SCIENTIFIC & PRIVATE CLOUD INFRASTRUCTURES

Albert van Eck

Director: eResearch & HPC

T: +27(0)51 401 9111 | info@ufs.ac.za | www.ufs.ac.za

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

UFS UV

OUTLINE

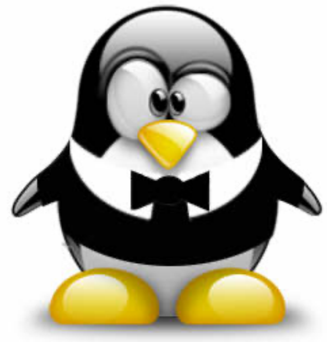
- Introduction
- Infrastructure
  - Hardware concepts, the importance of certain components
- Installing a GNU Linux system
  - Installation procedure installing Rocky Linux
- TCP/IP concepts and Firewalling
  - Services, configuring and debugging
  - Domain Services such as DNS, DHCP, NTP, NAT, LDAP
- Networking and network segmentation using VLANs
- Distributed File Systems
- Hypervisors and Cloud Infrastructure installation

INTRODUCTION

WHAT INDIVIDUALS SHOULD KNOW

- Basic understanding of IP networking
  - (DNS, LDAP, DHCP, IPv4, UDP/TCP, Gateway, Firewall, Switches)
- Comfortable using the terminal/console/command line
- Intermediate knowledge of Linux (Using bash)
- Use basic text editors (vi, nano, emacs)
- Able to understand and write bash scripts
- Understand environment variables
- Understand and use of Makefiles & Compilers
- Some debugging techniques

THIS PRESENTATION



- If you see a Tux, it means there is an activity that you should perform on your own at a later stage
- Most images were downloaded from the Internet (like Tux) and could have copyright disclosures
- Although most information contained in these documents are accurate, somethings are open to interpretation (simplified for better understanding) and it is good to form your own opinion

CONTACT

- If you have specific questions, you are welcome to contact me
  - But please, try and get to a solution on your own first, it is good practice

- This is an open session, and you are welcome to ask questions and even point out if something is incorrect

- You may feel overwhelmed but keep pushing yourself.
  - Learning new skills takes some time
  - The author of this presentation has over 20 years of GNU Linux experience and still learns several new things every week or so
  - Just take it as it comes and practice on your own, you'll be amazed what you will accomplish with some practice

ENCOURAGEMENT

- We encourage critical- and logical thinking
  - Always ask yourself why and if you don't know why, find out....don't just simply take my or anyone’s word for it
  - There are several ways to accomplish a certain task in GNU Linux, we only demonstrate one or maybe even a few
    - You may find a simpler way that works better for you
- Engage in conversation with us or each other, even if it's controversial

INTRODUCTION TO HPC

UNIVERSITY OF THE
FREE STATE
UNIVERSITEIT VAN DIE
VRYSTAAT
YUNIVESITHI YA
FREISTATA

UFS
UV

HPC BACKGROUND

- High Performance Computing is the use of a computer system (usually a distributed system, working together) to solve a magnitude and diverse set of real-world problems

- The concept of HPC started in 1994 with Beowulf Clusters (Distributed Memory Computers)

- The UFS HPC currently serves researchers from numerous departments:
  - Animal & Plant Genetics, Bio-Informatics, Chemistry, Computer Sciences, Engineering, Human Genetics, Mathematics, Medical Microbiology, Medical Physics, Physics, Statistics and Virology…..

UFS
UV

WHY...JUST WHY

- As an IT professional, **why** care about HPC/Linux?

- Get exposure to the bleeding edge technology
  - NDA with Intel and other vendors
  - The Internet was developed by CERN in 1989 due to demands of experiments

- Trends
  - Cloud Computing, Virtualization, Big Data, Data mining, AI, Computer Rendering, Machine Learning, Block Chain technologies
    - All these services make use of concepts contained in these sessions, and almost all these technologies makes use of GNU Linux

- HPC shape the Datacenters of tomorrow
  - Building blocks for any datacenter

- Give yourself a competitive advantage, know the lower levels and how they function

WHAT TO CONSIDER, WHEN BUILDING YOUR OWN HPC

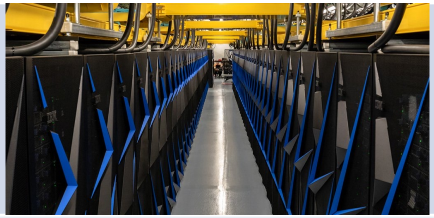
- Effectively solving problem sets in the shortest possible time, taking into consideration:
  - Cost
    - Hardware, Software Licenses, Power, UPS, Cooling etc.
  - Complexity/Ease of use
    - HR required to manage, user base, support
  - Resource Management
    - Focusing on hardware resources available
  - Storage
    - Requirements & Speed

TECHNICAL SIDE OF THE DESIGN

- Assuming: You can afford to, and need to build your own HPC
  - Cost
    - Hardware, Software Licenses, Power, UPS, Cooling etc.
  - Complexity/Ease of use
    - HR required to manage, user base, support
  - Resource Management
    - Focusing on hardware resources available
  - Storage
    - Requirements & Speed

SUPER COMPUTERS

- From the past to present

	1987	1994	2019
			
Name	Cray 2	Beowulf	IBM Summit OLCF-4
Performance	~ 1.9 GFlop	1.1 GFlop	200 PFlop
CPU Speed	244 MHz	16 Cores Pentium MMX	9 216 x Power9 22-Core CPUs 27 648 x nVidia Tesla V100 GPUs Total: 2 414 592 Cores
Price	$32m	$50k	$325m

- A **petaflop** is 1,000 **teraflops**; or one quadrillion (10^{15}) floating-point operations per second
- An Intel Core i9-9900K CPU = $ 650.00 @ 84.33GFLOPS ~ 2 271 635
- More @ Top500.org

FLOPS – MEASURING UNIT

- FLOP is a measuring unit for compute performance
- It can be quarter(a bit rarer measurement)-FP8, half-FP16, **single-FP32**, or **double precision-FP64**
- For example, a computer that does 1 Petaflop in FP8 will do 500 TeraFLOPs in FP16, 250 TeraFLOPs in FP32 and 125 TeraFLOPs in FP64
- So, you need to determine what kind of precision is measured when determining power in FLOPs
- You can use some benchmarking tool such as BOINC to see how many FLOPs your PC can do
  - Windows, Mac, Linux or Android devices

HARDWARE COMPARED

https://pages.experts-exchange.com/processing-power-compared

UFS
UV

PBS JOB LIST EXAMPLE

ID	Username	SessID	NDS	TSK	Memory	Req Time	S	Elap Time
183787	cl	11143	1	60	480gb	1000:00:	R	833:55:35
185944	co	22786	2	32	800mb	500:00:0	R	315:26:40
185954	va	24119	1	1	--	600:00:0	R	315:09:09
185968	va	42387	1	1	--	600:00:0	R	315:09:07
185977	va	42776	1	1	--	600:00:0	R	315:09:07
185996	va	7132	1	60	500gb	300:00:0	R	246:30:07
186031	va	44393	1	60	--	480:00:0	R	220:53:15
186114	la	45839	4	1080	750mb	100000:0	R	148:14:54
186172	ad	171404	1	16	--	125:50:0	R	79:14:06
186175	co	174538	1	16	--	195:50:0	R	78:34:07
186176	co	174550	1	16	--	195:50:0	R	78:34:07
186177	co	174565	1	16	--	195:50:0	R	78:34:07
186178	co	174586	1	16	--	195:50:0	R	78:34:07
186180	co	174633	1	16	--	195:50:0	R	78:34:06
186181	co	174653	1	16	--	195:50:0	R	78:34:06
186182	co	174685	1	16	--	195:50:0	R	78:34:06
186183	co	174718	1	16	--	195:50:0	R	78:34:06
186185	co	174951	1	16	--	195:50:0	R	78:34:05
186186	co	175001	1	16	--	195:50:0	R	78:34:05
186187	co	175058	1	16	--	195:50:0	R	78:34:05
186188	co	44279	1	16	--	195:50:0	R	78:34:05
186191	co	176302	1	16	--	195:50:0	R	78:33:58
186192	co	44447	1	16	--	195:50:0	R	78:33:58
186260	ma	10182	1	12	96gb	48:00:00	R	33:27:04
186261	20	46180	1	16	128gb	48:00:00	R	31:53:00

STATISTICS FOR 2022

Indicator	Value
Number of jobs	46 560
CPU Hours	403y 172d 10:11
Estimated Amazon Pricing*	R 19 268 215.77
Storage Used (Homes)	131.09 TB
Storage Used (Scratch)	10.76 TB
Users	499
Scientific Software Packages (modules)	510
Total CPU Hours since 2010	4 737y 363d 09:19

* Only cost of CPU hours. This excludes:
    - Time of configuration, support, installation
    - Storage Costs
    - Data Transfer rates etc.

