

# DRESS: Scalable Multi-Node Data Series Indexing

Manos Chatzakis[1][2], Panagiota Fatourou[1][2][3], Eleftherios Kosmas[1], Themis Palpanas[3] and Manolis Papadospiridakis[1]

[1] Computer Science Department, University of Crete

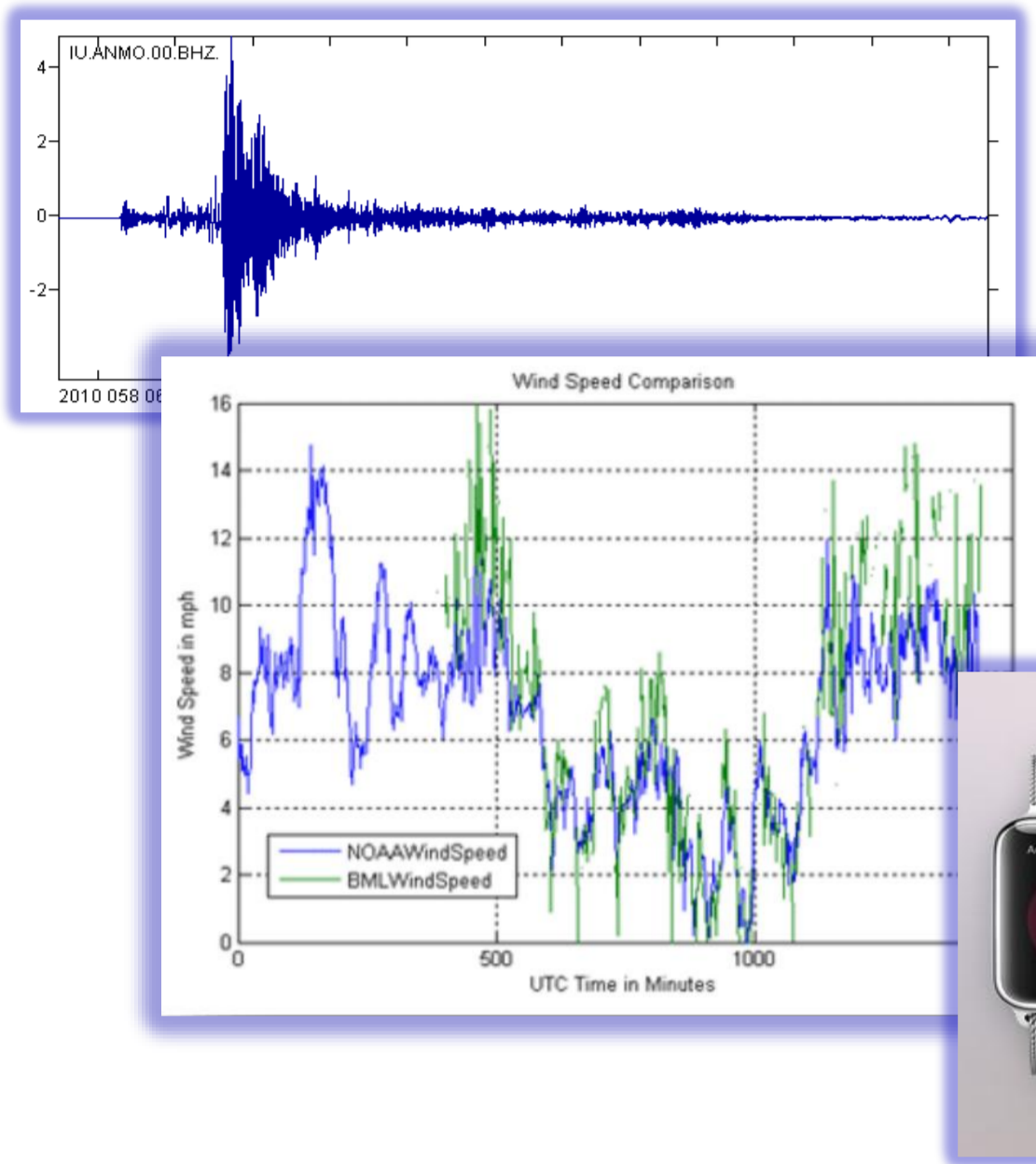
[2] Institute of Computer Science, Foundation for Research and Technology (ICS-FORTH)

