

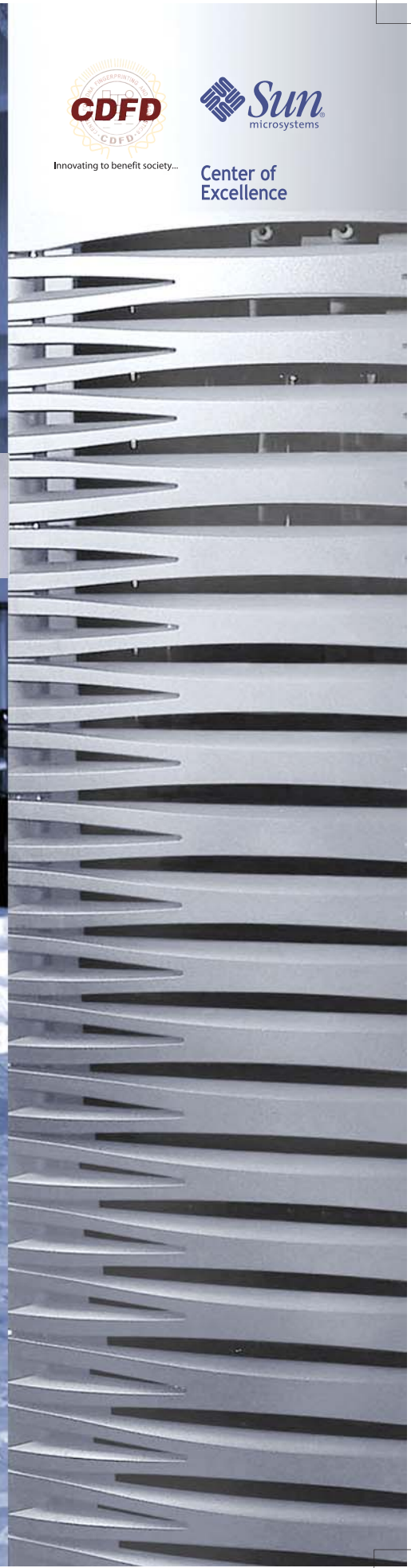
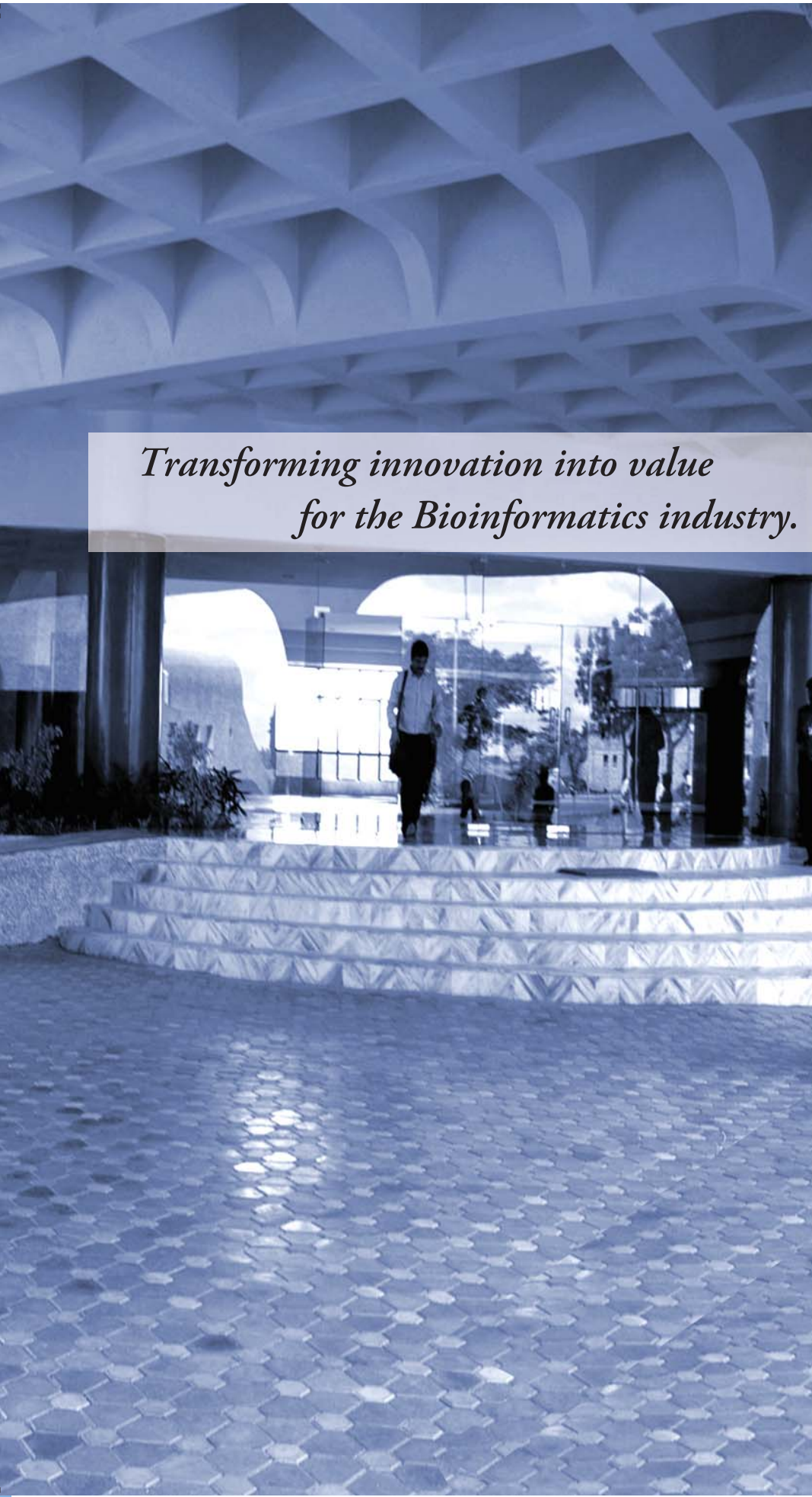