THEORETICAL PERFORMANCE

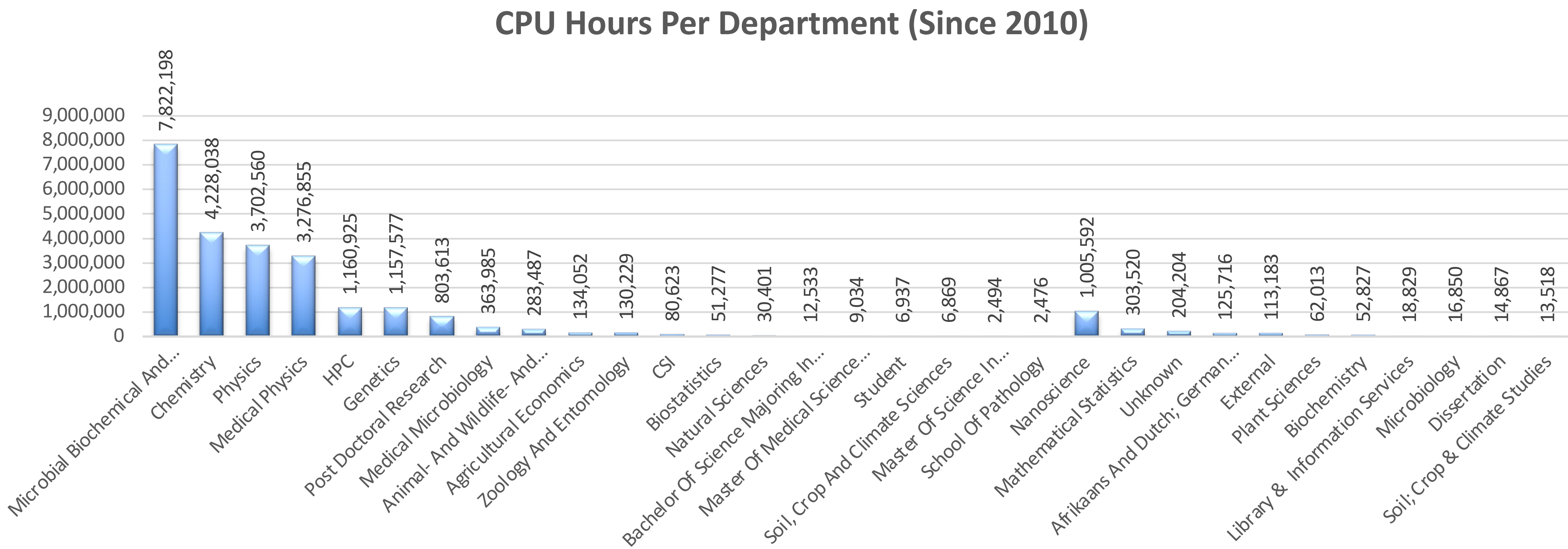
Hardware	# Nodes	Units Per Node	Tera FLOPS per Unit	Total TFLOPS
NVidia V100 Tesla Cards	2	4	7.000	56.000
Intel Xeon E5-2697A	18	2	0.750	27.000
Intel Xeon Phi 7250	8	1	2.662	21.296
Intel Xeon Phi 7290	8	1	2.995	23.960
Intel Xeon Gold 6142	2	2	1.147	9.176
Total Peak Performance				**137.432***

Equivalent to 333 Intel Core i7-8700K Laptops (412.5 GFLOPS ea.)

or

Half of the performance of the world's largest cluster in 2005 (BlueGene/L. 131 072 Cores)

STATISTICS SINCE 2010



HPC HARDWARE

THE OLD UFS GEN 3 & 4 CLUSTER

PAST: GEN 3 & GEN 4 AMD NODES

17 x Generation III Nodes (816 Cores)
AMD Opteron 6174 CPU 2.2GHz (48 Cores)
64 Gb 1333MHz DDR3 ECC Registered Memory
Mellanox ConnectX-2 VPI Dual Port InfiniBand QDR
2 x 1400W PDU

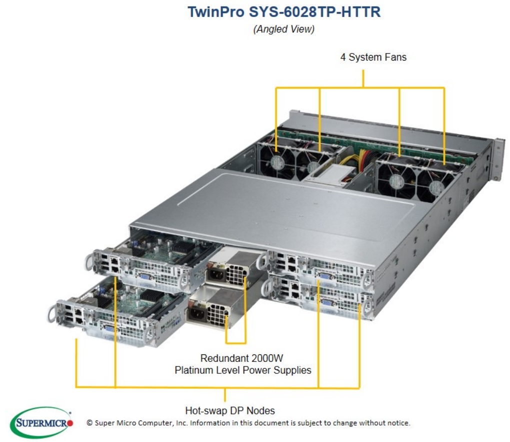
8 x Generation IV Nodes (512 Cores)
AMD Opteron 6274 CPU 2.2GHz (64 Cores)
128 Gb 1600MHz DDR3 ECC Registered Memory
Mellanox ConnectX-3 VPI Dual Port InfiniBand QDR
2 x 1400W PDU

THE UFS GEN 5 CLUSTER

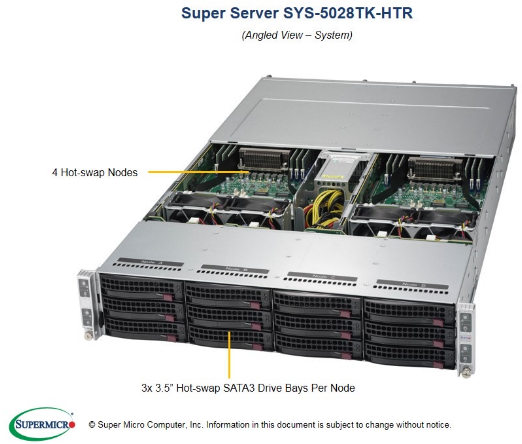
UFS
UV

DECOMMISSIONING: GEN 5 & KNL INTEL NODES

16 x Generation V Nodes (1 024 Cores)
Intel Xeon E5-2697A v4 2.6GHz (64 Cores)
512 Gb 2400MHz DDR4 ECC Registered Memory
Intel Omni-Path 100GB PCIE
2 x 2000W PDU



16 x Generation V Nodes (4 480 Core Threads)
Intel Xeon CPU 2.2GHz (72 Cores, 288 Threads)
256 Gb 2400MHz DDR4 ECC Registered Memory
Intel Omni-Path 100GB PCIE
2 x 2000W PDU



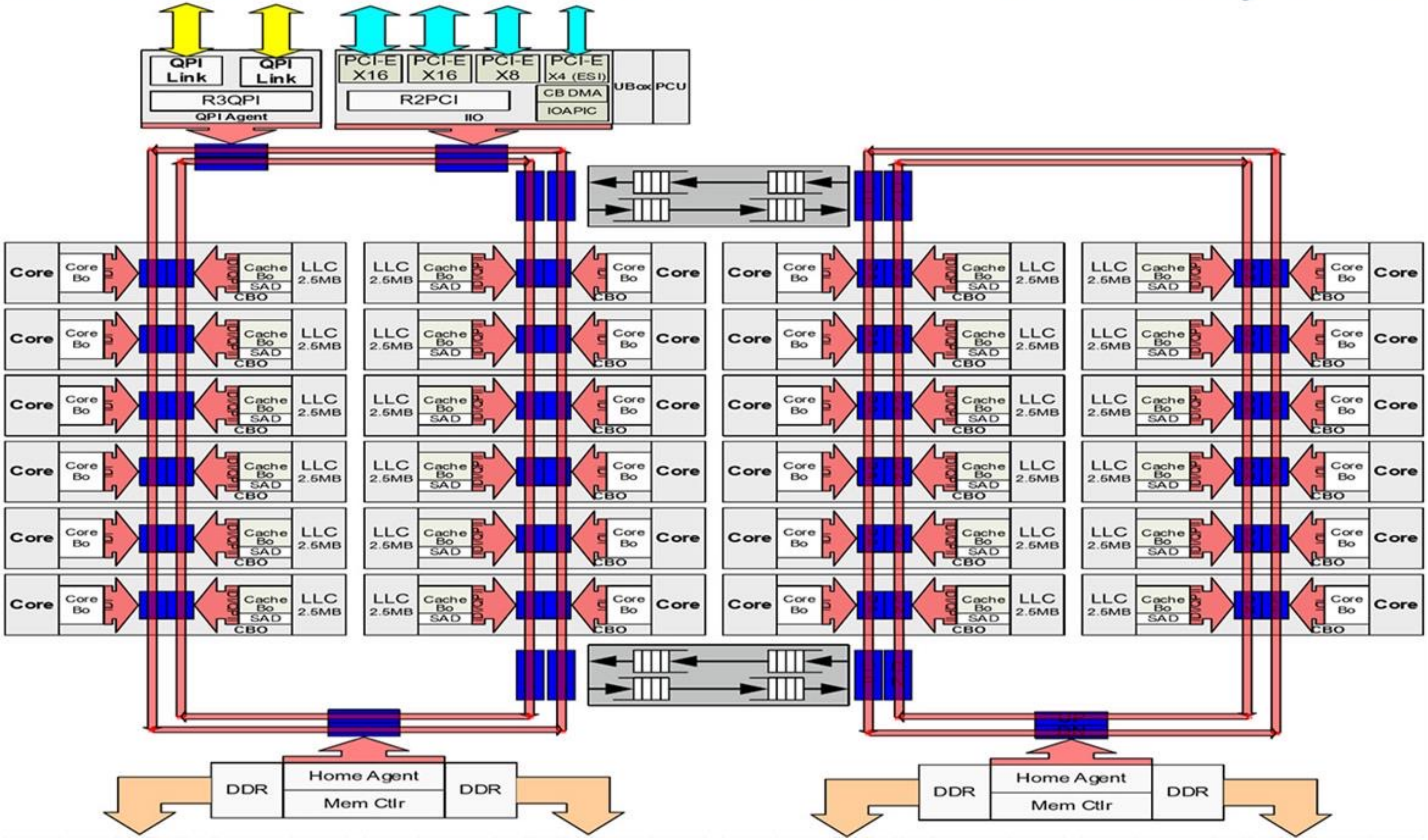
CPU

- We have all seen specs like this:

Description	Value
Clock Speed	2.60GHz
Number of Sockets	2
# Cores	16
L1 Cache	32K
L2 Cache	256K
L3 Cache	40 960K
Threads per Core	2
# Mem Channels	4

But what does this mean....is it good, is it bad, is it high is it low?

