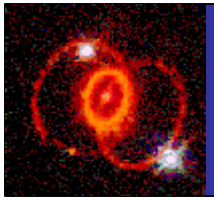


NOVA

A Networked Object-Based EnVironment for Analysis

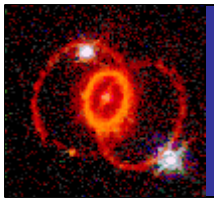
"Development of a State-of-the-Art
Object Oriented Analysis Framework"
LDRD Project

LDRD Mid-Year Review
April 30, 1999

Torre Wenaus
BNL Physics

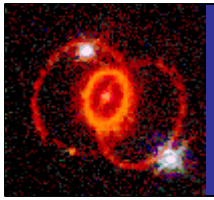
http://duvall.star.bnl.gov/nova

BROOKHAVEN
NATIONAL LABORATORY



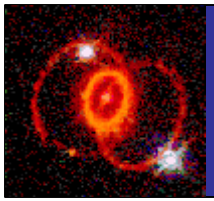
Outline

- Project Participants
- Background
- Motivations and Goals
- NOVA Architecture
- Project Components and Current Status
- Milestones and Deliverables
- Budget



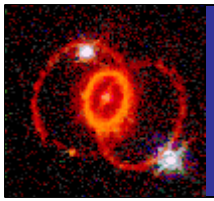
Project Participants

- PIs: Torre Wenaus, BNL and Tim Hallman, BNL
- Project leader and part-time design/development: Torre Wenaus
- Full time design/development: Sasha Vanyashin, visiting (April - end September) from Royal Institute of Technology, Stockholm
- Contributing developers from the BNL STAR computing group:
  - Valery Fine
  - Pavel Nevski
  - Victor Perevoztchikov
  - Jeff Porter



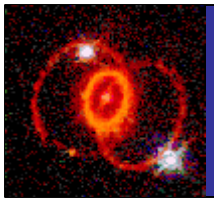
Background

- Analysis frameworks
  - manage the data analysis process from algorithm development through data selection and retrieval, filtering based on physics features, to final results analysis, in interactive and production environments
  - are typically experiment-specific, developed with scarce resources and focused on local needs rather than reuse
  - are highly prized when they are reusable
- This LDRD project gives us the capability to
  - Go beyond narrowly focused tools to provide a generic, reusable analysis framework designed for wide use in similarly complex, data intensive computing environments



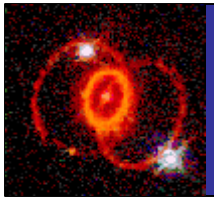
Motivations

- Unprecedented data volume and software complexity in new large HENP experiments (RHIC,LHC…)
  - ⇨ New approaches to analysis and data handling software
  - ⇨ Distributed computing vital and increasingly powerful
- BNL experienced in developing such solutions
  - ⇨ Build upon RHIC experience to develop new analysis tools for wide use in similarly challenging environments
  - ⇨ Focus on new capability; not reinventing the wheel
- Improve depth & visibility of BNL contributions to software, within and beyond HENP community
  - Better positioning for major computing roles



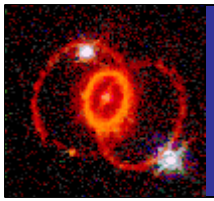
NOVA Goals

- Develop software tools -- a Networked Object-based Environment for Analysis (NOVA) -- for
  - coordination and control of widely distributed analysis development and physics analysis activity
  - distributed management and analysis of very large datasets
  - enhanced robustness, reusability and maintainability of analysis software
- Applicable in many global computing environments (RHIC, LHC, muon collider, climate, environment, …)
  - generic tools not tied to specific implementation choices
  - select, templatable implementations provided such that NOVA can be used as a baseline framework



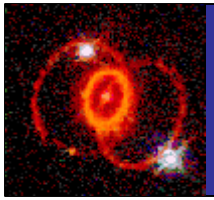
MOKA Design Approach and Architecture

- Small, modular components; application-neutral interfaces
  - Can be used as a coherent framework or in isolation to extend existing analysis systems
- Focused on support for C++ based analysis
  - Widest use for compute-intensive object oriented analysis
- Emphasis on user participation in iterative development; real-world prototyping and testing (STAR, ATLAS)
- Extensive use of existing tools and technologies
  - Must be readily available, true or de facto standards, well supported, widely used or showing good growth



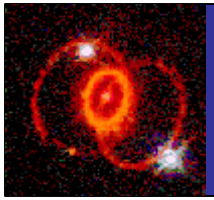
Tools and Technologies

- Third party tools and technologies used in NOVA:
  - ROOT: analysis tools
  - MySQL: relational database for catalogs and state information
  - XML: low-volume data exchange, software source distribution
  - Apache: customizable web server for communication and monitoring
  - CORBA: low-volume interprocess data exchange



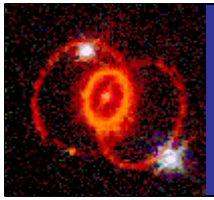
Components and Status

- NOVA components fall into four domains
  - Mobile Analysis: the end-user analysis environment
    - A remote or local analysis environment supporting the complete analysis process
      - Analysis software development
      - Data set selection, data retrieval
      - Bulk data filtering, event selection
      - Production or real-time analysis
      - Results analysis
      - Iteration of the whole process
    - Implementation about to start



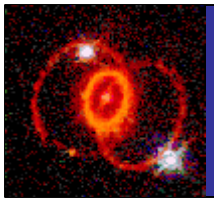
Components and Status (2)

- Data Management
  - Data repository, catalogue, and interface
  - Prototype implemented, except for generic interface
- Analysis Server
  - Central management and execution of analysis
  - Server implementation over existing ROOT foundation underway
- Web Middleware
  - Data exchange and communication tools
  - Most components prototyped



Milestones and Deliverables

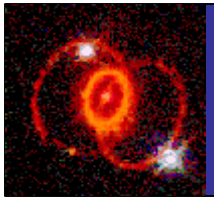
Milestone	Activity	Deliverable	Schedule
1.)	Design, prototyping	Status report	April
2.)	Implementation, testing	Status report and year two plan	August
3.)	Documentation, refinement	Delivery, final report, users manual	September



Budget

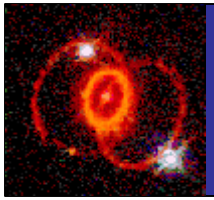
- Project allocation for 1999: $100k

- Sasha Vanyashin ~6mo visit including fringe (36%) and burden (12.3%): $55k
- Torre Wenaus salary fraction: $25k
- Other STAR BNL computing group salary fraction (Fine, Perevoztchikov, Porter): $10k
- Computing facilities and software licenses including 6.5% materials burden: $4k
- Space: $1.4k
- Travel: $4.6k



Other Potential Funding/Leverage

- We are participating in other externally funded developments for which NOVA provides leverage for a BNL role
  - SSI: Scientific Simulation Initiative
    - Participating in ~$2M HENP proposal to SSI focused on distributed data management; strong overlap with NOVA
  - US ATLAS Computing
    - A likely US focus in ATLAS computing is the analysis framework; NOVA strengthens BNL as a participant



Preliminary Year Two Plans

- We anticipate a similar funding request and budget usage profile with a program directed at
  - Identifying and working with users outside HENP to exercise, test and extend the capability of NOVA in non-HENP environments
  - Scalability testing with multi-TB data sets
  - Integration and further development of NOVA components with external projects such as SSI