
NIMBIX
Supercomputing made super human™

# The New Age of Accelerated Computing:
## A History of Innovation and Optimization in Computing

Steve Hebert, Cofounder and CEO, Nimbix

NIMBIX
supercomputing made super human™

# The First "Supercomputers"

## A History

- 1880 census had taken eight years to process
- 1890 Census used the Hollerith "Tabulator" reducing census processing by 2 years
- Innovation: Electro-mechanical relay
- Became core of IBM

# Supercomputing Defined

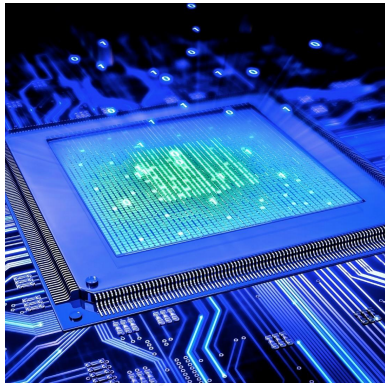
## A History

- “Supercomputer” referred to large tabulators that IBM had made for Columbia University in the late 1920s
- In 1935, punch card systems processed data for Social Security for 26 million workers

# Control Data CDC6600

## A History

- In 1964, the CDC 6600 ran 10 times faster than the next fastest machines
- 1 MegaFLOPS
- Innovation: Germanium to Silicon transistors, refrigeration