Intel® Xeon® Processor E5 v4 Product Family HCC

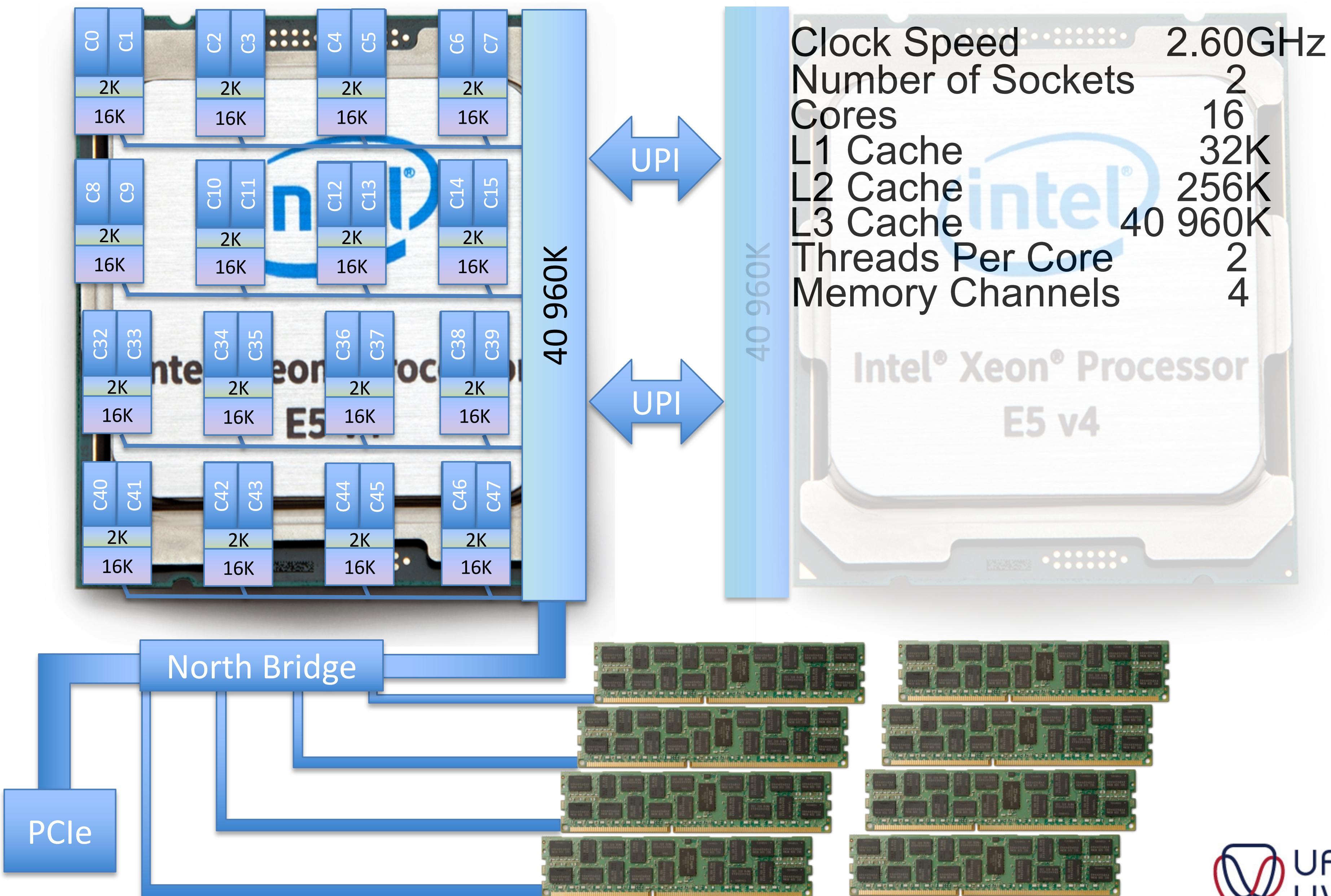


https://software.intel.com/en-us/articles/intel-xeon-processor-scalable-family-technical-overview

UFS UV

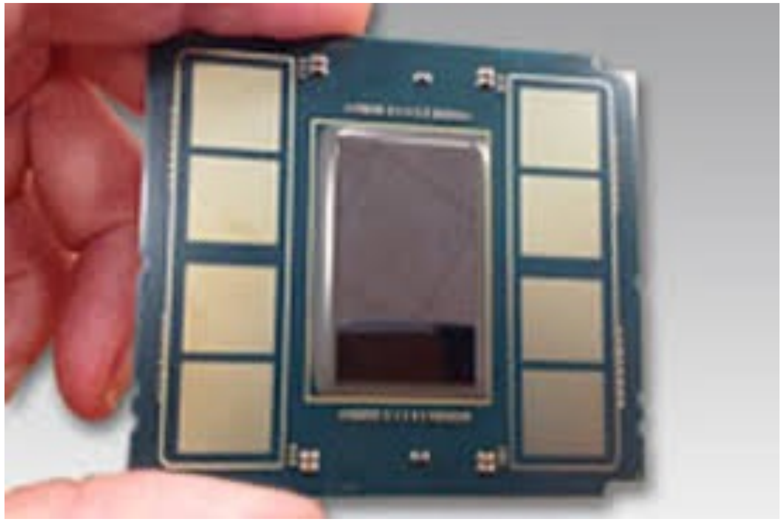
CPU ARCHITECTURE

*** Own representation, actual design differs

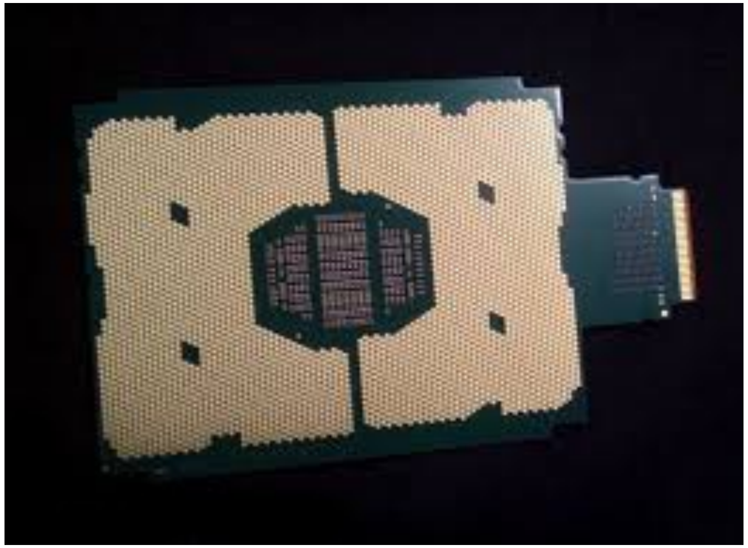


UFS UV

INTEL KNIGHT’S LANDING HOST PROCESSOR



Xeon Phi



Xeon Phi with Integrate Omni-Path

Also known as:

Xeon Phi or

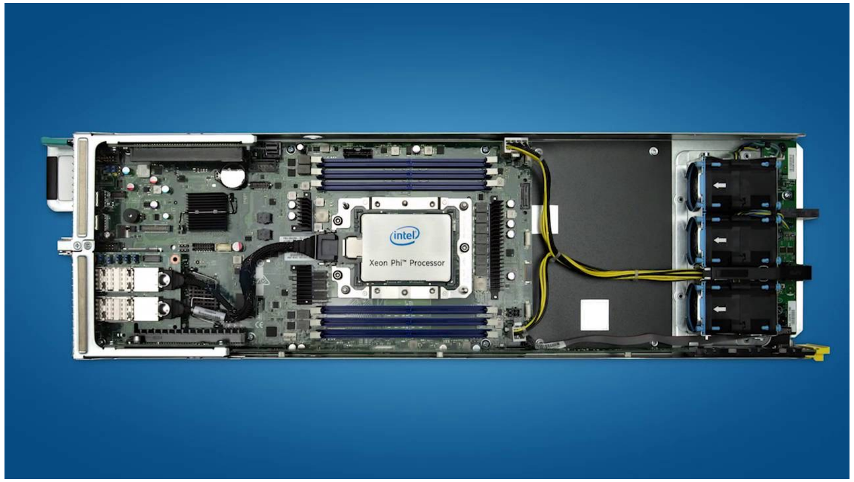
Xeon Many Integrated Cores (MIC)

