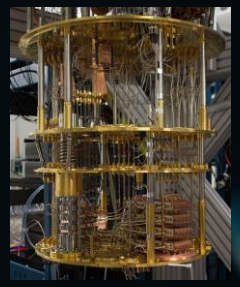
Nejvýkonnější SuperPOČÍTAČe dneška a budoucnosti od IBM

Otevřené inovace IBM pro EXASCALE systémy a kvantové počítače – CORAL/ORNL Summit, IBM Q



Radek Špimr, IBM

LinuxDays Praha, 2018

Agenda

- **IBM AI/DL/ML Policy**
- DoE CORAL EXAscale projekt – dnes, zítra
- **EuroHPC Program**
- Historie HPC od PetaFlopsu po EXAFlopsy
- **Svět HPC dnes s velkými daty a umělou inteligencí**
- Válka o superpočítačovou dominanci ve světě, TOP500
- **Server IBM POWER AC922 - základ nejvýkonnějšího superpočítače na světě**
- CORAL Summit – nejvýkonnější a ***nejinteligentnější*** počítač na světě
- **IBM a OpenPOWER inovace aneb úspěšně vzdorujeme Moorovu zákonu**

- ***Kvantový ekosystém IBM*** *a výhled do budoucna - IBM Experience a IBM Q*

IBM Policy for Trust and Transparency

THE PURPOSE OF AI IS TO AUGMENT HUMAN INTELLIGENCE

The purpose of AI and cognitive systems developed and applied by IBM is to augment – not replace – human intelligence.

Our technology is and will be designed to enhance and extend human capability and potential. At IBM, we believe AI should make ALL of us better at our jobs, and that the benefits of the AI era should touch the many, not just the elite few. To that end, we are investing in initiatives to help the global workforce gain the skills needed to work in partnership with these technologies.

NEW TECHNOLOGY, INCLUDING AI SYSTEMS, MUST BE TRANSPARENT AND EXPLAINABLE

For the public to trust AI, it must be transparent. Technology companies must be clear about who trains their AI systems, what data was used in that training and, most importantly, what went into their algorithm's recommendations. If we are to use AI to help make important decisions, it must be explainable.

DATA AND INSIGHTS BELONG TO THEIR CREATOR

IBM clients' data is their data, and their insights are their insights. Client data and the insights produced on IBM's cloud or from IBM's AI are owned by IBM's clients. We believe that government data policies should be fair and equitable and prioritize openness.

Od Peta k ExaFLOPSům

Supercomputing at IBM

1954
The Naval Ordnance Research Calculator helped forecast weather and performed other complex calculations.

1961
The **IBM 7030** was capable of 2 million operations per second.

1966
The **IBM 360** and its successors helped power NASA's Apollo program.

1997
Deep Blue wins its match with chess grandmaster Garry Kasparov.

2004
Blue Gene ushers in a new era of high-performance computing as it helps biologists explore gene development.

2008
Built for Los Alamos National Laboratory, **Roadrunner** is the first supercomputer in the world to reach petaflop speed.

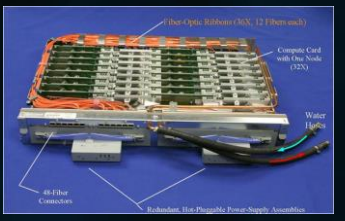
2011
Watson beats human competitors on *Jeopardy!*, earning a million-dollar jackpot for charity.

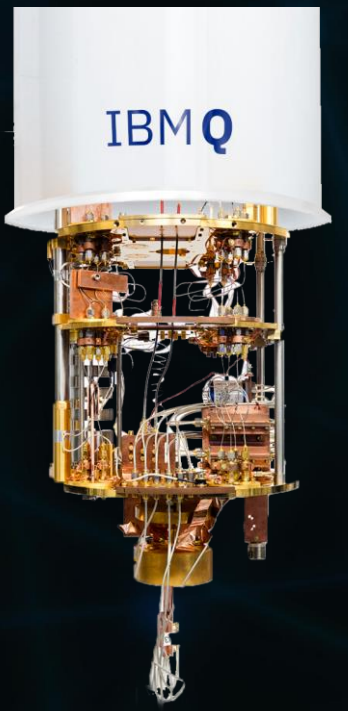
2012
Sequoia, the third-generation **Blue Gene** system, reaches speeds of 16.32 petaflops.

2018
Summit begins work at Oak Ridge National Laboratory; a sister machine, **Sierra,** launches at Lawrence Livermore National Laboratory.

