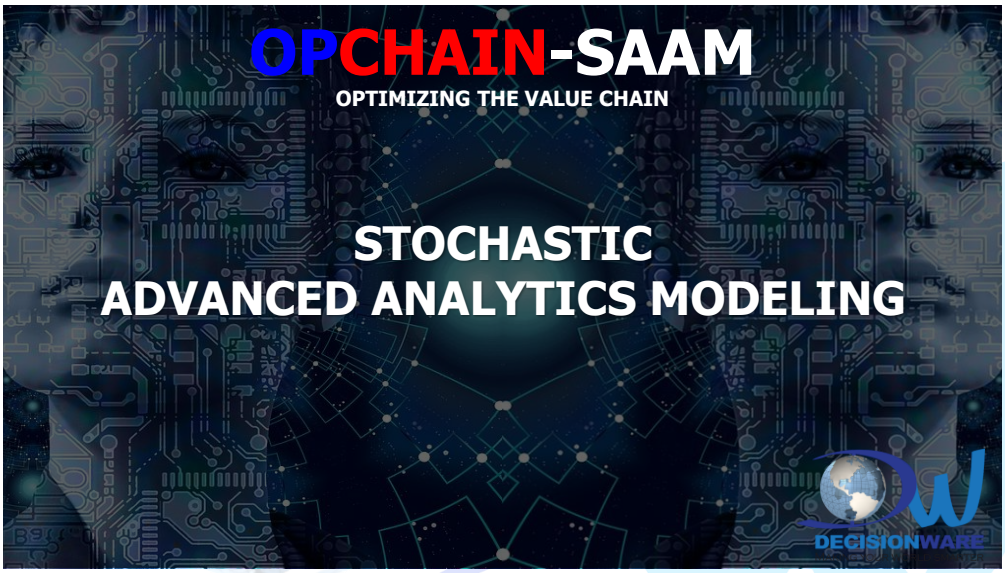
OP-CHAIN

OPTIMIZING THE VALUE CHAIN



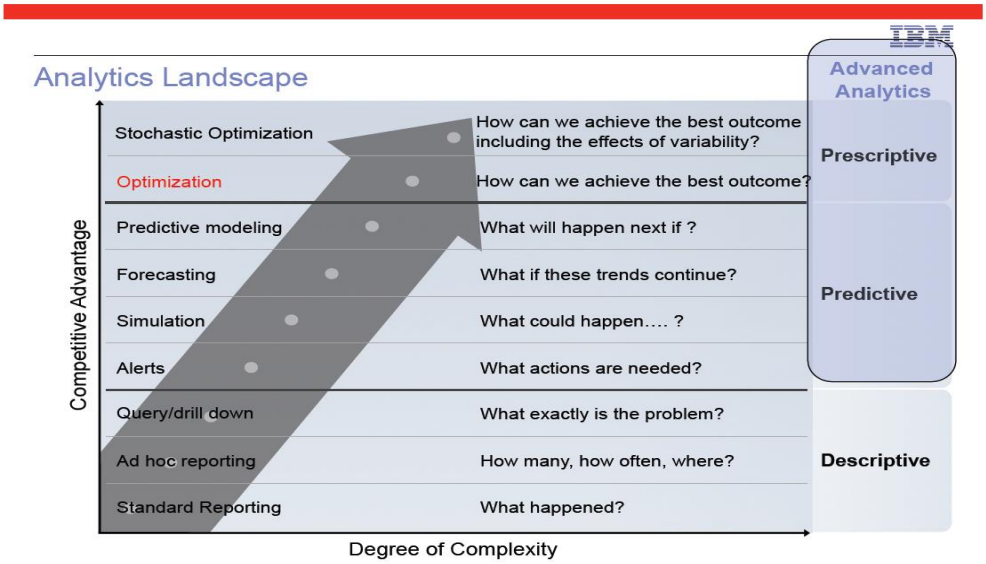
DECISIONWARE

MAKING YOUR WORLD SMARTER

Bogotá, March 2019

1. PREDICTIVE ADVANCED ANALYTICS

Following the publication of the book **"Competing on Analytics"**, write by Tomas Davenport, the applications related with Operations Research, Data Mining, Artificial Intelligence and Advanced Probabilistic Modeling have reorganized about the concept of **Advanced Analytics** (**AA**), which considers three types of analytics: descriptive, predictive and prescriptive. The following image presents the concept given by IBM for the types of applications of AA.