Innovating to benefit society...



Center of
Excellence

*Transforming innovation into value
for the Bioinformatics industry.*





Bioinformatics in India

India is on the verge of taking the global leadership in genomic studies. Analysis shows that pure cost benefits for biotech companies will drive the bioinformatics industry in the country. Slated as the next big opportunity for India in the global arena after the success of IT Services, Bioinformatics and Life Science research is gaining attention from government, industry and the academia.

In today's fast paced scientific innovations, Bioinformatics plays a very crucial role in the areas of contemporary biomedical research. With the introduction of Bioinformatics, the focus of the pharmaceutical companies has shifted from the trial and error process of drug discovery to a rational structure based drug design.

India has several ethnic populations that are valuable in providing information about disease predisposition and susceptibility, which in turn will help in drug discovery. Many eminent research and development institutes are in the process of developing a database of genetic profiles among diverse Indian population in terms of ethnicity, demographics and ancestral roots.

However, as there is lack of records and data management capabilities, the biotech and pharma companies need tremendous software support. With computers biological problems that can be addressed are in the areas of DNA / Protein sequence characterization, Protein Structure Predictions, Protein Interaction Prediction, i.e., Drug Discovery and Pathway Modeling.

The need for the hour is High Performance Computing Infrastructure by development of new advanced software and optimization of existing software.

The CDFD – SUN Centre of Excellence will help in developing local competencies in Computational Biology and High Performance Computing. As a Sun CoE, CDFD joins Sun's community of research institutions developing advanced technology in Computational Biology and Bioinformatics in the post-genome age.



SUN Microsystems Inc. – The Journey so far...

SUN Microsystems Inc. with strengths in HPC has been a key technology provider to the life sciences industry for more than a decade. SUN has more than 14,000 software vendor partners, which support multiple industries but have a special focus on life sciences. In addition, SUN also has more than 400 software partners who are focused solely on life science research and development, along with 12 academic Centers of Excellence (CoE) engaged in bioinformatics, life science research projects ranging from systems biology to virtual surgery and bioinformatics on a super computing scale.

The Sun Center of Excellence – An Overview

The SUN Center of Excellence (CoE) is a formalized, documented relationship between a Global Education and Research organization, an Educational institution and a Third party. All the three, involved in the Center of Excellence brings the partnership of a special expertise of strategic importance to Sun. The CoE offers a scalable, reliable and secure backbone infrastructure for life sciences organizations to capture, organize and analyze vast amounts of data and to convert this data into information and knowledge useful for designing improved medical, agricultural products and processes.

