What is
the Sandbox

Contact info

UNECE Secretariat
support.stat@unece.org

Hardware Specifications

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
— Supports virtual machines for user access and management services
— 10Gbit connection to Internet

Installed Software

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

and growing…

the
Sandbox

A collaboration platform for the statistical community

BIG DATA

DATA DELUGE

DATA SCIENCE

DATA REVOLUTION

Huge opportunities for statistical organisations

Is your organization ready?

- New and more detailed statistics
- New and faster ways of production
- Additional sources of data
- Very timely statistics
- Benchmarks

the Sandbox

"I want to know what other organisations are doing in the field"

- Community guidance about the path to Big Data and Data Science
- Shared datasets instantly available
- A training facility: tools and material for capacity building

"I know I have to work on Big Data but I don't know where to start"

- An infrastructure for big data processing ready for being used
- Software already installed and proved/tested

Big data infrastructures and software are **expensive** and **complex** to deploy and maintain
With **the Sandbox**:
- you are ready to go in zero time
- you don't have to support the full cost of the infrastructure
- you don't have to care about maintenance
- you can test alternative software

"I have data but limited skills and/or backing IT infrastructure"

- A collaboration community built around shared working environment
- Worldwide network of organizations, with a catalogue of initiatives

"I already have skills and experience, I want to share them"

- Not an exclusive club! Open to every organization that wants to join
- A space to teach and to learn
- The possibility to shape the new frontiers of statistics

The Sandbox enables the **sharing of skills and knowledge** among organisations

A place for collaboration on methods and products, *not only related to Big Data*

power consumption