Big Data Analytics Platforms Architecture Requirements and Analysis Techniques

> Spring 2019 NASPI Sean Patrick Murphy

> > April 16th, 2019 San Diego, CA

SPRING 2019 NASPI

Habit #2

"Start with the end in mind"

25TH ANNIVERSARY EDITION OVER 25 MILLION COPIES SOLD

THE HABITS OF HIGHLY EFFECTIVE PEOPLE

POWERFUL LESSONS

WITH A FOREWORD BY JIM COLLINS, author of Good to Great and co-author of Great by Choice

Stephen R. Covey

Ping**Things**

SPRING 2019 NASPI

W PEO (Proc

(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

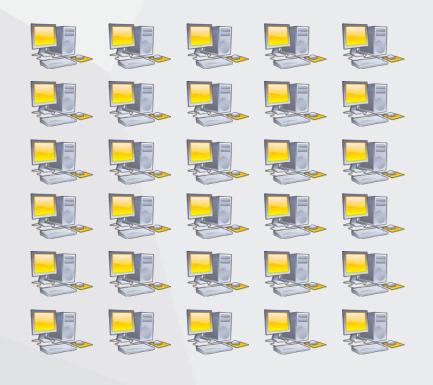
Annual Data Volumes

100000	4.6 TB	45.9 TB	459.2 TB	4.5 PB	44.9 PB	448.5 PB	4.4 EB	43.8 EB	437.9 EB	4.3 ZB	42.8 ZB
10000	470.2 GB	4.6 TB	45.9 B	459.2 TB	4.5 PB	44.9 PB	448.5 PB	4.4 EB	43.8 EB	437.9 EB	4.3 ZB
1000	47.0 GB	470.2 GB	4.6 TB	45.9 TB	459.2 TB	4.5 PB	44.9 PB	448.5 PB	4.4 EB	43.8 EB	437.9 EB
Streams	4.7 GB	47.0 GB	470.2 GB	4.6 TB	45.9 TB	459.2 TB	4.5 PB	44.9 PB	448.5 PB	4.4 EB	43.8 EB
	481.5 MB	4.7 GB	47.0 GB	470.2 GB	4.6 TB	45.9 TB	459.2 TB	4.5 PB	44.9 PB	448.5 PB	4.4 EB
J0 #	48.1 MB	481.5 MB	4.7 GB	47.0 GB	470.2 GB	4.6 TB	45.9 TB	459.2 TB	4.5 PB	44.9 PB	448.5 PB
	0.1 Hz 1 Hz 10 Hz SCADA PI			100 Hz PMU	1 KHz	10 Khz DFR	100 Khz	1 MHz	10 MHz	^{100 мн} г Sampling	1 GHz Rates
						Megabytes	Gigabytes	Terabytes	Petabytes	Exabytes	Zettabytes

Ping**Things**

Solving Large Compute Problems





Ping**Things**

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



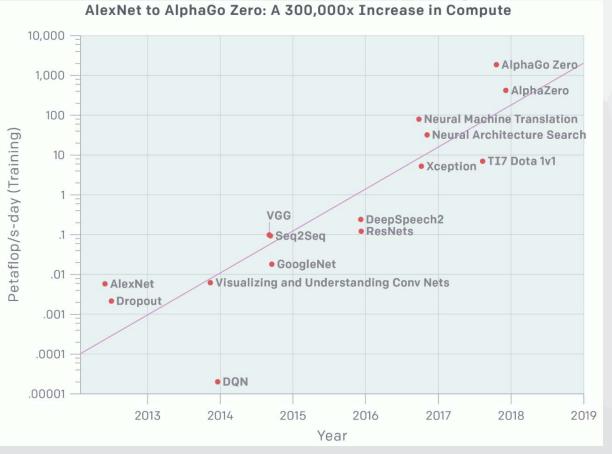
Watson beats champions 2011



AlphaGo beats Lee Sedol 2016

Ping**Things**

Never Enough 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