Distributed Database Systems

Main Goal

Theoretical aspects and practical applications of NoSQL DBMS.

Threefold Motivation

To business:

In today's enterprises not all use cases lend themselves intuitively to the relational DBMS, neither they need the strictness of ACID property (especially the Consistency and Isolation). But what happens to the large volume of uncontrolled, unstructured, information oriented data explosion happened in enterprises in the last 5 years with the advent of web, digital commerce, social computing etc? Enterprises really don't need RDBMS to store and retrieve them, as the core characteristics of RDBMS do not fit with the nature and usage of this data. Businesses need to find new ways to store and scale large amounts of data. These storage solutions differ quite significantly with the RDBMS model and are known as the NoSQL DBMS.

The students should be able to:

understand the underlying principles of distributed DBMS;
develop specific issues in geographically or architecturally distributed databases;
apply the theory of NoSQL distributed DBMS to build and manage distributed database applications.

To applied science: Research and development in the area of column-oriented databases, distributed algorithms in NoSQL databases, NoSQL data modeling techniques.

Expertise

- Linux system administration;
- Hypertable DBMS;
- Hadoop Map/Reduce techniques;
- C++ programming language;
- Microsoft SQL server;
- Microsoft WCF.

Key Partners

Hypertable Inc., USA

Hildebrand Technology Ltd., UK

TXT e-solutions, Italy

Main Results

This R&D direction was established in 2008 when we were invited to be a member of the following FP7 EU-funded International Research projects:

1. DEHEMS - Digital Environment Home Energy Management Systems Project no. 224609
2. TIPSS - Tools for Innovative Product-Service-Systems for Global Tool and Die Networks, Grant Agreement Number 214794

DEHEMS - Digital Environment Home Energy Management Systems

The Digital Environmental Home Energy Management System (DEHEMS) is an EU funded Framework 7 project investigating ways in which state of the art technologies can be used to improve domestic energy efficiency and hence reduce CO_2 emission. The main objective of DEHEMS is to support households to reduce their energy usage through better management and analysis of their energy consumption. The first and fundamental step in achieving this goal is to have devices for monitoring the energy consumption of household appliances. In DEHEMS project we get energy consumption reading of appliances every six seconds. These readings are then sent to server, which then stores these readings in central database. The database maintains energy consumption record of hundred of households. The server receives six hundred readings per hour for each household.

In order to assess the energy consumption of household appliances, heating and cooling systems we need to have access to their energy consumption data. DEHEMS uses monitored energy consumption data to provide live feedback to household on their energy consumption and provide advice to household on efficient use of energy. Such functionality requires a large amount of energy consumption data of households to be stored and analyzed.

The main issue developed by Ruse University team is

"Extension of Hypertable Query Language with aggregate operations"

One of the goals of DEHEMS project is to work with a great number of households in order to preserve a vast amount of data while providing a possibility for formulating requests to the accumulated data with the purpose of extracting information from them. There exist three key moments, having an effect on the problem of data storage and processing:

1) the data model,
2) the type of requests, and
3) the amount of data.

The data processing system should be able to write down data flowing in with high intensity, to store vast amounts of measured data values, to provide chronological information and to register special requests to data currently stored and filed.

The data extraction service is a module from a comprehensive web service developed by DEHEMS and is part of the raw data processing module (Data Extraction Engine). Its purpose is to extract data from the server system in Manchester, to transfer them to the cluster in Ruse University (consisting of 4 nodes), to write them down under the Hypertable DBMS control, while expanding the language capacity for requests with data aggregation operations (fig. 1). The service has been developed under the ReSTful standard, using advanced innovative technologies.

Many of the NoSQL DBs are based on the DHT (Distributed Hash Table) model, which provides a hash table access semantics. To access or modify any object data, the client is required to supply the primary key of the object, then the DB will look up the object using an equality match to the supplied key. Although query processing and indexing technique is pretty common in RDBMS world, it is seriously lacking in the NoSQL world because of the very nature of the "distributed architecture" underlying most of NoSQL DBs.