ibm.com/summit

IBM



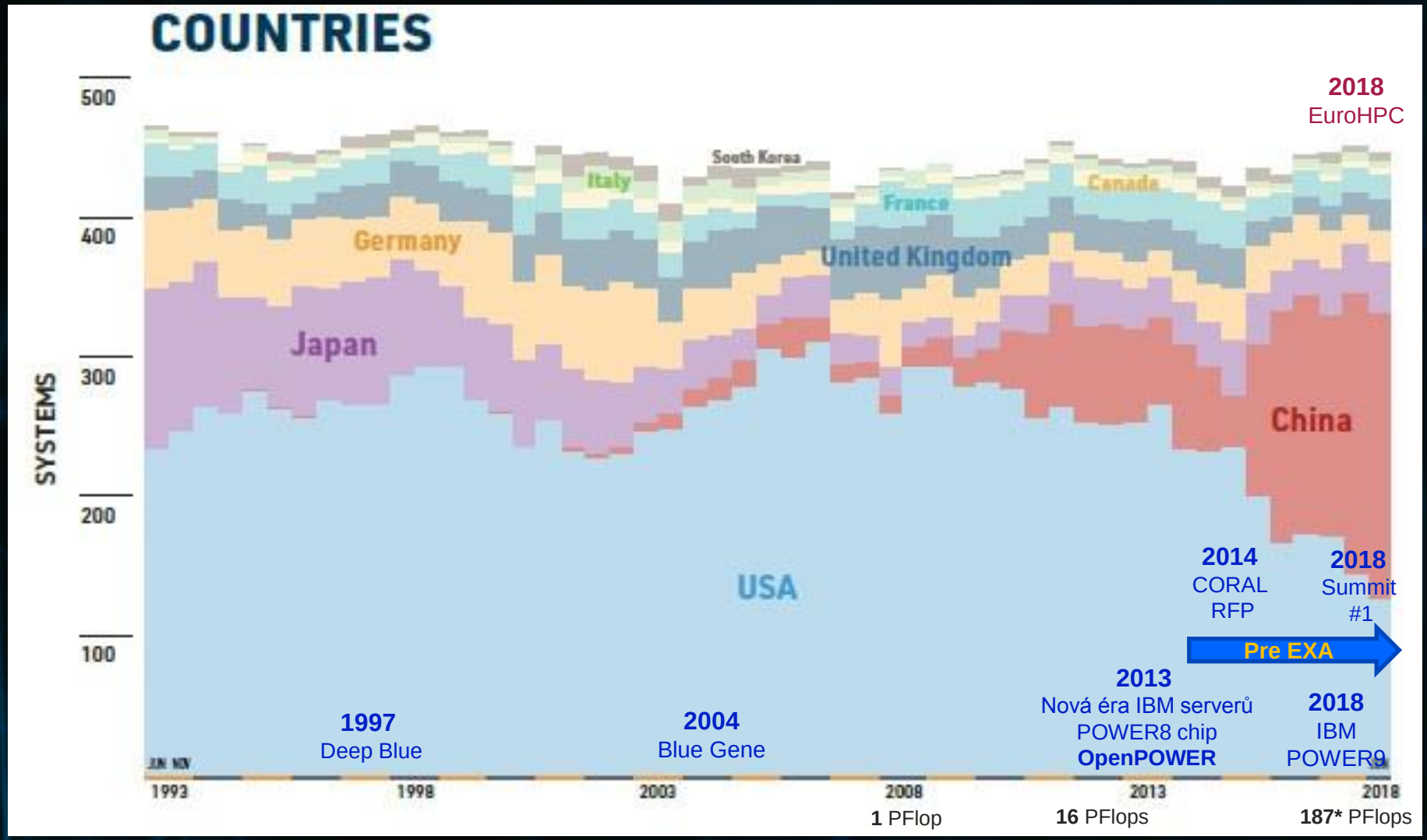


Art of SuperComputing War (TOP500.org)

Pre exaSCALE

TOP500 10/2018
IBM podíl
20% výkon
1% instalace

Komerční systémy TOP500
Pouze IBM
Power v TOP5



ExaScale Computing Future – CORAL, Aurora

Projekt : ExaScale HPC Systems **RFP 2018**
Rozpočet : cca. 1,8B USD
Dodávka : 2 -3 ExaScale Computers
Cíl : 50-100x performance of current TOPs

First Exascale Authority CORAL (ORNL, LLNL, ANL)
Účel/zaměření : AI/ML, simulace, HPC.
Příklady :

- Next GEN materials
- High-energy physics data
- Industrial product design reducing cost to market
- Options of nuclear security
- https://aurora.alcf.anl.gov/

USA/DoE/ORNL, Superpočítač Summit : #1

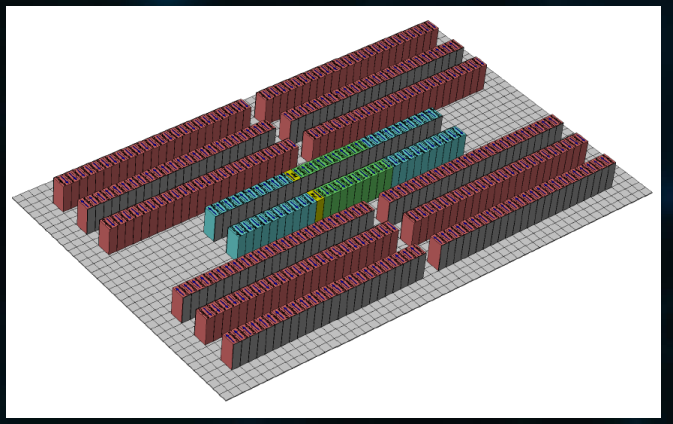
Nejvýkonnější a *nejchytřejší* superpočítač na světě (10/18) : 5-10x výkon Titan
>4600 uzlů
Spolupráce IBM, Nvidia, RedHat, Mellanox, ORNL
CAAR – Center of Accelerated Application

