Mass data storage at PDC

Many research projects and organisations have large volumes of research data that they need to store. As well as providing high performance computing systems for research, PDC also offers long-term storage for research data – currently this is primarily via our Mass Storage System (MSS) which is managed by IBM Spectrum Protect software (which was previously known as the Tivoli Storage Manager or TSM).

If you are involved in a research project that needs to store data long-term, you are welcome to contact PDC Support to discuss purchasing storage from PDC. As the PDC mass storage can be extended fairly easily and cheaply (by buying more tapes for the MSS and extra licenses for the software), this can be a more economical solution than other storage alternatives such as setting up and then maintaining a tape storage system dedicated just for a single project, or buying storage from commercial companies.

Some projects storing data at PDC

CENTER-TBI is a large European project "Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injuries" that aims to improve the care for patients with traumatic brain injuries and identify the most effective clinical interventions for managing such injuries.

The *Odin* satellite combines two scientific disciplines on a single spacecraft in studies of star formation and the early solar system (astronomy) and the mechanisms behind the depletion of the ozone layer in the Earth´s atmosphere and the effects of global warming (aeronomy). The Swedish Space Corporation, on behalf of the Swedish National Space Board and the space agencies of Canada (CSA), Finland (TEKES) and France (CNES), has developed the satellite for astronomers and atmospheric researchers in the participating countries.

Prisma is a Swedish-led satellite project that aims to develop and qualify new technology necessary for future science missions in space. Many of the future projects comprise formation flying and rendezvous, so several spacecraft need to communicate and interact with each other with high precision. That requires exceptional accuracy in measuring and controlling the inter-satellite orientation.

SNIC-SENS is a Swedish project that uses high performance computing resources for analyzing sensitive data. PDC is a partner in this project and provides a backup resource for the National Genomics Infrastructure (NGI), which includes backup of sensitive personal data. The system is based on the IBM Spectrum Protect software and provides backup for the NGI facilities at the KTH Royal Institute of Technology and Uppsala University, and also acts as the backup of the NGI production systems which are operated by the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) at Uppsala University.

The *Human Proteome Atlas* (HPA) is a large Swedish program that provides knowledge-based analyses from different defined transections of the human tissue proteome. A basic description of a defined proteome is made available publicly and includes gene lists, analyses of expression patterns, and examples of protein expression on a cellular level. This data is archived and backed up at PDC (see also under IBM Spectrum Protect to the right).

iRODS-based storage at PDC

In collaboration with other partners from the Swedish National Infrastructure for Computing (SNIC), PDC is operating the Swestore iRODS (Integrated Rule-Oriented Data System) data storage service. This service makes it easier to manage research data and is a significant step towards providing open access to Swedish research data. You can find more information about iRODS here: https://irods.org.

SNIC users who need medium-term data storage (say for up to 4 or 5 years) are advised to apply for a Swestore allocation. Note that research data can be moved from PDC's Lustre file system to Swestore using PDC transfer nodes. For research data that is stored in Swestore, two copies of the data are stored at two different geographical locations, unless specified otherwise. More information about the SNIC Swestore iRODS storage service is available here: http://snic.se/resources/swestore/irods (or via the QR code).

A similar iRODS-based storage service is available at PDC for researchers who do not come under the auspices of SNIC, and for those researchers in need of longer term storage. For further information about the PDC iRODS storage solution, please get in touch with PDC Support directly by sending email to support@pdc.kth.se.

Mass Storage System (MSS)

PDC's Mass Storage System is essentially a large library of magnetic tape cartridges that are accessed using a tape robot. This is a very efficient way to archive data that is not accessed frequently but that needs to be stored for a long time.

