Parallel Inception

MPP Databases    GPGPU

Kyle Dunn
Me

Data nerd for Pivotal

Pythonista

Recovering HPC/GPGPU researcher
Poll
Edmund Cartwright (1784/5)

Scale vs. Capability
GPGPU

Code blueprint

Magic

Results
SELECT foo(columnA), bar(columnB) FROM someTable;
The inception

MPP Node

100s-1000s of GPU cores
100s-1000s of GPU cores
Ship it.
So what?
\[
\frac{Data^{BIG}}{Nodes_{MPP} \cdot Cores_{GPU}}
\]
x1 SQL query

xN - MPP Nodes

xP - GPU Cores

Shard 1

Shard 2
Free lunch?

1. IO
   - Depends on the restaurant...

2. SIMD
   - And the entrée...

3. Concurrency
   - Hoard your Pi[e]
“Find the related measurements in full flight data.”
“Find a set of plants with resemblance to this one.”
Gnetophyta

Magnoliophyta

Cycadophyta

x1 SQL query

xN - MPP Nodes

xP - GPU Cores
Geospatial

“Find locations with high historical solar insolation.”
x1 SQL query

xN - MPP Nodes

Florida

Colorado

California

xP - GPU Cores
tl;dr

• Turn-key >> nuts and bolts

• JIT your code

• SQL lives (seriously though)
and then...

kdunn@pivotal.io

@kdunn926