[3] LIPADE, Université de Paris

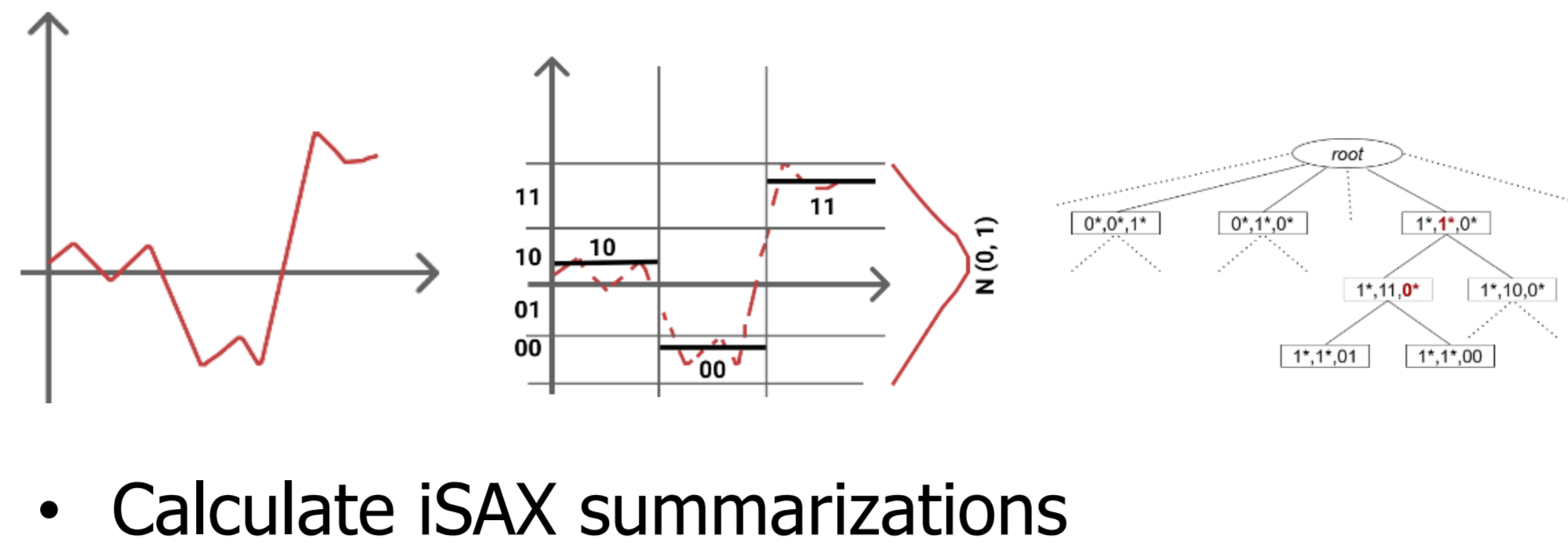
## Motivation and Challenge

### Data Series

- Many Applications:
  - Seismology
  - Astrophysics
  - Neuroscience
  - Engineering etc.