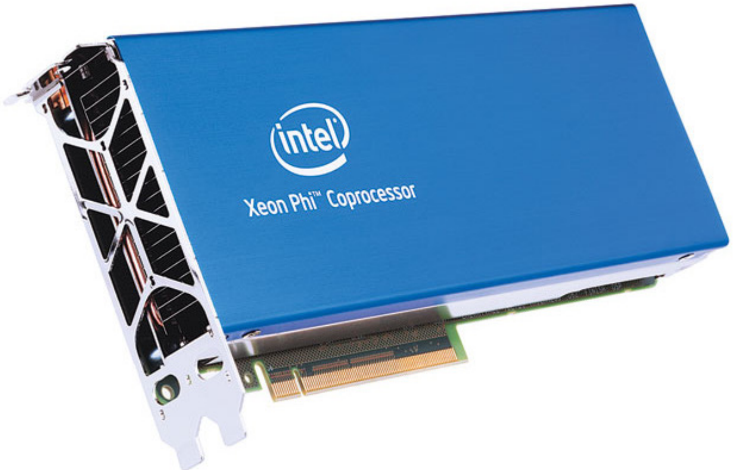
Based on the Intel Atom Processors (1.3 ~ 2.2 GHz)

64 to72 Cores, with 4 Threads each

Respectively providing between 256 and 288 Threads

INTEL KNIGHT'S LANDING COPROCESSOR





Like Host Processor but can add up to 8 coprocessors in a server

HARDWARE

Intel (Gen5) vs AMD (Gen6)

Description	Capacity
Clock Speed	2.60GHz
Number of Sockets	2
# Cores	16
L1 Cache	32Kb
L2 Cache	256Kb
L3 Cache	40 960Kb
Threads per Core	2
# Mem Channels	4
RAM (2600MHz) DDR 4 ECC Registered	512GB

Description	Capacity
Clock Speed	2.20GHz – 3.5GHz
Number of Sockets	2
# Cores	64
L1 Cache	64Kb (per core)
L2 Cache	512Kb (per core)
L3 Cache	768 000Kb
Threads per Core	2
# Mem Channels	8
RAM (2933MHz) DDR 4 ECC Registered	2 048GB

UFS UV

HARDWARE

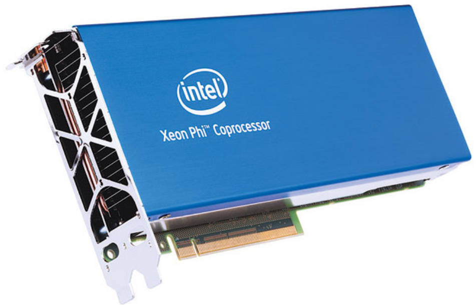
AMD (EPYC 7773X) vs Cloud (Intel Gold 6326)

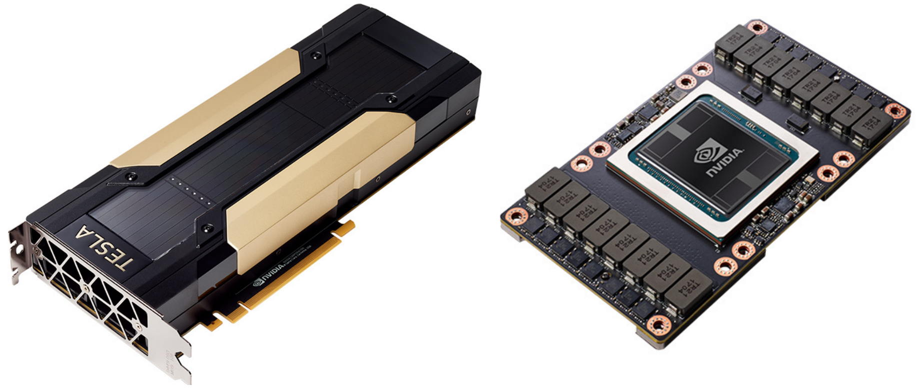
Description	Capacity
Clock Speed	2.20GHz – 3.5GHz
Number of Sockets	2
# Cores	64
# Threads	128
# Mem Channels	8
RAM (2933MHz) DDR 4 ECC Registered	2 048GB
GPU	8 x RTX A4000 or 4 x RTX A5000

Description	Capacity
Clock Speed	2.90GHz – 3.5GHz
Number of Sockets	2
# Cores	16
# Threads	32
# Mem Channels	8
RAM (3200MHz) DDR 4 ECC Registered	1 024GB

UFS UV

INTEL VS NVIDIA



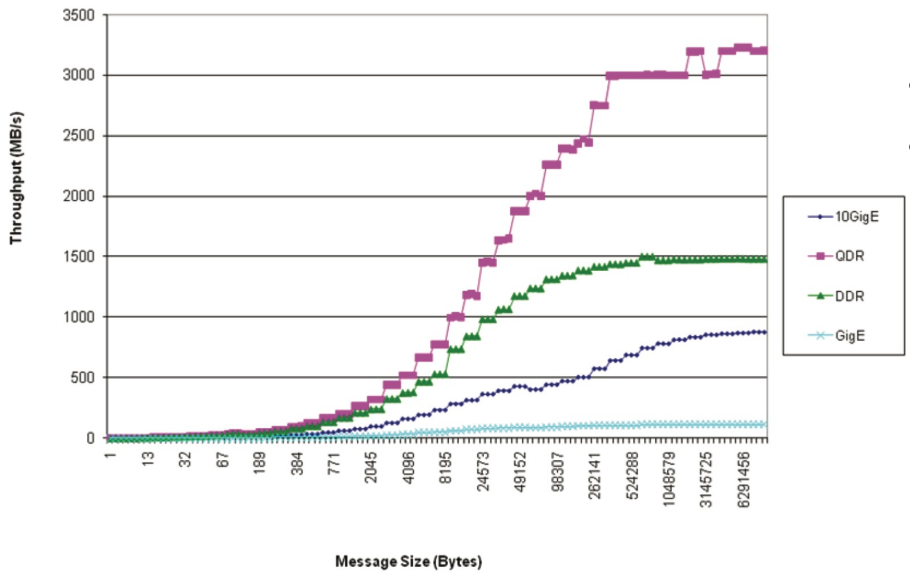


Intel® Xeon Phi™ coprocessor 7200 Family	NVIDIA Tesla™ V100 Architecture NVLink
Normal 64-Bit operation	GPU accelerated computation
Coprocessor runs a minimalistic GNU Linux	No direct access to processors
Code can be easily recompiled GCC/Intel	Code rewritten in CUDA
2.828* TeraFLOPS double-precision	7.8 TeraFLOPS double-precision
$2 214 to $6 294	$15 800 to $19 200
http://www.intel.com/content/www/us/en/products/processors/xeon-phi/xeon-phi-coprocessors.html	https://www.nvidia.com/en-us/data-center/tesla-v100/

* Claim based on calculated theoretical peak double precision performance capability for a single coprocessor. 32 Double Precision FLOPS/clock cycle/core * 68 cores * 1.3 GHz = 2.8 TeraFLOPS

UFS UV

NETWORKING: ETHERNET 1GB vs 10GB vs OMNI PATH



- Omni Path not featured in Graph
- Scale too high for comparison

Feature	Ethernet 1Gb	Ethernet 10Gb	Intel Omni Path
MTU (Max. Transport Unit)	1 500 – 9 000	1 500 – 9 000	65 520
* N/2 (bytes)	12 300	98 300	10 000
* Latency (µs)	47.61	12.51	0.93
Max Bandwidth (MB/s)	112	875	12 500

* Lower value is better

ETHERNET VS OMNI PATH

- Omni Path has a lower latency and higher throughput
- Omni Path costs less over time, due to time saved/power consumed
- As the number of nodes increases, the value of Omni Path becomes evident
- Omni Path may not be "cheapest" therefore the "best" solution for a cluster with very few (less than 4) nodes
- Some applications can run up to 450+% faster when nodes are increased beyond 24 nodes, using Omni Path vs. 10Gb Ethernet
- Omni Path is created by Intel and intended for Intel CPUs

HPC HARDWARE

- Design Considerations
  - Problem Specific
    - **Bioinformatics**
      - Large datasets ~ Large Memory (512GB – 1 TB RAM)
      - Low(er) CPU requirement (Some jobs 1-8 CPU Cores)
    - Physics
      - Usually Embarrassing Parallel ~ More Nodes
      - Monte Carlo Simulations ~ Random Data
      - High CPU utilization
    - Fluid Dynamics (Engineering, Chemistry, Physics, Weather etc.)
      - High CPU utilization
      - High Communication between CPUs (Latency important)
  - Is redundancy Important?
  - Is expansion possible/required?