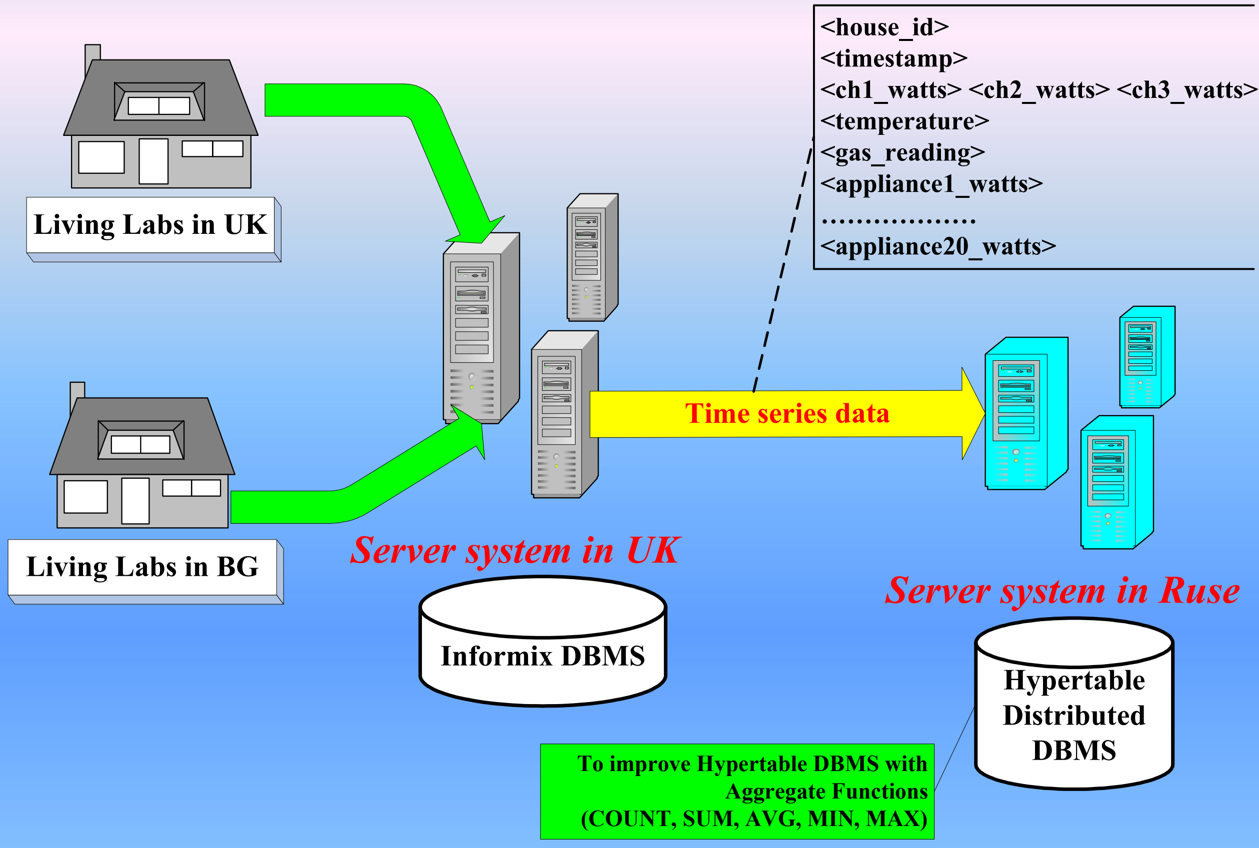

Fig.1. Main flows in the DEHEMS data extraction system

TIPSS - Tools for Innovative Product-Service-Systems for Global Tool and Die Networks