Mass Storage System
PDC's IBM TS4500 tape library currently has
- 14 TS1150 tape drives
- 1250 IBM 3592 JD tape cartridges
- ~3500 available tape slots (which gives a total uncompressed data capacity of ~35 PB, extendable to 17,550 slots and 175 PB)
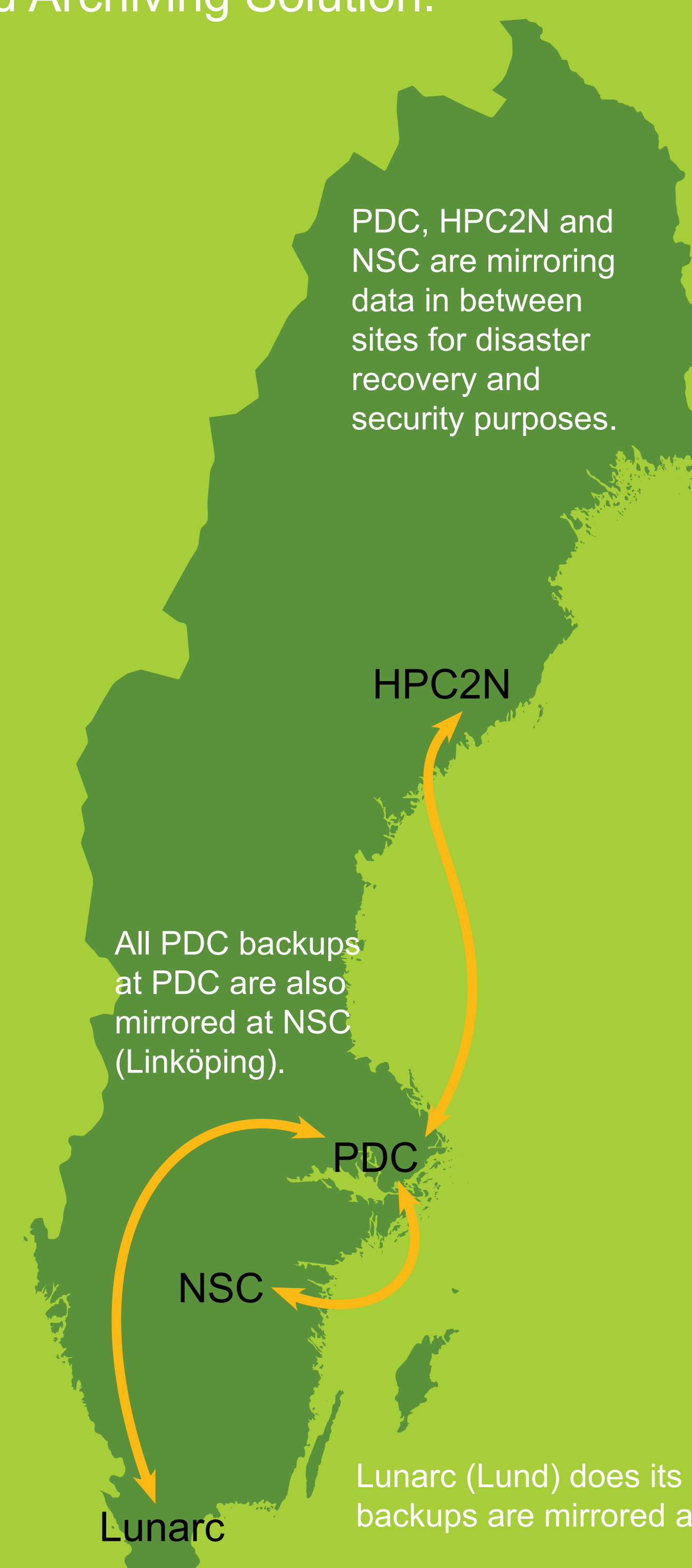
Each IBM TS1150 tape drive has
- built-in compression, up to 3:1
- dual 8 Gbit/s fibre channel interface
- 360 MB/s native speed
- up to 700 MB/s with compression

Each IBM 3592 JD cartridge has
- 10 TB native data capacity
- up to 30 TB compressed data capacity

IBM Spectrum Protect

PDC's IBM Spectrum Protect is essentially software that is used to manage the data archived in PDC's MSS - this includes storing the data, backing up the data, and recovering damaged or lost data. This system is sometimes referred to as PDC's Backup and Archiving Solution.

PDC, HPC2N and NSC are mirroring data in between sites for disaster recovery and security purposes.

All PDC backups at PDC are also mirrored at NSC (Linköping).

Lunarc (Lund) does its backup at PDC and their backups are mirrored at NSC

Different projects and organizations are using PDC's Backup and Archiving Solution which is connected to the mass storage system or MSS. Data lands in the disk storage pools and, after a certain time, it is migrated to PDC's MSS. Off-site backup of the system is performed twice a day to the National Supercomputer Centre (NSC) in Linköping. For example, the Swedish Human Proteome Atlas (HPA) program is using archiving at PDC and has stored about 440 TB of data which is mirrored off-site, in this particular case to the High Performance Computing Center North (HPC2N) in Umeå.

Sweden - Single Color by FreeVectorMaps.com

Access QR codes or visit www.pdc.kth.se for more information.

PDC Center for High Performance Computing ▪ SNIC