



omni • sci

Productivity at Scale:

Data Science at the Speed of Curiosity

o m n i • s c i

What we'll talk about

- Who we are
- A little bit about 'curiosity' in data science
- The idea of *productivity at scale* for data science
- Are we there yet? How can we get there?
- Q&A

Who we are



Venkat Krishnamurthy

VP Product, OmniSci
Formerly - GS, Oracle, Cray (HPC)



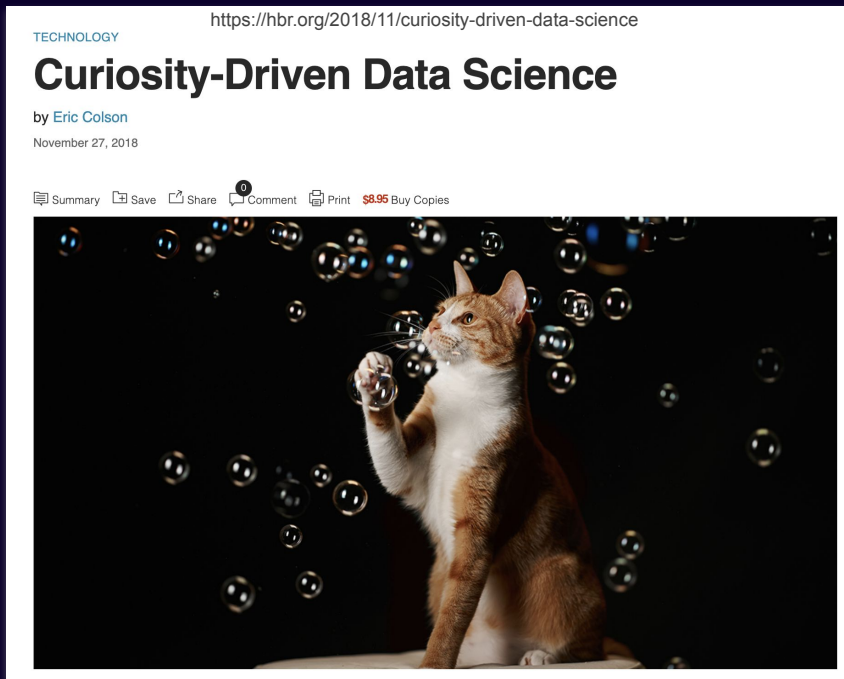
Alex Baden

Technical Director, OmniSci
Formerly - PhD Johns Hopkins, Allen Institute

What are we going to talk about?

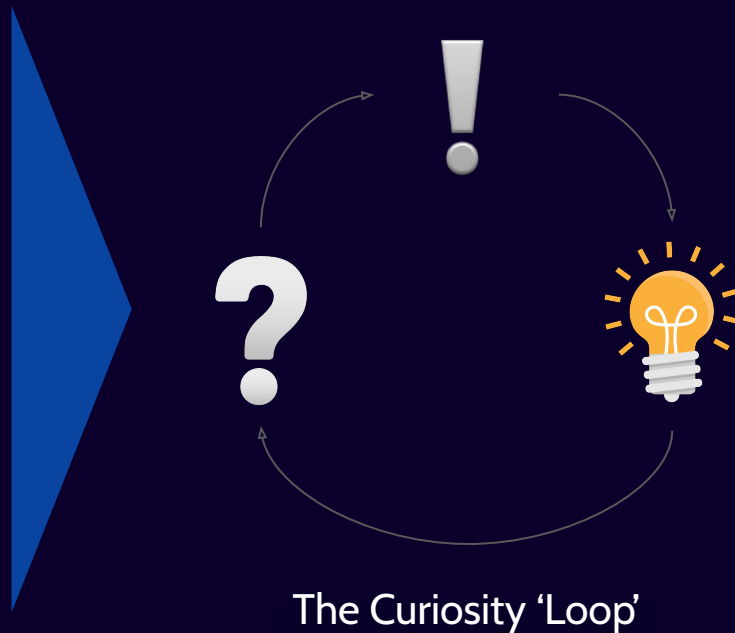


Why care about curiosity?