Located worldwide, the CoEs' deploy a full range of horizontal and vertical scaled servers, from high density Opteron, Xeon and Sparc based clusters, to highend Sparc-based SMP servers with very large memory for both compute- and data intensive applications. They also offer robust software stacks, from compilers and debuggers, through middleware, on to extensive data center administration functions, and productivity software.



SUN-Centers of Excellence around the world and their focus areas:

1. University of Calgary, Alberta, Canada: Visual Genomics
2. Columbia University, New York, USA: Neuro imaging research
3. Genome Institute of Singapore, Singapore: Digital Biology
4. University of California, Davis, USA: Public Health and Safety informatics
5. Samuel Lunfield Research Institute, Mount Sinai Hospital, Canada: Systems Biology
6. Delaware Biotechnology Institute, Delaware, USA: High Performance Computational Biology
7. Canadian Bioinformatics Resource, Halifax, Canada: Distributed Bioinformatics
8. Switzerland's Federal Institute of Technology, Zurich, Switzerland: Biomedical Research
9. Dartmouth College, California, USA: Neuroscience
10. Cheng Kung University(NCKU), Taiwan: Bioinformatics
11. North Carolina Research Triangle, USA: Computational Biology

The SUN Center of Excellence at CDFD – Hyderabad

Hyderabad now stands out to be amongst the privileged few, establishing itself to be a part of the worlds leading destination for Computational Biology and Bioinformatics. The recent unveiling of initiatives by Sun Microsystems Inc., indicates India's growing strength in these unique areas. Sun Microsystems Inc., has established the world's first state-of-the-art **Center of Excellence (CoE) for medical bioinformatics** at Hyderabad, in collaboration with the Center for DNA Fingerprinting and Diagnostics (CDFD), a premier, autonomous institute of the Department of Biotechnology (DBT), Ministry of Science and Technology, Government of India.

CDFD is engaged in providing services and carrying out research in the frontline areas of modern biology. The major service components of CDFD involve DNA fingerprinting, diagnostics, genome analysis and bioinformatics. Basic research in overlapping frontier areas of modern biology, especially in the post genomic scenario, is an integral component of this institute. A technological first, the CoE in Medical Bioinformatics has been an initiative of the Government of Andhra Pradesh, when it signed a MoU with Sun Microsystems Inc., through Locuz Enterprise Solutions, a Hyderabad based leading system Integrator and Global Education and Research partner in India for SUN Microsystems Inc.



The CDFD – SUN Center of Excellence, with a substantial investment, by three partners – SUN Microsystems Inc., the Government of Andhra Pradesh and the CDFD was officially launched on February 11th, 2004.

The CDFD – SUN Center of Excellence comprises of a data warehouse that links genomic data to patient genetic and clinical data generated by researchers at CDFD, its consortia and network partners; providing an efficient means of processing the biological and clinical data; and develop tools to mine the data in biologically and medically meaningful ways. SUN CoE program promotes open standards and collaboration to help build new technologies that advance academic research.



The CDFD - SUN Center of Excellence – Infrastructure

Groundbreaking Research & Development facilities

Technical Architecture: CoE follows a tiered architecture, with the following components –

- Web / Firewall /Access Tier
- Portal / Grid Tier
- Compute Farm, High Performance Computing / SAN Each tier is scalable to handle the individual load, independent of the other tiers.

The CDFD – SUN COE Hardware:

- SUN E20K with 1.2GHz Ultra 36 SPARC IV (dual threaded with chip multithreading) CPUs and 128GB memory
- SE6320 Enterprise storage with 10TB expandable to 20TB
- Two L25 LVD LTO Tape library each with two tape drives
- SUNFire V440 server with 4X1.28GHz UltraSPARC IIIi
- Two SUNFire V240 2X1.28GHz UltraSPARC IIIi
- SUN V20z Linux Grid Cluster with 17 nodes each with 2X2.2GHz AMD opteron processors and 4X512MB memory
- Twenty five SUN Ray thin clients
- Ten SUNBlade 1500 3D graphics workstations

Computational Environment

The environment offers scalability from desktop to teraflop, with binary compatibility across architecture offering mainframe-class like availability and superior balanced performance. The servers support features like fault-isolated Dynamic System Domains (Dynamic Hard partitions), Dynamic Reconfiguration, full hardware redundancy, and Hot CPU Upgrades.

