



ICHEC
Irish Centre for High-End Computing



INNOVATIVE APPROACHES TO CAPACITY BUILDING IN THE USE OF BIG DATA - PLATFORMS



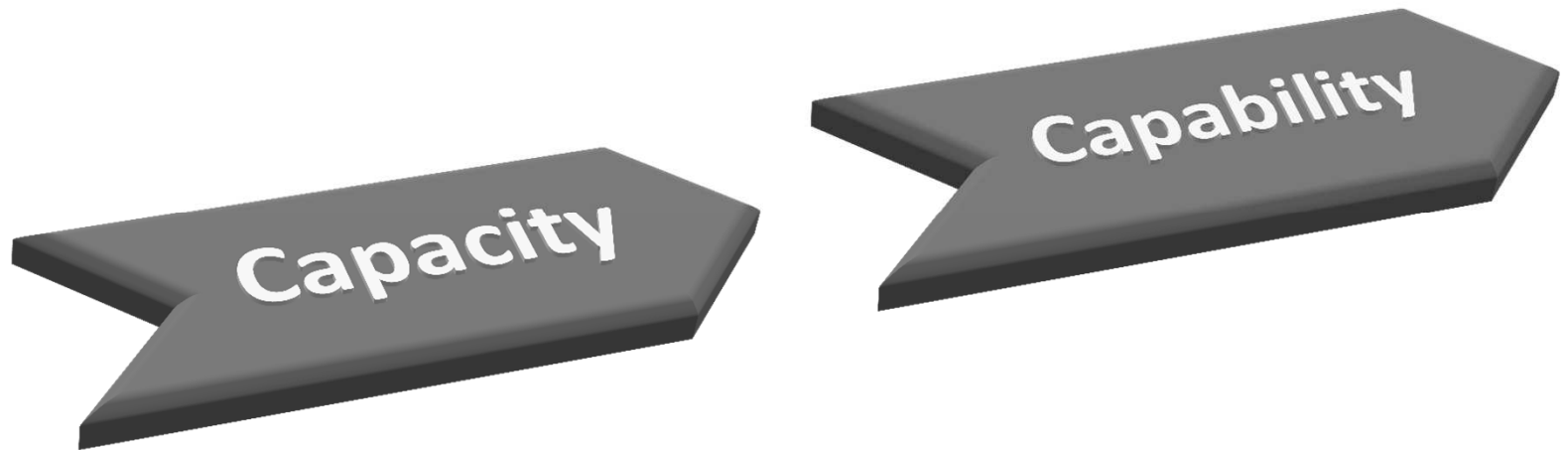
An Roinn Post, Fiontar agus Nuálaíochta
Department of Jobs, Enterprise and Innovation



A.S. ROINN
OIDEACHAIS AGUS SCILEANA
DEPARTMENT OF
EDUCATION AND SKILLS



www.ichec.ie



- **Capacity**
 - ❖ Resources (central, distributed and/or public cloud)
 - ❖ Datasets (size, diversity, accessibility)
 - ❖ Community and Knowledge
- **Capability**
 - ❖ Increases in Capacity creates Capability for research:
 - Tackle larger “grand challenge” problems
 - Facilitate linkage and novel methodologies

Current Sandbox

Installed Software

- Hadoop (Hortonworks Data Platform)
- R – Rstudio
- RHadoop
- Spark
- ElasticSearch

and growing...



4 Data/Compute nodes

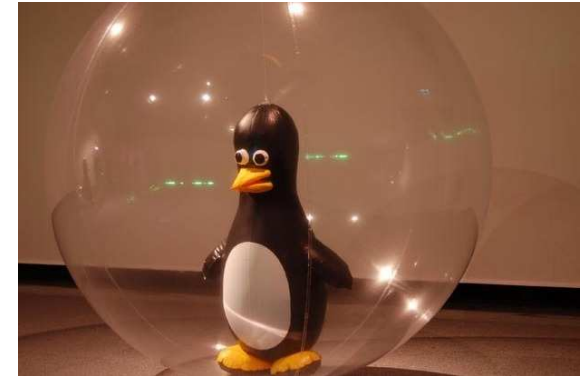
- 2 x 10 core Intel Xeon CPUs
- 128 GB RAM
- 4 x 4TB disks
- 56 Gbit InfiniBand network