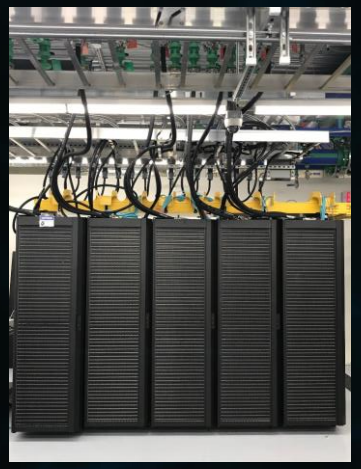
Hybridní akcelerované uzly IBM Power AC922

- **2x POWER9/6x Nvidia V100** (**10PB MEM** Unified NVLink Memory)
- Interconnect **: NVLink** (MEM/CPU/GPU, MEM/GPU/GPU), Mellanox **EDR Ifiniband** (100Gb/s), CAPI
- Storage : **250PB IBM Spectrum Scale** filesystem **Alpine** (2,5TB/s sIO, 2,2TB/s rIO)
- Burst Buffer Cache : **NVMe storage**
- Compilers : IBM XL, PGI, LLVM, GCC, NVIDIA CUDA Stack
- Perf. tools : MAP, Open|SpeedShop, TAU, HPCToolkit, VAMPIR/Score-P, Parallel Performance Toolkit, nvprof, gprof, OpenMP, ...
- Peak* power consumptions : 15MW

OAK RIDGE National Laboratory
IBM
Mellanox TECHNOLOGIES
NVIDIA
Lawrence Livermore National Laboratory

IBM Summit – Datacentrický přístup

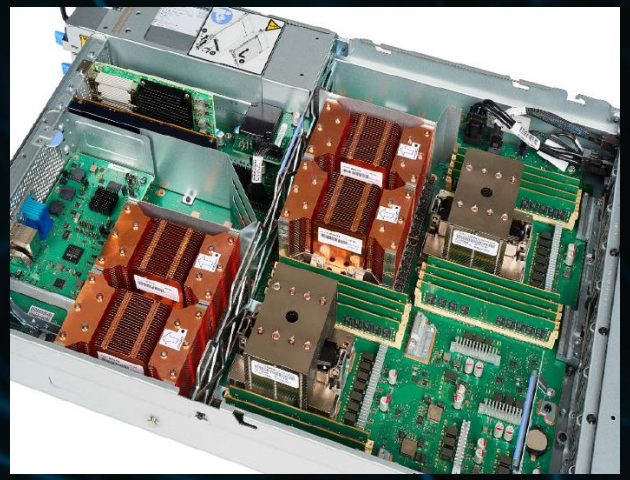


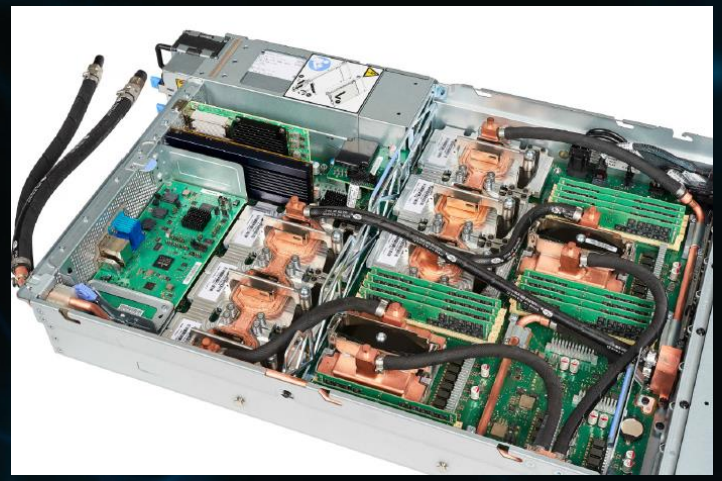


IBM Power System AC922 server

OpenPOWER Linux Scale-out server

- 2-Socket, 2U Packaging
- IBM POWER9 Turbo,
- Nvidia NVLink2+Volta Boost,
- PCIe Gen4, OpenCAPI3, NVMe





IBM POWER9

CPU : 2x 16/20c (max. 3,3GHz), 18/22c (max 3,45GHz)
GPU : 4-6x NVIDIA Tesla V100 SXM2 16/32GB (15,7TFLOPs SP/7,8TFLOPS DP)

- 5120 CUDA Cores
- 640 TensorCores

RAM : 2TB RAM DDR4 (16x 128GB)
Storage : 2x SSD/HDD (1-4TB)
IO storage adapters : NVMe 1,6/6,4TB
IO : LAN/EDR/FC (4x PCIe4/3x CAPI)
OS : Rhel 7le, Ubuntu 18
FW: OpenPOWER FW, EnergyScale
Packaging : 30kg, 2U

**Minimum Features Rule*

Varianty chipu IBM POWER9

SMP scalability / Memory subsystem

Up to 24 Cores / Chip / socket
max of 4 threads per core
Linux Ecosystem Optimized

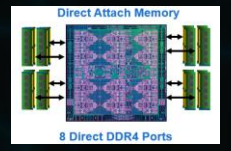
Up to 12 Cores / Chip / socket
max of 8 threads per core
PowerVM Ecosystem Continuity

