

APPENDIX A  
STATEMENT OF WORK

**TITLE:** CWS4DB: A Customizable Web Service for Efficient Access to Distributed Nuclear Physics Relational Databases

**TECHNOLOGY AREA:** ST - Science and Technology and Other R&D

**BROOKHAVEN P.I.(s):** Dr. Jerome Lauret  
RHIC/STAR Experiment, Physics Department  
Phone: (631) 344-2450  
Fax: (631) 344-3257  
EMail: jlauret@bnl.gov

**INDUSTRY PARTICIPANT:** Tech-X Corporation

**INDUSTRY PARTICIPANT P.I.:** Dr. Mark L. Green  
5621 Arapahoe Avenue, Suite A  
Boulder, CO 80303  
Phone: (716) 634-3834  
Fax: (716) 634-2481  
Email: mlgreen@txcorp.com

**PROJECT OBJECTIVE:** Tech-X proposes to develop a system providing an efficient access for the High-Energy and Nuclear Physics data stored in distributed heterogeneous relational databases. The system will consist of a generic Web Service, customized client APIs, and tools that facilitate generation of custom client APIs. The access to the data will be mediated by mechanisms that provide data caching and adaptive scheduling of the query plans.

<b>INDUSTRY CONTRIBUTION:</b>	<b>Year 1</b>	<b>Funds-in:</b>	<b>\$75,000.</b>
		<b>In-kind:</b>	<b>\$75,000.</b>
	<b>Year 2</b>	<b>Funds-in:</b>	<b>\$75,000.</b>
		<b>In-kind:</b>	<b>\$75,000.</b>
		<b>Total Funds-in:</b>	<b>\$75,000.</b>
		<b>Total In-kind:</b>	<b>\$75,000.</b>

Attached is a budget breakdown for BNL's work for each year.



# APPENDIX A

## STATEMENT OF WORK

### **PROJECT DESCRIPTION**

An increasing fraction of the data generated in Nuclear and High-Energy Physics is managed in distributed and relational databases. As the size of this data grows and the collaborative nature of these experiments increases, the ability to access differently organized relational databases remotely, efficiently and yet in a user-friendly and interoperable manner is becoming very important. This community lack tools addressing this need and accommodating related challenges.

Tech-X therefore proposes a system to overcome the outlined challenges by bridging relational databases with high-level APIs through Web services. In particular, the distributed and heterogeneous nature of the databases will be addressed by creating a Web service on top of OGSA-DAI, which provides mechanisms coordinating access to diversified data resources. Such a Web service would not be easily integrated into legacy code and applications or present a user-friendly environment for configuration and management. Therefore, the challenge of allowing for high-level queries will be overcome by providing a means to generate customized interfaces on top of the Web service client and provide tools and infrastructure for load balancing, efficient caching, on-demand services and tiered deployment and management. Additionally, we intend to address the challenge of efficiency of data access in the situations when there are many queries of different types over distributed data sources, so that the number of data transfers is minimized. Other factors may also be taken into account when making query plan include workload of database servers, network bandwidth, pattern of requests, and priority of different operations.

### **REPORTING:**

The Tech-X Corporation will prepare the project reports and the final report.

## **ROLES AND RESPONSIBILITIES:**

The STAR Software and Computing (S&C) team will share the responsibilities with Tech-X for the determination of Phase II specific CWS4DB system and load balancing additional requirements and properties definition, the design and implementation of an auto-caching infrastructure, the development of a prototype on-demand data resource node, and a prototype pre-cache capability for production job workflows. This collaborative work is essential for providing a useful, robust, and tested software infrastructure.

The STAR S&C have the capability of supplying a realistic testbed for the CWS4DB framework, along with the the operational experience to critically assess the system performance and efficient data access that is essential for the project success. Furthermore, STAR S&C will provide the a deployment and integration use case for delivering the CWS4DB framework in a smooth and non-disruptive manner by deploying the prototype framework within their development and eventually (after a hardened product is developed) production environments.

## **BENEFITS TO BNL AND PARTICIPANT:**

The Phase I project proved the feasibility and a working prototype of the OGSA-DAI data resource presentation of STAR relational databases. Phase II focus is to migrate the infrastructure and knowledge develop in Phase I into a robust framework for satisfying Nuclear Physics, and therefore STAR, specific needs and requirements into a production environment. While accessing of data in such databases is convenient and available for local users who are familiar with a particular database, the situation becomes more complicated when the databases are distributed and highly heterogeneous. For example, the Calibration and Geometry databases in the STAR experiments are replicated and distributed on fourteen computers at seven major sites around the world. In addition to being distributed, these databases are heterogeneous in nature. They have different content types, data structures, conditions data, and event data etc., and are supported by different vendors, such as MySQL and Oracle.

The importance of this project comes from the fact that a large fraction of the ever-growing data generated by Nuclear Physics experiments is stored in relational databases. For example, relational databases (such as Condition databases, Calibration databases, and Geometry databases) are heavily

used in the STAR experiment. The popularity of relational databases is the reflection of the fact that relational databases are well suited for storing and managing multiple kinds of data, and relational databases are widely supported by the industry. They are robust and have tools and APIs in most popular programming languages.

Tech-X proposes to develop a customizable Web service for efficient access to distributed Nuclear Physics (NP) databases. The proposed system will consist of a generic Web service for accessing arbitrary distributed relational databases, a reference client implemented for the Solenoidal Tracker at the RHIC (STAR) experiment at Brookhaven National Laboratory.

The benefit for Tech-X will be the development of a hardened commercially viable solution for efficient large-scale data access. More specifically a web-grid based service framework utilizing GSI-PKI role-based authorization with scalable on-demand enterprise database and resource hosting for Service Oriented Architecture and/or Virtualization based Information Technology customers is a major benefit. The value added commercial variant of the acquired knowledge, algorithms, and approach are also part of the commercialization strategy.

## **APPLICATION TO DOE PROGRAM(S):**

The proposed application could be used in various virtual organizations where users need to access data from large distributed and heterogeneous relational database efficiently. Nuclear physics, high-energy physics, space science, bioscience, and climate modeling would greatly benefit from the tools developed in the project. Business-to-business applications where customers require easy and efficient access to distributed relational databases are possible commercial applications for the proposed system.

## **ENVIRONMENTAL, HEALTH, AND SAFETY ISSUES:**

None foreseen for this project as this is strictly focused on software development.