'**Data scientists** are a **curious** bunch, especially the good ones. They work towards clear goals, and they are focused on and accountable for achieving certain performance metrics.

But they are also easily distracted, in a good way. In the course of doing their work they stumble on **various patterns, phenomenon, and anomalies** that are unearthed during their data sleuthing'



The Curiosity 'Loop'

Data Science in the age of plenty

2009

The Unreasonable Effectiveness of Data

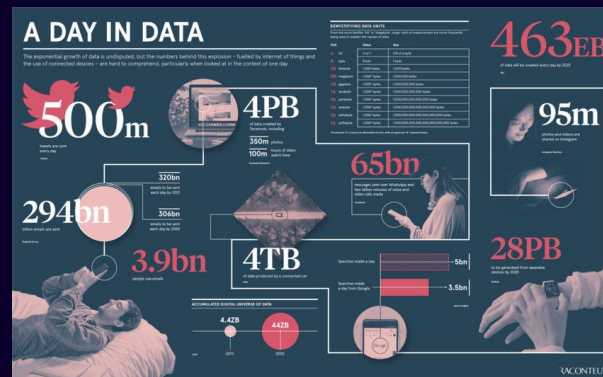
Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

An informal, incomplete grammar of the English language runs over 1,700 pages.² Perhaps when it comes to natural language processing and related fields, we're doomed to complex theories that will never have the elegance of physics equations. But if that's so, we should stop acting as if our goal is to author extremely elegant theories, and instead embrace complexity and make use of the best ally we have: the unreasonable effectiveness of data.

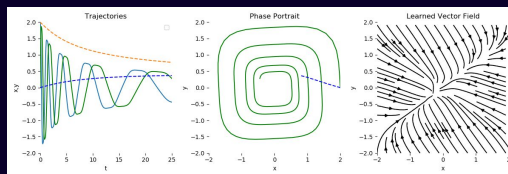
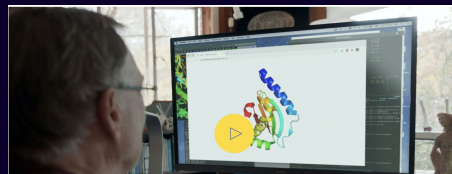


AKA the 'More data beats better algorithms' article

Today

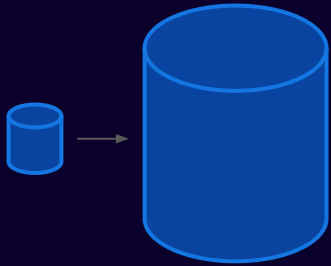


More Data...



...AND Better algorithms

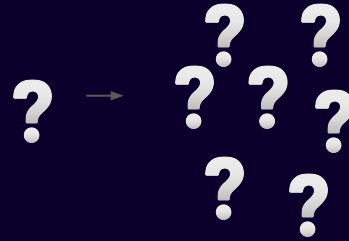
Yet...



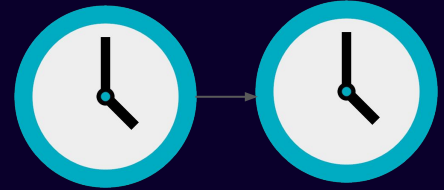
More data...



Better methods...



More questions...



of useful insights

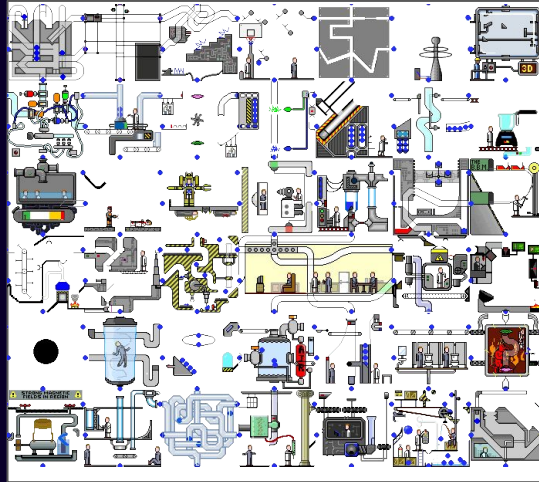
time

Data science *productivity* hasn't scaled

Why is this?