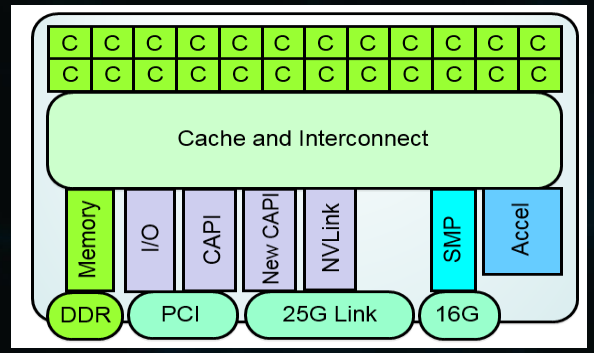
Scale-Out – 2 Socket Optimized

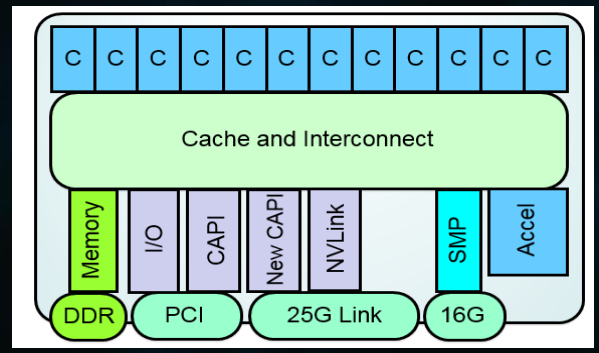
Robust 2 socket SMP system

Direct Memory Attach

- Up to 8 DDR4 ports
- Commodity packaging







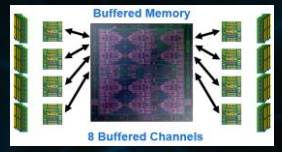
Scale-Up – Multi-Socket Optimized

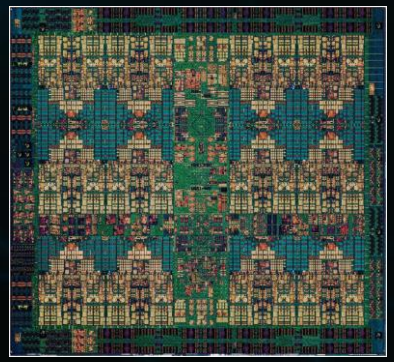
Scalable System Topology / Capacity

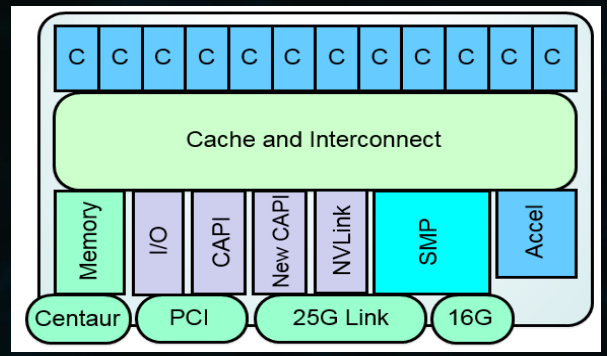
- Large multi-socket

Buffered Memory Attach

- 8 Buffered channels







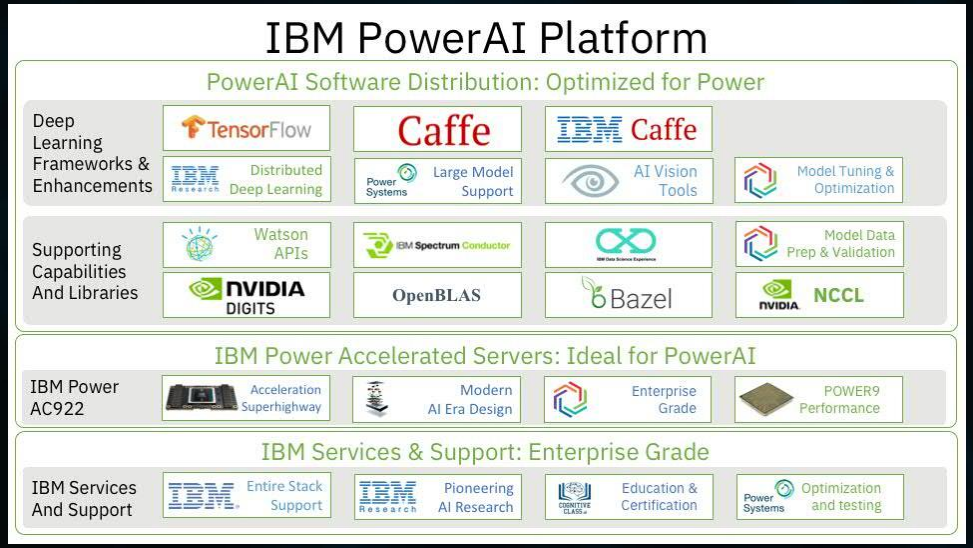
IBM PowerAI/Vision software

PowerAI použití

Computer Vision – obj. detection, classification, ..
Pattern Recognition/Anomaly Detection – img/vid. Series, data trends, ...
Human Computing Interface – language processing, sentiment analysis, intelligent agents.

AI Framework pro rychlou adopci AI a zpracování velkých dat

Up&Run – templates, preloads, preprocessing.
Time Reduction – 10x rychlejší příprava dat (CPU/GPU tuning, hyperparameters, LMS, HiRes)
Fast Training – DDL až 256GPU, 58x rychlejší,
Auto optimization – hyper parameters, data processing design and processing, ..