UFS UV

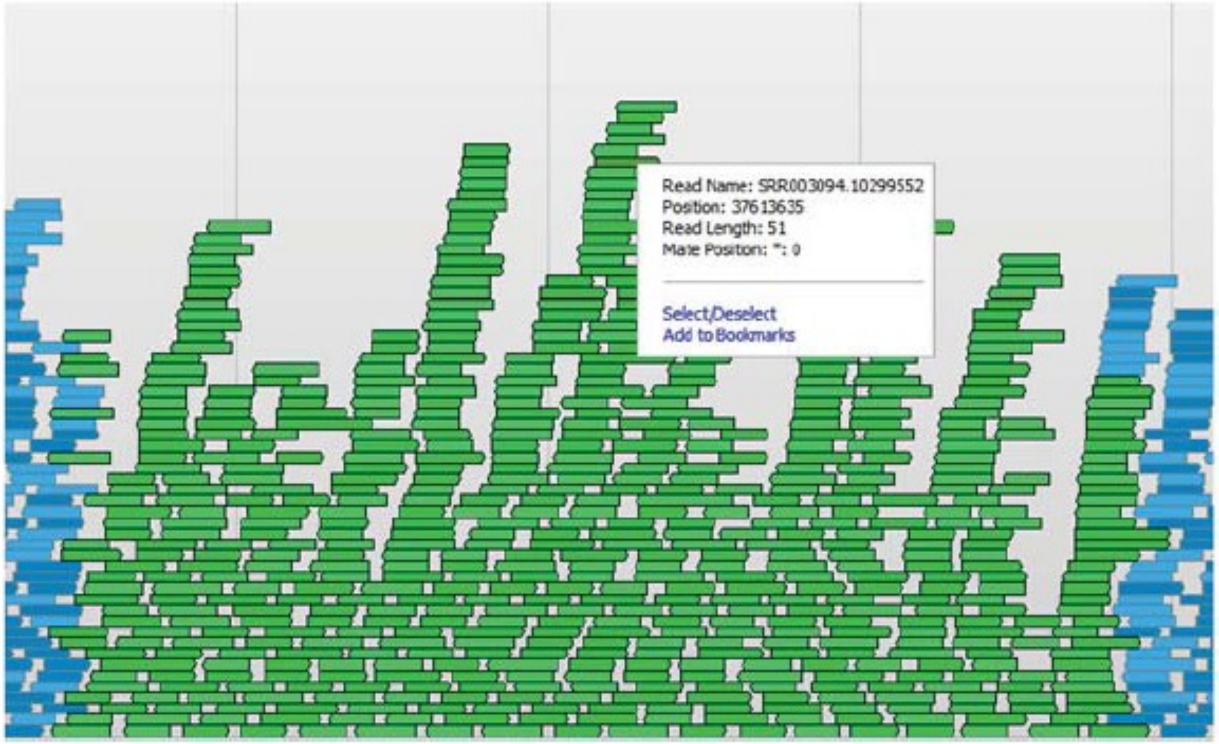
NEXT-GEN SEQUENCER

Illumina Next-gen sequencer

DNA or RNA sequence

HOW DO WE BUILD (ASSEMBLE) A GENOME?

- A Next-Gen Sequencer (NGS) produces millions of short (100-300 base-pair) reads of the sheered DNA in one go
- The algorithm then looks for overlaps and assembles the consensus genome



Read assembly

ATGCGCGGAAAAGGAATGGTTGTCATGCTACGGTACACTGCTATCTCGATCACTGTACTCAGTTGACGCTATCTACATCTACGTATGC Genome

WORKINGS OF THE ALGORITHM

"GCGGAAAAGG"

"ATGCGCGG"

"ATGCGCGG"

"ATGCGCGG"

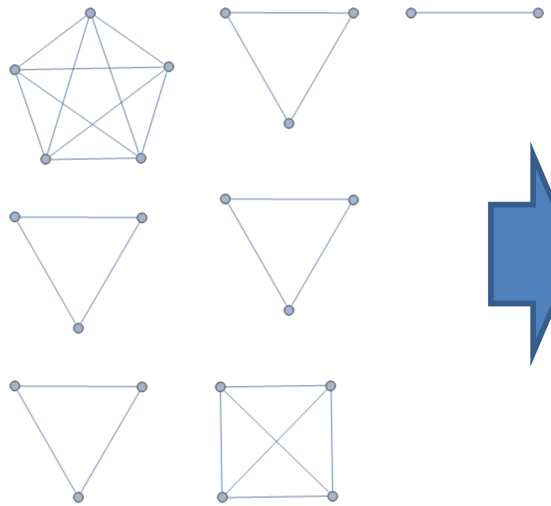
"GCGCGGAA"

"CGCGGAAA"

kmers (multiplication of data)

```
ATGCG
ATGCG
ATGCG
 TGCGC
 TGCGC
 TGCGC
  GCGCG
  GCGCG
  GCGCG
  GCGCG
   CGCGG
   CGCGG
   CGCGG
   CGCGG
   CGCGG
    GCGGA
    GCGGA
    GCGGA
     CGGAA
     CGGAA
     CGGAA
      GGAAA
      GGAAA
       GAAAA
        AAAAG
         AAAGG
```

