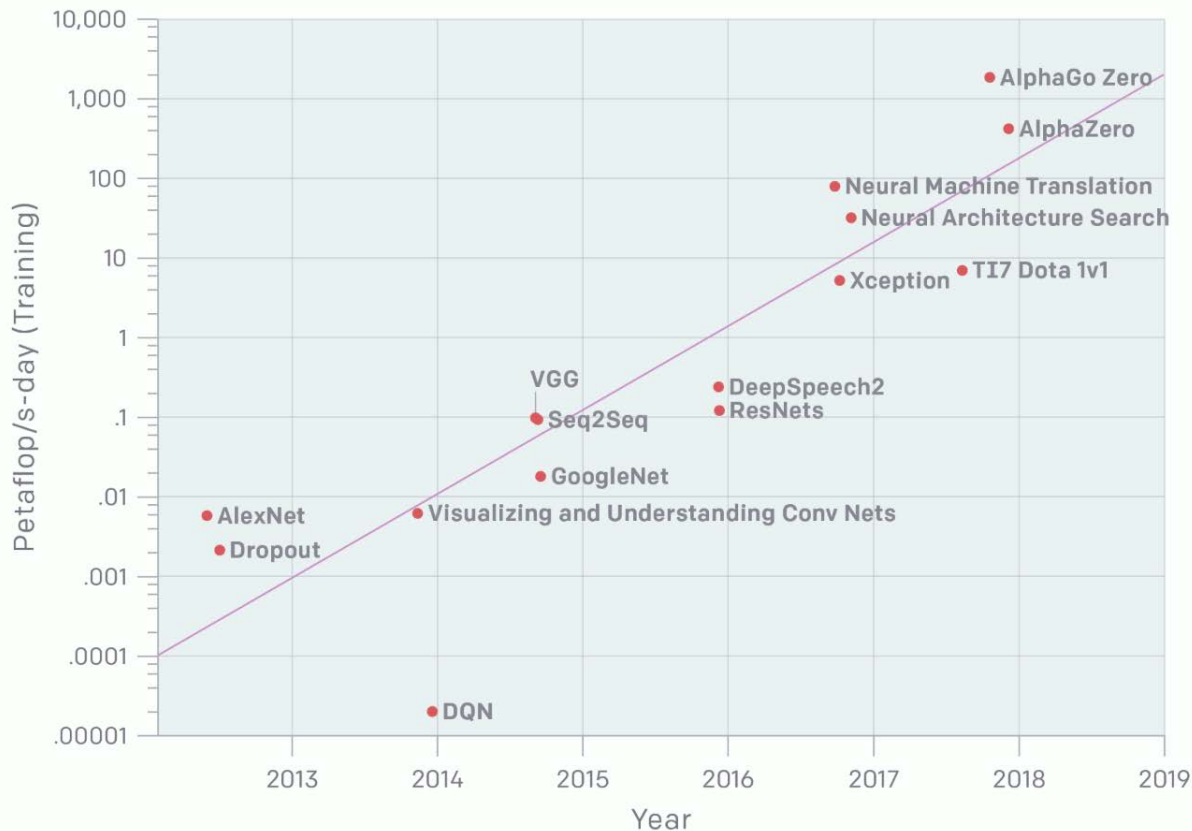
Watson beats
champions
2011



AlphaGo beats
Lee Sedol
2016

Never Enough Compute

AlexNet to AlphaGo Zero: A 300,000x Increase in Compute



"We're releasing an analysis showing that since 2012, the amount of compute used in the largest AI training runs has been increasing exponentially with a 3.5 month-doubling time (by comparison, Moore's Law had an 18-month doubling period). Since 2012, this metric has grown by more than 300,000x (an 18-month doubling period would yield only a 12x increase). Improvements in compute have been a key component of AI progress, so as long as this trend continues, it's worth preparing for the implications of systems far outside today's capabilities."

Misplaced Myth of the Killer App



Questions?

sean@pingthings.io