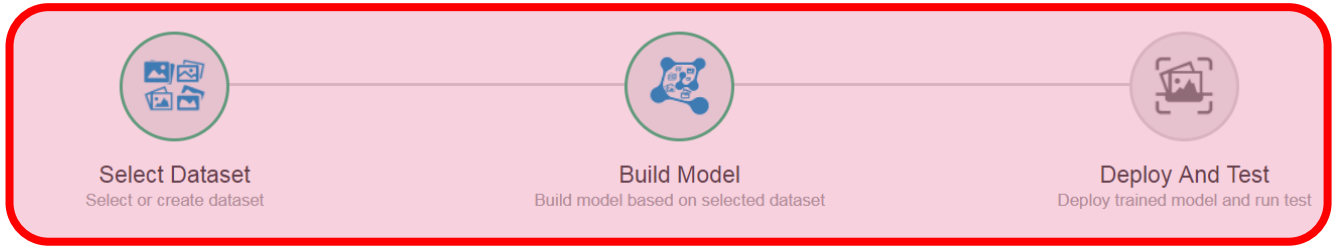


Result:

Result:

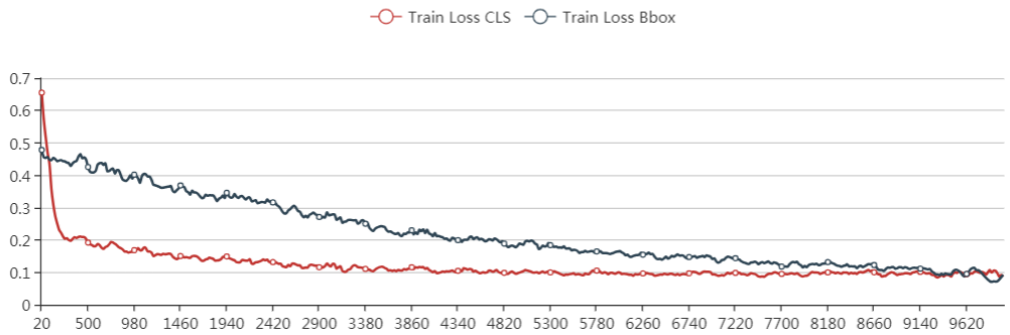
Result:

Task Status - def_det_custom



Latest Status: deployed

Total Iteration: 10000
Train Iteration: 10000
Train Loss CLS: 0.08680
Train Loss BBox: 0.09202



Result:

DataSet Quality Impact : Obj. Detection

Data 100%, Accuracy 82%, T:90

Result: No valid object detected!

Data 50%, **Accuracy 100%,** T:90

Result: No valid object detected!

This Ecosystem of Innovators

Creates True Differentiation in Performance and Cost

Growing Ecosystem Of **OpenPOWER Servers**

Micron, Rambus, BROADCOM, HGST, IBM, POWERCORE, redhat, ubuntu Supported by Canonical, kinetica, redislabs, XILINX, Nallatech, SAMSUNG, Microsemi, Mellanox TECHNOLOGIES, ALPHA DATA, NVIDIA, NEC

Growing Ecosystem Of **OpenPOWER Innovation**

Cirrascale, 中太服务器 ZOOM SERVER, Neu Cloud Oriental, NIMBIX, inspur 浪潮, TYAN, HITACHI Inspire the Next, TL 英特力, PENGUIN COMPUTING, YADRO, E4 COMPUTER ENGINEERING, msi, SUPERMICRO, Mark III Systems, RTDS Technologies, IBM, Inventec, wistron

