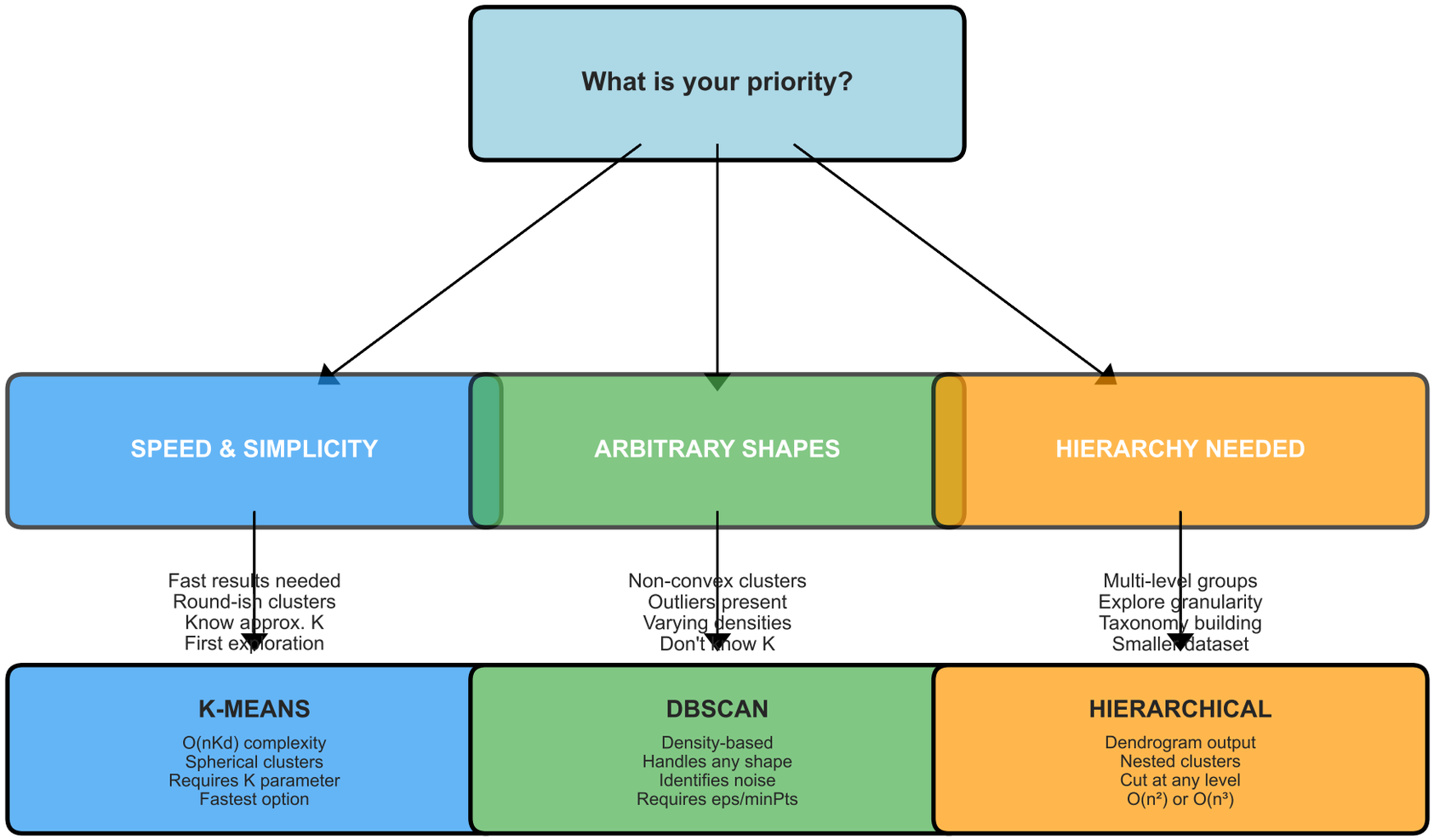
When to Use Which Clustering Algorithm: Decision Framework



Additional Considerations

```
Dataset Size: Very large (>100K points) → MiniBatch K-means; Small (<10K) → Hierarchical feasible
Outliers Critical: Fraud detection, anomaly detection → DBSCAN preferred
Soft Assignments Needed: Mixed populations, uncertainty quantification → GMM (Gaussian Mixture)
High Dimensions: d>20 → Curse of dimensionality affects distance; Consider dimensionality reduction first
Reproducibility: Random init sensitivity → Use K-means++ or fixed seed; DBSCAN/Hierarchical deterministic
Production Deployment: Streaming data → BIRCH; Real-time → K-means; Batch → Any algorithm suitable
```

Principle: Start simple (K-means), upgrade if needed (DBSCAN for shapes, Hierarchical for structure)