



File Formats in Hive

File Formats in Hive

```
1$8$31397$R$F$2$Emergency$2$51-60$4911.0$0-10$radiotherapy$Extreme
56$8$33340$R$F$4$Emergency$2$31-40$5130.0$21-30$radiotherapy$Moderate
179$8$117334$Q$F$4$Trauma$4$31-40$4244.0$21-30$gynecology$Moderate
510$8$52406$R$F$2$Trauma$2$71-80$4782.0$11-20$gynecology$Moderate
514$8$52406$Q$F$4$Urgent$2$71-80$4912.0$21-30$gynecology$Moderate
535$8$90761$R$F$3$Trauma$2$41-50$4757.0$21-30$gynecology$Minor
598$8$92488$R$F$4$Emergency$6$51-60$4876.0$51-60$radiotherapy$Extreme
616$8$100741$R$F$2$Trauma$2$31-40$4739.0$21-30$gynecology$Moderate
1000$8$29799$Q$F$3$Emergency$3$51-60$4694.0$31-40$gynecology$Extreme
1107$8$28680$R$F$3$Trauma$2$31-40$5108.0$21-30$gynecology$Minor
1133$8$95836$Q$F$2$Trauma$4$41-50$6454.0$31-40$gynecology$Moderate
1354$8$9873$R$F$2$Trauma$3$51-60$4721.0$31-40$gynecology$Extreme
```

File Formats in Hive

```
1$8$31397$R$F$2$Emergency$2$51-60$4911.0$0-10$radiotherapy$Extreme
56$8$33340$R$F$4$Emergency$2$31-40$5130.0$21-30$radiotherapy$Moderate
179$8$117334$Q$F$4$Trauma$4$31-40$4244.0$21-30$gynecology$Moderate
510$8$52406$R$F$2$Trauma$2$71-80$4782.0$11-20$gynecology$Moderate
514$8$52406$Q$F$4$Urgent$2$71-80$4912.0$21-30$gynecology$Moderate
535$8$90761$R$F$3$Trauma$2$41-50$4757.0$21-30$gynecology$Minor
592488$R$F$4$Emergency$6$51-60$4876.0$51-60$radiotherapy$Extreme
60741$R$F$2$Trauma$2$31-40$4739.0$21-30$gynecology$Moderate
60799$Q$F$3$Emergency$3$51-60$4694.0$31-40$gynecology$Extreme
628680$R$F$3$Trauma$2$31-40$5108.0$21-30$gynecology$Minor
62895836$Q$F$2$Trauma$4$41-50$6454.0$31-40$gynecology$Moderate
1354$8$9873$R$F$2$Trauma$3$51-60$4721.0$31-40$gynecology$Extreme
```