### Data Series Processing



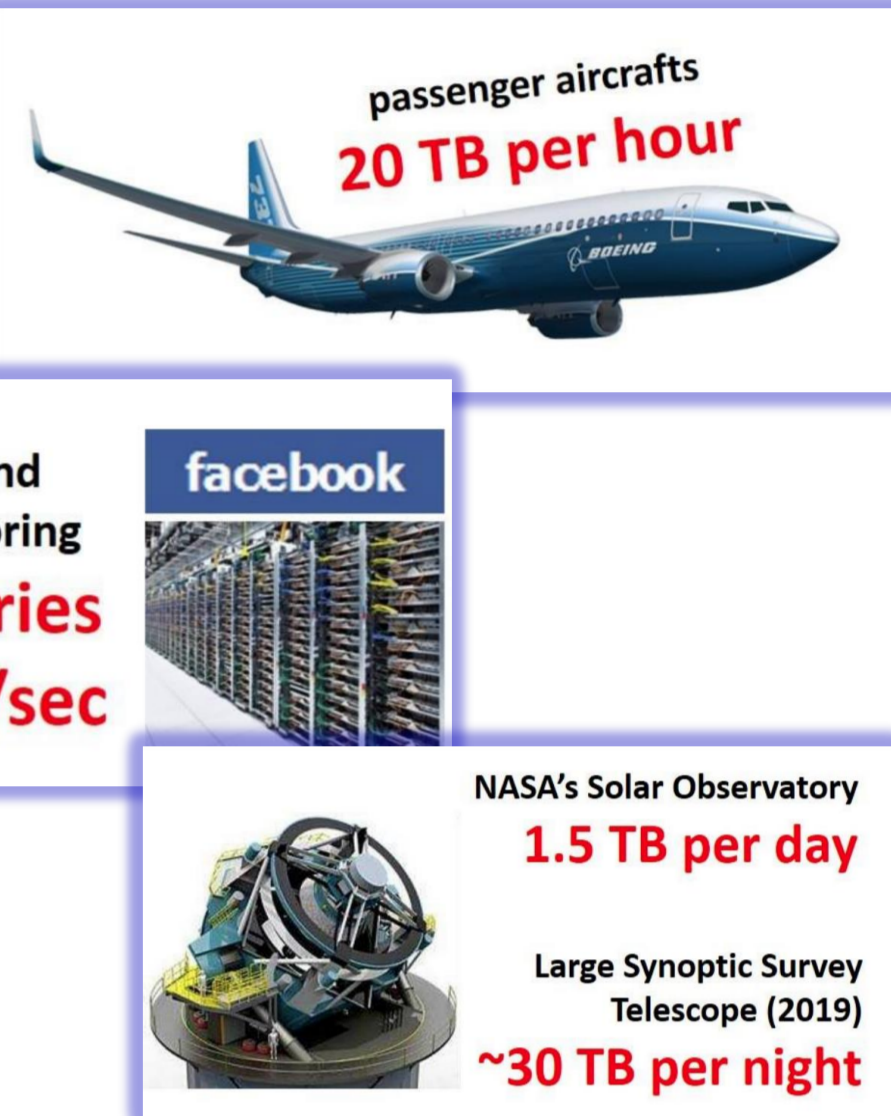
- Calculate iSAX summarizations

### Similarity Search

Find the most similar series of a collection to a query series

### Main Challenge

- Massive Data Series collections!
- High Dimensionality Data!

