



Data Science for Scientists

Jim Dowling
SCS, KTH

SeRC Faculty Meeting, Nov 2016

SeRC 1.0

- Growing prevalence of computational models in the natural sciences and engineering disciplines.

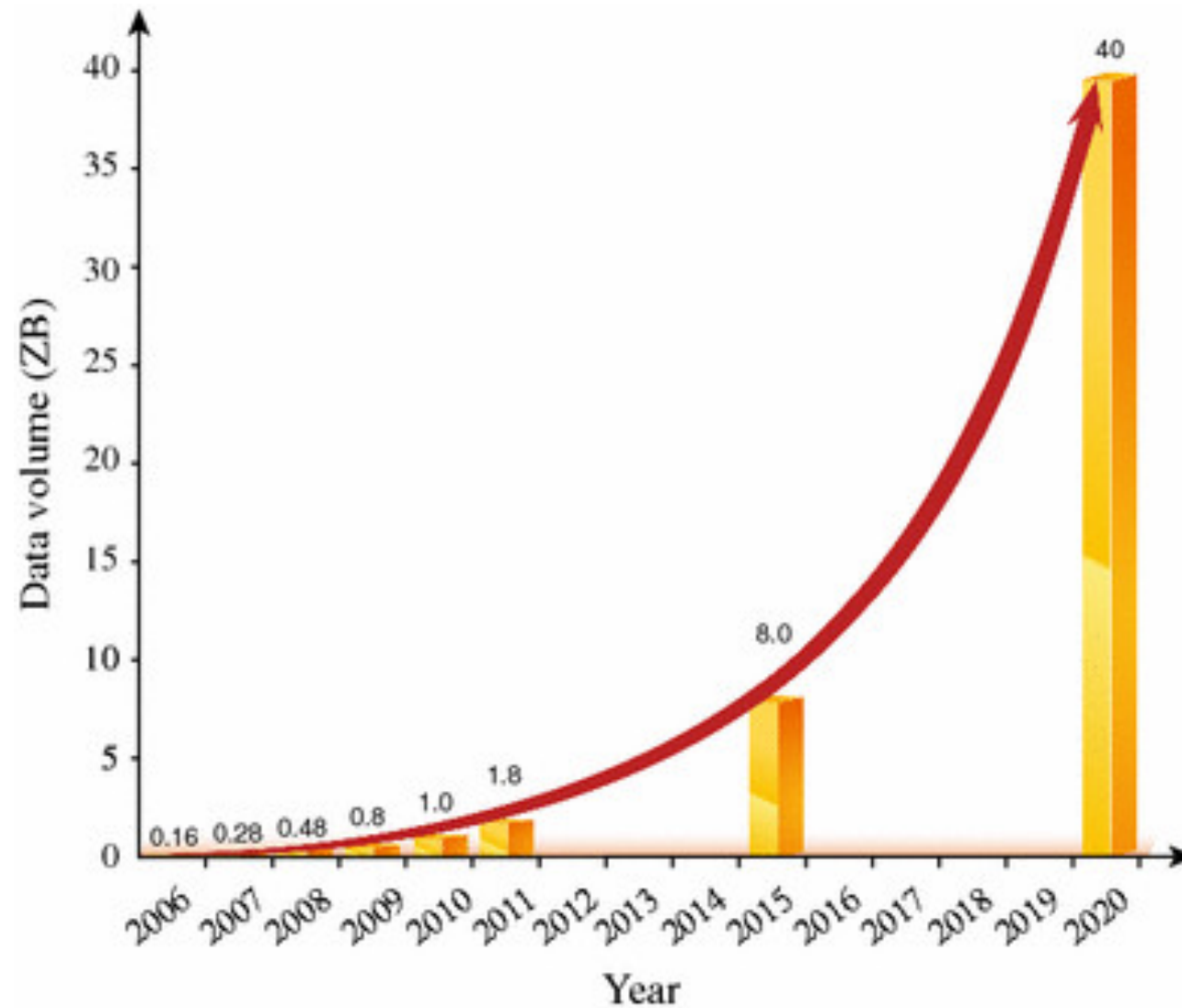
$$\frac{dx}{dt} = x(\alpha - \beta y)$$
$$\frac{dy}{dt} = -y(\gamma - \delta x)$$

circa 1900 (equation-based)

```
procedure ACO_MetaHeuristic
  while(not_termination)
    generateSolutions()
    daemonActions()
    pheromoneUpdate()
  end while
end procedure
```

circa-2000 (algorithm-based)

Growth in Data Volumes



What is Data Science?



