Task Pipelining 
 

Adding Receivers 
 

Discretized Stream (DStream) 
• Core Spark Streaming abstraction 
• Micro-batches of RDDs 
• Operations similar to RDD – o...

Execution Optimizations 
 

Adding Workers 
 

Spark + AWS + Lambda 
 

Spark Streaming + Kinesis Architecture 
 

Lambda Architecture Overview 
• Batch Layer 
– Immutable, 
Batch read, 
Append-only write 
– Source of truth 
– ie. HDFS 
...

Spark + Lambda + GraphX + MLlib 
 

Throughput and Pricing 
 Stream Worker/Processor Failure 
• No problem! 
• DStream RDD partitions will be recalculated from lineage 
• Causes blip ...