Later, another type of analytics has been included in **AA** which are called as cognitive.

1.1. OPCHAIN-SAAM - STOCHASTIC ADVANCED ANALYTICS MODELING

OPCHAIN-SAAM (**Stochastic Advanced Analytics Modeling**) has been developed by **DW** to support operations of **Predictive Advanced Analytics** in: i) consulting and software companies, ii) analytical departments in large companies and iii) SMEs (Small and Medium Enterprises) on the web.

OPCHAIN-SAAM complements the services offered by **OPTEX Expert Optimization System (OPTEX)** in the development of the algorithms of the so-called **Prescriptive Advanced Analytics**.

1.2. PREDICTIVE ADVANCED ANALYTICS FRAMEWORK

There are many applications that can be implemented using the mathematical models included in **OPCHAIN-SAAM**; in general terms, these applications make up what is known as **Data Mining** (**DM**), today better known as **Advanced Analytics** (**AA**).

Data Mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, artificial intelligence, statistics, optimization and database systems. Typically, **DM** it is integrated by the following conceptual applications (Wikipedia, https://en.wikipedia.org/wiki/Data_mining):

- **Anomaly Detection** (outlier/change/deviation detection): The identification of unusual data records, that might be interesting or data errors that require further investigation; for example, hacker attacks, error data, …

- **Association Rule Learning** (dependency modelling): Searches for relationships between variables. For example, a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket analysis.

- **Clustering** (unsupervised learning)**:** is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data. The goal is to discover the internal pattern in a data set using one, or various, data mining algorithms, that can group entities (samples) based on similarity, and forms thousands of data objects into organized tree to help people view the content. For example: segmentation clients according its commercial behavior.

- **Segmentation & Classification** (supervised learning)**:** Segmentation is the task of discover a mathematical structure to segment entities (samples) in a cluster/segment that may pre-defined or may defined in a clustering process. This process is based on supervised learning models with associated learning algorithms. Its results, mathematical rules and/or equations, are used for classification and regression analysis. Classification is the task of to apply a known mathematical structure to new data to classified in a cluster/segment previously defined. For example, based on a sample of experiments (successes or failures) defined the rules to classify entities don't included in the sample.

- **Identification** (regression): attempts to find a mathematical function (response function) which models the behavior of a system based on continuous state space that is function of exogenous and endogenous variables (data). Its goal is identified the response function that if the "best" based in a key indicators of the goodness of the model. Normally the objective function is to minimize a penalization function, for example to minimize the sum of the square errors. This methodologies are classified as "supervised learning". For example, forecast the demand of a product.

- **Behavior Analysis.** A behavior model is a stochastic model used to represent a stochastic system based on discrete state space; its goal is to forecast the future states. Examples of this type of study are the analysis of the behavior of a: i) client to customer management purposes, and ii) an equipment to establish maintenance policies.

- **Summarization**: is a key data mining concept which involves techniques for finding a compact description of a dataset. providing a more compact representation of the data set, including visualization and report generation. It is used in the analysis of unstructured databases; for example, analysis of email databases.

- **Pattern Recognition**: is the automated recognition of patterns and regularities in data. Pattern recognition is closely related to artificial neural nets and machine learning to discover knowledge

in databases. However, also algorithms of mathematical programming may be used for this purpose. Example images identification

- **Efficiency:** Efficiency is about making the best possible use of resources. Efficient firms maximize outputs from given inputs, and so minimize their costs. By improving efficiency, a business can reduce its costs and improve its competitiveness. The study of operational efficiency is closely related with:
  - Data Envelopment Analysis (**DEA**) and **DEA** is a nonparametric method for the estimation of Pareto production frontiers. It is used to empirically measure productive efficiency of decision-making units (**DMU**s).
  - Stochastic Frontier Analysis (**SFA**) is a method of economic modeling based on regression models, its starting point in the identification of the stochastic production frontier.