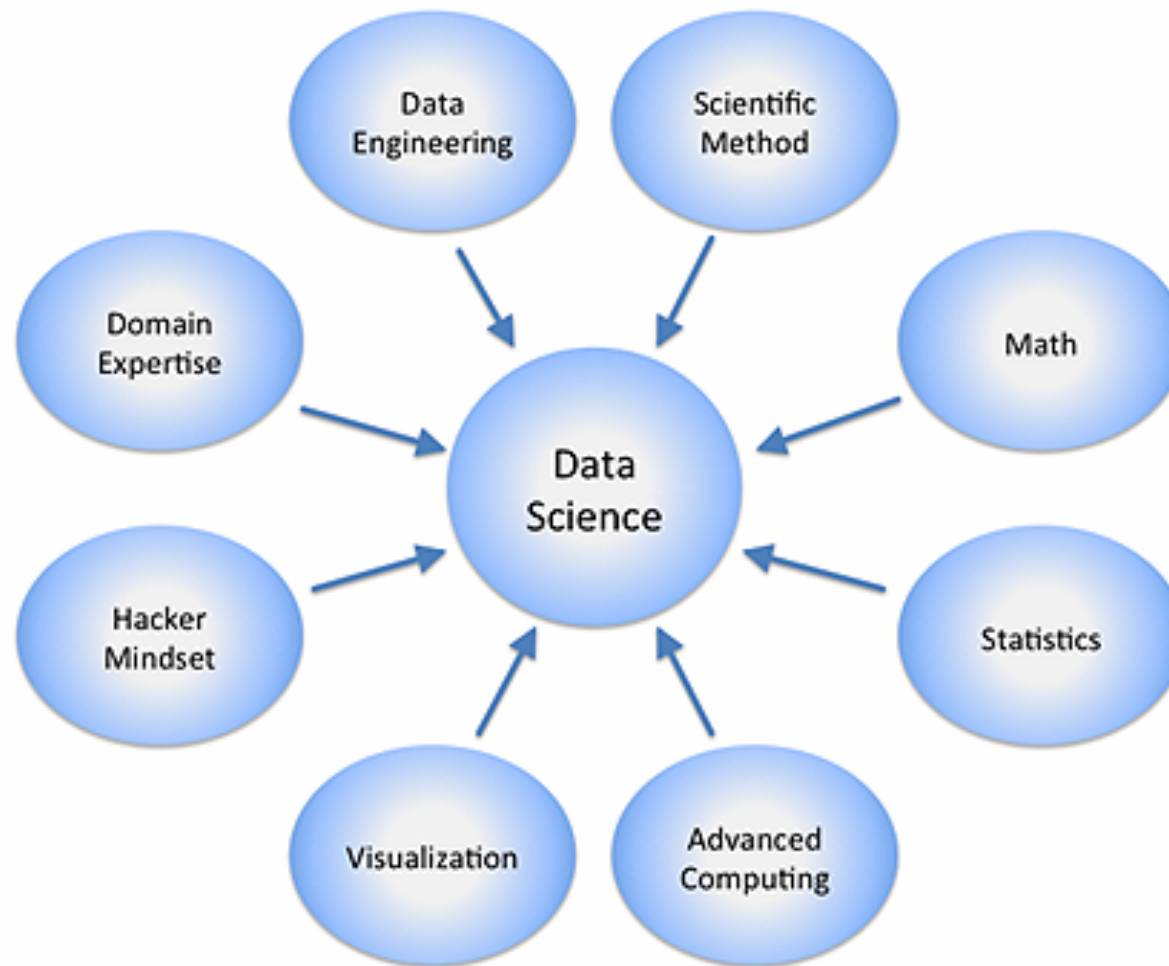
Software

Data

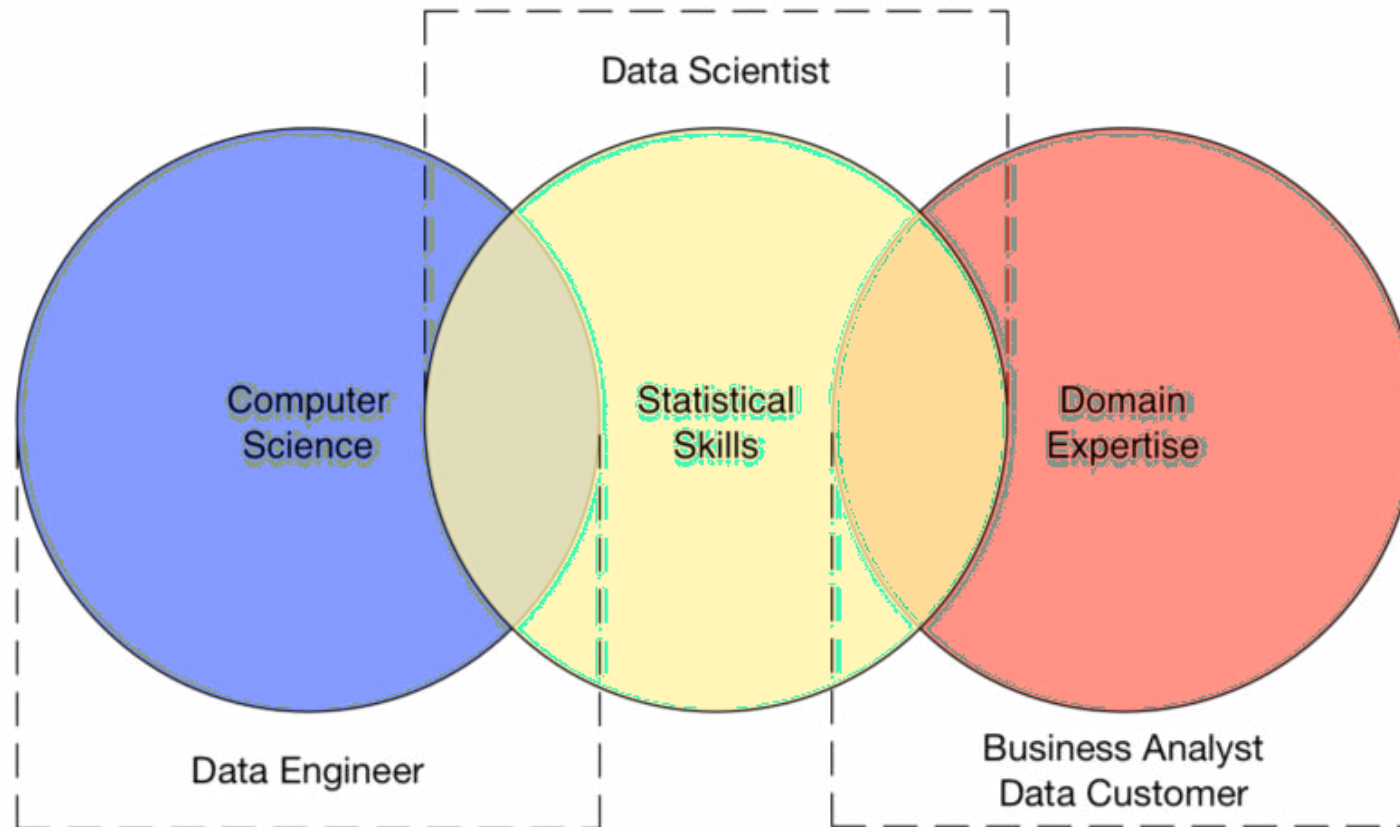
Gain insights of maximal value from
data, while securely minimizing
the cost needed to acquire
said insights

