XENON High Performance Computing

25 YRS OF HPC

Is (Your) eResearch Ready for HPC/DL/AI-as-a-Service?

Dr. Werner Scholz

CTO and Head of R&D, XENON Systems

20 Oct. 2022

25 Years of High Performance Computing, Storage, Networking Solutions

1990 — 2000 — 2010 — 2020

1998
Telstra Digital Video Network
Supplied Video Server equipment

2000
Animal Logic
Render Farm delivered for the "Matrix" movie trilogy

2004
TataSky DTH India
Supplied Video Logging solution

2005
Thales Defence
Supplied Image Generators for the ASLAV simulators

2007
Victorian Partnership for Advance Computing
Supplied HPC Cluster for Advance Computing

2009
CSIRO
(10th in Green500 ranking)
Supplied Australia's first GPU Cluster ("Bragg")

2011
Australian Stock Exchange (ASX)
Supplied Infiniband equipment

2011
Fox Broadcasting USA
Supplied Video Logging solution

2012
Fujitsu
Supplied Infiniband Network for Australia's largest Supercomputer ("Raijin")

2013
Digital Cinema Systems
Upgraded 150 Independent Cinemas in ANZ region

2014
University of Queensland
FlashLite HPC cluster system for high data throughput

2015
Thales Defence
Delivered High Fidelity Solution for Australian Army Tiger Helicopter Simulator

2015
Futuris
XENON drive high performance innovation

2016
Garvan Institute of Medical Research
NVMe solution that enables high data throughput

2016
Oil Search
Deployed Intel® CAS to speed reservoir modeling

2017
The Walter and Eliza Hall Institute
Private Cloud For Next Generation Cancer, Disease & Medical Research

2017
NCI
Supercomputer dedicated to scientific research & technology innovation

2018
Harrison.ai
Democratising reproductive healthcare through AI

2019
Pawsey Supercomputing Centre
New GPU Cluster for Pawsey Supercomputing Centre, Perth WA

2020
CSL Limited
Professional Services HPC Cluster

2020
Todd Energy
HPC Cluster for Todd Energy in Wellington, NZ

2021
XENON Cluster Stack
XENON launches containerised HPC solution

2021
Murdoch Children's Research Institute
Implementation of a backup solution

2021
Hong Kong Uni HPC
Deployed HPC Cluster with XENON Cluster Stack for HKU

Our customised technology solutions and products are designed to tackle the most data intensive and complex visualisation challenges, allowing our clients to do great new things, and break new ground in their respective fields.

25 YRS OF HPC | XENON® High Performance Computing

eResearch Challenges – User Expectations

User expectations

- I need more storage
- I want to store everything – forever
- I cannot delete any data – anything – ever
- I cannot lose any data
- I want to access all my data instantly
- I want to share all my data – except for the stuff I don't want to share
- I want to access everything – from anywhere – any time
- I want to process my data now
- I want to process my data faster
- I want to use this new code I found in a dark alley of the internet and I want it to run it now
- I want to get access to this new thingy, which is going to supercharge my research – now (and I will just pay with my credit card – thanks!)

...sounds familiar?

eResearch Challenges – User Expectations

More user expectations

- And no, I don’t want to...
  - write a long-winded justification
  - get multiple quotes
  - apply for a grant
  - wait for the grant to be awarded
  - go through the RFQ/procurement process
  - wait 6 months for the thingy to arrive

Image: iStock.com/wildpixel

eResearch Challenges – User Expectations

More user expectations

- And no, I don’t want to...
  - write a long-winded justification
  - get multiple quotes
  - apply for a grant
  - wait for the grant to be awarded
  - go through the RFQ/procurement process
  - wait 6 months for the thingy to arrive


Image: iStock.com/

- **The cloud is magic and the solution to all my problems!**

...sounds familiar, too?

Seems like your are ready for HPC/DL/AI-as-a-Service...

eResearch Challenges – What it really means...