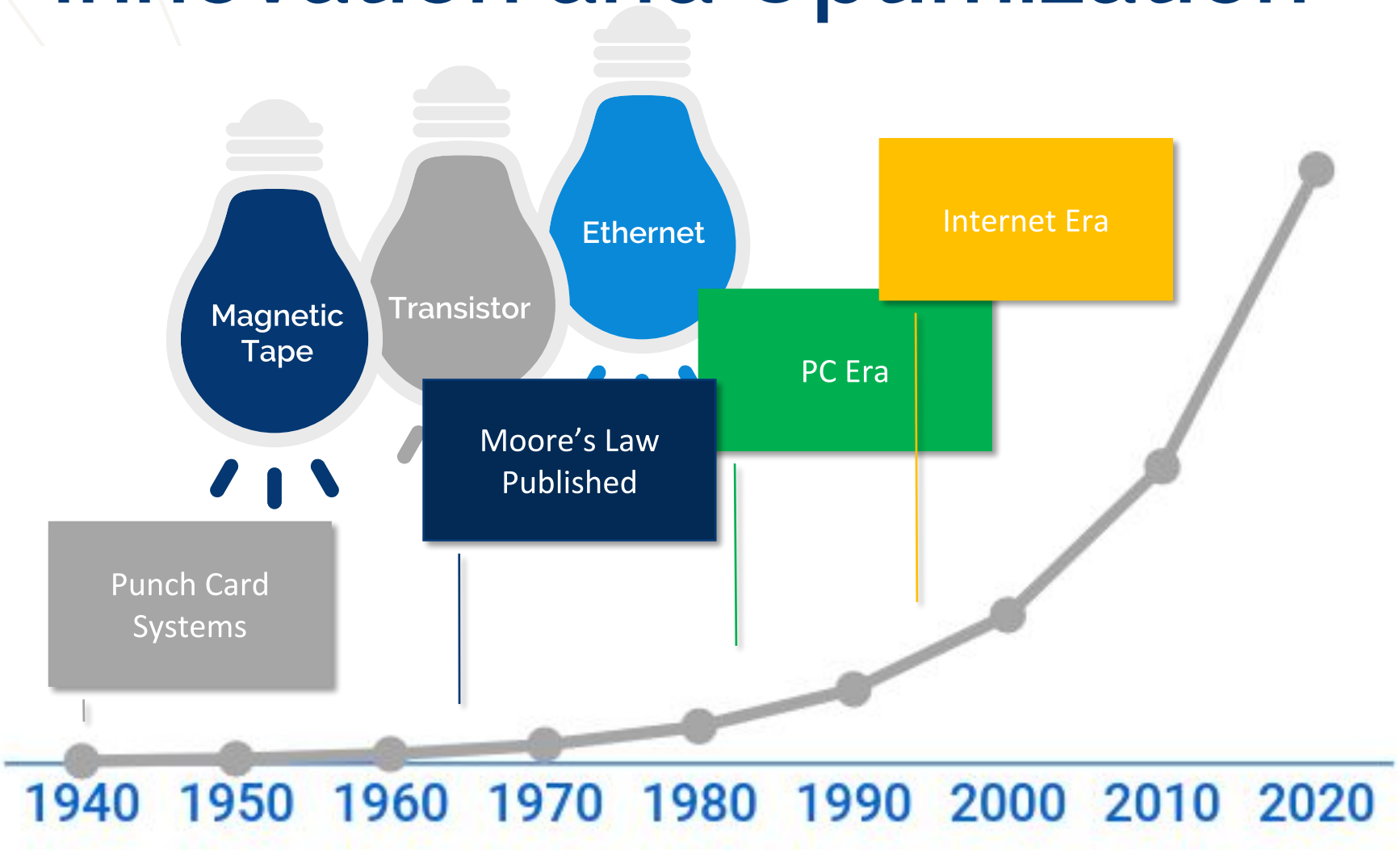


# Moore's Law

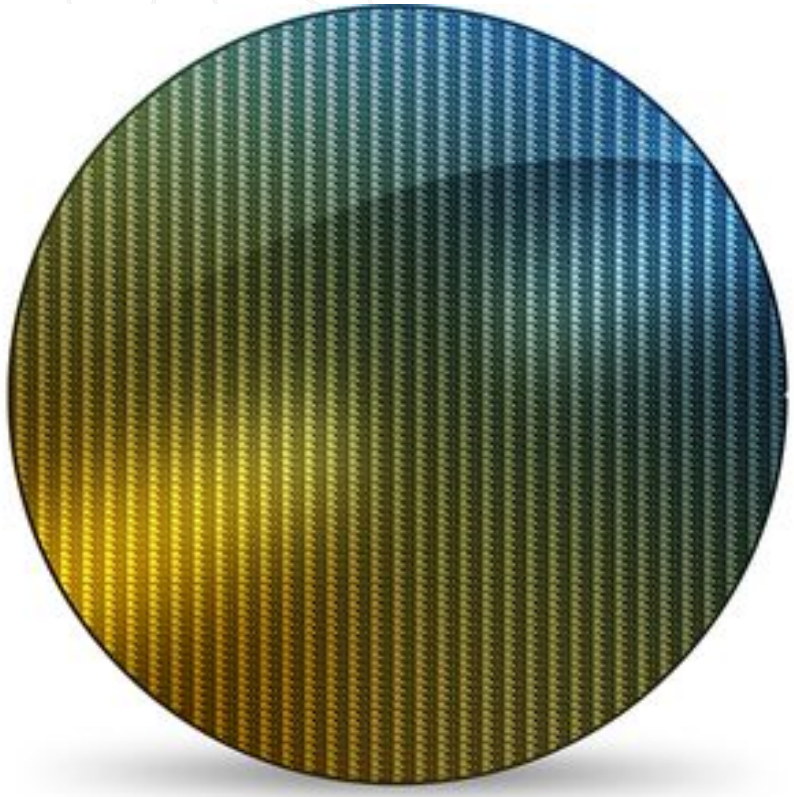
## Since CDC6600, 50 Years of Moore

"Moore's Law" is the observation that, over the history of computing hardware, the number of transistors in a dense integrated circuit doubles approximately every two years.

# Innovation and Optimization



# What has Moore's Law Enabled?



- Consumerization of Computing
- Democratization of Software Development
- Economic Predictability