The Tools for Innovative Product-Service-Systems for Global Tool and Die Networks (TIPSS) is an EU-funded Framework 7 project which main objective is to develop suitable tools to enable toolmakers to improve their local and global performance at internationally networked locations. All activities along the value-adding chain can be supported by a Web-based platform to coordinate all forms of forward and backwards collaboration (fig. 2). At the same time, the platform accesses the data collected by the smart tools and the customers' production facilities. Smart tools are injection moulds equipped with state-of-the-art sensor technology that delivers real time data from the production process. The on-line analysis of the gathered data enables an optimized mix of both condition-based and preventive maintenance services, which directly leads to an increase of the overall operational availability of the production cell. In order to monitor the production process, algorithms and methods for the intelligent interpretation of the gathered data are developed within TIPSS, which allow the simulation of different scenarios forecasting the tool's further behavior in operation (fig. 3). Based on those forecasts, errors or downtimes can be interpreted in real time by the toolmaker, thus minimizing the time to respond and the repair time, according to the service level that has been agreed upon the customer.

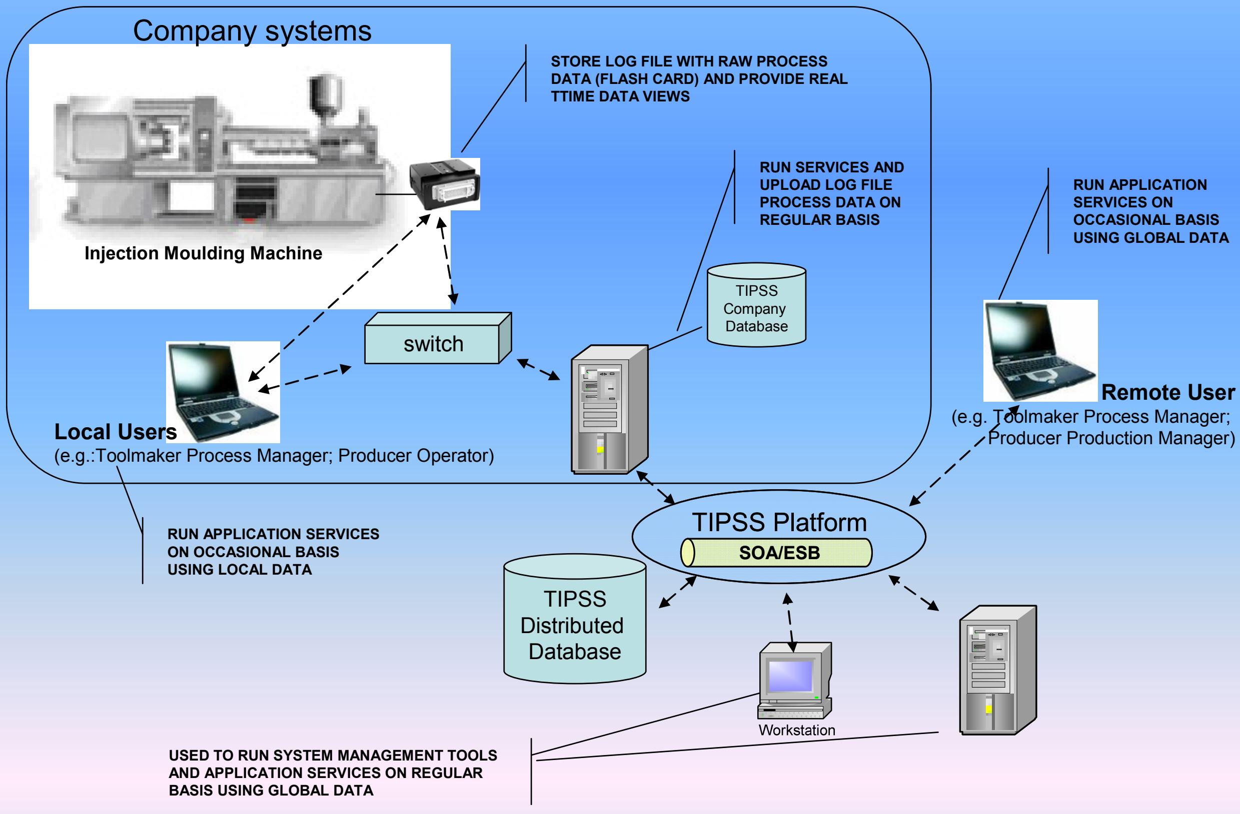

Fig.2. TPSS Platform

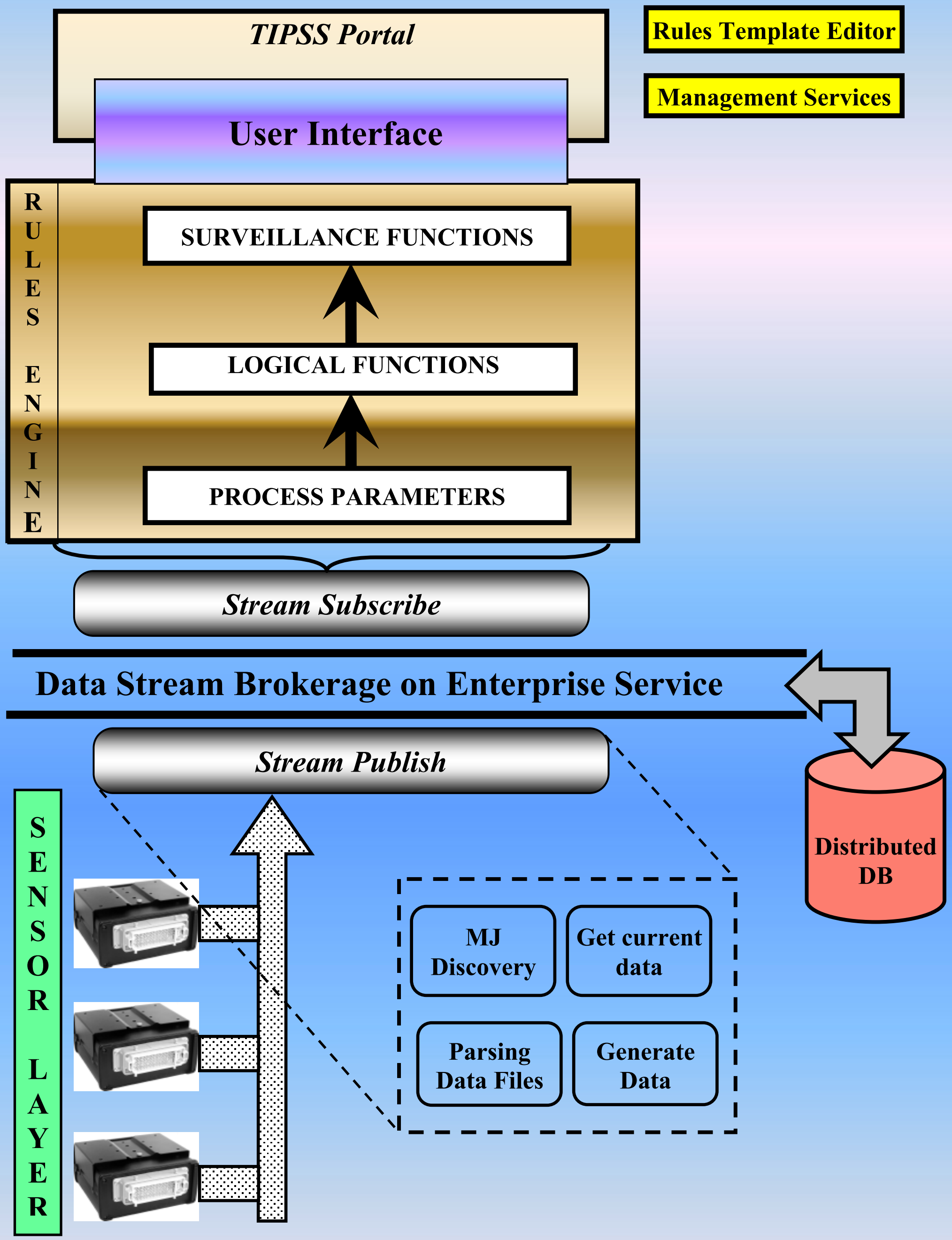

Fig.3. The Architecture of TIPSS Software Platform

The main issues developed by Ruse University team are

" Rule-based Decision Support Tools "

