

Özellik\Kütüphane	Dask	Vaex	Polars	Modin	Koalas	PySpark
API Uyumluluğu	Pandas API'ye büyük ölçüde uyumlu	Pandas API'ye kısmen uyumlu	Pandas API'ye kısmen uyumlu	Pandas API'ye büyük ölçüde uyumlu	Pandas API ve Spark Data Frame API'sine büyük ölçüde uyumlu	Spark DataFrame API ile uyumlu
Performans	Ortak, paralel ve dağıtık işlemlerle hızlandırılabilir	Çok hızlı özellikle büyük veri setleri için	Çok hızlı, Rust dilinde yazıldığı için yüksek performans sunar	Orta, Dask veya Ray backend'leri ile hızlandırılabilir	Yüksek, Spark tabanlı olduğu için büyük veri üzerinde iyidir	Yüksek, büyük veri setlerinde en iyi performans sunar
Bellek Kullanımı	Optimize edilebilir, ancak tüm veri bellekte tutulabilir	Belleği verimli kullanır, disk tabanlı işlemler yapabilir	Belleği verimli kullanır, paralel işlemlerle optimize edilebilir	Optimize edilebilir, Dask veya Ray backend'leri ile	Orta, Spark'ın bellek yönetim özelliklerine bağlıdır	Yüksek, ancak dağıtık sistemlerde iyi yönetilir
Paralel İşleme	Evet, yerel veya dağıtık olarak	Evet, ancak sınırlı ve genellikle tek makine üzerinde	Evet, çok çekirdekli işlemcilerde etkili	Evet, Dask veya Ray ile yerel veya dağıtık	Evet, Spark tabanlı olduğundan güçlü paralel işleme sahiptir	Evet, Spark'ın temel özelliği
Büyük Veri Desteği	Dağıtık yapıda büyük veri setlerini işleyebilir	Çok büyük veri setleriyle çalışabilir	Var ancak genellikle bellek içinde çalışır	Dask veya Ray ile büyük veri setleri işlenebilir	Spark'ın büyük veri işleme özelliklerini kullanır	Evet, büyük veri için optimize edilmiştir

Kolaylık	Orta, Pandas'a benzer ancak dağıtık işlemler için ekstra adımlar gerekebilir	Kolay, büyük veri setleri için optimize edilmiş kullanımı kolay araçlar sunar	Orta, Pandas'a benzer ancak bazı farklılıklar mevcut	Kolay, Pandas kullanıcıları için çok benzer	Kolay, Spark'ın gücünü Pandas kullanıcılarına sunar	Orta, ancak büyük veri işleme konusunda deneyim gerektirir
Yazılım Desteği	Python, NumPy, Pandas, Scikit-learn	Python	Python, Rust	Python, Pandas	Python, PySpark, Pandas	Python, Java, Scala, R
Dağıtık İşlem	Evet, Dask ile doğal olarak desteklenir	Hayır, tek makine üzerinde çalışır	Hayır, yerel paralel işlemler	Evet, Dask veya Ray ile dağıtık işleme yapılabilir	Evet, PySpark ile dağıtık işleme sağlar	Evet, dağıtık işleme için en güçlü araçlardan biridir
En İyi Kullanım Alanı	Büyük veri setlerinde Pandas işlemlerini dağıtık olarak yapma	Çok büyük veri setleriyle tek makine üzerinde hızlı analizler	Yüksek performans gerektiren veri işlemleri	Pandas işlemlerini hızlandırma	Pandas ve Spark arasında köprü olarak büyük veri işleme	Büyük veri setlerinin dağıtık işlenmesi