**Cheat Sheet: Improving SVM Efficiency on Large Datasets**

**1. Utilizing Distributed Compute**

* **Description**: Leverage distributed SVMs that parallelize processing across multiple processors or nodes.
* **Example**: Use frameworks like **Apache Spark’s MLlib** to handle large-scale data efficiently.

**2. Using Kernel Approximation**

* **Description**: Instead of using computationally expensive kernels directly, approximate them to maintain efficiency while allowing for non-linear decision boundaries.
* **Techniques**:
  + **Random Fourier Features** for approximating RBF kernels.
  + **Polynomial kernel approximations** for polynomial decision boundaries.

**3. Subsampling the Data**

* **Description**: Instead of random sampling (which may lose critical support vectors), apply techniques to select a smaller, representative dataset.
* **Method**:
  + **Core-set selection techniques**: Identify a smaller set of points that can approximate the decision boundary effectively.