Insights

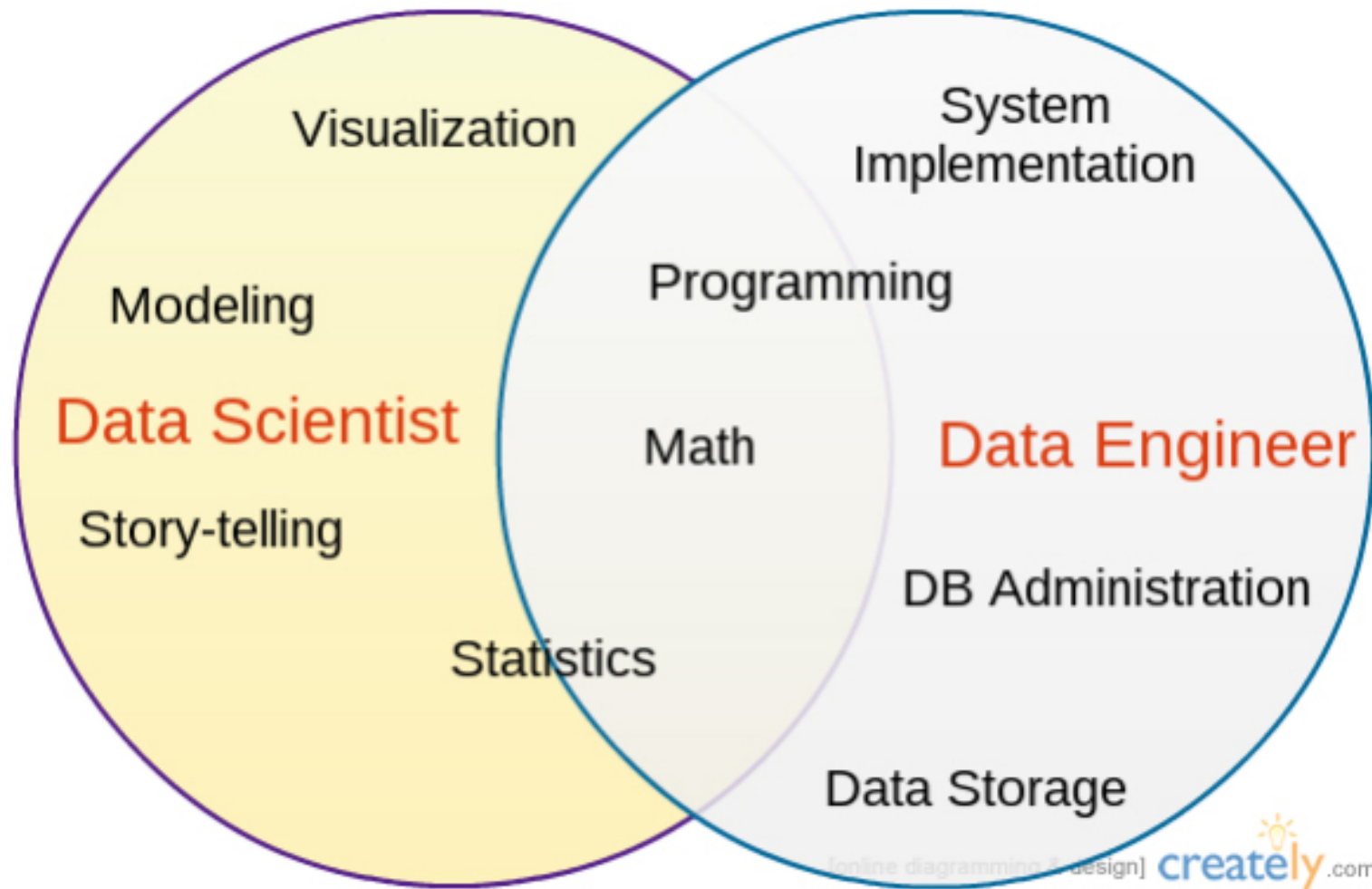
The Ideal Data Scientist



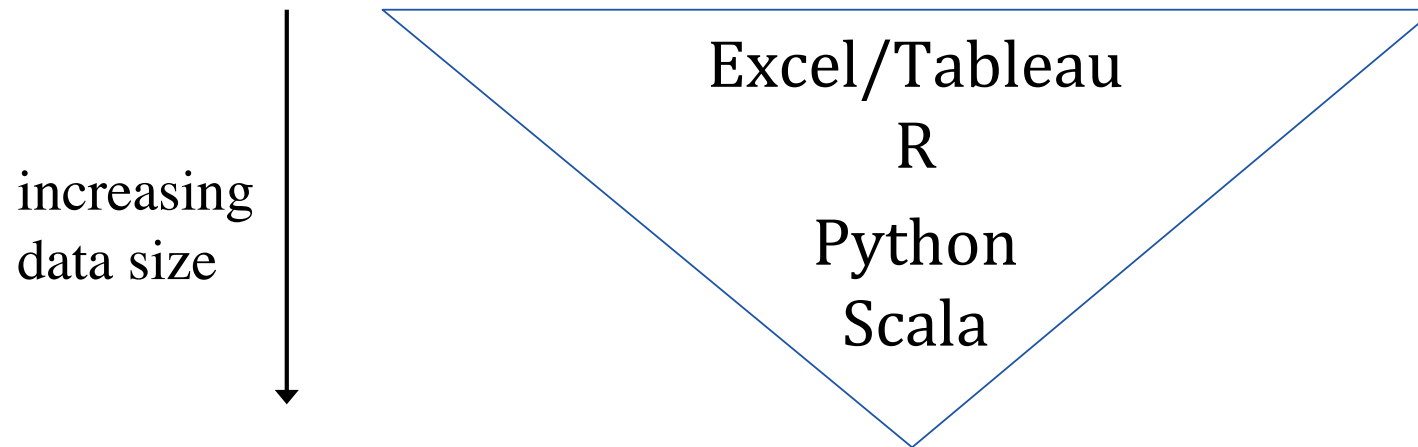
Industry's View on Being a Data Scientist