Data Architecture and Management

The HPC SAN is configured to scale up from general-purpose to extremely high-performance environments. The HPC SAN will offer support for heterogeneous connectivity, not only from a storage perspective but also from a data perspective. The HPC SAN allows for the sharing of storage devices, file systems and individual files between clients in either a homogeneous or heterogeneous environment.

User Interface – CDFD Portal

The Portal eliminates the complexity by giving users a single secure access point through a point-and-click browser interface, for all applications, jobs, and projects. The Portal provides HPC users a secure Web-based single point of delivery to access services, content, and complex applications.

The architecture is built with Distributed Resource Management (DRM/GRID) software that can aggregate compute power and deliver it as a network service. The Grid Software will be able to provide dependable, consistent, and pervasive access to highthroughput computational capabilities, and also provide the job accounting and statistics information that are needed to monitor resource utilization and determine how to improve resource allocation.

The architecture includes HPC Cluster software that includes the highperformance, multi-protocol implementation of Message Passing Interface (MPI), a full implementation of the MPI I/O protocol, and tools for debugging, performance analysis, and tuning of technical computing applications. The Software also includes various mathematical tools, performance libraries & scientific subroutine libraries to enhance the productivity of the HPC user.





Scalability

The architecture is a scalable Compute and Storage subsystem. The servers are based on the SMP architecture. The processors and I/O interfaces have a load/store access to a large, cache coherent memory space. The interconnect transfers cache blocks (~64-bytes) between processors and memory. The SMP interconnect protocol is built into the interface of the processor chip. A single instance of an operating system controls the SMP. This SMP-based system provides a high degree of concurrency and parallelism that multi-threaded applications can use to scale linearly, and thus deliver high throughput.

The storage architecture is designed to move very large data objects between high performance computers, disks, and tape libraries. In order to ensure high speed transfer and low latency also manages parallel data transfers from multiple network-connected disk arrays at hundreds of megabytes per second

Integrated

The architecture can seamlessly serve to the needs of multiple computers and clusters attached into a single, integrated storage system.

The major goals of the SUN-CoE:

- APBioNET node in India for open source codes integration and testing.
- APBioNET, via S* alliance, providing free web learning for Bioinformatics.
- CDFD as one of APSTC (Asia Pacific Science and Technology Council) / APBioNET Academies, and hub for India BioGrid.
- Part of 45 worldwide CoE's in areas of Computational Biology, HPC, GRID etc.
- APSTC providing reference architecture for Grid & High performance computing infrastructure for Computational Biology.
- APSTC with other bioinformatics Centers of Excellence in Computational Biology forming collaborative network with CDFD.
- HPC & GRID Utility centre in for regional Research Institutes.
- Rock Solid IT Infrastructure with scalable and highly available components.
- Sponsor students / researchers.

(A) **Forecasting diseases:** The CoE will have a major focus on Medical Bioinformatics – an area which is at the junction of two knowledge domains viz., Bioinformatics and Medical Informatics. Bioinformatics helps scientists to get insights into the details of diseases at the biomolecular level whereas Medical informatics gives details of diseases in individuals, their relationship to socioeconomic backgrounds, familial disease histories, response to various drugs etc. Medical Bioinformatics thus, links the molecular level knowledge to the health or disease states of individuals and gives us a favorable platform to accrue a knowledge base which would help us to do disease forecasting.

(B) **In diagnostics and healthcare:** The CoE provides necessary infrastructure to create a knowledge base pertaining to its priority area of diagnostics. CDFD proposes to link collaborating hospitals and diagnostic centers to create a network and a national database of genetic disorders. The database can be made available to doctors and researchers to help in their work and is expected to leave an impact on way we manage our health care system in the country.

(C) **Way to help fellow scientists in the country:** CDFD has been operating as the National node for EMBnet and a regional node of India Bio- Grid and has been providing bioinformatics services to the Indian sub-continent. The CoE infrastructure will enable CDFD to extend its services in the forms of access to larger number of tools and databases. With CoE it would also become a national node for Asia Pacific Bioinformatics Network (APBioNet) through which it has plans to provide bioinformatics education and training in the region.

The CDFD – SUN Center of Excellence
Gandipet, Hyderabad. INDIA.

Ph: +91-(0)8413-235463. 235353 Fax: +91-(0)8413-235462

Email: director@cdfd.org.in

