# Spark vs Athena/Trino: Performance Analysis with Window Functions

Test Environment

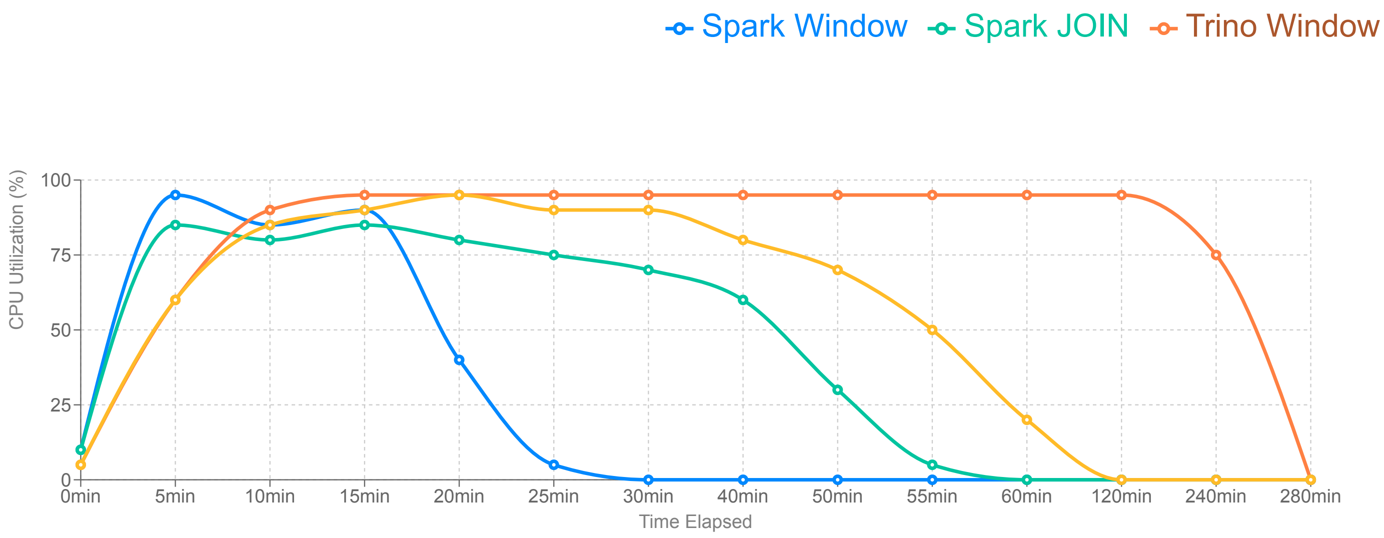
**Dataset Size:** 2.4 billion rows (35M staging + 2.37B main table)

**Infrastructure:** EMR cluster with 1 core node and 1 task node

**Node Configuration:** 4 CPUs and 15GB memory per node

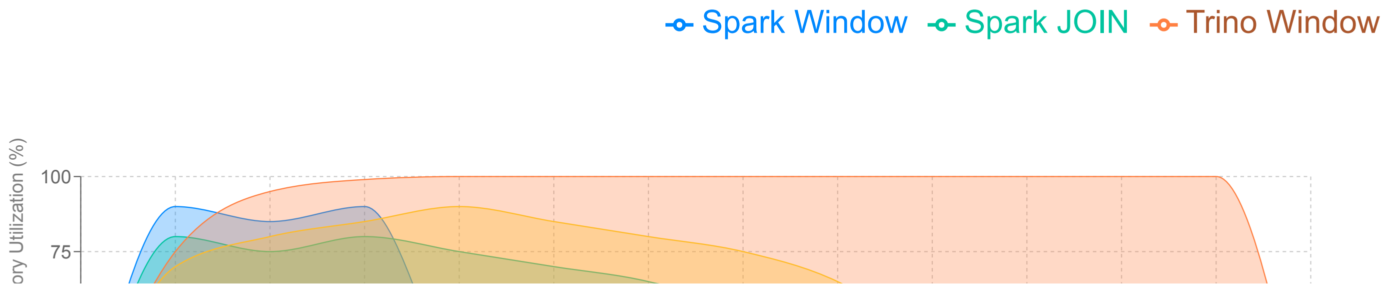
**Query Type:** Window function using row_number() to find most recent record per ID

## CPU Utilization Over Time



*Spark shows clear execution phases with efficient resource release, while Trino exhibits sustained high CPU utilization*

## Memory Utilization Over Time


Spark Window
Spark JOIN
Trino Window
ory Utilization (%)