Industry's View on Being a Data Scientist

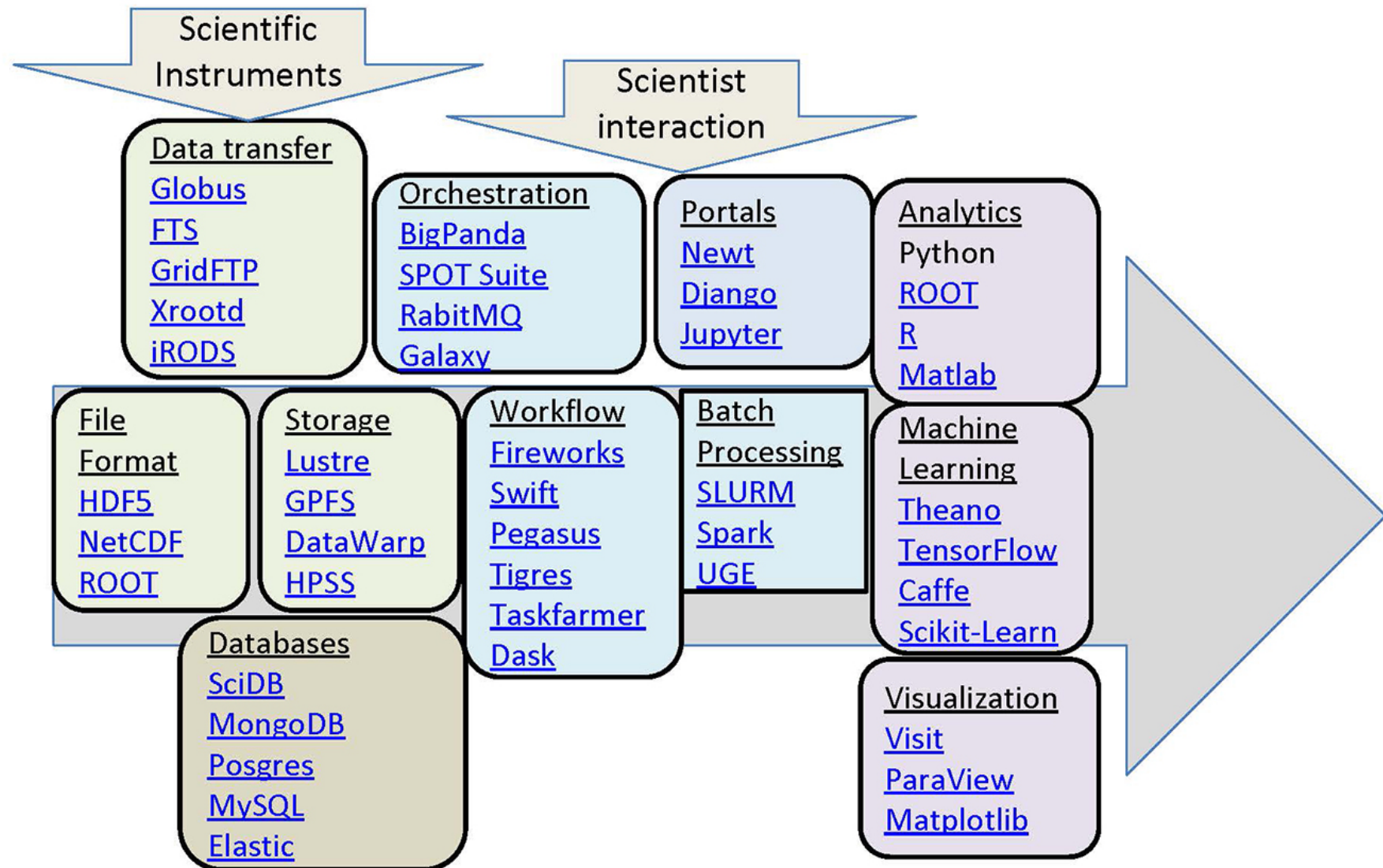


Programming for Data Science



Tools: R/Studio, Jupyter, Zeppelin, Tableau

Tools for the E-Science Data Scientist



Tools for the Industrial Data Scientist

IDE/Visualization (Jupyter, Zeppelin, Tableau)

Parallel Data Processing
(Spark, Tensorflow, Flink, SQL, MapRed)

Resource Mgmt (YARN, Mesos, Kubernetes)













Storage (HDFS, S3, WAS, Collosus)