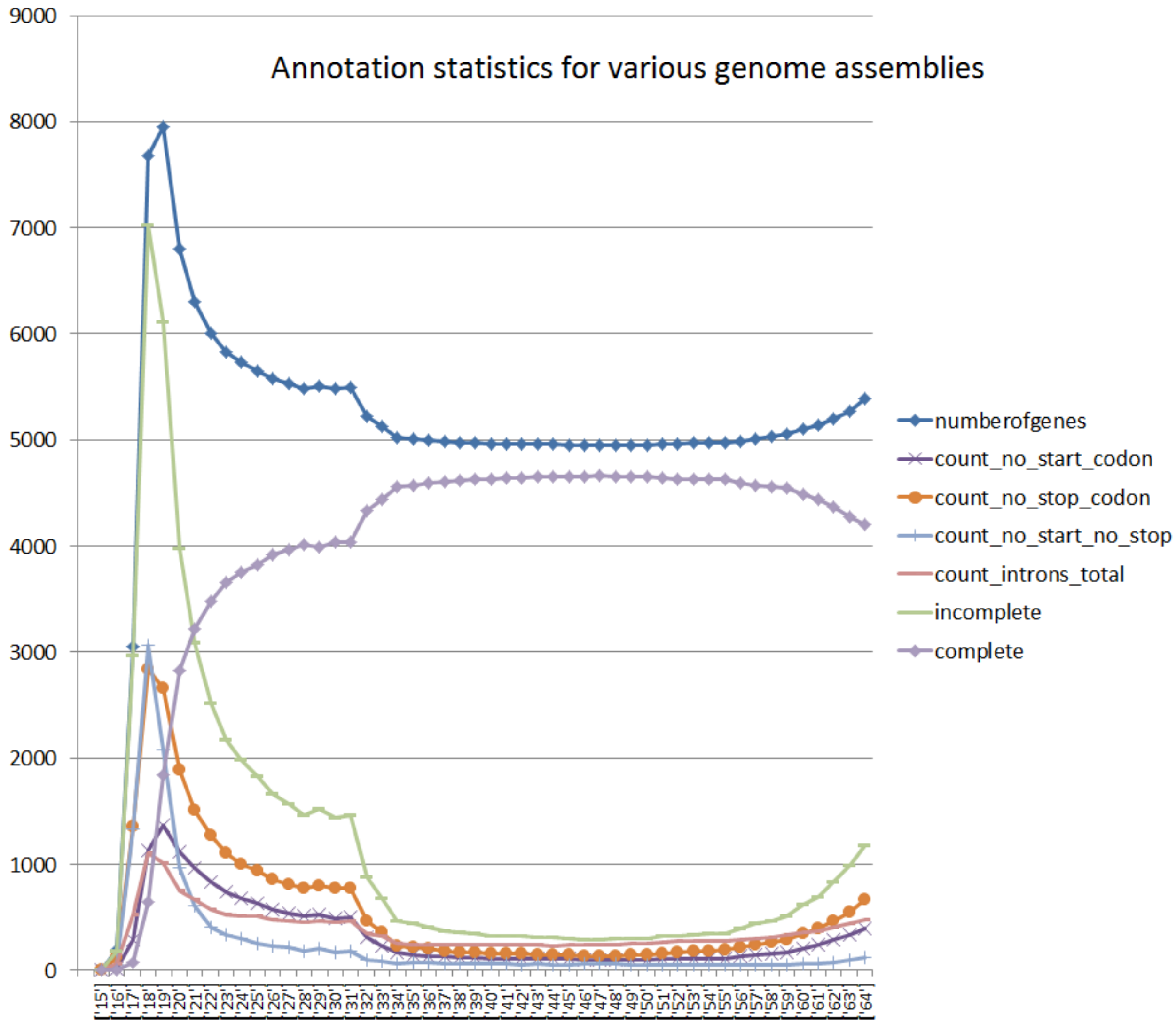
puzzle out links



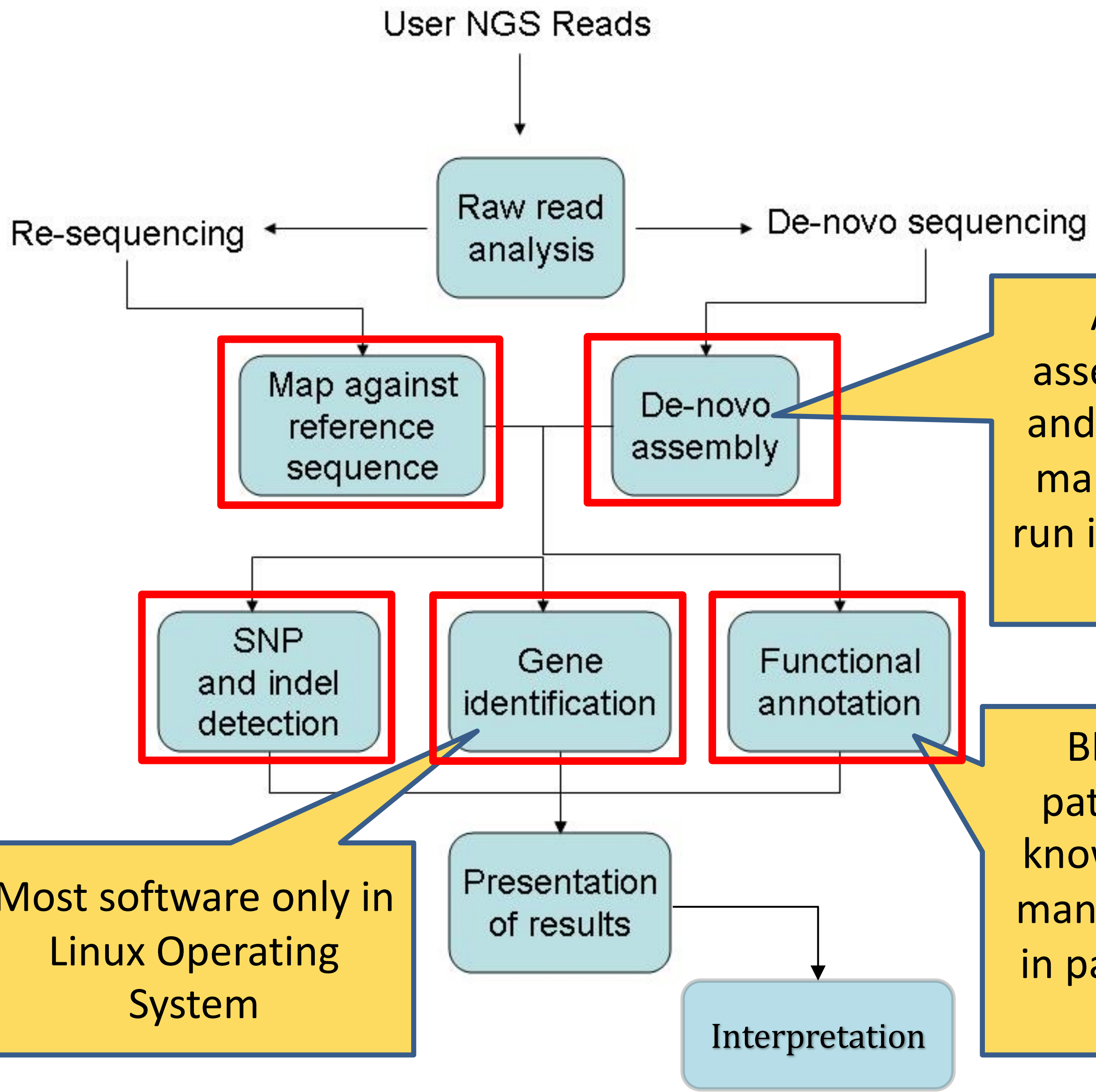
Assemble consensus

```
1|ATGCGCGG
2|ATGCGCGG
3|ATGCGCGG
4|  GCGCGGAA
5|   CGCGGAAA
6|    GCGGAAAAGG
```

ATGCGCGGAAAAGGAATGGTTGTCATGCTACGGTACACTGCTATCTCGATCACTGTACTCAGTTGACGCTATCTACAT


Annotation statistics for various genome assemblies
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
numberofgenes
count_no_start_codon
count_no_stop_codon
count_no_start_no_stop
count_introns_total
incomplete
complete

NGS PIPELINE



ABYSS/VELVET/MIRA: assembling whole genomes and transcriptomes requires many medium sized jobs to run in parallel (needs to all be done in one go)

BLAST: homology pattern matching to known genes requires many small jobs to run in parallel (completely separate jobs)

Most software only in Linux Operating System

WHY DOES NGS REQUIRE AN HPC?

- Sequencing of a genome, produces large datasets
  - A yeast: 10 000 000 bp genome.
    - At 100×coverage: 1000 000 000 = 1 Gigabase base pair (± 10 Gigabyte of data on the hard drive)
  - A human genome: 3 000 000 000 bp genome.
    - At 100×coverage: 300 000 000 000 = 300 Gigabase base pair (± 3 Terabyte of data on the hard drive)
  - Many plants have genomes 100 times larger than a human

UFS
UV

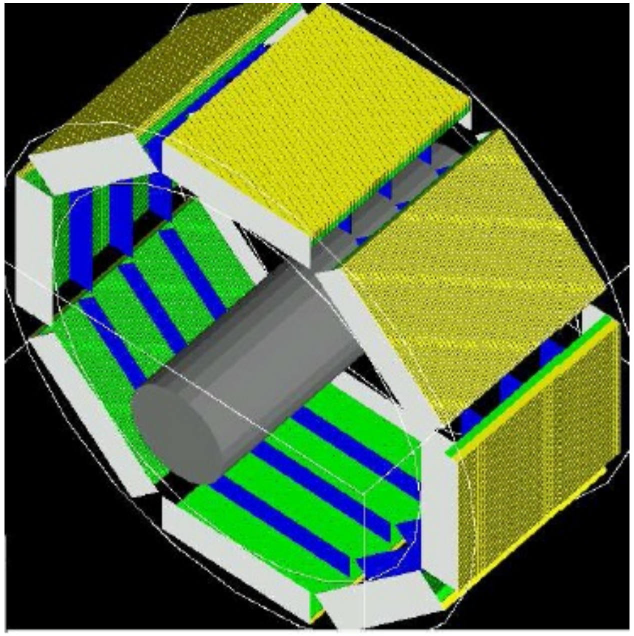
HPC HARDWARE

- Design Considerations
  - Problem Specific
    - Bioinformatics
      - Large datasets ~ Large Memory (512GB – 1 TB RAM)
      - Low(er) CPU requirement (Some jobs 1-8 CPU Cores)
    - **Physics**
      - Usually Embarrassing Parallel ~ More Nodes
      - Monte Carlo Simulations ~ Random Data
      - High CPU utilization
    - Fluid Dynamics (Engineering, Chemistry, Physics, Weather etc.)
      - High CPU utilization
      - High Communication between CPUs (Latency important)
  - Is redundancy Important?
  - Is expansion possible/required?