- **Data Storage**
  - Capacity
  - Growth
  - Performance: Throughput, Latency
- **Data Management**
  - Ownership
  - Lifecycle
  - Governance
  - Protection (Archive and Backup)
- **Compute performance**
- **Connectivity between**
  - Data generators (e.g. instruments)
  - Data caches
  - Primary data storage
  - Data processors (e.g. compute for data analysis)
  - Archive for long term preservation, e.g.
    - 3-5 years (research project)
    - 5-10 years (center of excellence)
    - forever (medical research data for eventual TGA review and approval)
- **Security**
- **Platforms**
  - Bare-metal on-prem or off-prem
  - Cloud platform

Image: /wallpapersafari.com/w/HOZrL6

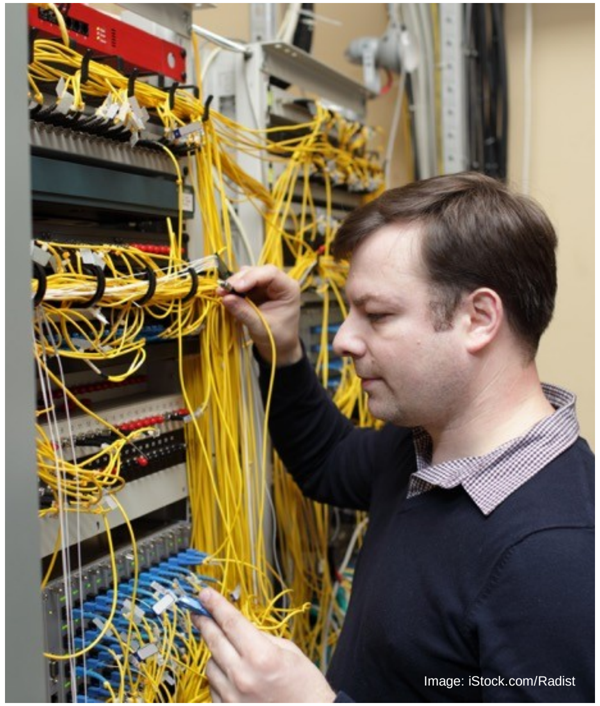
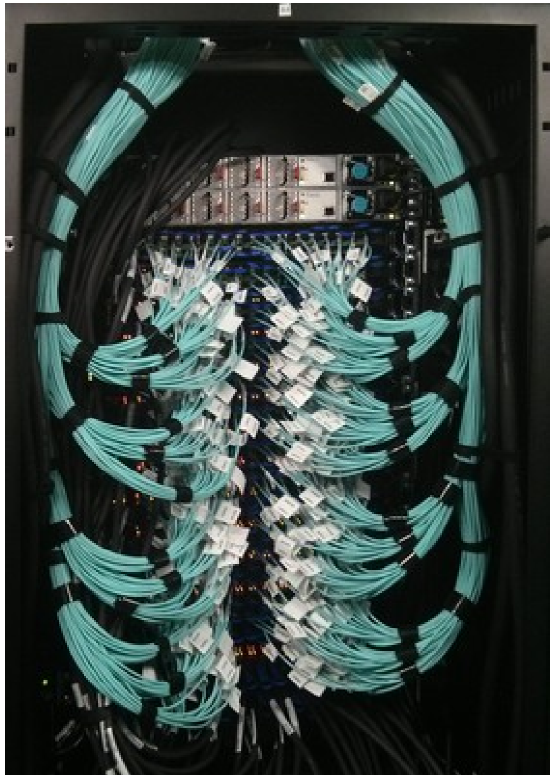


Image: iStock.com/Radist



XENON CLUSTER STACK

25 YRS OF HPC XENON High Performance Computing

XENON Cloud

What is it?

XENON's Cloud service for HPC/DL/AI/Big Data/Analytics

How is it different?

It is your cloud solution, customised and optimised for

- your workloads and workflows
- your compute, storage, network requirements
- your budget
- support for a variety of platforms

Where does it run?

- On-prem (in your on-site data centre)
- Off-prem (in your off-site data centre)
- Partner serviced (in our data centre)
- In cloud – multi-cloud (on a 3rd party cloud service)

...with the option to migrate between platforms

Who manages it?

- XENON can manage it (fully managed service)
- You can manage it (in your DC or cloud tenant)
- Collaboration models available

XENON CLOUD

How do I access it?