Non-scalable tools



Operational overheads





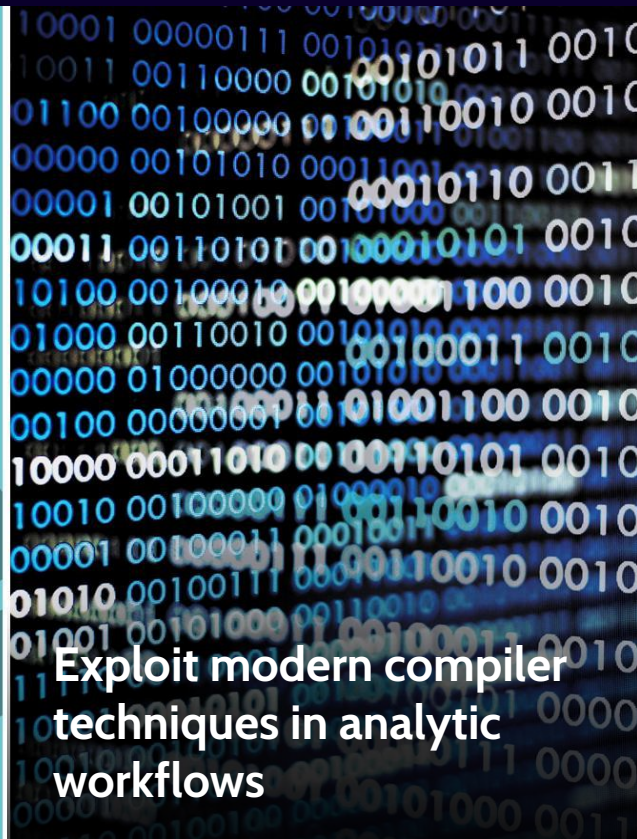
Demo



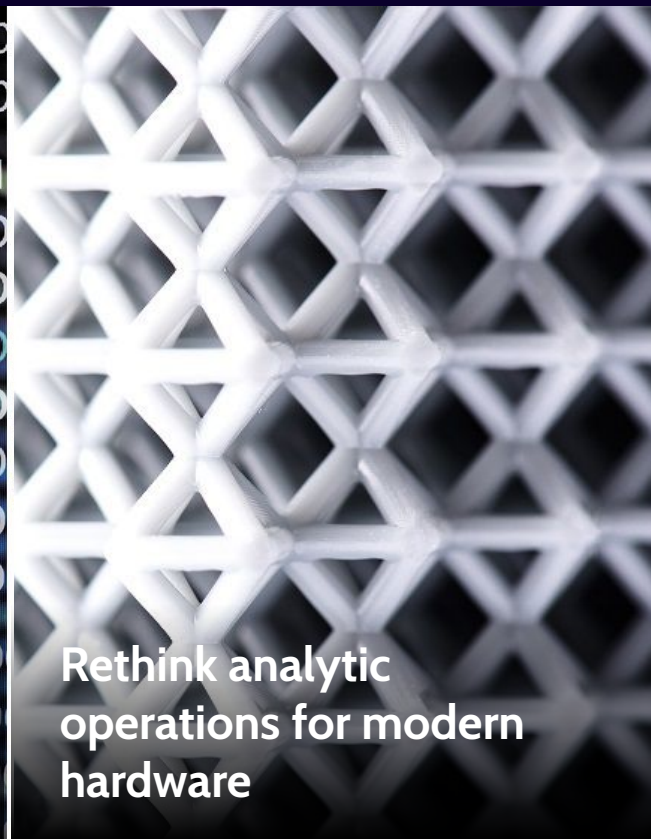
How do we do it?



