ByteDance 字节跳动

BERT-TF-FP32 REPORT (STC)

Graph Compilation Results		
Compile Duration	Compile Precision	Subgraph Coverage
348.77504	fp16	99.69

Performance Results			
BS	AVG_Latency	P99_Latency	QPS
2	19.5	20.13	820.73

Accuracy Results (OPEN_SQUAD)			
Data Percent	100	Mean Diff	0.02251
Exact Match	79.45901		
F1 Score	86.45179		