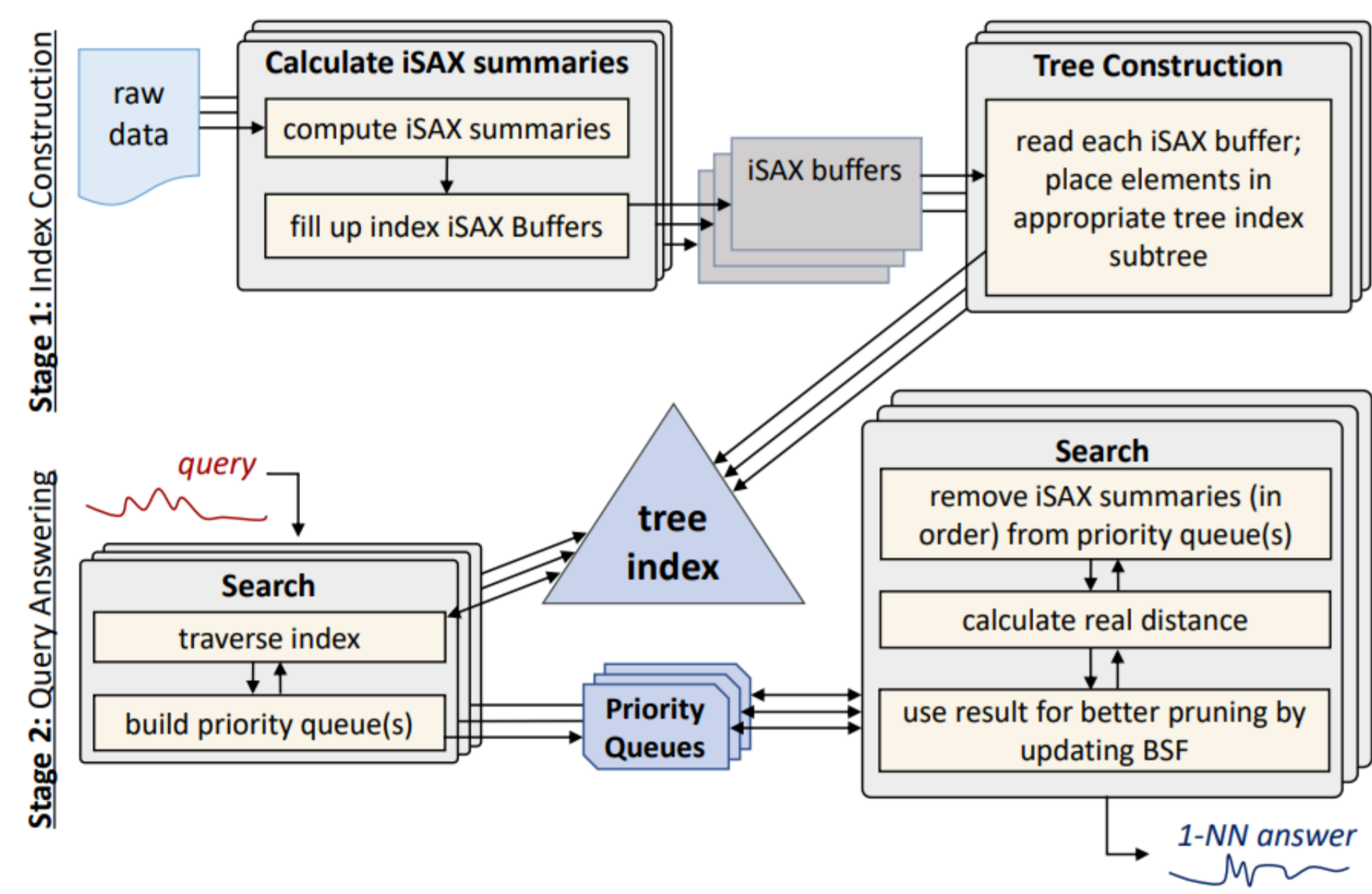


### Our Approach

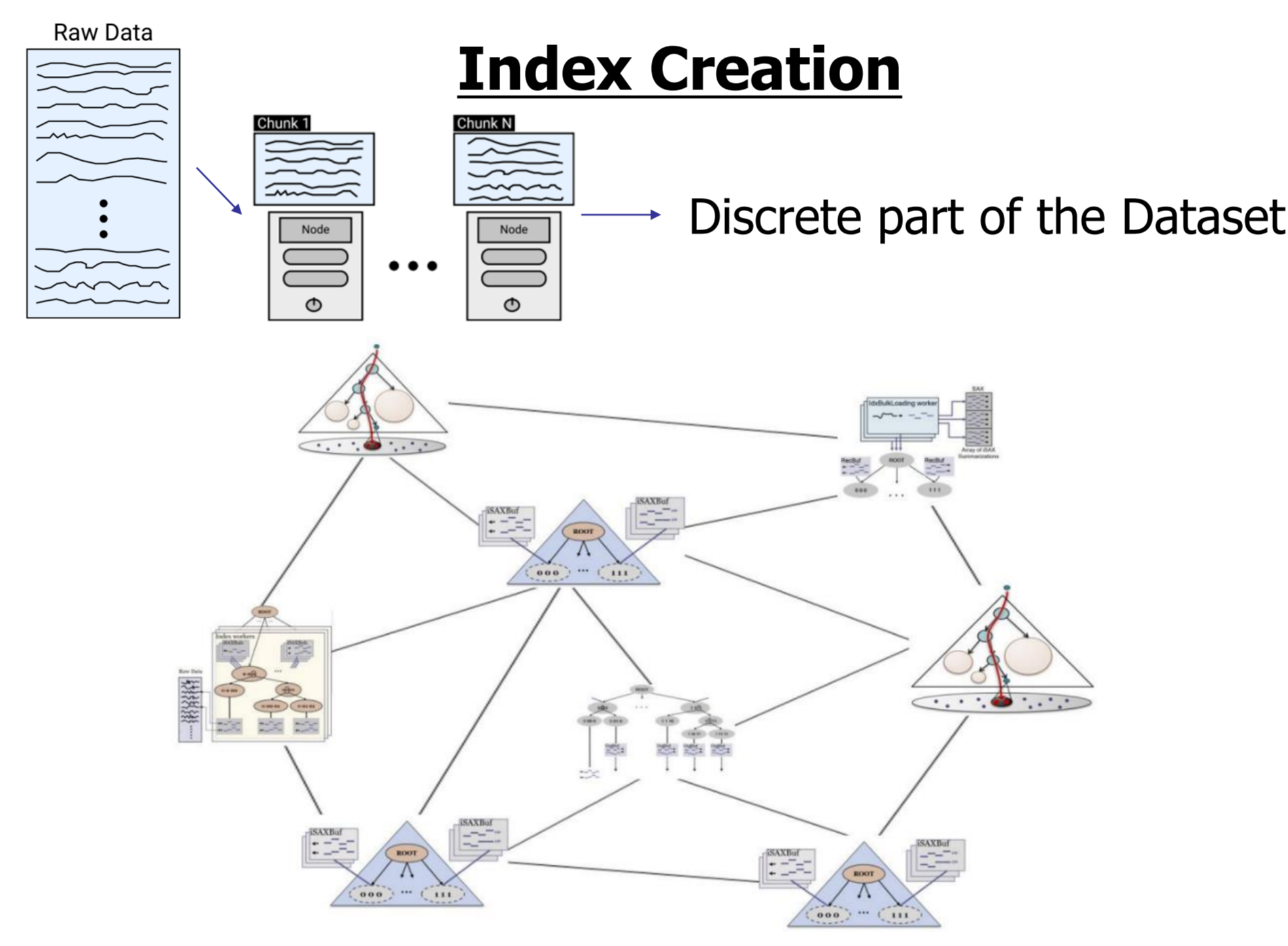
Use distributed systems to address the challenge