Ecosystem Driven **Customer Choice**

This is What A Revolution Looks Like

Implementation / HPC / Research

Software

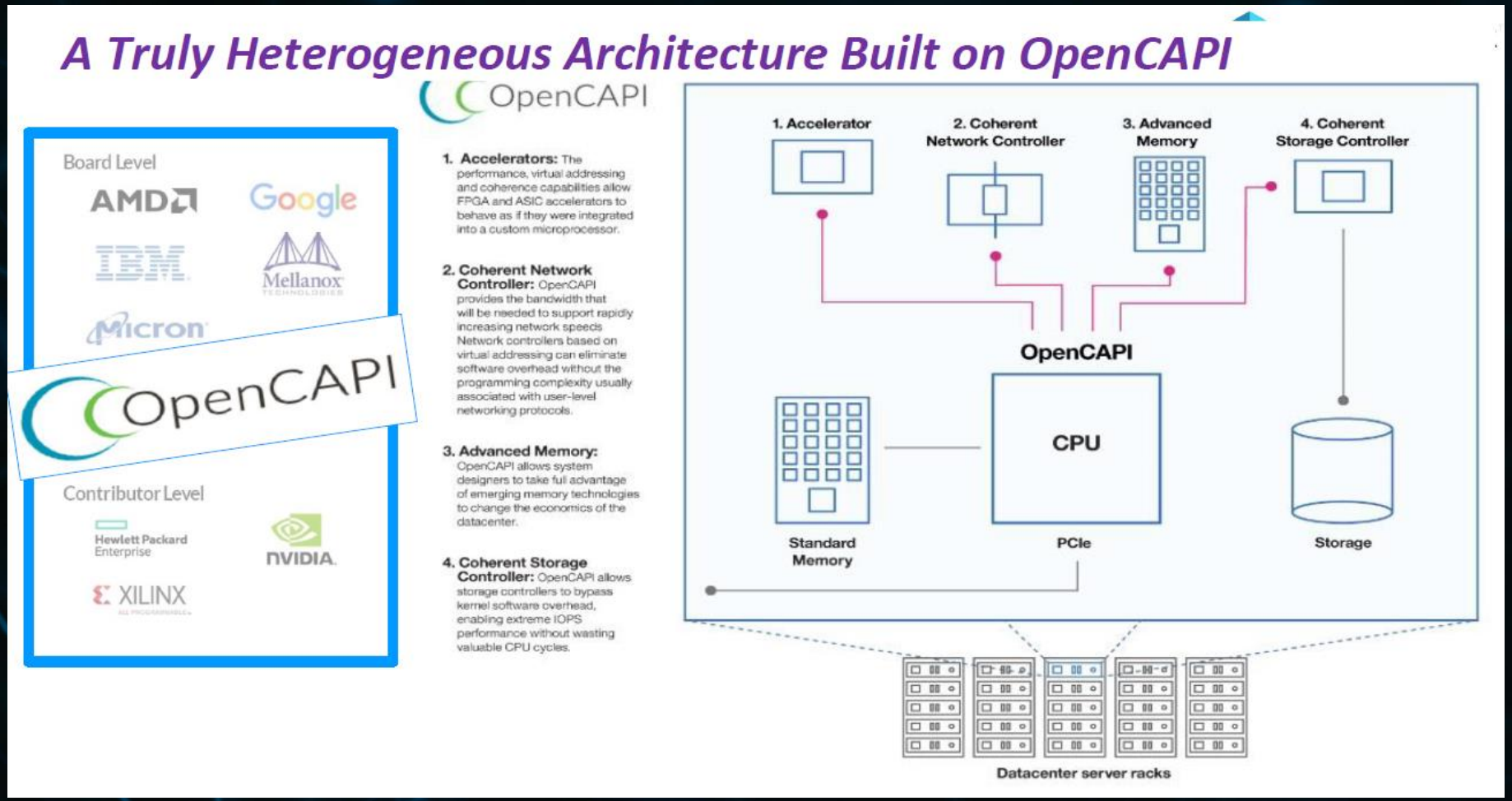
System / Integration

I/O / Storage / Acceleration

Boards / Systems

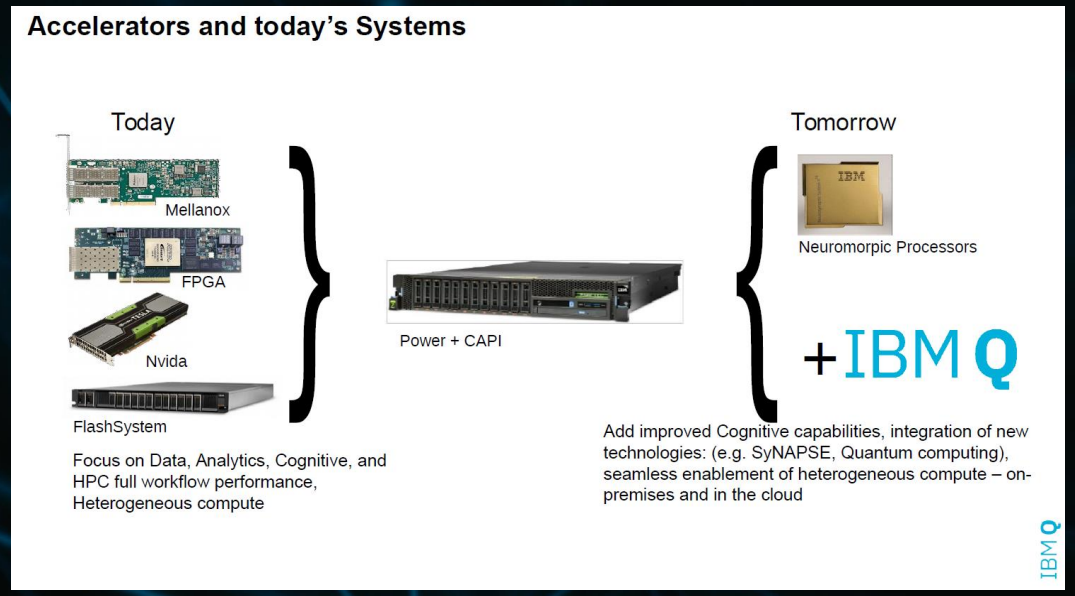
Chip / SOC

Budoucnost HW akcelerace HPC a serverů

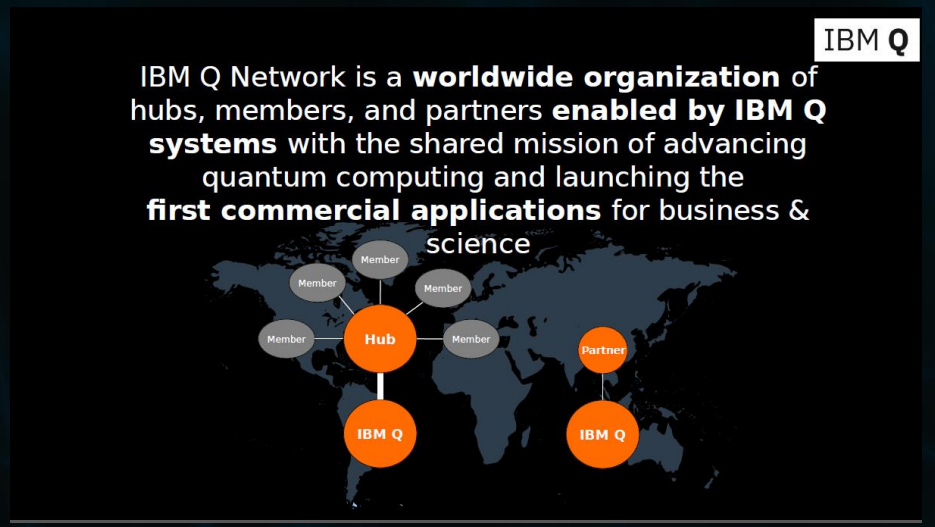