- ssh or through a secure web portal from any web browser from anywhere any time with 3D accelerated graphics

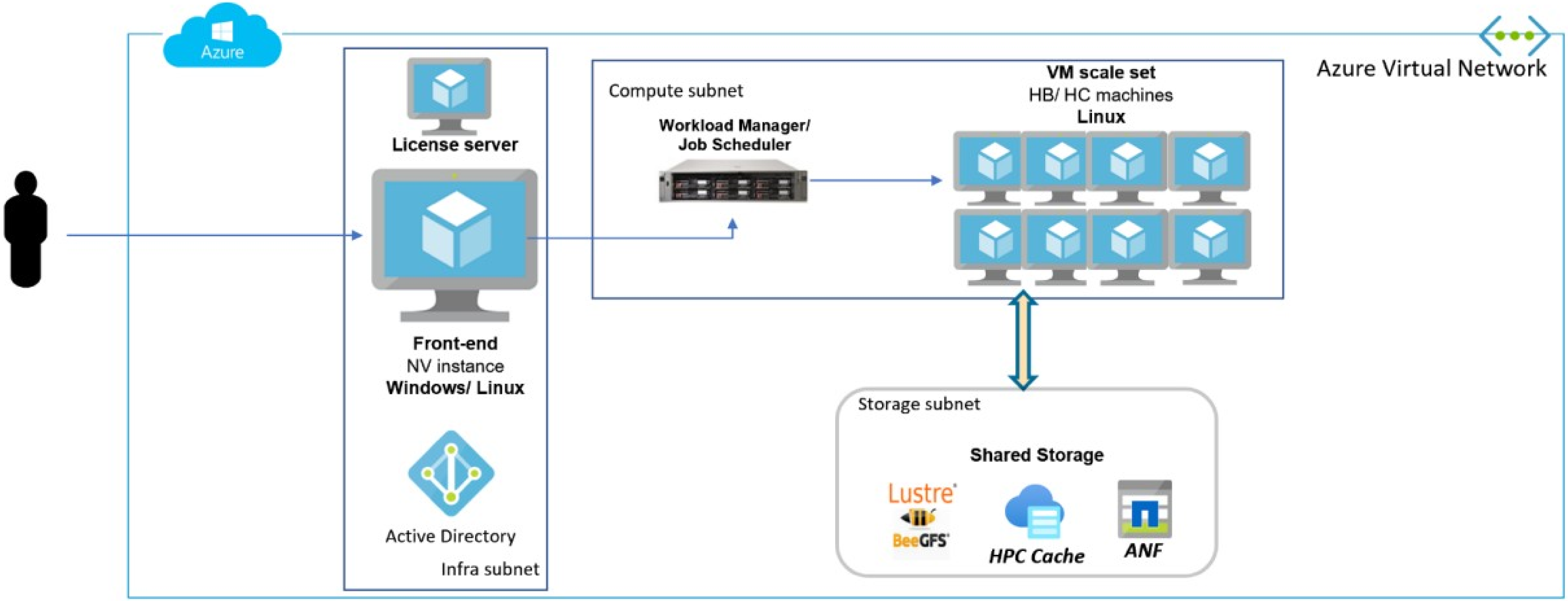
What is the cost model?

- Fixed commitment with upfront capex
- As-a-service with monthly billing based on utilisation

When is it ready?