2 Service/login nodes

- Similar hw as data nodes
- 10Gbit connection to Internet

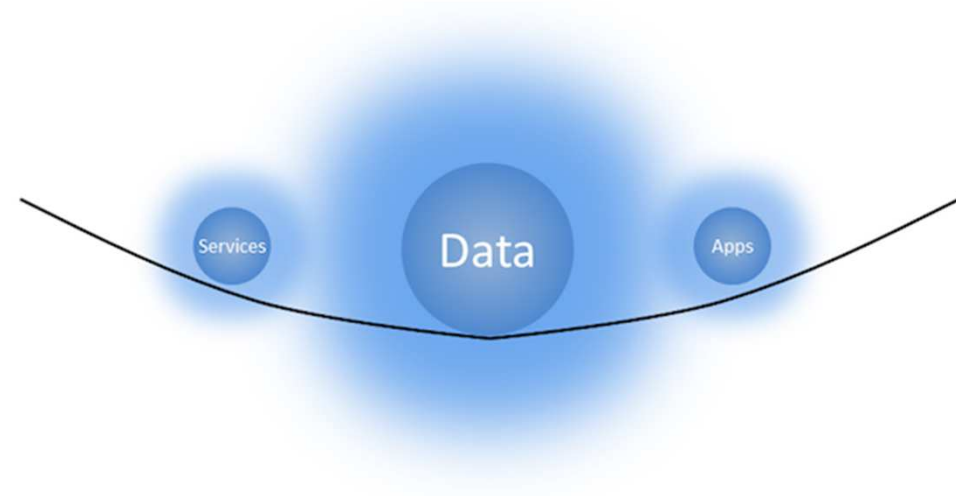
Applications and Compute Platforms

- Increasing software application complexity and interdependence
- Hardware requirements are diverse
- Repeatability => software is part of methodology
- Application containers promise the possibility of:
 - ❖ Encapsulation & isolation
 - ❖ Portability

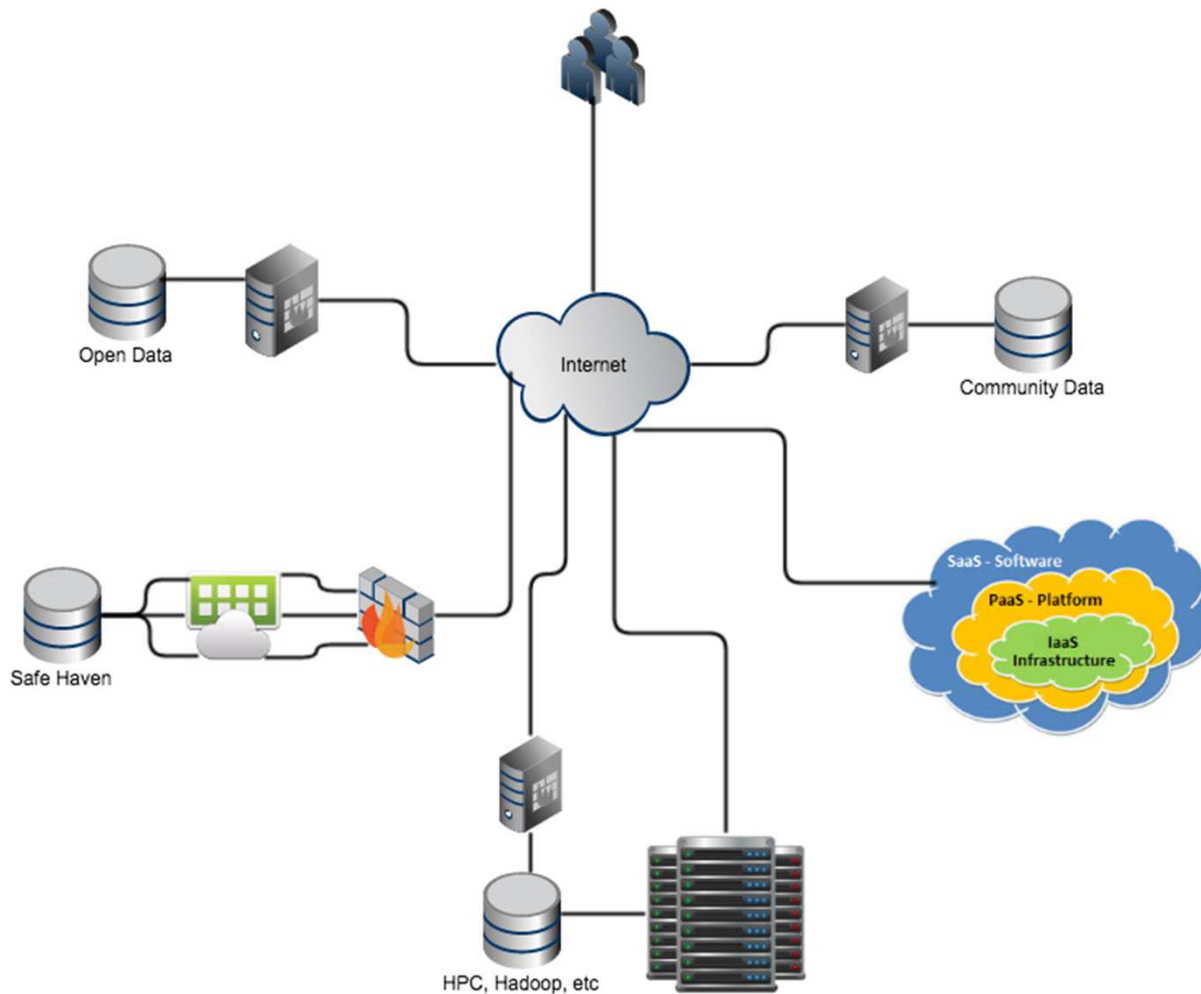


The Data

- Large datasets difficult to move =>
 - Centralised stores
and/or
Distributed, capacity built through federation and linkage
- F_{indable} A_{ccessible} I_{nteroperable} R_{eusable}
- Identifiable data
 - ❖ Safe Havens / data processors
 - ❖ Trusted Third Parties



Future Platform Architecture



- Distributed
- Non-Uniform
- Interoperable
 - Standards based
 - Lightweight
 - Portable apps