Deep Learning

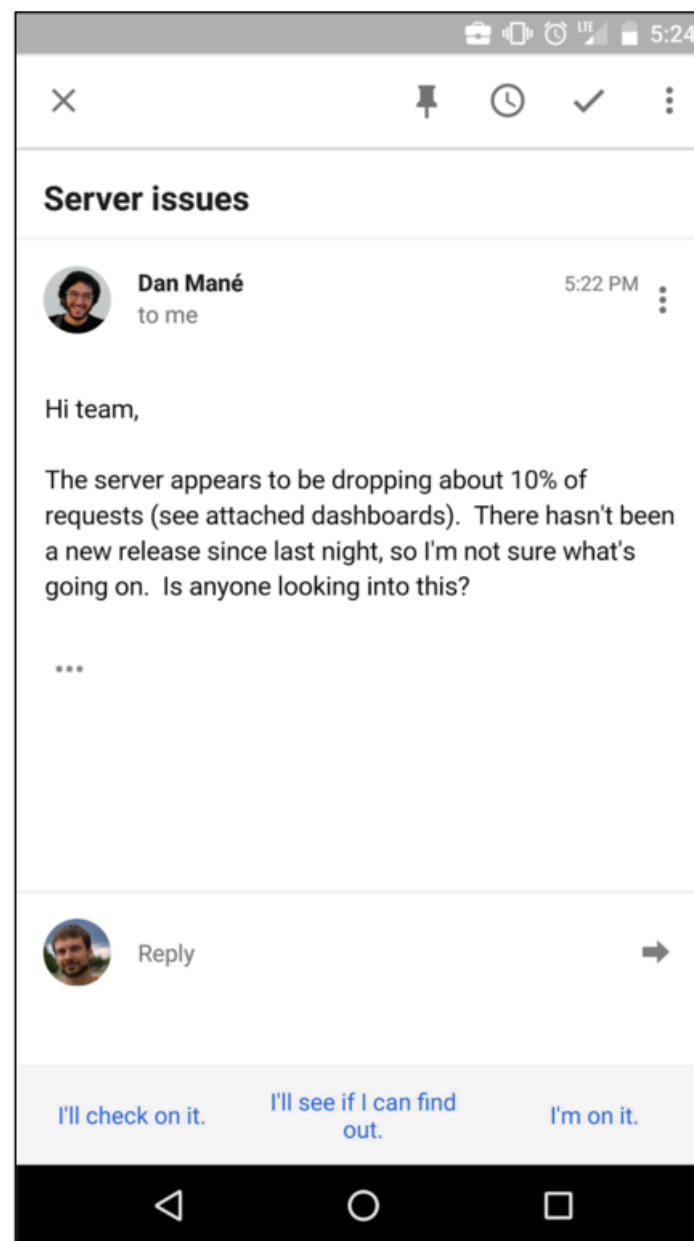
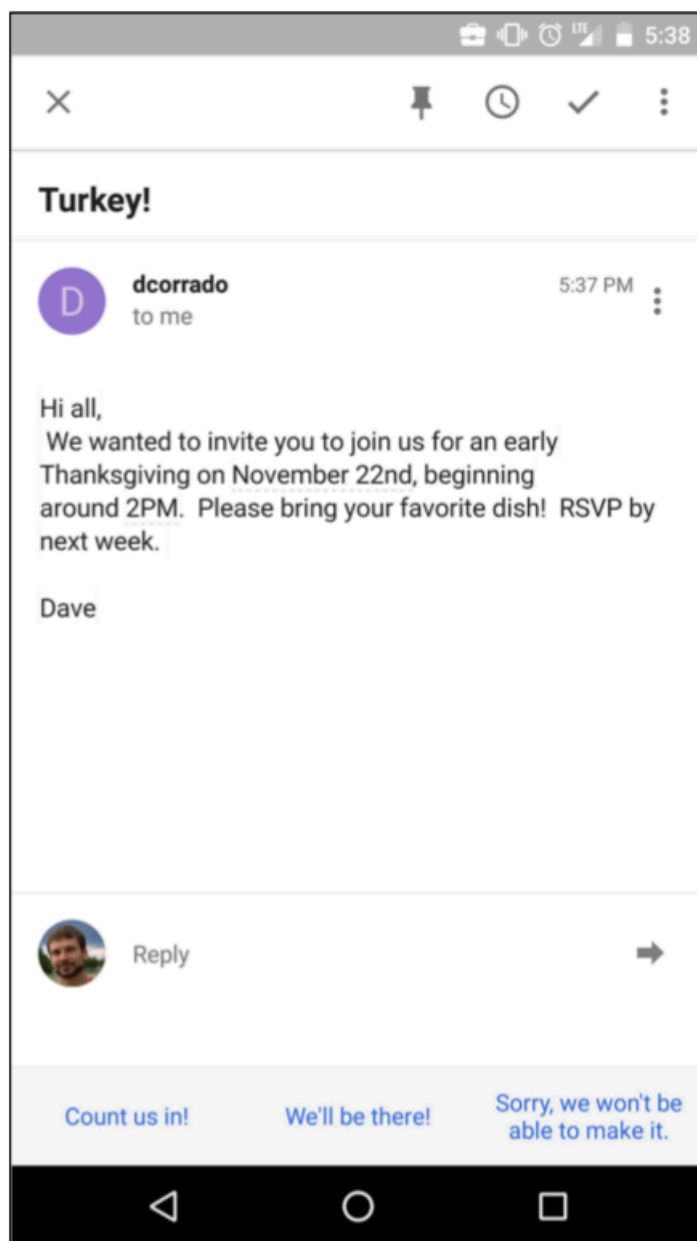


2017-05-26

Automated Image Captioning

Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image
 <p>A person riding a motorcycle on a dirt road.</p>	 <p>Two dogs play in the grass.</p>	 <p>A skateboarder does a trick on a ramp.</p>	 <p>A dog is jumping to catch a frisbee.</p>
 <p>A group of young people playing a game of frisbee.</p>	 <p>Two hockey players are fighting over the puck.