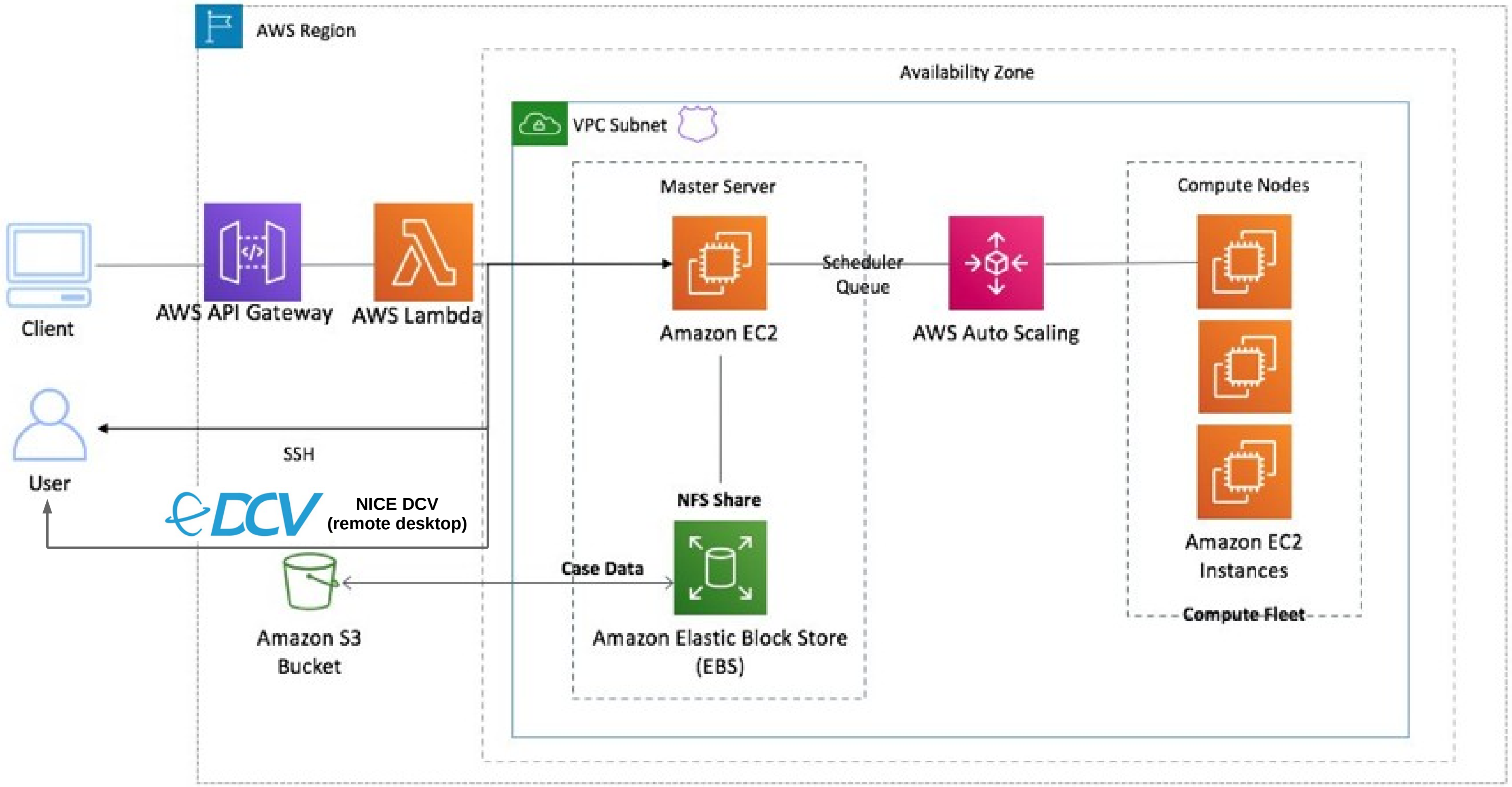
- Ready now and running in several projects

XENON HPC Cloud – Classic Architecture in Azure



Ref: Azure HPC Technical Onboarding Guide: https://azure.microsoft.com/en-au/resources/azure-hpc-technical-onboarding-guide/