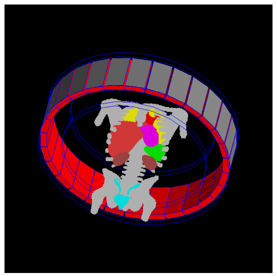
UFS
UV

REAL WORLD vs MODELLED

GE PET/CT Scanner
± R 1 600 000
+ R 4 000p/use



GATE modelled PET Scanner
"Free"
CPU Time

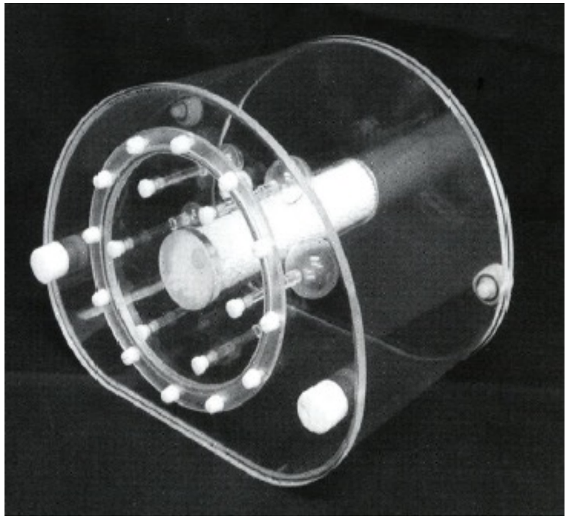


http://www.opengatecollaboration.org/

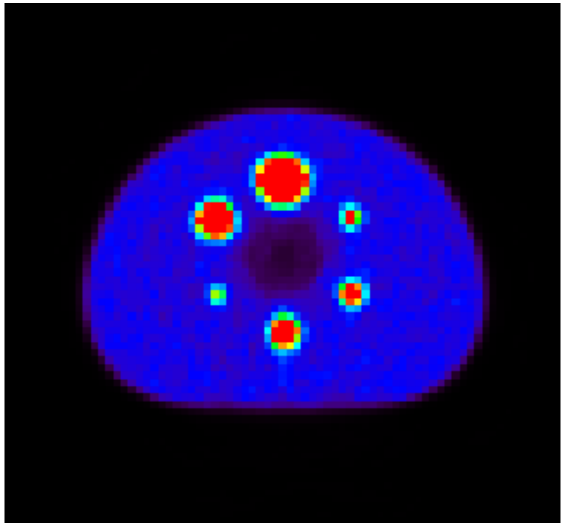
UFS UV

...PHANTOM STUDIES: PET/CT

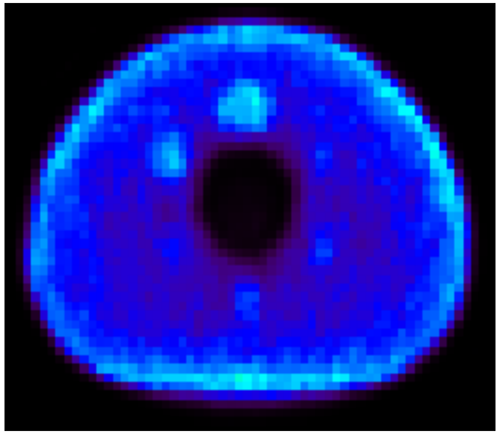
IEC Body Phantom (physical)



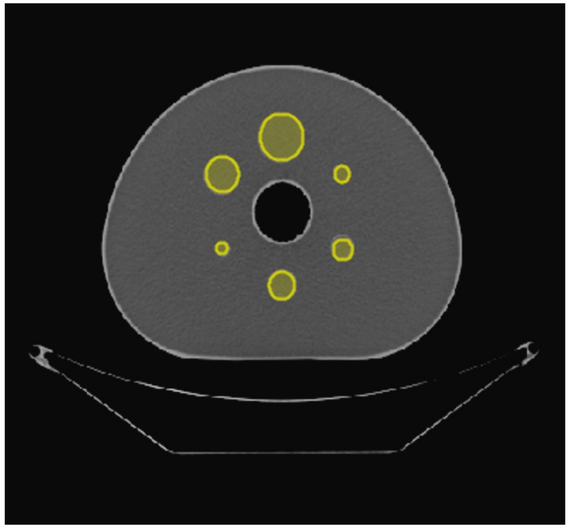
PET Scan (10:1)



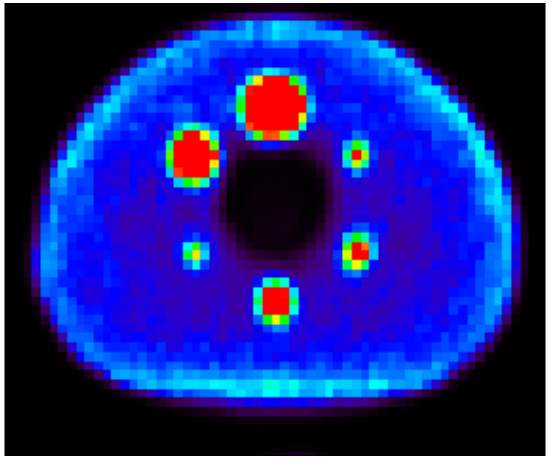
Simulated (2:1)



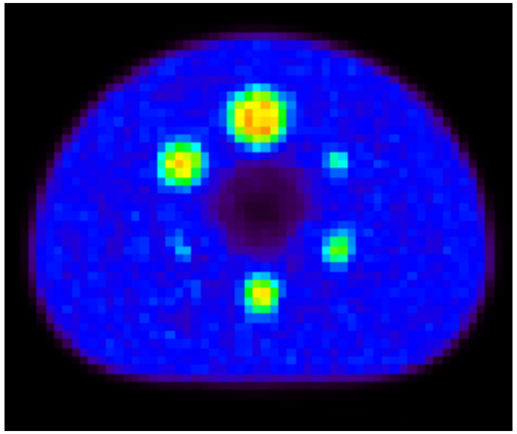
IEC Body Phantom (voxellised)



Simulated (10:1)



Simulated (5:1)



HPC HARDWARE

- Design Considerations
  - Problem Specific
    - Bioinformatics
      - Large datasets ~ Large Memory (512GB – 1 TB RAM)
      - Low(er) CPU requirement (Some jobs 1-8 CPU Cores)
    - Physics
      - Usually Embarrassing Parallel ~ More Nodes
      - Monte Carlo Simulations ~ Random Data
      - High CPU utilization
    - **Fluid Dynamics** (Engineering, Chemistry, Physics, Weather etc.)
      - High CPU utilization
      - High Communication between CPUs (Latency important)
  - Is redundancy Important?
  - Is expansion possible/required?

ASTRO PHYSICS

Active Galactic Nuclei – M87

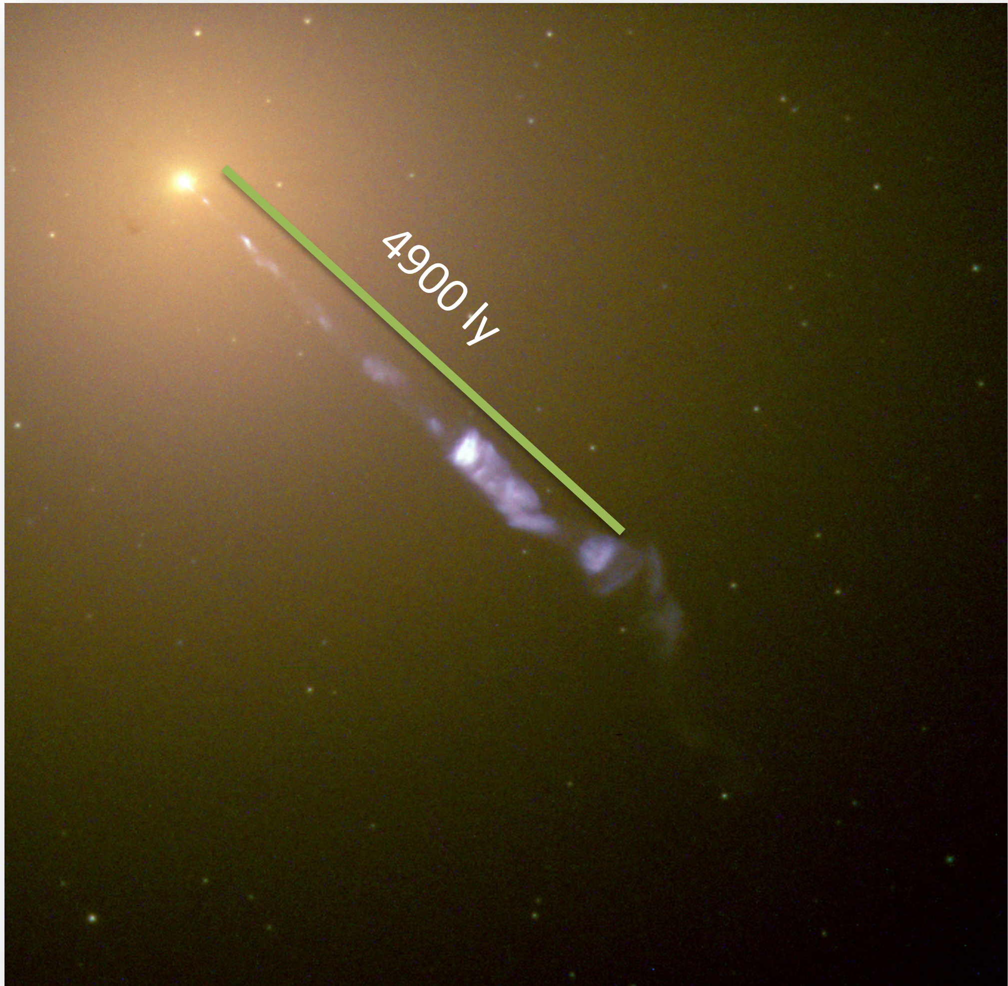
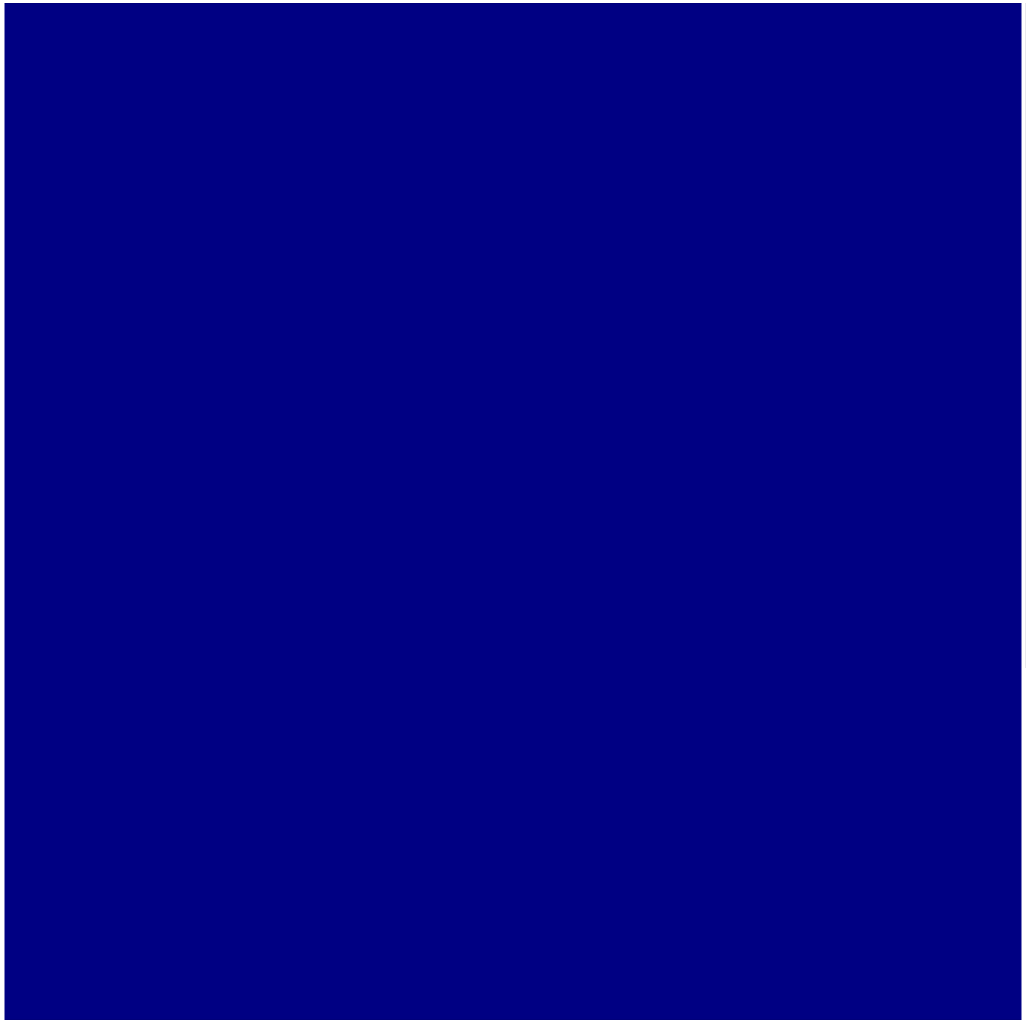
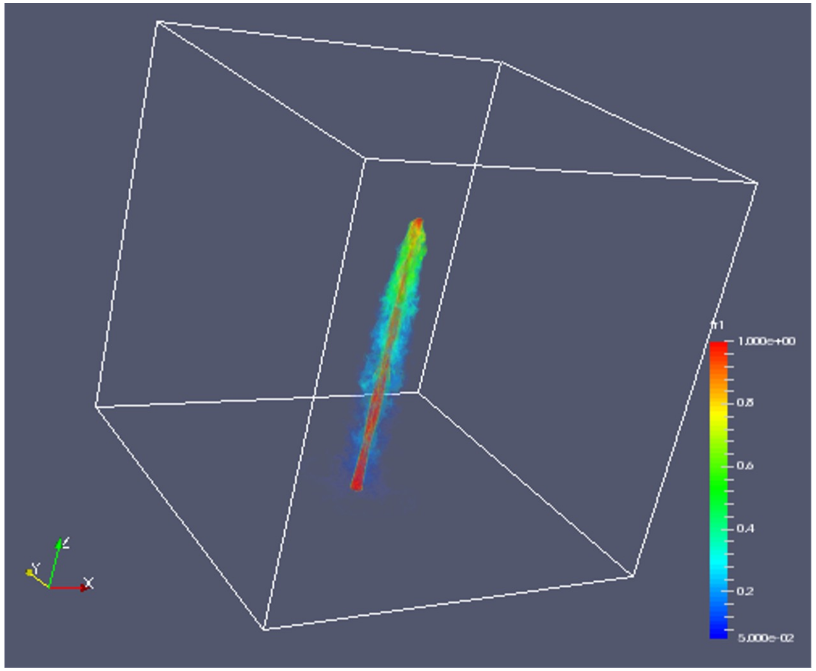