  For example: Analysis of efficiency of hospitals

1.3. MATHEMATICAL METHODOLOGIES

OPCHAIN-SAAM is a set of mathematical models oriented to analyzed historic and on-line data sets using advanced analytical tools based on the following mathematical methodologies:

- Machine Learning
  - Generalized Support Vector Machines (**GSVM**)
  - Generalized Support Vector Regression (**GSVR**)
- Artificial Neural Nets (**ANN**)
  - Deep Neural Networks (**DNN**)
- Markovian Process
  - Markovian Nets
  - Markovian Decision Process (**MDP**)
- Clustering Analysis (**CA**)
- Advanced Probabilistic Models (**APM**)
  - Simultaneous Equations Regression
  - S-ARIMAX-GARCH Models
  - Discriminant Analysis (**DA**)
  - Logistic Regression (**LR**)
  - Bayesian Methods (**BM**)
- State Estimation (**SE**)
  - Kalman Filter
    - Standard Kalman Filter (**KF**)
    - Multi-State Kalman Filter (**MS-KF**)
    - Dual Kalman Filter (**D-KF**)
- Bayesian Ensemble of Models (**BEM**)
- Data Envelopment Analysis (**DEA**)

OPCHAIN-SAAM is primarily oriented to support predictive analytical applications. The results of these models are used as input parameters of prescriptive analytical models. All **OPCHAIN-SAAM** models are based on algebraic formulations own of mathematical programming.

1.4. ADVANCED ANALYTICS APPLICATIONS

Below, the algorithms implemented in **OPCHAIN-SAAM** are listed and an overview of its main features:

1. **System Identification:**
   Mathematical models oriented to identify how respond a system (response functions), considering the variables, endogenous or exogenous, which affect it. The results of these models are:
   - Forecasting of variables that represent a system..
   - Estimation of parameters of functions that represent the operation of a system; For example:
     - Markets (demand, prices, market share, elasticity,...).

- Physical systems (variable climatic hydro,...)

The mathematical methodologies used are:

- Advanced Probabilistic Models
- Generalized Support Vector Regression
- State Estimation
- Bayesian Ensemble of Models

2. Segmentation & Classification Models:

The process of classifying entities, known or not known, consists of three steps, seeking to define:

- Clusters (segments/classes/categories) in which it must grouped the.
- Mathematical rules that must meet each training sample to be part of a segment
- Processes to classify new entities in defined clusters

2.1 Cluster Analysis (Unsupervised Learning):

Given a set of entities, Cluster Analysis aims at finding clusters, which are homogeneous and/or well separated. As many types of clustering and criteria for homogeneity or separation are of interest, this is a vast field. Homogeneity means that entities within the same cluster should resemble one another and separation that entities in different clusters should differ one from the other. The following questions may be answered:

- The number of clusters
- The classification of known entities/samples in these clusters.

The mathematical methodology used is:

- Mathematical Programming
  Reference: Hanssen P. and Jaumard, B. "Cluster Analysis and Mathematical Programming". Mathematical Programming"· October 1997.

2.2 Segmentation (Supervised Learning):

Given a set of training samples, each marked as belonging to one cluster of a multiple categories. The actual version of **only** include the segmentation for two categories. A binary segmentation algorithm builds a model that assigns training samples to one category or the other. The following questions may be answered:

- The rules to classified new entities

The mathematical methodologies used is:

- Non-Probabilistic Binary Classifier
  A **GSVM** model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. Mathematical methodology:
  - Generalized Support Vector Machine
- Probabilistic Binary Classifier
  In statistics, binomial regression is a widely used to model a binary dependent variable; there are at least three methodologies to face the problem:
  - Discriminant Analysis
  - Logistic model (Logit model)
  - Probit model.

  These models compute maximum likelihood and it differs in their probabilistic assumptions.

2.3 Classification:

Given a set of new samples, each is classified in a segment using the mathematical rules identified in the segmentation process.

3. Behavior Analysis:

A behavior model is a stochastic model used to represent a stochastic system based on discrete state space; its goal is to forecast the future states. The methodology used is based on Markov Chains (**MC**) which are supported on probabilistic concepts to define state transitions. In addition,

Markov chains have been studied as part of the decision-making processes giving rise to the so-called Markov Decision Process (**MDP**, reinforced learning)
This process involves the following steps:

- Definition of the system states. To do so may be use clustering algorithms
- Estimation of Markov Chain transition probabilities between periods.
- Estimation of state transition probability matrices between periods considering the decision-making process

The following questions may be answered:

- Markov chain transition probabilities
- Optimal policies for management of entities

The mathematical methodology used is:

- Mathematical Programming
  - Markov Decision Process (**MDP**, reinforced learning)

4. **Pattern Recognition**:
It is the automated recognition of patterns and regularities in data. Examples of pattern are: images, voices, sounds, …
The mathematical methodology used is:

- Machine Learning
- Mathematical Programming
- Artificial Neural Nets (this service is on implementation.

