



# Building and Scaling FPGA Accelerated Applications for Cloud Computing

Presented By

**NIMBIX**  
*supercomputing made super human™*

Steve Hebert

CEO

October 2, 2018



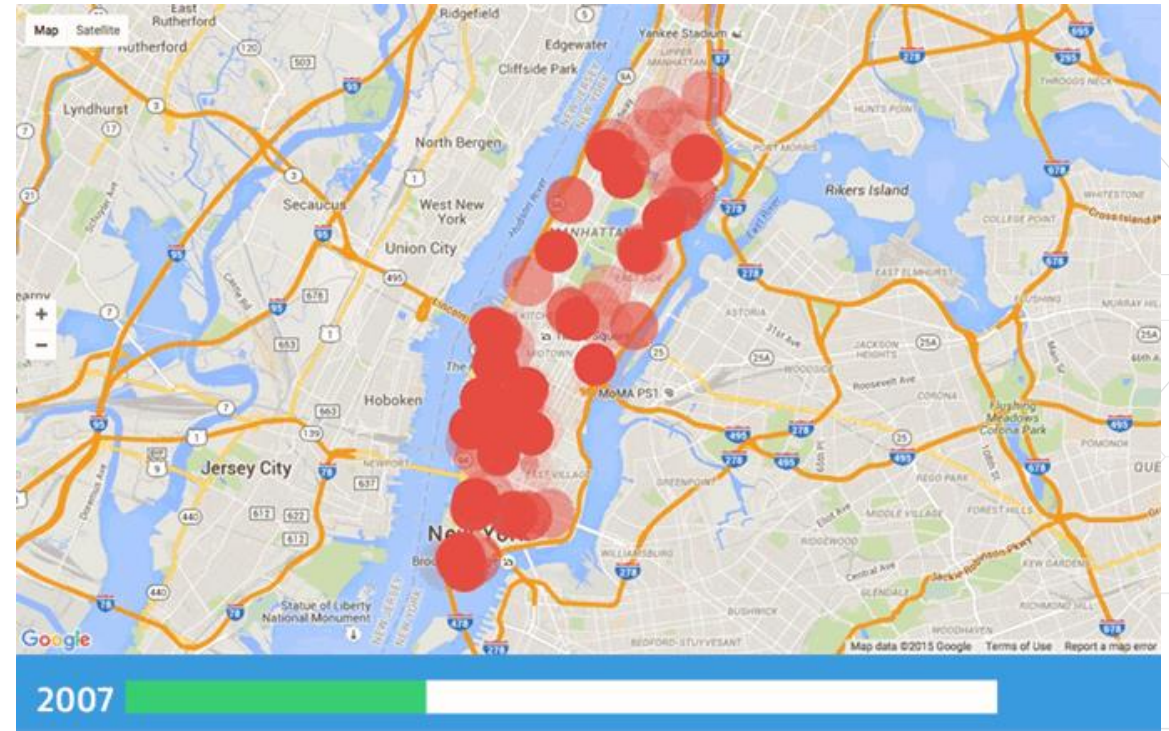


# Why FPGAs in Computing Matter Now

- > **Moore's Law horizon is in view**
- > **When we are done shrinking transistors where do we innovate?**
  - >> 1) Invent a new kind of chip technology
  - >> 2) Iterate, Optimize, Reconfigure
- > **Blank slate of logic gates that can be repurposed a workload at a time**
  - >> Economics resulting from Moore's Law finally enabling FPGAs to exist in the datacenter in high volume
- > **Exploding data to process while reducing energy and driving down processing latency**
- > **AI an emerging tool to optimize and automate reconfigurability in the near future**

# Moore's Law and Manhattan Island

- > **Limits and economics drive reconfiguration**
  - >> Manhattan from 1600-1900: Room to develop
  - >> Silicon from 1965-2020: Room to shrink transistors
  - >> Limits: Land mass / Transistor size and power envelope
- > **Visualizing Reconfiguration**
- > **FPGAs in the datacenter become canvas for accelerated reconfigurable processing power**



Diane Pham, 6SQFT.com

# FPGAs in Cloud Computing

- > The new way to extend a “Moore-like” law
- > Compute node growth rate \* number of transistors @ sub 10nm
- > Datacenters become a new substrate



# Where do we start with FPGAs in Cloud?

The Nimbix Cloud and  
JARVICE™



# Democratize Access for FPGA App Development

- > Access tools in a browser from anywhere on any device
- > Access to FPGA Hardware from anywhere on any device

