

招待講演

State of the art data-intensive sources and tools evaluated for domain scientist and analyst productivity and serendipity

Eng Lim Goh, Ph.D.

Senior Vice President & Chief Technology Officer, SGI

We will be highlighting, for domain scientists and analysts, some of our recent data-intensive evaluations that range from:

- high data-rate instruments to extreme capacity PCs; - looking for needle in a haystack to relationships between hay in a stack;
- Social media, crowd-sourcing to massive-scale graph analytics; - current high dynamic range computers to the emerging quantum computers; - the mobile to wearable; is the embeddable next?

ここでは、各分野の研究者およびアナリストのために、最近のデータ処理の動向について、以下の内容で講演を行います。

- ・ 高速データ転送能力を持ったセンサー機器と大容量メモリ搭載 PC;
- ・ 干し草の山の中の干し草同士の関係と干し草の山の中から針を探すこと;
- ・ ソーシャルメディア、クラウドソーシングと大規模グラフ解析;
- ・ 現在の高いダイナミック・レンジ・コンピューターと量子コンピューター;
- ・ 携帯電話とウェアラブル端末; その次は、埋め込み型へ?

1. Summary

Recent data-intensive or Big Data sources, software tools and hardware platforms are evaluated for the domain scientist and analyst.

Volume sources include massive archives such as global weather and climate data collected and simulated over the decades. Velocity instruments range from desktop DNA sequencing machines to the planned Square Kilometer Array. Variety sources are relationships rich such as high-dimensional data from the internet and social media or about the human brain and genetic interactions. Associatively, as we observe the migration from the personal desktop to the portable, the mobile and the wearable, we consider the implications of interconnected digital exhaust if the embeddable is next.

Software tools for the data analyst and scientist range from those for search, discovery to serendipity. At one end of this spectrum, you know the question and also the answer but you still want to run your query to obtain validation or better precision. At the other end, you may not even know what specific question to ask yet; all you may have is just a hunch desiring insight. That is, you go from looking for needle in a haystack to relationships between hay in a stack.

Just as tools vary with the state of the data analyst or scientist, different computer systems are also employed. When you are in a more specific state, scale-out clusters built with the Hadoop File System running MapReduce is a growing example. However, when you are in a less specific, more qualitative or even at an early gut-feel state, extreme scale-up memory systems for real-time analyst-in-the-loop visualization of graph and machine learning algorithms are becoming a competitive advantage. Additionally, the emerging classes of quantum computers are also evaluated.