

# Unified Data & Compute Intensive Platform

UiT The Arctic University of Norway  
Thursday, 25 April 2019

# The Need for Speed

C/Assembly MPI OpenMP UPC++

# Development Costs

Hadoop Spark Flink Storm

# Programmability

Chapel Julia Legion Pregel

# Performance meets Programmability

Husky Ray PlinyCompute HPAT

# The Secret Sauce?

Memory & Serialization

# The Secret Sauce?

Asynchronous Contention-Free

# Rich Interface

BSP SSP Actor Model

# The Art of Abstraction

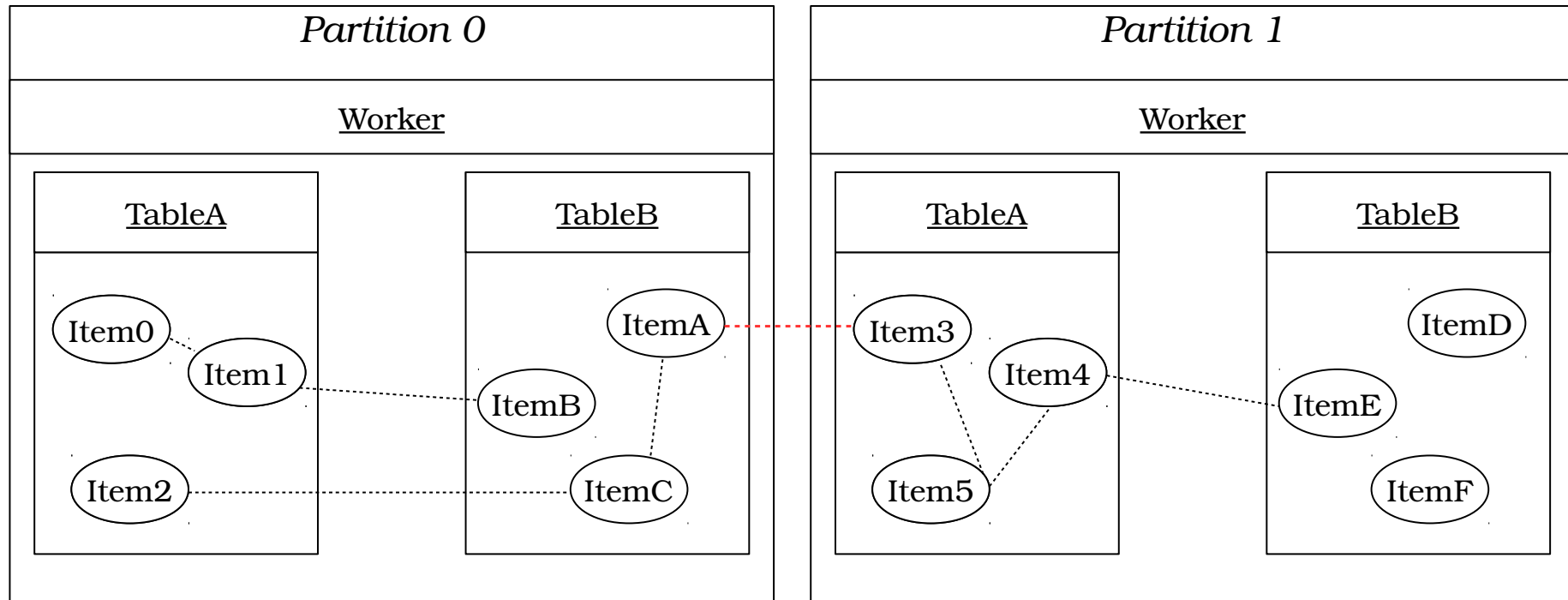
Making Communication Explicit

# Saddlebag

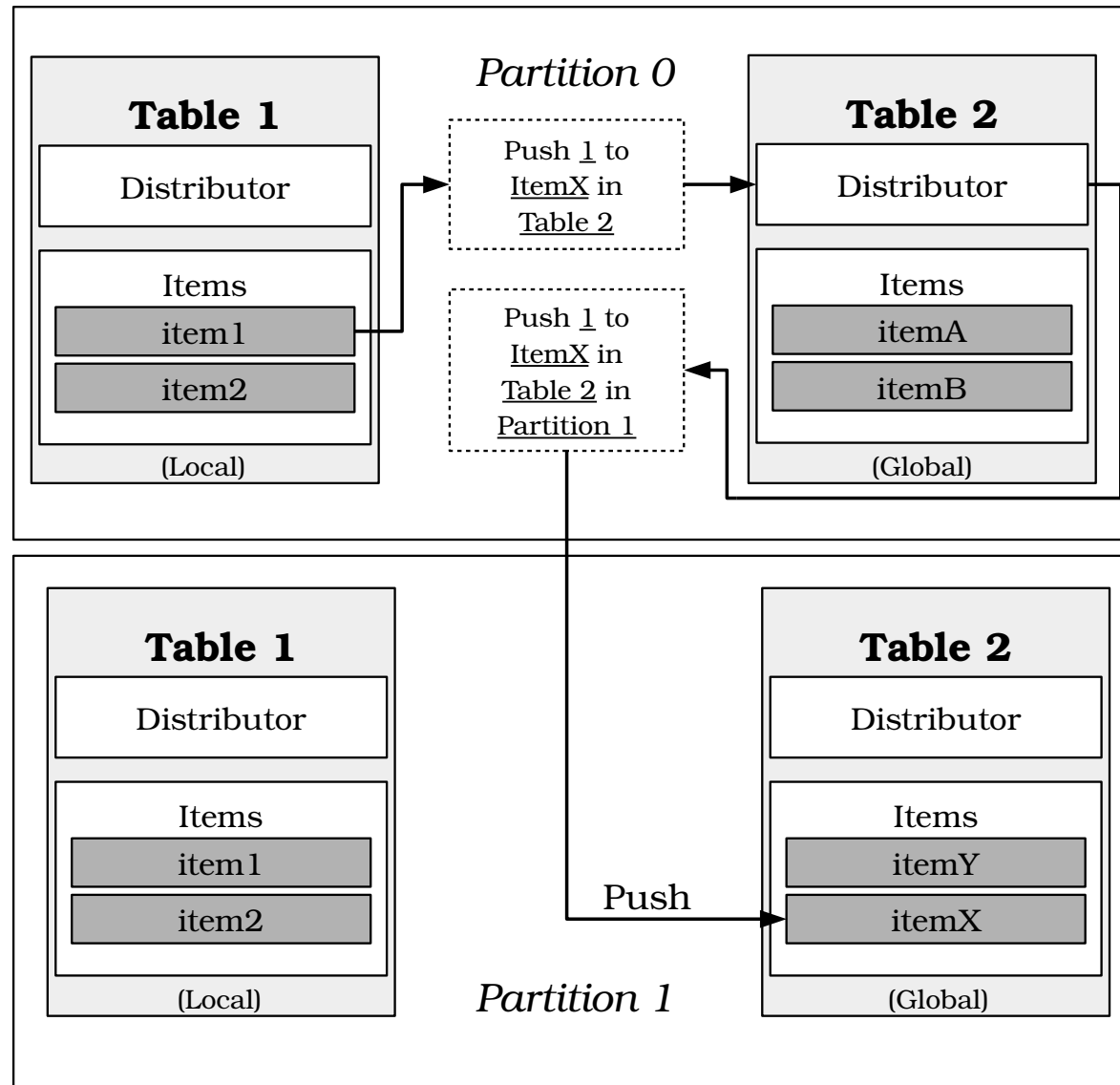
Data-Intensive & Compute-Intensive

# Ingredients?

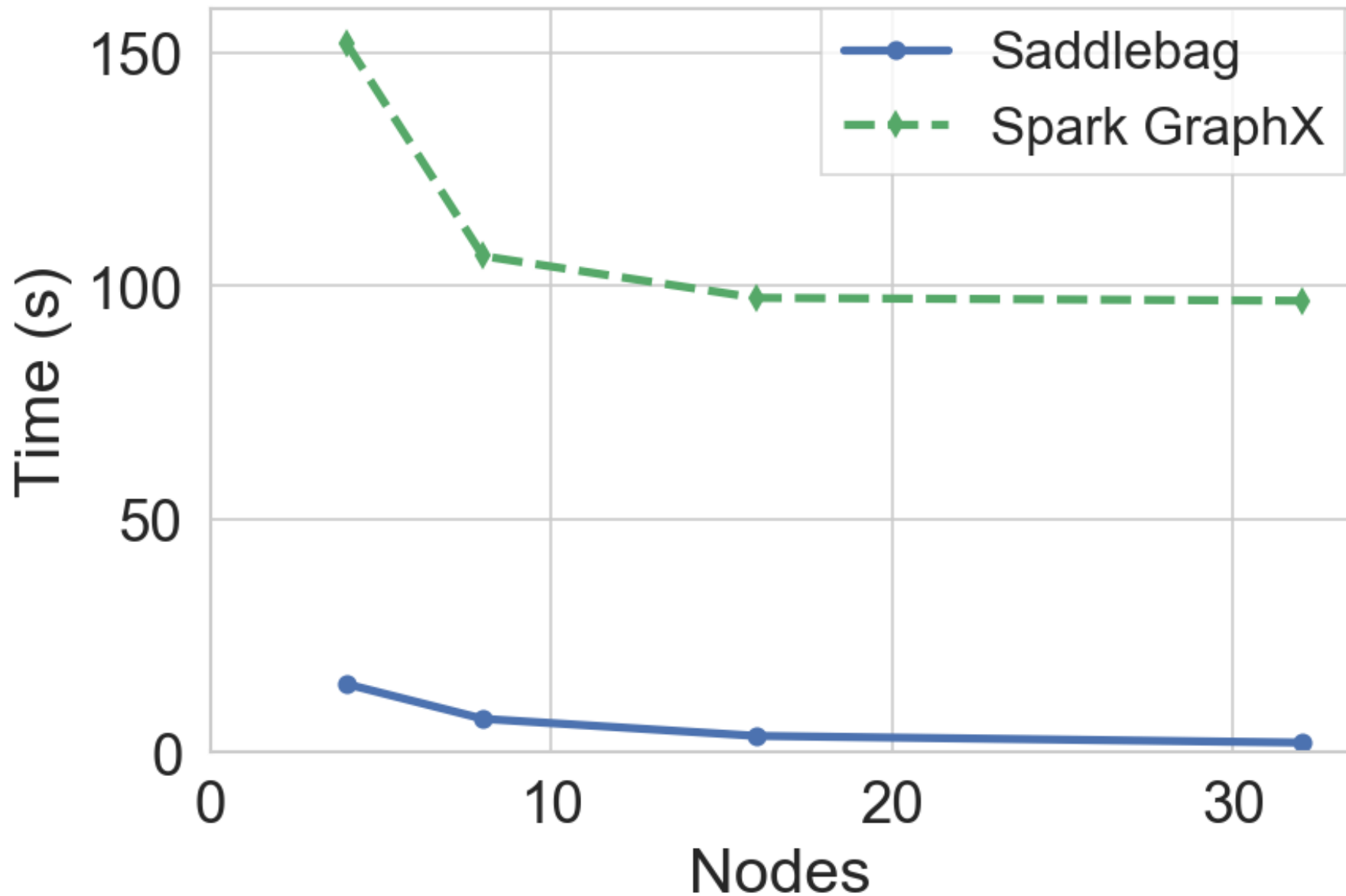
