**CSV, Parqeut ve Orc Arasındaki farklar.**

Comma Separated Values (CSV) – Virgülle Ayrılmış Değerler  dosyası, bir veri listesi içeren düz metin dosyasıdır.Verileri ayırmak için çoğunlukla virgül karakterini kullanırlar, ancak bazen noktalı virgüller gibi diğer karakterleri kullanırlar. Karmaşık verileri bir uygulamadan CSV dosyasına aktarabilir ve daha sonra bu CSV dosyasındaki verileri başka bir uygulamaya aktarabilirsiniz. Uygulamalar arası veri transferi için de kullanılır.

Parquet maksimum şema esnekliğine izin vermek yerine, sorgu hızlarını artırmak ve disk i/o’sunu azaltmak için kullanabileceğiniz şema türlerini optimize etmeye çalışır. Parquet, sütunlarda iç içe tiplere izin vererek geleneksel sütun bazlı tutmanın bazı zayıflıklarının üstesinden gelmeye çalışır. Böylece teknik olarak bir dizi olan bir sütuna ya da aslında birkaç sütun olan bir sütuna sahip olabilirsiniz.

ORC , Parquet’nin sütun biçimini paylaşır ancak bazı farklılıklar vardır. Başlıca farklılıklardan biri iç içe geçen veri türlerine izin veren listeler ve haritalar gibi karmaşık türleri destekler. Sıkıştırma söz konusu olduğunda, ORC’nin verileri Parquet’den daha verimli bir şekilde sıkıştırdığı söylenir, ancak bu verilerinizin nasıl yapılandırıldığına bağlıdır.

**1 TB civarı veriniz olsaydı hangisinde saklardınız? Neden?**

ORC’yi seçerdim. Sebebi ise;

Daha iyi sıkıştırma sunması bu bize daha fazla alan verimliliği katar, ne kadar alan ayırırsak o kadar hızlı sonuç alırız.

 Aynı zamanda daha iyi şema gelişimi desteği de sağlar