5. **Anomaly Detection**:
It is referred to the identification of items or events that do not conform to an expected pattern or to other items present in a dataset, that might be interesting or data errors that require further investigation.
The following questions may be answered:

- Identification of unusual data records that may be: error in data, medical problems, errors in a text, hacker intrusion, … .
- Anomalies are also referred to as outliers, novelties, noise, deviations and exceptions.

The mathematical methodologies used are:

- Machine Learning
- Multi-State Kalman Filter (**MS-KF**)
- Bayesian Optimization

6. **Ensemble Algorithms**
They are meta-algorithms for combining multiple learning algorithms together)
The mathematical methodology used is:

- Bayesian Combination

7. **Operational Efficiency:**
Study of operational efficiency of production units (**D**ecision-**M**aking **U**nits, **DMU**s). Examples of DMUs are: bank branches, professional football, ATM, players, departments, production plants, … ; in general any entity included in a production process.
The following questions may be answered:

- Identification of the most productive DMUs Anomalies are also referred to as outliers, novelties, noise, deviations and exceptions.
- Ordering of the DMUs in accordance with a measure of efficiency
- Borders of production frontiers as a function of the resources used.

The mathematical methodologies used are:

- Data Envelopment Analysis (**DEA**)
- Stochastic Frontier Analysis (**SFA**).

2. TECHNOLOGY PLATFORM

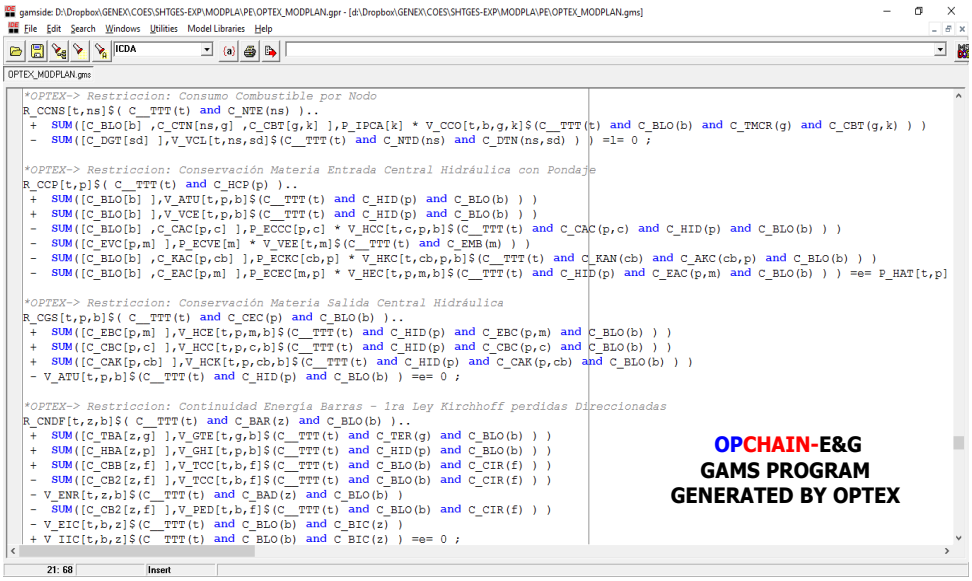
OPCHAIN-SAAM provides the solution to a wide range of strategic, tactical and operational problems considering many aspects of the different types of problems and its components. **OPCHAIN-SAAM** can be used to support consultancy work, end customer studies, or in the "cloud" on an optimization server to directly support decision-making processes.

OPCHAIN-SAAM was implemented using **OPTEX Optimization Expert System**, it can produce program algorithms in various optimization technologies. This implies that it inherits all the characteristics of **OPTEX.** For more information:
https://www.linkedin.com/pulse/optex-optimization-expert-system-new-approah-make-models-velasquez/