The screenshot shows the NIMBIX web interface. At the top, there is a navigation bar with a hamburger menu icon, the word 'Compute', and the NIMBIX logo. Below this is a search bar containing the text 'Xilinx'. Underneath the search bar, the text 'Search For "Xilinx"' is displayed. The search results are presented in a grid of eight cards. Each card contains the product name, the Xilinx logo, and the starting price per hour. The two cards in the top row are highlighted with a red border. The first highlighted card is 'Xilinx SDAccel Development 2017.4\_op' for \$1.50/hr. The second highlighted card is 'Xilinx SDAccel Development 2018.2' for \$1.25/hr. The other cards include 'Xilinx GEMX' (\$3.00/hr), 'Xilinx ML Suite' (\$3.00/hr), 'Xilinx SDAccel Runtime 2017.1' (\$3.00/hr), 'Xilinx SDAccel Runtime 2017.4\_op' (\$3.00/hr), 'Xilinx SDAccel Runtime 2018.2' (\$3.00/hr), and 'Xilinx WEBP' (\$3.00/hr).

Product Name	Price (from)
Xilinx GEMX	\$3.00/hr
Xilinx ML Suite	\$3.00/hr
Xilinx SDAccel Development 2017.4_op	\$1.50/hr
Xilinx SDAccel Development 2018.2	\$1.25/hr
Xilinx SDAccel Runtime 2017.1	\$3.00/hr
Xilinx SDAccel Runtime 2017.4_op	\$3.00/hr
Xilinx SDAccel Runtime 2018.2	\$3.00/hr
Xilinx WEBP	\$3.00/hr



# Build and compile your designs in the Nimbix Cloud

The image displays the Nimbix Cloud interface, which is used for managing and executing FPGA designs in the cloud. The interface is divided into several sections:

- Dashboard:** Shows the user's profile (Steve Hebert) and navigation options like Compute, Dashboard, and PushToCompute™.
- Jobs:** A sidebar menu with options for Current, History, Reports, and Team Jobs.
- Current Job:** Displays details for a job named "Xilinx SDAccel Development 2018.2(439626)". It shows the command "server", status "Processing", address "NAE-165-254-189-72.jarvice.com", and a password field. A timer indicates the job has been running for 00:01 on 1 node.
- Stats:** A summary section showing "JOBS RUN" as 11, a "COMPUTE" cost indicator, and a total time of 1056:01:30.
- Project Window:** A window titled "Projects - SDx - Acceleration/project.sdx - Xilinx SDx" showing the project structure in the Project Explorer. A "Build Project" dialog is open, indicating the creation of a resource and the invocation of the command "make -j20 incremental".
- Assistant:** A panel showing the project's configuration, including "Acceleration [OpenCL]" and "Emulation-SW [Software Emulation]".
- Console:** A "Problems" and "Console" panel showing the build progress, including the message "SDx Build Console [Acceleration, System] INFO: [XOCC 60-423] Target device: xilinx\_vcu1525\_dy INFO: [XOCC 60-242] Creating kernel: 'openc1\_sw\_maxscoc'".