</p>	 <p>A little girl in a pink hat is blowing bubbles.</p>	 <p>A refrigerator filled with lots of food and drinks.</p>
 <p>A herd of elephants walking across a dry grass field.</p>	 <p>A close up of a cat laying on a couch.</p>	 <p>A red motorcycle parked on the side of the road.</p>	 <p>A yellow school bus parked in a parking lot.</p>

Email Smart Reply with RNNs



Convnets for Art



DeepDream reddit.com/r/deepdream

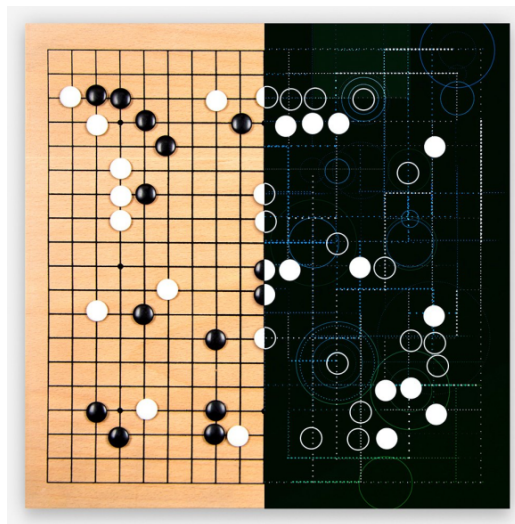


NeuralStyle, Gatys et al. 2015
deepart.io, Prisma, etc.