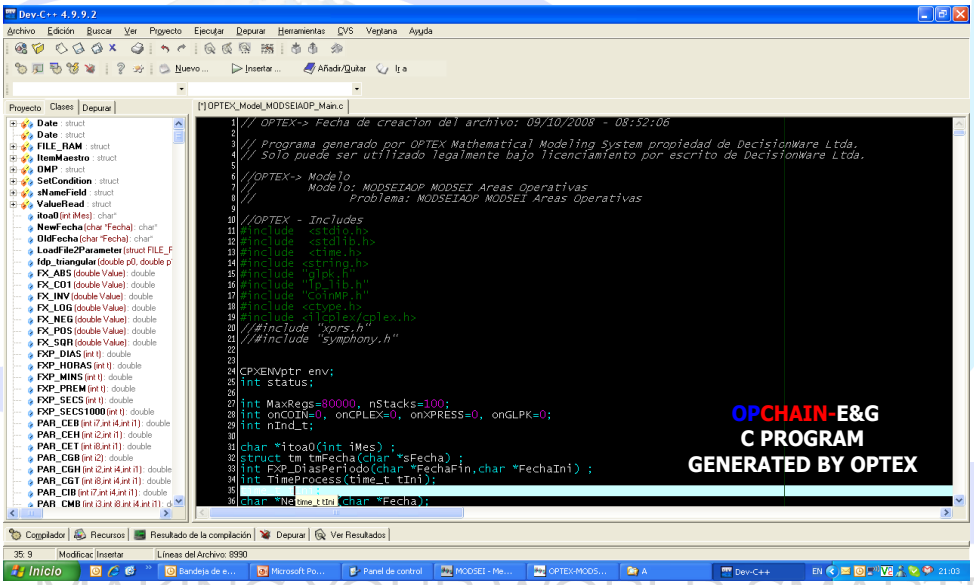
2.1. BASIC ALGORITHMS

OPTEX separates completely the mathematical models of optimization technologies, storing the models in normalized database, allows the user to interact with **OPCHAIN-SAAM** models to update the equations according to the changes in its modeling environment. **OPCHAIN-SAAM** may include services to design, to implement, to support and to solve quickly optimization models in multiples optimization technologies like **GAMS**, **IBM ILOG CPLEX OPTIMIZATION STUDIO**, **MOSEL**, **C**, **AMPL**, **GMPL** … .

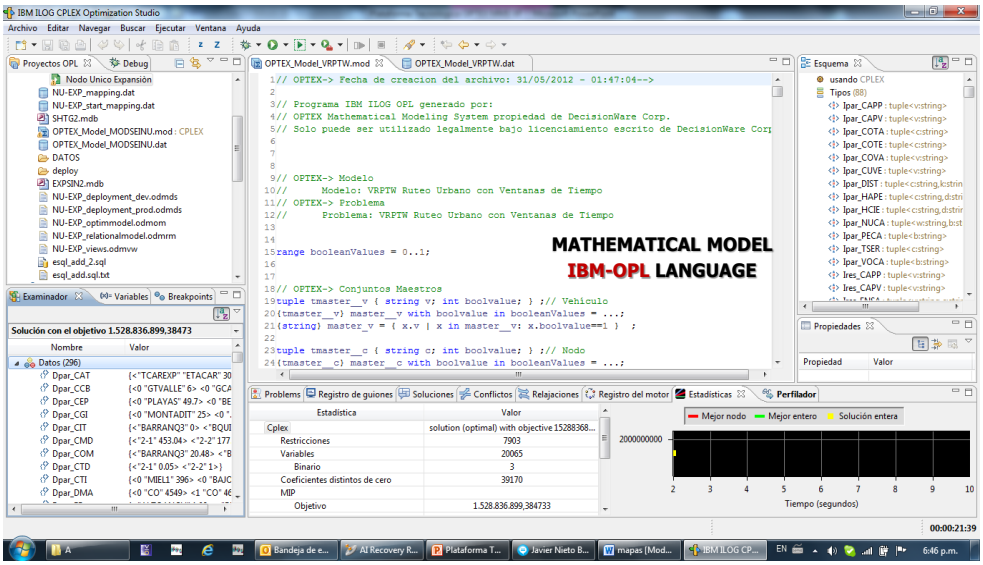
Next image presents a **GAMS** and a **C** computer programs generated by **OPTEX**.



Next image presents a **C** computer programs generated by **OPTEX**.