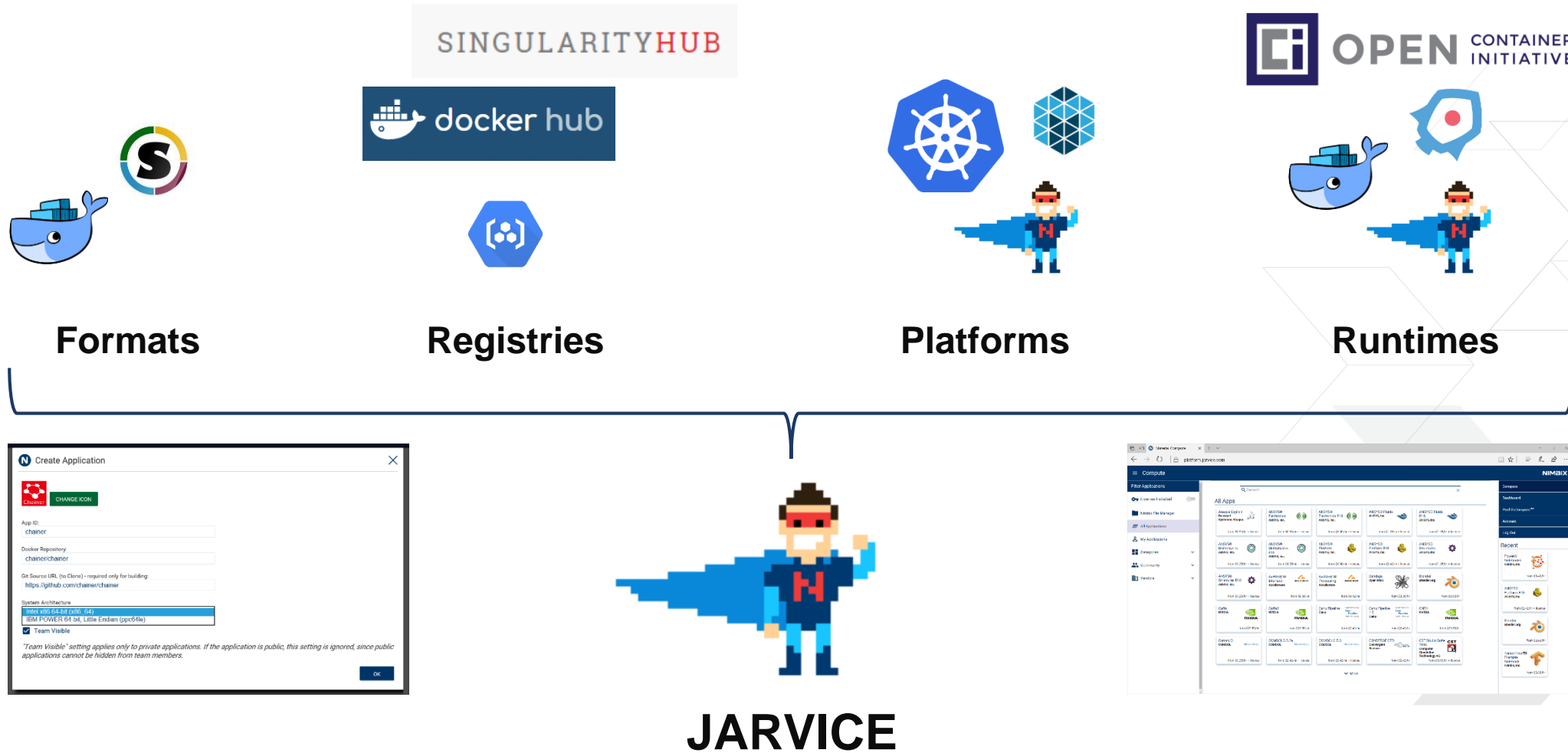
# Deploy FPGA Applications on a Scalable Platform

The screenshot displays the PushToCompute™ web interface. On the left, there is a sidebar with a 'PushToCompute™' header, a 'DOCKER Registry' section with login options (USE PASSWORD, USE JSON KEY), and a 'Filter Applications' section with a 'License Included' toggle and 'All Applications' selected. The main area shows 'All Apps' with a search bar and a grid of application cards. One card for 'caffe\_app' by 'hebert' has a context menu open with options: Edit, Delete, Build, Build+Pull, Abort Build, Pull, History, and Download AppDef. A 'Create Application' dialog box is overlaid on the right, containing fields for App ID, Docker or Singularity Repository, Git Source URL, System Architecture (Intel x86 64-bit), and a 'Team Visible' checkbox. A note at the bottom of the dialog explains the 'Team Visible' setting.

> <https://www.nimbix.net/pushtocompute/>

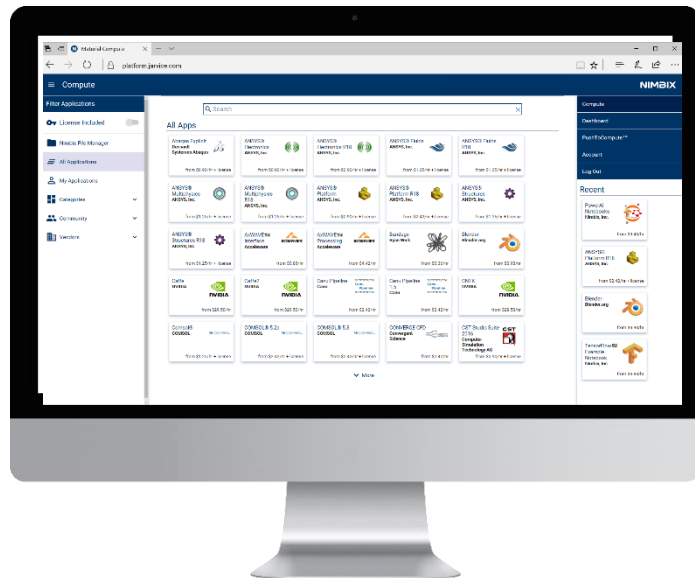
> <https://www.nimbix.net/pushtocompute-work-flow-deployment-guide/>

# JARVICE with PushToCompute: Container PaaS to Distribute & Scale FPGA Applications Globally



# Monetize and Run Your FPGA Accelerated Apps

- > Add FPGA runtimes to private team or the public Nimbix Cloud Application Marketplace
- > Leverage JARVICE API to launch apps and FPGA workloads or integrate into your own web service



