

Distributed Scoring Using PMML

Bill Hosken and Bernard Scherer

Abstract:

As modeling and direct marketing continue to converge, many companies are faced with managing multiple models and executing them in batch across multiple titles and companies. SPSS worked with a division of Experian that provides marketing solutions for the business-to-consumer catalog industry. Experian needed a system that would satisfy the following requirements:

- The ability to deploy billions of scores.

- The ability to manage multiple models from hundreds of titles.

- An open system that could utilize multiple machines.

- The solution must be flexible in that it was scalable and could handle high volume at high speed.

In this case, the Clementine data mining application was used to create models, which generated PMML. The PMML was then processed by the High Speed Scoring Engine (HSSE) for distributed processing. This talk discusses the architecture of the system and lessons learned.