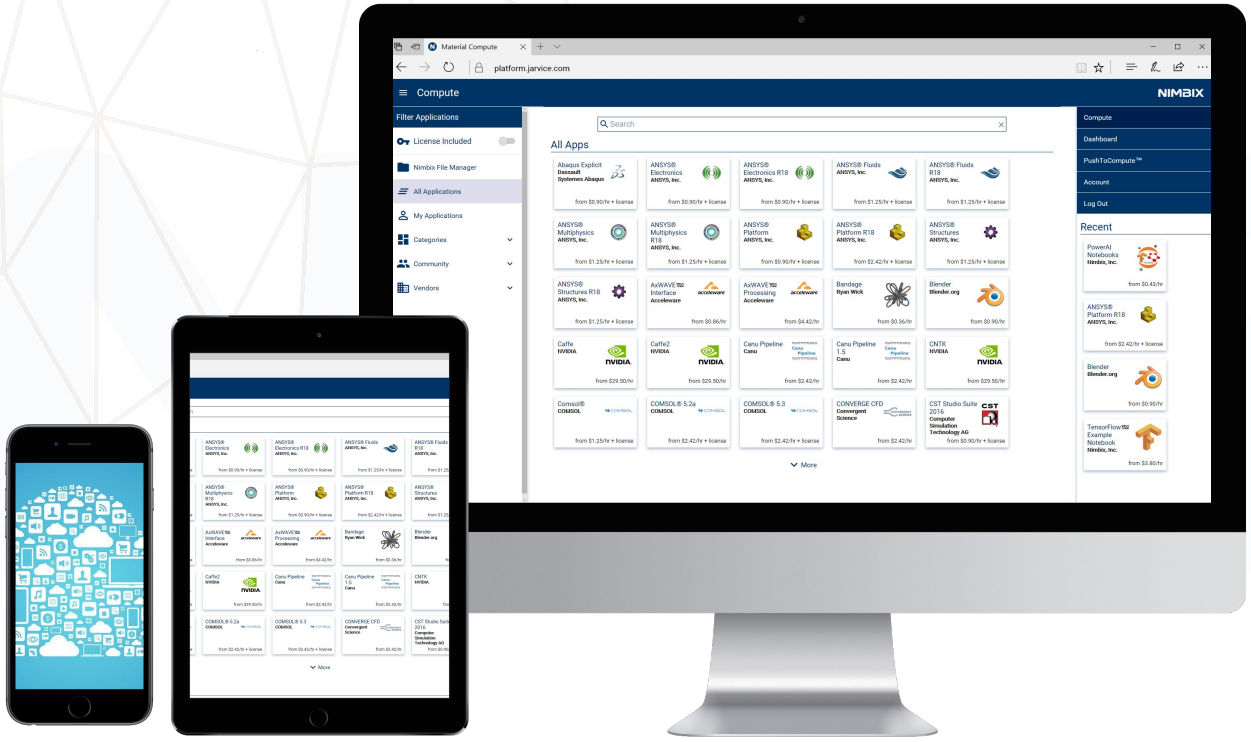
# Consumerization of Computing

- < 5000 PCs in 1975
- By 2004, > 130M units shipped annually
- PCs give way to notebooks, tablets and smartphones

# Democratization of Software Development


- Platforms for writing code and applications
- FORTRAN, COBOL, BASIC
- C, Python