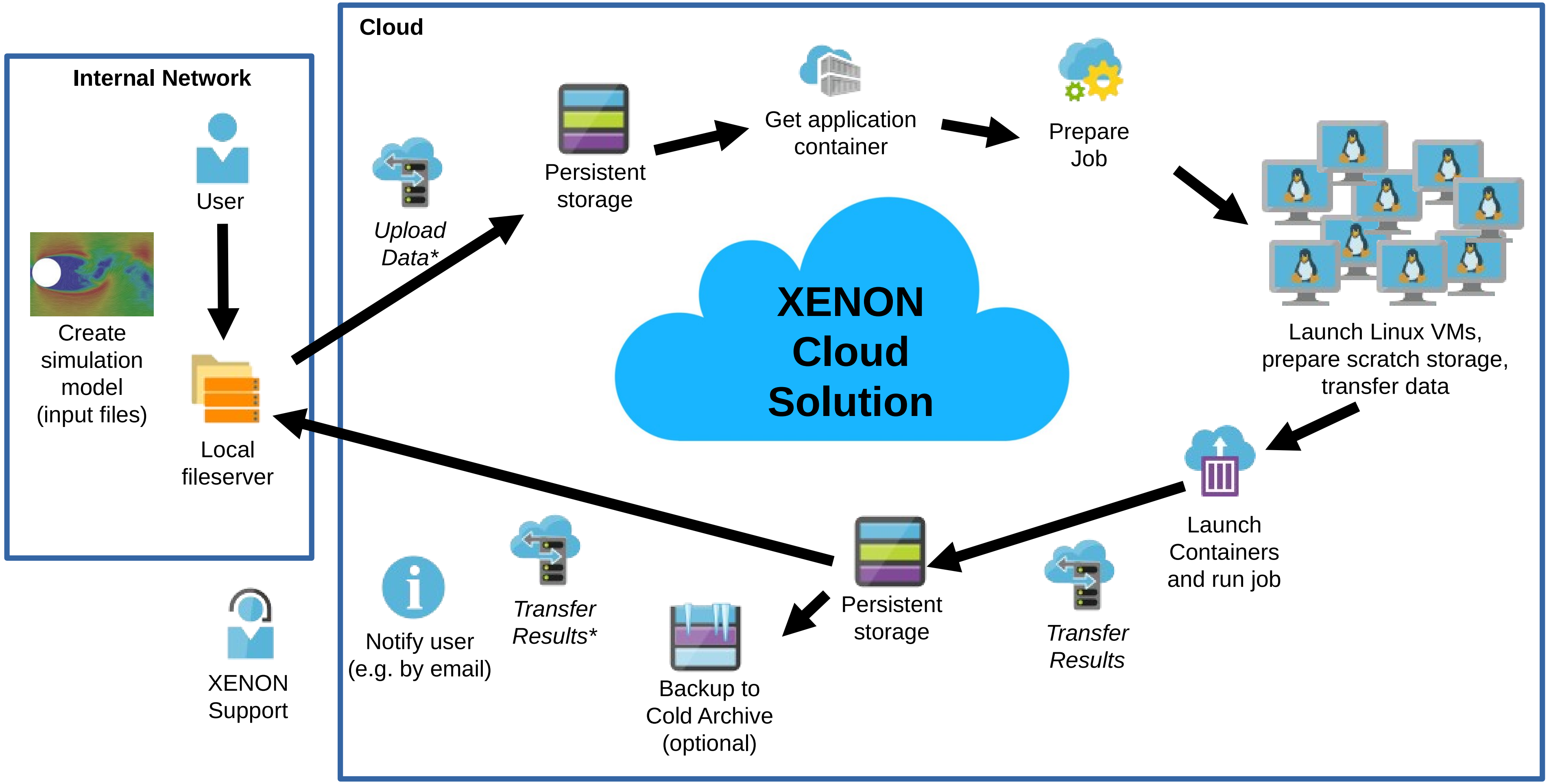
XENON HPC Cloud – Classic Architecture in AWS



Ref: Azure HPC Technical Onboarding Guide: https://aws.amazon.com/blogs/compute/using-aws-parallelcluster-with-a-serverless-api/

XENON HPC Cloud – Headless Design

Architecture Overview



XENON HPC Cloud - Benefits

- **Quick setup:**
  ~20 min from zero to an autoscaling HPC cluster in the cloud
- **Autoscaling:**
  Jobs automatically spin up as few or many compute nodes as needed and shut them down again – minimising cost
- **Hardware options:**
  Optimal compute node configs for any application:
  few fast cores, many cores, large memory, GPUs, FPGAs, Eth or IB networks, ARM, storage
- **Access to latest hardware:**
  Instantly use the latest hardware as soon as it is available on the platform
- **Workflows:**
  XENON engineers can install, configure, optimise applications and migrate complete workflows
- **Different architecture options:**
  Classic HPC or cloud-native headless
- **Data storage, management, protection:**
  Variety of options from fast parallel filesystems to low cost deep archive; easy integration of data protection/backup/archive
- **Resilience:**
  Migrate compute and data between regions
- **No vendor lock-in:**
  Flexibility to migrate workflows and complete HPC environments from on-prem to cloud to another cloud to on-prem
- **Convenience:**
  Fully managed by XENON engineers
- **Opex cost model:**
  Pay-as-you-go only for consumed resources

XENON High Performance Computing

Thank You

Dr. Werner Scholz
werners@xenon.com.au
https://www.xenon.com.au