

# /arrow + parquet



	session_id	timestamp	source_ip
Row 1	1331246660	3/8/2012 2:44PM	99.155.155.225
Row 2	1331246351	3/8/2012 2:38PM	65.87.165.114
Row 3	1331244570	3/8/2012 2:09PM	71.10.106.181
Row 4	1331261196	3/8/2012 6:46PM	76.102.156.138



Traditional Memory Buffer			Arrow Memory Buffer		
Row 1	1331246660		session_id	1331246660	
	3/8/2012 2:44PM			1331246351	
	99.155.155.225			1331244570	
Row 2	1331246351			1331261196	
	3/8/2012 2:38PM		timestamp	3/8/2012 2:44PM	
	65.87.165.114			3/8/2012 2:38PM	
Row 3	1331244570			3/8/2012 2:09PM	
	3/8/2012 2:09PM			3/8/2012 6:46PM	
	71.10.106.181		source_ip	99.155.155.225	
Row 4	1331261196			65.87.165.114	
	3/8/2012 6:46PM			71.10.106.181	
	76.102.156.138			76.102.156.138	

- `{arrow}` evaluates lazily by default
- Verbs: `filter`, `select`, `mutate`, `join`, `distinct`, `group_by` + `summarize`, and `across`
- execution only runs on `dplyr::collect()`
- massive performance gains using parquet files and smart partitioning



# /shiny architecture

