The **Department of Theoretical Bioinformatics** at the German Cancer Research Center develops computer-aided procedures for the interpretation of complex genomic data as well as procedures for the modelling and simulation of biological processes.

While, during the past few years, most of the bioinformatic proce-du-res were aimed at the analysis of genome sequences, these procedures are now primarily used to clarify the correlations bet-ween the superior organisation of the genome and its multiple functions.
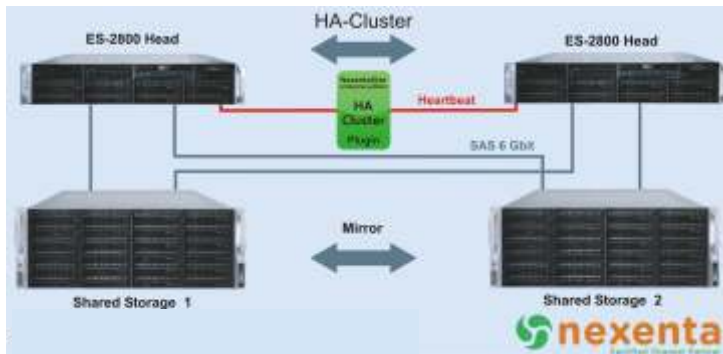
The objective is to achieve a holistic understanding of complex bio-logical processes. Therefore a highly-available NFS-Server was re-quired for this department, to be used primarily for storing the home directories of Unix users.

In addition, the server was intended to allow the safe exchange of sensitive research data between the research groups and the co-operation partners

as well as hosting the centrally installed bioinformatics programs and molecular-biological reference data required for a Linux cluster.

The Nexenta HA cluster was purchased since it provides for an uninterrupted operation and simplified administration at the same time, ensuring data safety by using NFSv4 and Kerberos.

Up to 50 users may access a Linux HPC cluster with 50 nodes and 1600 cores via the system, processing different tasks from the fields of next-generation-sequencing, network modelling and systems biology.

Sequenzer systems in DFKZ labs

Being the largest biomedical research institution in Germany and a member of the Helmholtz community of German research centers, the **German Cancer Research Center (DKFZ)**, in accordance with its articles of asso-ciation, is exclusively dedicated to the task of cancer research. The research center's staff - more than 3.000 employees, more than 1.000 of them being scientists, are researching the causes of cancer in more than 90 departments and study groups, they record risk factors for cancer and are seeking strate-gies with the aim to prevent people from developing cancer. They also deve-lop new approaches for a more precise tumor diagnosis and for a more successful treatment of cancer patients. The modern procedures which allow an analysis of complete genomic sequences (genomes) or proteins (prote-omes), generate enormous data volumes that cannot be dealt with by human beings alone.

Solution:

EUROstor offers cluster systems with Nexenta software providing especially high available NFS series. To keep the complete hardware redundant, the two cluster nodes are connected to two JBODs containing the hard disk drives. The disk drives are being mirrord in pairs so that even a failure of a complete JBOD does not interrupt availability of data.

As a filesystem ZFS guarantees a high data integrity with checksums, which also make it easy to recognize and correct "Stealth Errors" on single disk drives. ZFS also allows to create a nearly infinite number of snapshots to reconstuct a former status of the data. On this file system NFS volumes can be made accessible to the clients, transparent over virtual IP addresses.

To provide high speed data access, SAS disks are being used together with RAM disks for ZIL cache and SSDs for L2ARC (both coming from Stec/HGST).

ES-2800 Nexenta

- **2 HE storage server head for NAS, iSCSI and optional Fibre Channel**
- **Quad Xeon processor, optional two processors**
- **64 GB RAM, optional up to 384 GB**
- **2 mirrored SAS boot disks**
- **6 Gbit SAS expansion ports for shared storage**
- **2 x 1 Gbit iSCSI/Ethernet port, optional more, also 10 Gbit**
- **ZFS file system and NexentaStor software**
- **Nexenta HA cluster plugin, Gold Level Support**
- **NAS (NFS, CIFS, FTP, WebDAV, RSync), iSCSI, opt. FC**
- **RAID level: mirror, raidz, raidz2, raidz3 (1-3 fold parity)**
- **practically unlimited snapshots for more data security**
- **thin provisioning for flexible investment into of storage capacity**
- **automatic error correction (scrubbing)**
- **manageable over WEB front-end, extensive monitoring**
- **8 GB RAM SSD for ZIL (up to 130K I/Os)**