# Economic Predictability

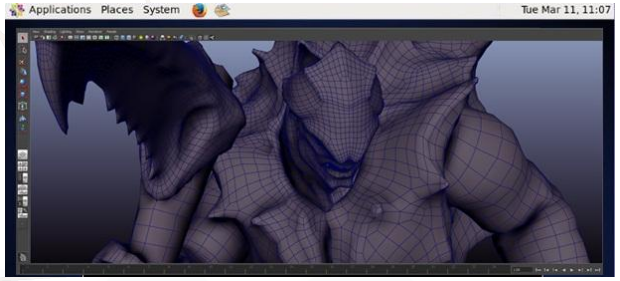


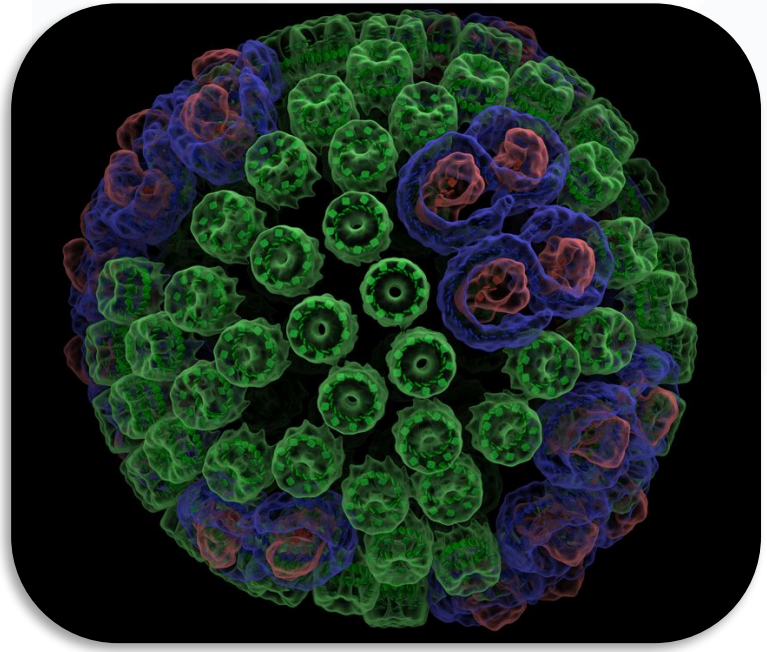

- We live in a tech world of planned Release Cycles
- Moore's Law -> The next semiconductor process node, is the clock cycle of optimization (and certainly incremental innovations)

NIMBIX
supercomputing made super human™

# What Happens Next?

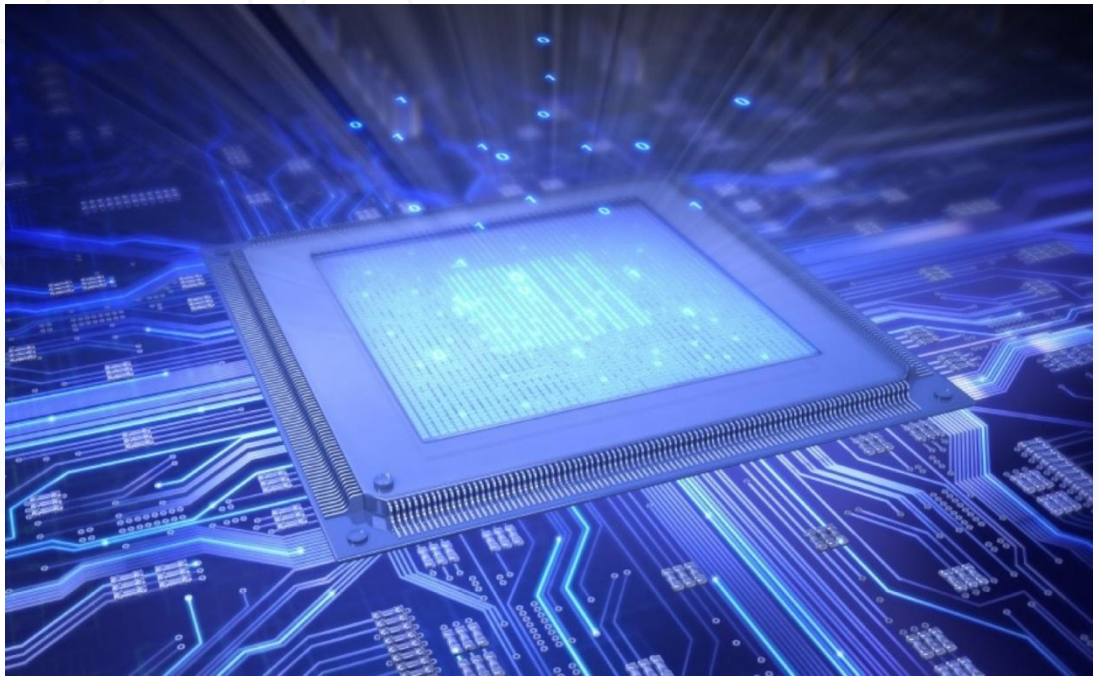
# Graphics Processors

- From Gaming to Supercomputing
- Highly Parallel Structure
- Moore's Law at work: Thousands of thread cores per device
- Significant performance gains over CPU for certain classes of problems and algorithms

# The Field Programmable Gate Arrays



- A blank slate of logic gates that can be reconfigured with different functionality
- Moore's Law at work: significant growth in gate count
- Important devices in the communications industry