PGAS UPC++ Zero-Copy RMA



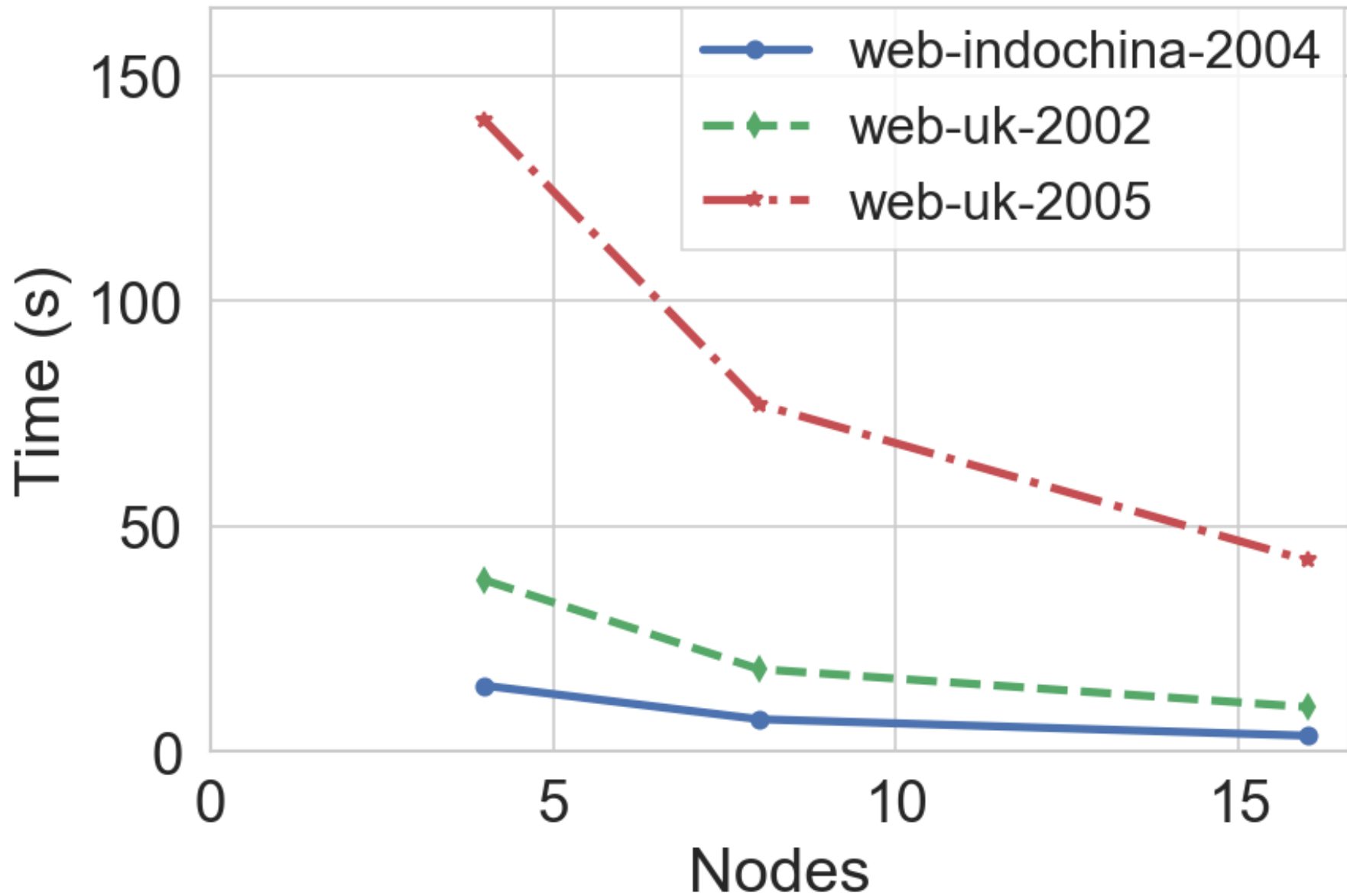
# Intuitive Rich Interface



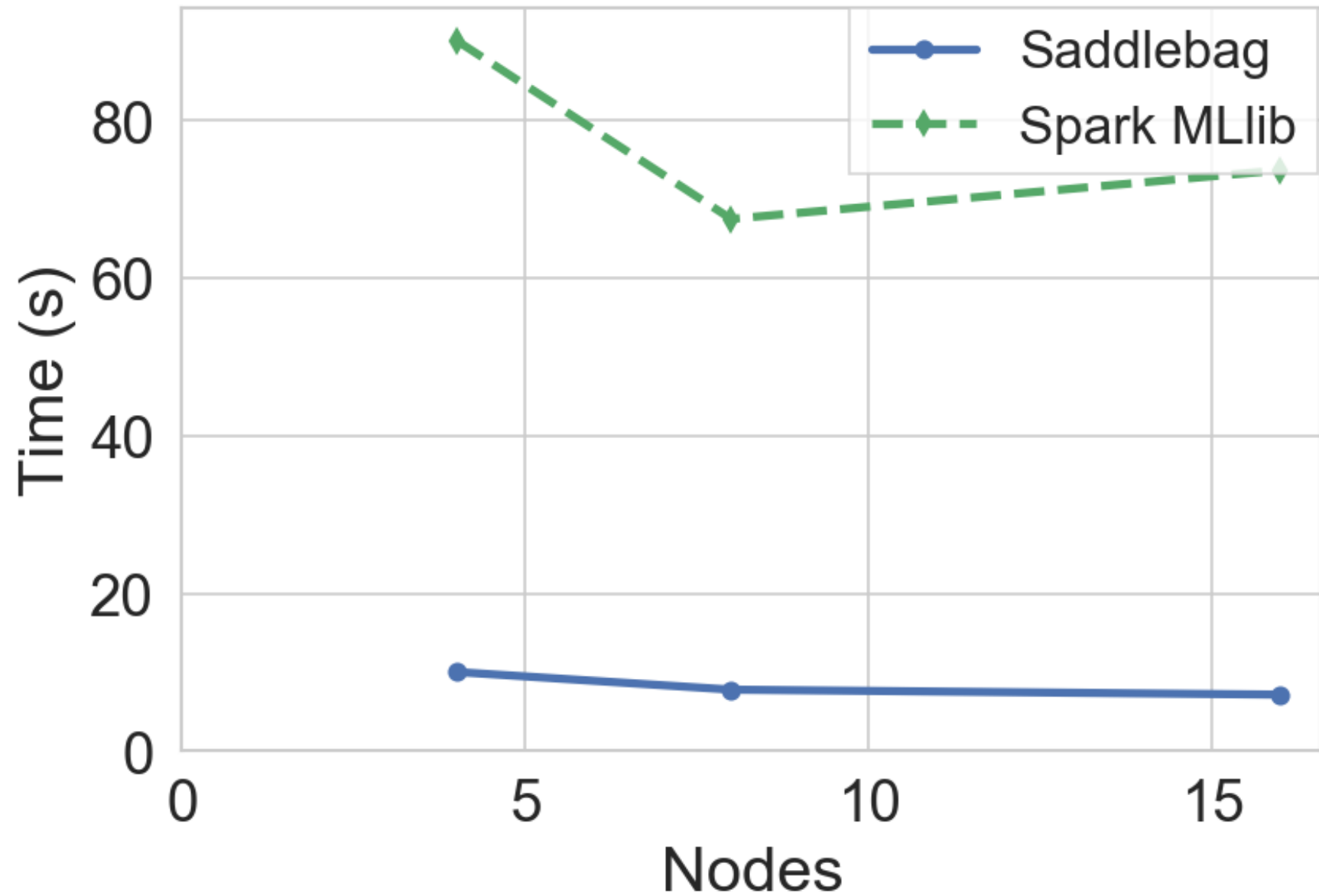
# Communication Model



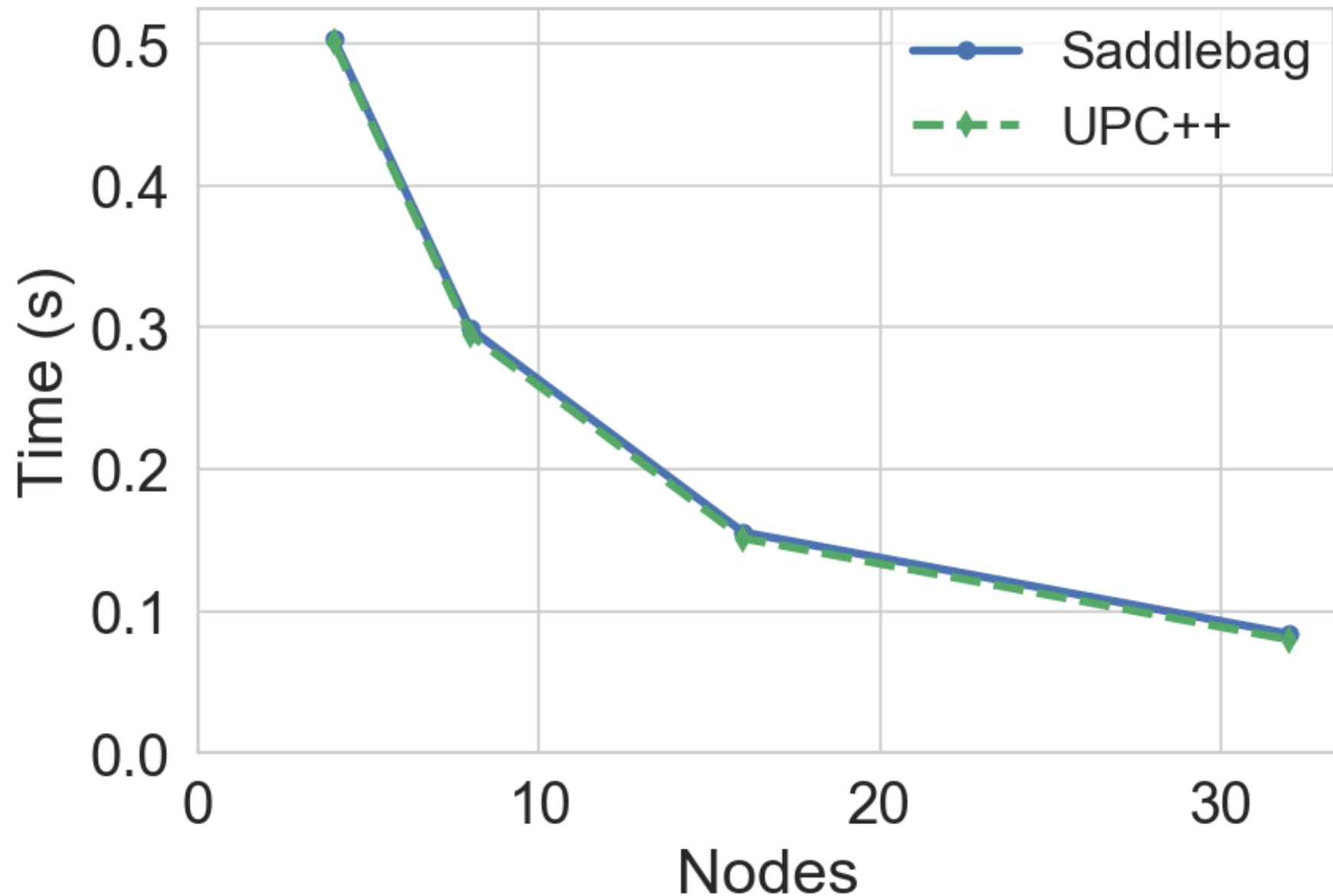
Performance vs Spark on PageRank



Scaling with Data Size on PageRank



Performance vs Spark on TF-IDF



Performance vs UPC++ on SpMV

# To UPC++ or Not?

Sockets   ZeroMQ   MPI   Brokers

# Local Optimization

Processes    Threads    Shared Memory

# Remote Optimization

Serialization RMA RPC Buffers

# Optimization

Stragglers Async Balance Data Flow

Thank you!

<https://AminMKhan.com>

# Further Reading

- Husky
  - + (SIGMOD 2017, VLDB 2015)
  - + (The Chinese University of Hong Kong)
- Ray
  - + (OSDI 2018)
  - + Berkley (from AMP Lab (Spark), now RISE Lab)
- PlinyCompute
  - + (SIGMOD 2018)
  - + (DAPRA's MUSE Project from RICE University)
- HPAT
  - + Tatiana Shpeisman (Intel, now at Google)
  - + HPAT: high performance analytics with scripting ease-of-use
  - + ISC 2017

# Further Reading

## ## Project Websites

- <http://www.husky-project.com/>
- <https://rise.cs.berkeley.edu/projects/ray/>
- <http://plinycompute.rice.edu/>

## ## Source Code on GitHub

- <https://github.com/husky-team/husky>
- <https://github.com/ray-project/ray>
- <https://github.com/riceplinygroup/plinycompute>