Efficiently use the modern
memory hierarchy

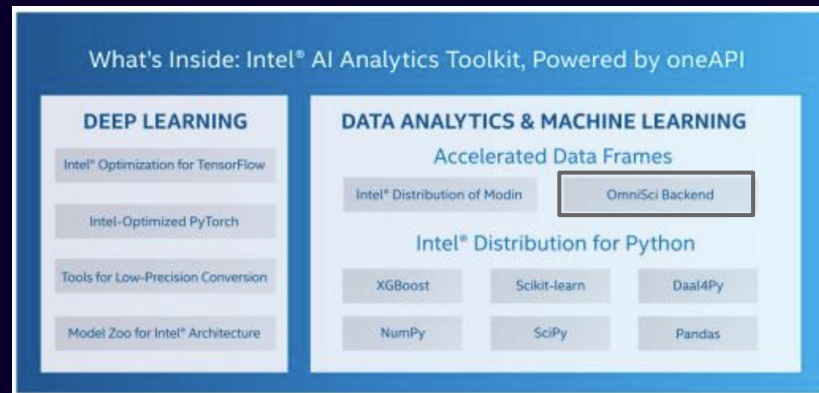
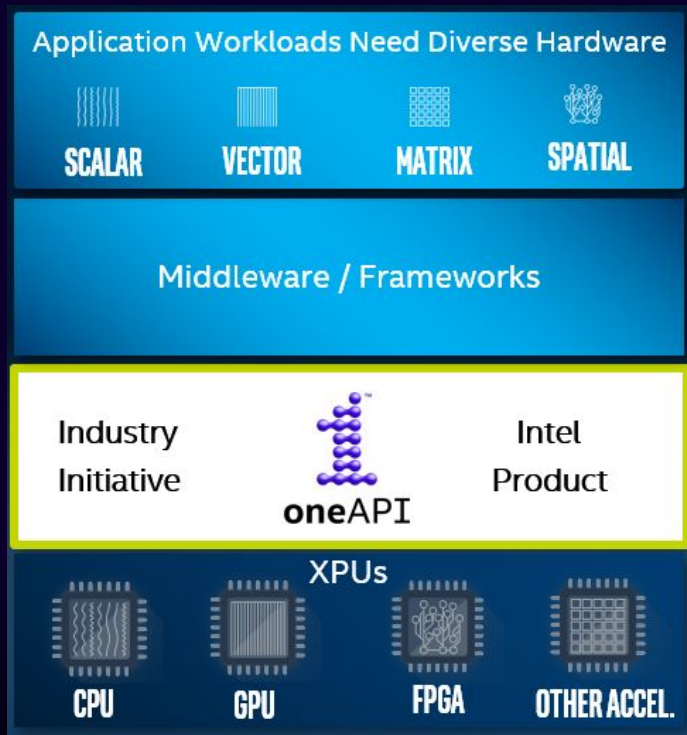


Exploit modern compiler
techniques in analytic
workflows



Rethink analytic
operations for modern
hardware

Intel and OmniSci together



OmniSci transparently enables greater scalability and performance as part of Intel's AI Analytics toolkit

A satellite night view of North America, showing the United States and parts of Canada and Mexico. The landmasses are dark blue, and the city lights are bright yellow and white, creating a dense pattern of stars across the continent. The text is overlaid on the central part of the image.

**OmniSci's mission is to make analytics instant,
powerful, and effortless for everyone**

A satellite night map of North America, showing the United States and Mexico. The landmasses are dark blue, and the city lights are represented by numerous bright yellow and white dots of varying sizes, indicating population density. The text "THANK YOU" is centered over the United States.

THANK YOU

The background is a dark blue gradient. It features several abstract elements: a large, multi-segmented circular arc on the left side; a horizontal dotted line in the upper center; a smaller horizontal dotted line to the left of the center; a line graph with two circular markers and an upward trend on the right; and a bar chart with seven bars of increasing height at the bottom right.

Q+A



o m n i • s c i