# Deep Learning

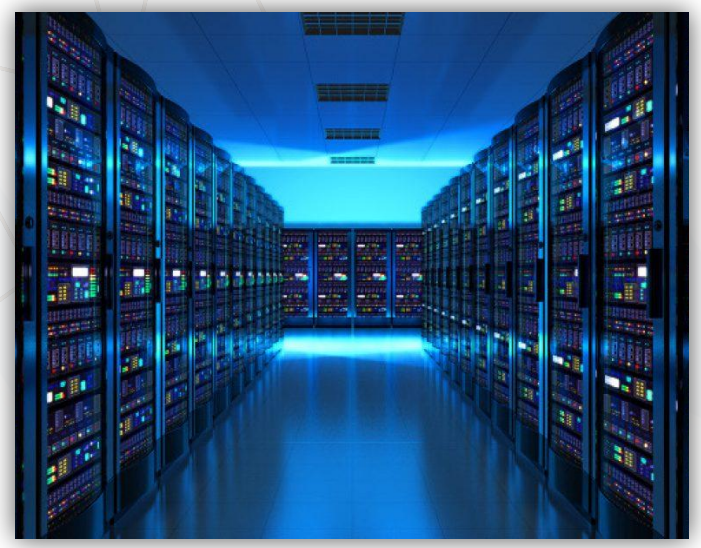
- Create automation to deal with massive amount of data
- Use massive amounts of labeled data to train machines ->Make them smart, fast
- Apply unsupervised learning to assemble unstructured data into groups -> New Insights

# Go and AlphaGo

"After humanity spent thousands of years improving our tactics, computers tell us that humans are completely wrong... I would go as far as to say not a single human has touched the edge of the truth of Go.“

– Ke Jie, #1 Worldwide in Go

# Moore's Law and Cloud Computing

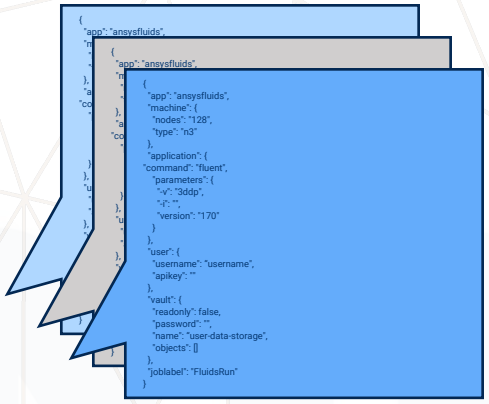


"What Moore's Law has delivered for Computing, Cloud Computing will deliver for Deep Learning."

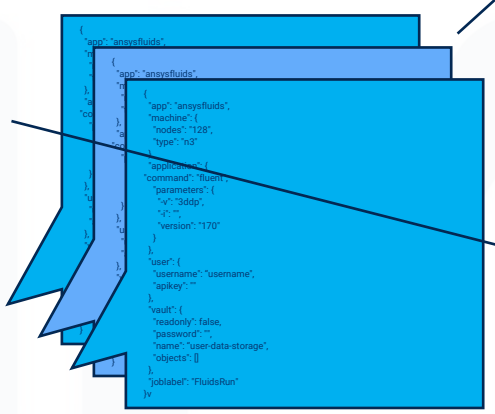
- Consumerization of Deep Learning
- Democratizing software and model development
- Economic Predictability -> Cloud Machine Release Cycles

# Machines Defined at Run Time

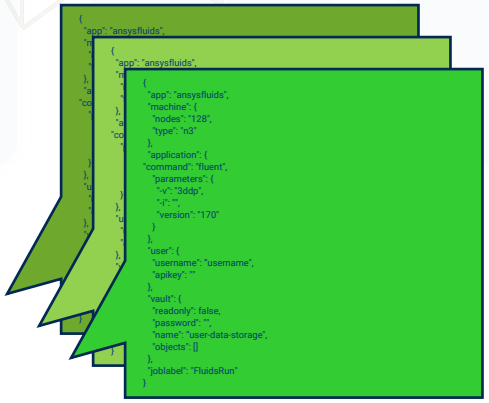
Deep Learning



NGS Sequencing



Simulations



```
> Genome Assembly
> Allocate 2 Nodes, 1TB RAM, FPGAs
> Optimize code
> Synthesize,load bitstream
> Execute
```

```
> Parallel Tensorflow
> Allocate 4 CPUs, 32 GPUs
> Execute
```

