

## **MRI: Acquisition of *Futuro*: A Data Intensive and High Performance Computing Cluster for Integrated Research and Education**

The computing research instrument acquisition described here – of the *Futuro* computing cluster for interdisciplinary research projects and Computer Science (CS) education programs – seeks to lay the foundation for a strong research and education integration centered on CS at the University of Texas at Brownsville (UTB), a minority serving institution in the south Texas Rio Grande Valley, a historically underrepresented region with more than 90% Hispanics. The MRI investment will enable promising young investigators to greatly raise the stature of CS research at UTB and in the region, attract new sources of research support, and elevate CS education for Hispanic students to a competitive level.

The south Texas valley region has 9 higher education institutions, including research institutions UTB and UT Pan American (UTPA). However, the entire valley does not have a single up-to-date cluster. The current clusters at UTB and UTPA are decommissioned after years of intensive usage. Temporary “cluster” consisting of several PCs is set up to carry on our research projects, which is very limited and non-reliable. *Futuro*—to be acquired and operated by UTB and shared with UTPA—will serve as a laboratory in which: core research can be conducted at a world-class level; collaborations with on-campus domain sciences provide real-world test applications and expertise to drive significant advances in interdisciplinary research; and significant opportunities are made available to both undergraduates and graduates in CS and related disciplines.

**Intellectual Merit:** *Futuro* will enable research activities in data mining, pattern discovery, genetic data analysis, experimental astronomical physics, collaborative filtering, theory of computation, high dimensional visualization, and other computational areas. More precisely, it will enable the following activities with transformative goals: (1) terabyte scale data mining and pattern discovery in time series datasets obtained from heterogeneous sensor network; and applying this research to address important data analysis problems in Laser Interferometer Gravitational Wave Observatory (LIGO), one of the nation’s most advanced scientific endeavors on the frontiers of physics; (2) Genetic data analysis in complex human diseases to identify totally novel susceptibility factors; this research will enable new understanding of genetic causes of complex diseases such as schizophrenia and thus likely lead to development of new therapeutic strategies for them; (3) studying the dynamical systems and Stellar populations to model the behavior of black hole binaries in globular clusters and galactic nuclei; this research will create models of formation of stellar systems via intensive computation and they can provide important information for interpretation of results from currently operating gravitational wave detectors (such as LIGO and LISA detectors); (4) exploring and creating computing-effective, scalable, robust and intelligent learning algorithms for large recommender systems based on collaborative filtering by incorporating multispectral information; this research will lead to development of next generation recommender systems; (5) visualizing high-dimensional streaming data from heterogeneous sensors; this research will lead to development of new data reduction methodologies, new visualization methodologies that incorporate intelligent computation such as data mining, and thus more advanced visualization systems that has cross-disciplinary utility; and (6) benchmarking and developing algorithms for approximating the NP-hard subgraph isomorphism problem with best possible practical performance; this research will benefit several important application areas such as image recognition and bioinformatics.

**Broader Impact:** The computing services enabled by *Futuro* will benefit potential users from physics, bioinformatics, computational engineering, and environment engineering at UTB and UTPA. *Futuro* forms a nucleus for collaboration between computer scientists and researchers from other departments and other institutions, resulting in further funding opportunities. The project’s educational activities will train students in the Rio Grande Valley, in areas of computer science that are expected to have the greatest national impacts in the coming years. More precisely, the CS curriculum will be enriched by lab modules enabled by the cluster facility for providing hands-on experience in parallel programming and scientific computing. As part of the proposed activities, students from computer science, physics and other programs will be involved in exciting and engaging interdisciplinary collaborative research.