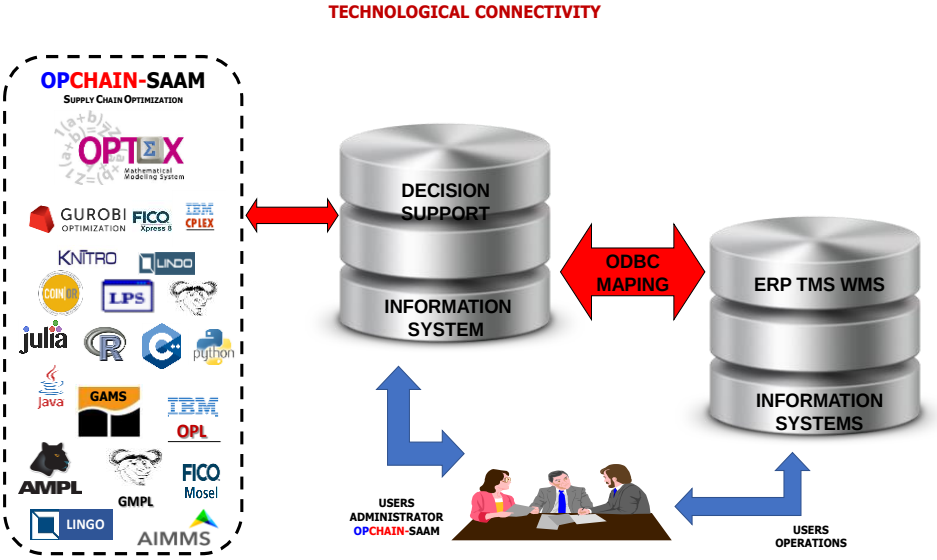


Next image presents a **IBM-OPL** computer programs generated by **OPTEX**.

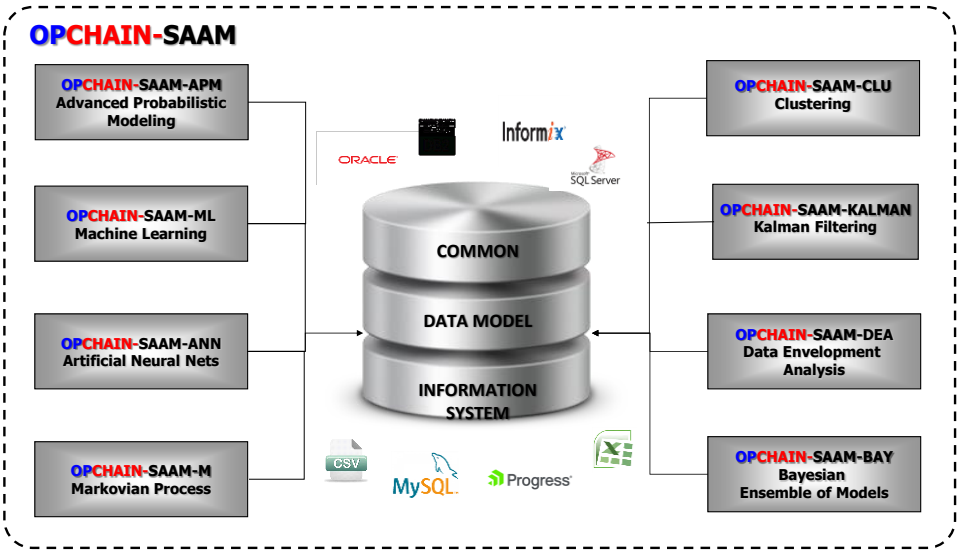


2.2. INFORMATION SYSTEMS

As a complement to other computer systems that require organizations, **OPCHAIN-SAAM** architecture is open (the user knows the data-model) to facilitate their integration with other solutions and/or existing technological tools in organizations (**ERP, TMS, WMS, GIS**).



OPCHAIN-SAAM integrate input-output data models around a common data-model, it permits to connect them automatically through the database.



The data base may be installed any **SQL** (**S**tandard **Q**uery **L**anguage) database, including **DBF**, **EXCEL** and **.csv** text files.

OPCHAIN-SAAM takes advantage of the facilities of visual environments found in personal computers, the ever-greater speed offered by the new family of servers and technological advances in mathematical optimization. The basic user interface (generate by **OPTEX**) provides an easy environment and the ability to integrate **OPCHAIN-SAAM** with the requirements of each client.

3. OPCHAIN-SAAM APPLICATIONS

Below, some applications developed using the **OPCHAIN-SAAM** algorithms are:

1. **OPCHAIN-DCO-DEM**: Scientific Marketing Advanced Demand Chain Optimization
   https://www.linkedin.com/pulse/scientific-marketing-advanced-demand-chain-jesus-velasquez/

2. **OPCHAIN-DCO-MKS**: Market Modeling Via Syndicated Databases
   https://www.linkedin.com/pulse/market-modeling-via-syndicated-databases-case-jesus-velasquez/

3. **OPCHAIN- MS-KF**: Dynamic Machine Learning using a Multi-State Kalman Filter
   https://www.linkedin.com/pulse/dynamic-machine-learning-using-multi-state-kalman-filter-velasquez/