Disk
Utilization

I/O
Performance



File Formats in Hive

- ORC
- SequenceFile
- Parquet



ORC File Format

- Optimized Row Columnar
- Binary format
- Compressed columnar format



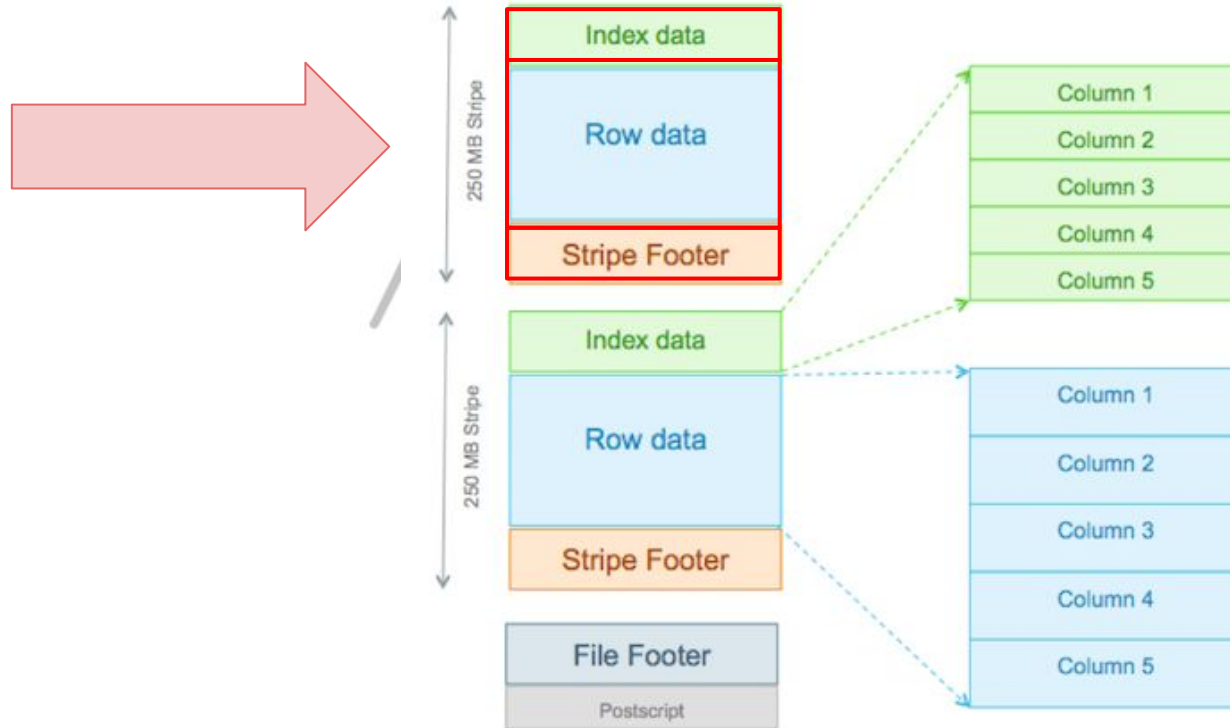
ORC File Format

Id	Name	Department	Salary
AV003	Jack	Technology	1000
AV053	Samantha	Admin	1200
AV062	Kim	Operations	900

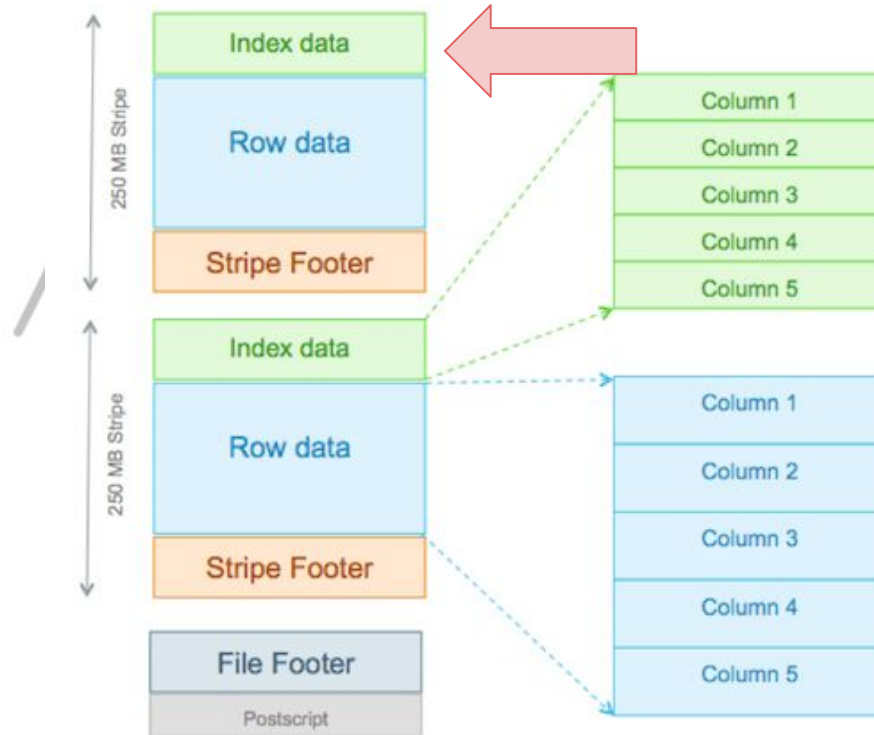
ORC File Format

Id	Name	Department	Salary
AV003	Jack	Technology	1000
AV053	Samantha	Admin	1200
AV062	Kim	Operations	900