DeepRL for Playing Games



ATARI game playing, Mnih 2013



AlphaGo, Silver et al 2016



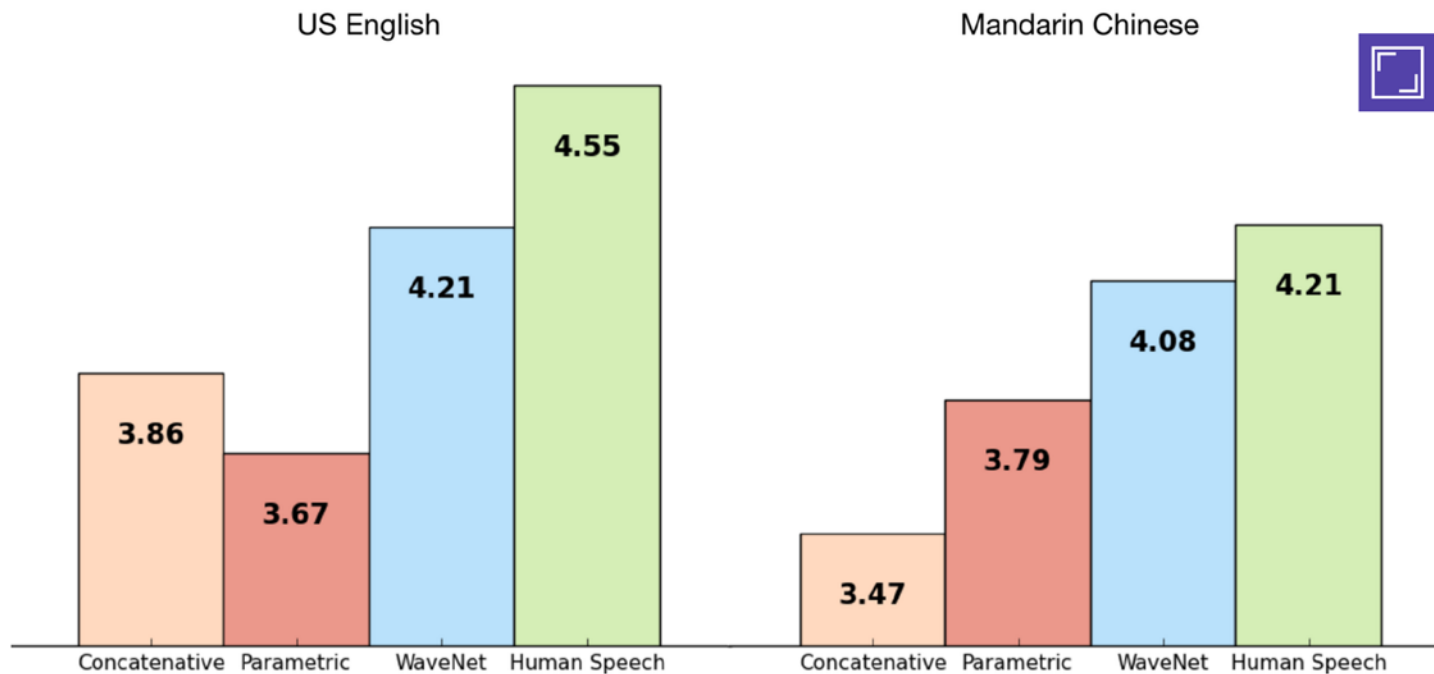
Self-Driving Cars



<https://www.udacity.com/course/self-driving-car-engineer-nanodegree--nd013>

DeepMind WaveNet

- A deep generative model of raw audio waveforms
 - Has to be heard to be believed



<https://deepmind.com/blog/wavenet-generative-model-raw-audio/>