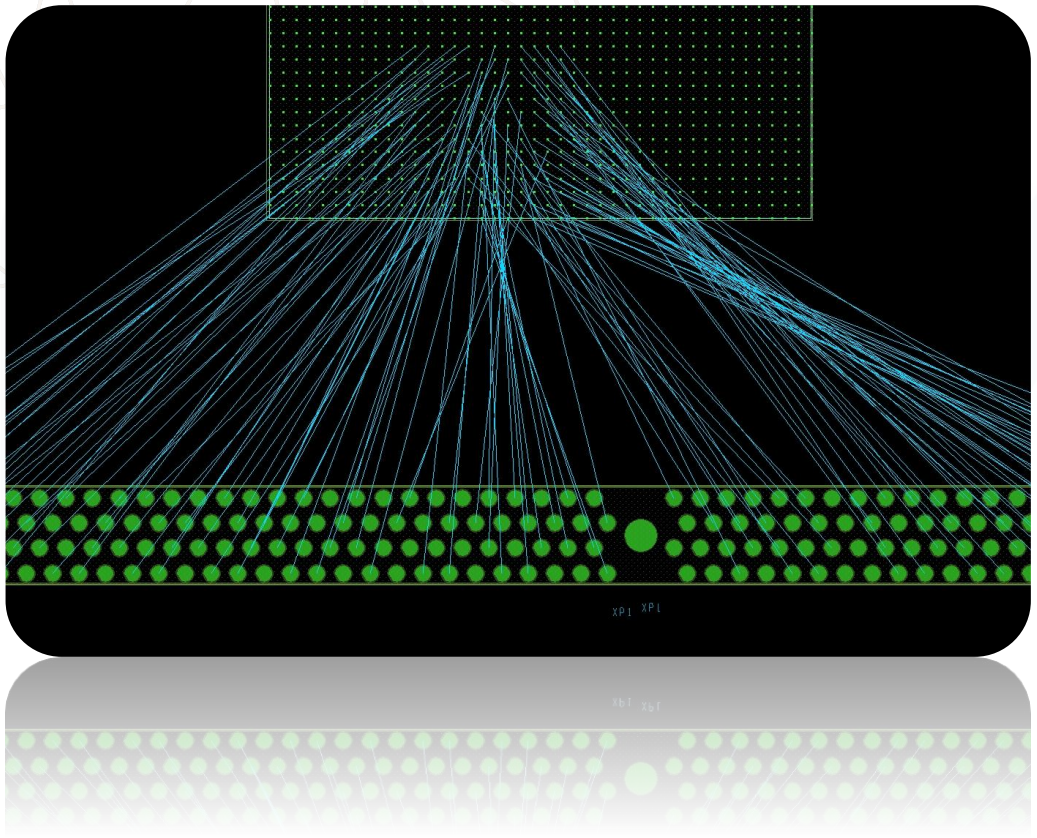
```
> Fluid Simulation
> Allocate 2000 CPU Cores
> Optimize code
> Synthesize,load bitstream
> Execute
```

# Deep Learning Applied to Cloud Computing


```
{
  "app": "ansysfluids",
  "machine": {
    "nodes": "128",
    "type": "n3"
  },
  "application": {
"command": "fluent",
    "parameters": {
      "-v": "3ddp",
      "-i": "",
      "version": "170"
    }
  },
  "user": {
    "username": “username",
    "apikey": ""
  },
  "vault": {
    "readonly": false,
    "password": "",
    "name": “user-data-storage",
    "objects": []
  },
  "joblabel": "FluidsRun"
}
```