Přirozené rozšíření HPC služeb

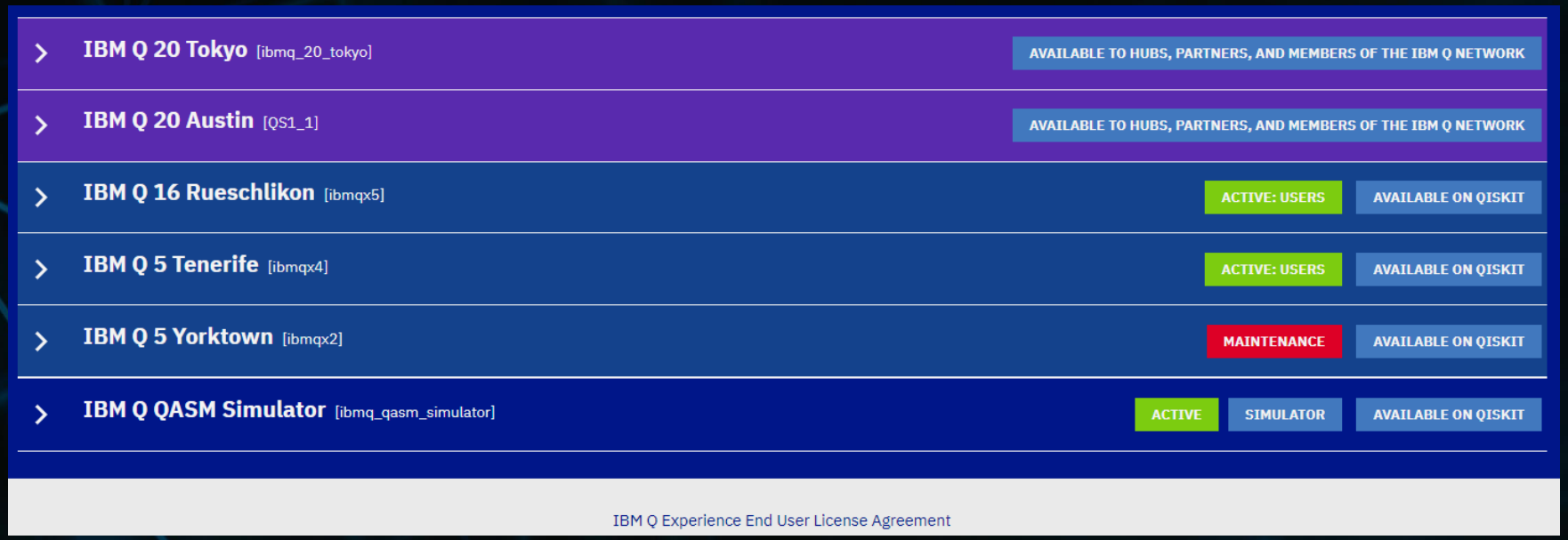
IBM Q is going to be a new kind of accelerators to the recent commercial HPC accelerators like GPUs, FPGAs, Neuromorphic processor (by IBM)

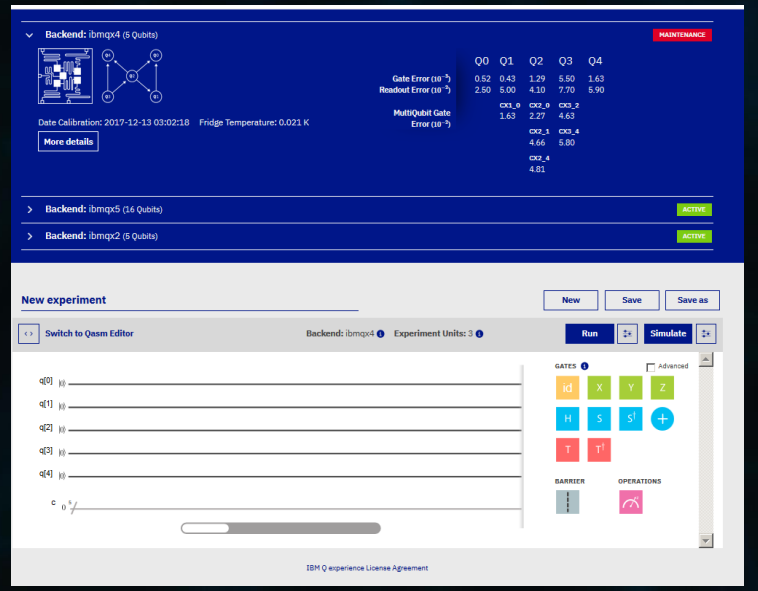


IBM Q services will be available as a part of HPC services by HPC providers to extend their current HPC services portfolio.