## Deep Learning

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

## Simulations

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

## NGS Sequencing

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

# Get Started with FPGA Cloud Computing Trial on Nimbix Cloud

- > 100 Hours of free compute time!
- > Use for developing an application using SDAccel or trying out accelerated run-time applications from ISV Partners



**FPGA Developer Trial**

Get started with **FPGA** accelerated applications today on the Nimbix Cloud

Nimbix has partnered with Xilinx to provide developers and engineers a trial account that provides up to 100 hours of free time on the Nimbix Cloud using Xilinx FPGA Tools and Accelerators

Register for access to the FPGA Developer Program to get started with **FPGA accelerated** applications today. Applicant must be part of a valid company or university. Free trial time is limited to FPGA enabled resources and Xilinx tools and applications. A valid credit card is required for access to non-FPGA resources.

# Industry-Leading Cloud FPGA Infrastructure

## > Introducing all new Nimbix FPGA-accelerated compute nodes featuring Xilinx Alveo:

### >> NX5u

Dual Intel Xeon CPU, 128GB Host Memory and Xilinx Alveo U200 FPGA Accelerator

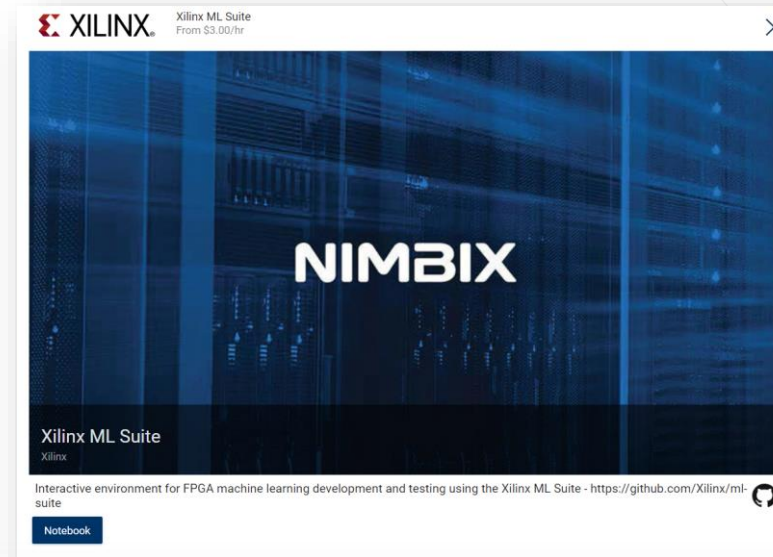
### >> NX6u

Dual Intel Xeon CPU, 128GB Host Memory and Xilinx Alveo U250 FPGA Accelerator



# Xilinx ML Suite on the Nimbix Cloud

- > Turnkey deployment of Xilinx optimized machine learning tools running on all new FPGA accelerators
- > Click-to-launch interactive Jupyter Notebook
- > Ready to run examples with support for major frameworks including: Caffe, TensorFlow, Keras, MXNet and Darknet



## Adaptable

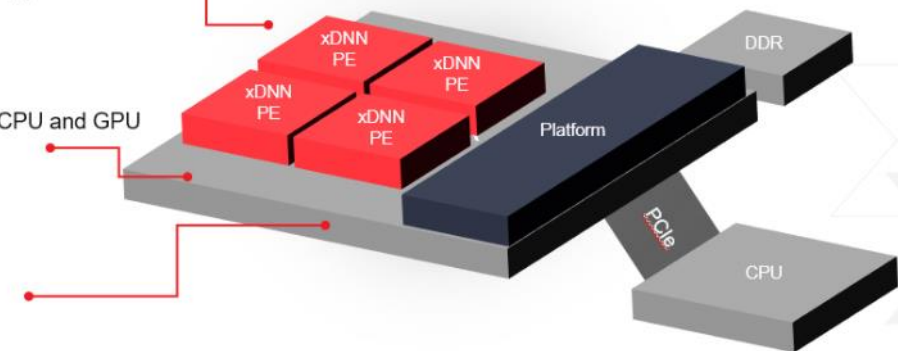
- > AI algorithms are changing rapidly
- > Adjacent acceleration opportunities

## Realtime

- > 10x Low latency than CPU and GPU
- > Data flow processing

## Efficient

- > Performance/watt
- > Low Power



# Summary

## FPGAs in the Datacenter

- > Hyperscale Reconfigurable Acceleration and Next Generation FPGA App Development Begins

## FPGA App Development Has Never Been Easier

- > Click and profile code in SDAccel in a browser in seconds

## FPGA App Deployments Scale Globally

- > The Nimbix Cloud and JARVICE provide a platform to deliver accelerated applications and workflows, anywhere

Adaptable.  
Intelligent.

**NIMBIX**  
*supercomputing made super human™*