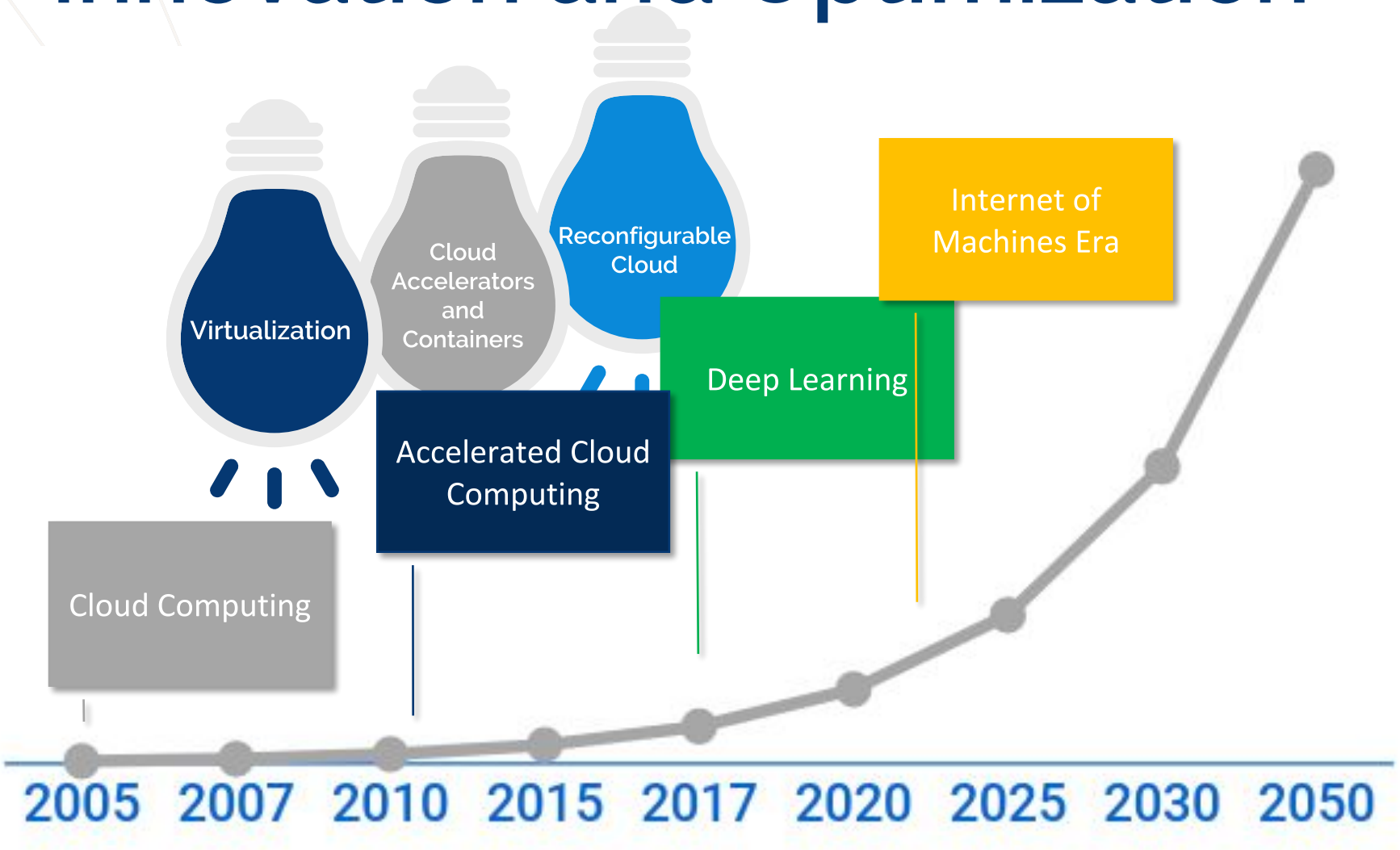

- Compute Workloads as API Calls
- Millions and billions of API calls define compute tasks and data payloads
- Rules to determine reward:
  - Optimize for Energy efficiency, throughput, run time, resource utilization

# Machine-Driven Innovation



- Millions and billions of API calls define compute tasks and data payloads
- Rules to determine reward:
  - Optimize for Energy efficiency, throughput, run time, resource utilization
- Moving from “machine-assisted” to “machine-driven”
- Evolve to Self Optimizing, Intelligient Systems

# Innovation and Optimization



# A New Kind of “Supercomputing”

“The *Internet of Machines* is the collection of intelligent systems that self-optimize in order to automate and accelerate the collection, distribution, analysis and transformation of zettabytes of data.”