The developed concepts for the decision support tools are based on distributed n-tier architecture. Figure 3 shows the distribution of the most important components across the tiers. The architecture has three tiers – data, service and client. The data tier is built on top of Microsoft SQL Server and takes advantage of the following technologies: Microsoft SQL Server Database Engine, Microsoft SQL Server Analysis Services, Microsoft SQL Server Reporting Services. The service tier includes management services, analysis services and rules engine service. Management services and analysis services are built on top of Windows Communication Foundation, .NET and Internet Information Services. The rule engine is an executable windows service. The client tier includes applications used to interact with the decision support system. The integration with the data management layer takes place in the service tier, respectively in the rule engine service.

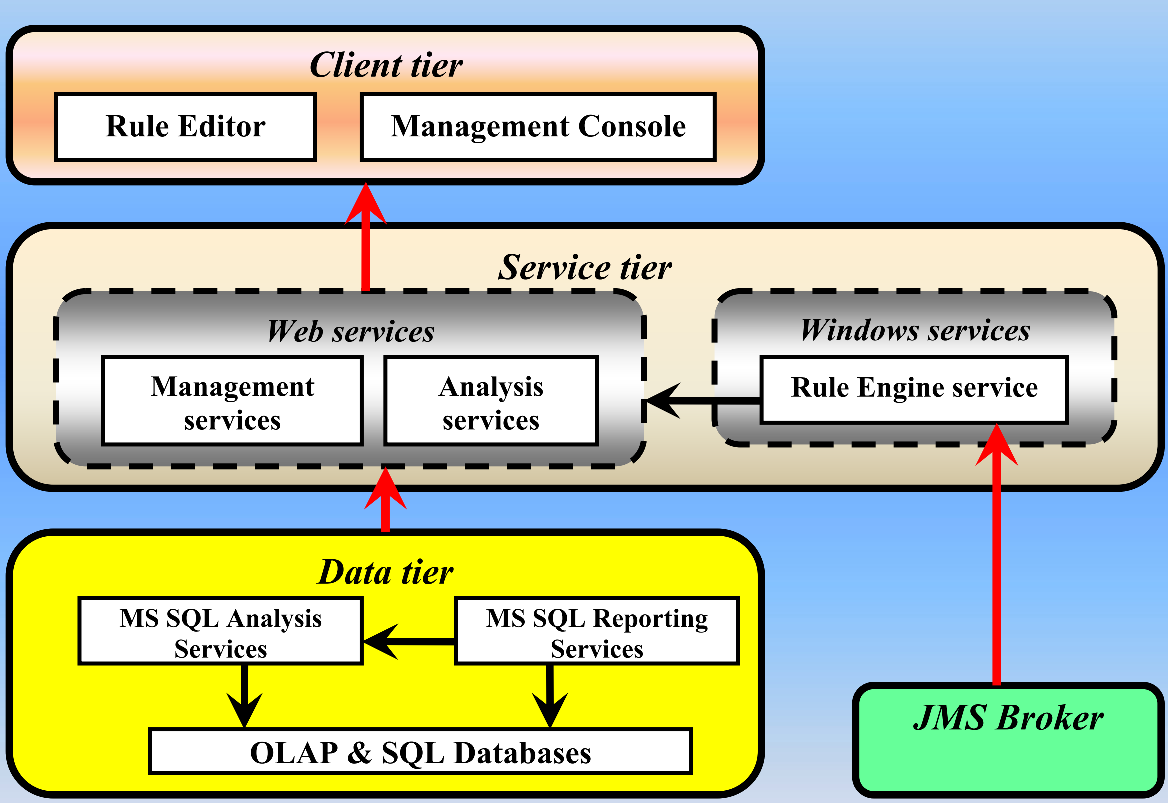

Fig. 3. Distribution of the Decision Support Tools Components across Tiers

Contact Information

Assoc. Prof. Milko Marinov, PhD, mmarinov@ecs.uni-ruse.bg

M.Sc. Eng. Tsvetelin Pavlov, cpavlov@uni-ruse.bg