Aktivní kvantová zařízení IBM Q





IBM Quantum Experience

The World's First Public Quantum Computer and Developer Ecosystem

IBM QX Features

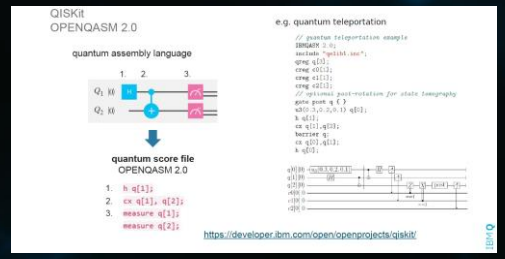
- Tutorial
- Simulation
- Graphical programming
- QASM language
- API (SDK coming soon)
- Active user community

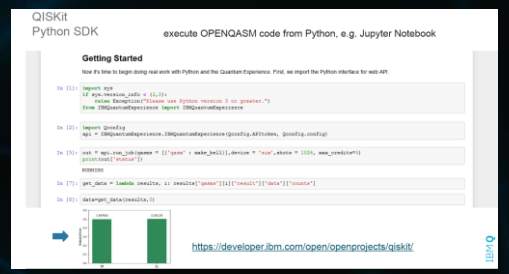
Experience quantum computing here:
www.research.ibm.com/ibm-q/qx/

Since launch

- > 60,000 users
- > 1,7 M executions
- 35 external publications
- > 10 professors using for quantum education
- 343 major media articles

IBM Q



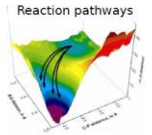


Applications and use cases

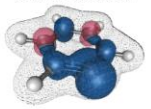
Initial applications will leverage algorithms that can tolerate or mitigate errors found in approximate quantum computers. Research & development for commercial use cases must be focused on selecting algorithms and determining how to best map problems to them.

Quantum Chemistry

Reaction pathways



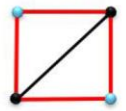
Molecule geometry



Optimization

Traveling Salesman

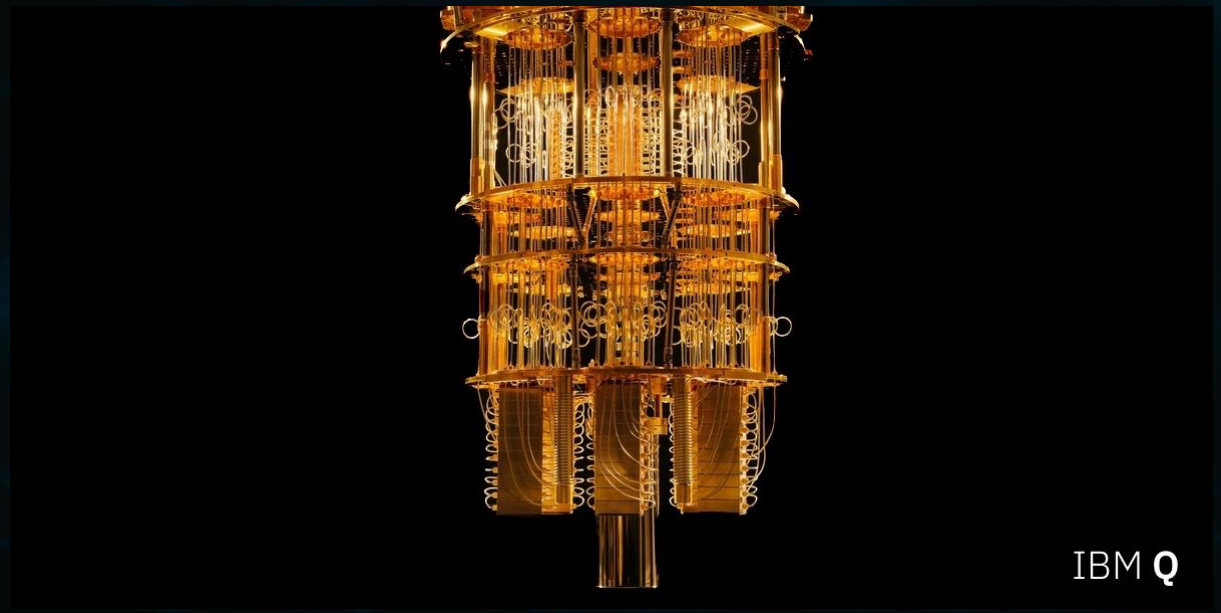
Max Cut



IBM Q

Dotkněte se kvantového počítače

Workshop na téma spolupráce v oblasti kvantových výpočtů v IBM ČR, Praha
Listopad 2018
Vývojáři z IBM Q laboratoří
radek_spimr@cz.ibm.com



Děkuji

radek_spimr@cz.ibm.com

Radek Špimr IBM (LinkedIn)