Image - Hubble Space Telescope

1 0000000000 0000000000 0000000000 0000000000 kg

DENSITY SIMULATIONS – ACTIVE GALACTIC NUCLEI – M87





18 000 CPU hours,
200 GB of memory
88 GB data files

$$\frac{\partial}{\partial t}\begin{pmatrix}\rho \\ \boldsymbol{m} \\ E+\rho\Phi\end{pmatrix} + \nabla\cdot\begin{pmatrix}\rho\boldsymbol{v} \\ \boldsymbol{m}\boldsymbol{v}+p\mathsf{I} \\ (E+p+\rho\Phi)\boldsymbol{v}\end{pmatrix}^T = \begin{pmatrix}0 \\ -\rho\nabla\Phi+\rho\boldsymbol{g} \\ \boldsymbol{m}\cdot\boldsymbol{g}\end{pmatrix}$$

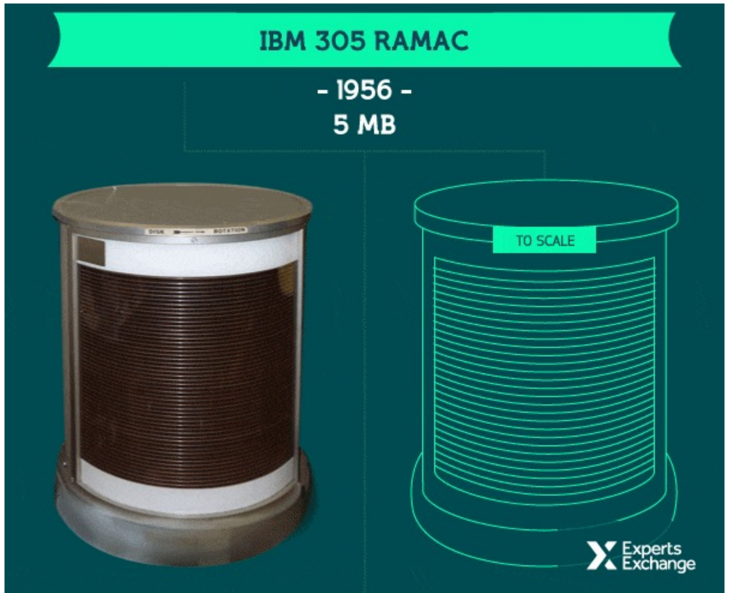
UFS
UV

HPC HARDWARE

- Design Considerations
  - Problem Specific
    - Bioinformatics
      - Large datasets ~ Large Memory (512GB – 1 TB RAM)
      - Low(er) CPU requirement (Some jobs 1-8 CPU Cores)
    - Physics
      - Usually Embarrassing Parallel ~ More Nodes
      - Monte Carlo Simulations ~ Random Data
      - High CPU utilization
    - Fluid Dynamics (Engineering, Chemistry, Physics, Weather etc.)
      - High CPU utilization
      - High Communication between CPUs (Latency important)
  - Is redundancy Important?
  - Is expansion possible/required?

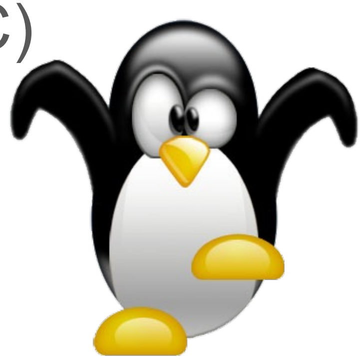
HPC HARDWARE

- Considerations when choosing:
  - **COST,** Power Usage, Speed/Throughput, Vendor
  - Size of Cluster:
    - Multiple nodes connected through network
    - Less nodes with more RAM and/or Cores
    - How many nodes can you afford
    - Are you going to consolidate Storage, Scratch
  - CPU
    - Highest common denominator CPU Flags (SSE4.2, VMX, AVX2 etc.)
    - Number of Cores per node
    - Clock Speed
  - GPU
    - Coprocessor vs. GP-GPU or non at all
  - Memory
    - Amount of RAM per machine
  - Network
    - Throughput vs. Latency
  - Storage
    - Size vs. Speed
    - Local vs. distributed



UFS UV

WHAT YOU SHOULD BE ABLE TO DO (CHPC SCC)



- Always keep within the budget (with as little as possible to spare)
- Choose compatible hardware
- Keep the design simple
- Choose wisely, don't focus on only one application
- Know why you choose a specific item; you should be able to motivate your choice in comparison to other options
- Make sure the hardware will fit in the physical space/motherboard
- Keep expansion and hardware failure in mind, but don't design specifically for that
- Give yourself leeway…don't choose one singular expensive component if good alternatives exist.

- Do your own research…be prepared … (why should or shouldn't you):
  - Use ECC RAM, OmniPath, InfiniBand, SSD, NVMe, GPUs etc. etc.
    - Compare AMD & Intel alternatives.
  - You will also notice we didn't discuss hard drives; see what is out there
    - Also, read up on RAID levels
  - There are also some recommended performance settings at the BIOS level that you can have a look at

UFS UV