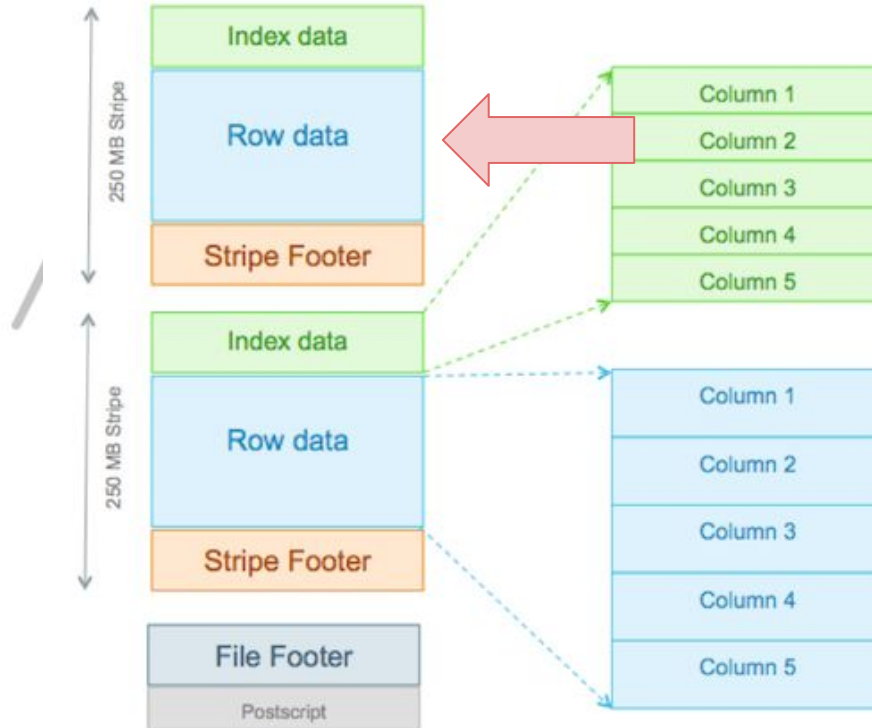
ORC File Format



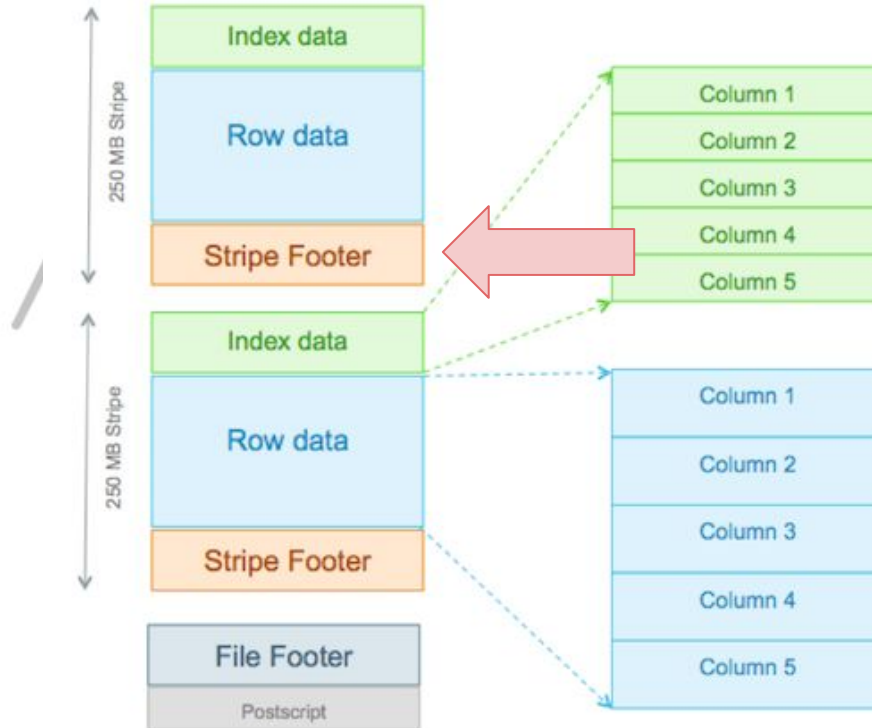
ORC File Format



ORC File Format



ORC File Format



Parquet File Format

- Columnar storage
- Compressed format



SequenceFile Format

- Binary key-value pairs
- Compressed format
- Row level storage



