Fundamental challenges to computation in Big Data Era

Irwin King

The 2nd HKUST-USC Joint Workshop
on Big Data Applications

7 Giants of Data

1. Basic statistics: means, covariances, etc.
2. Generalized N-body problems: distances, geometry, etc.
3. Graph-theoretic problems: discrete graphs
4. Linear-algebraic problems: matrix operations
5. Optimizations: unconstrained, convex, etc.
6. Integrations: general dimension
7. Alignment problems: dynamic programing, matching, etc.

Alexander Gray, GIT

7 General Strategies

- Divide and conquer, indexing trees
- Function transforms (series)
- Sampling (Monte Carlo, active learning), reduction, projections, etc.
- Locality (caching)
- Streaming/online techniques
- Parallelism
- Transformations between problems

Alexander Gray, GIT

Duality

- Macro vs. Micro
- Local vs. Global
- Parallelism vs. Distributed
- Theoretical vs. Applied
- Real vs. Virtual
- Precise vs. Approximation
- Dynamic vs. Static