## Methodology and Contribution

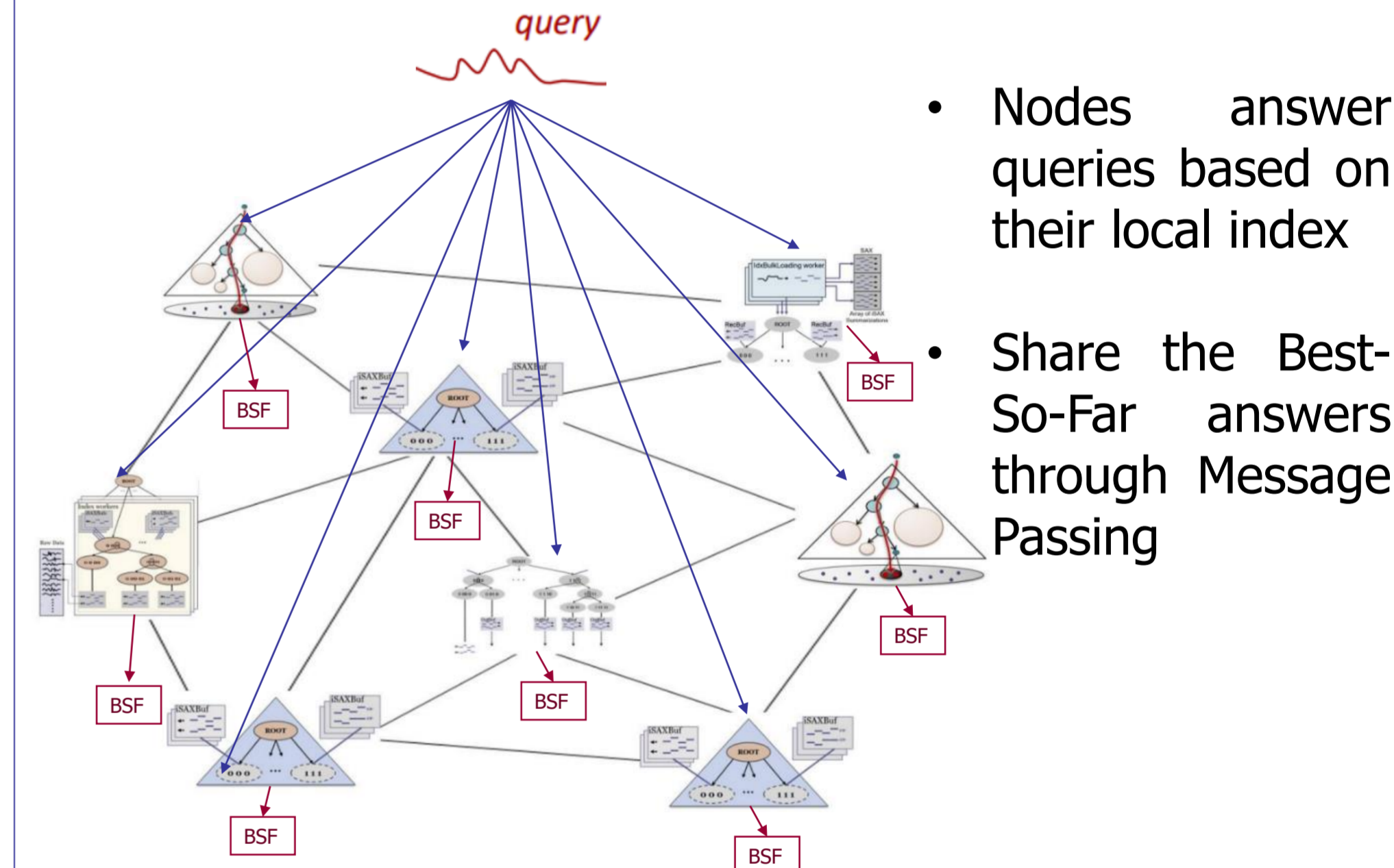
### Underlying Mechanism: MESSI



### Index Creation



### Query Answering



- Nodes answer queries based on their local index
- Share the Best-So-Far answers through Message Passing

## Evaluation

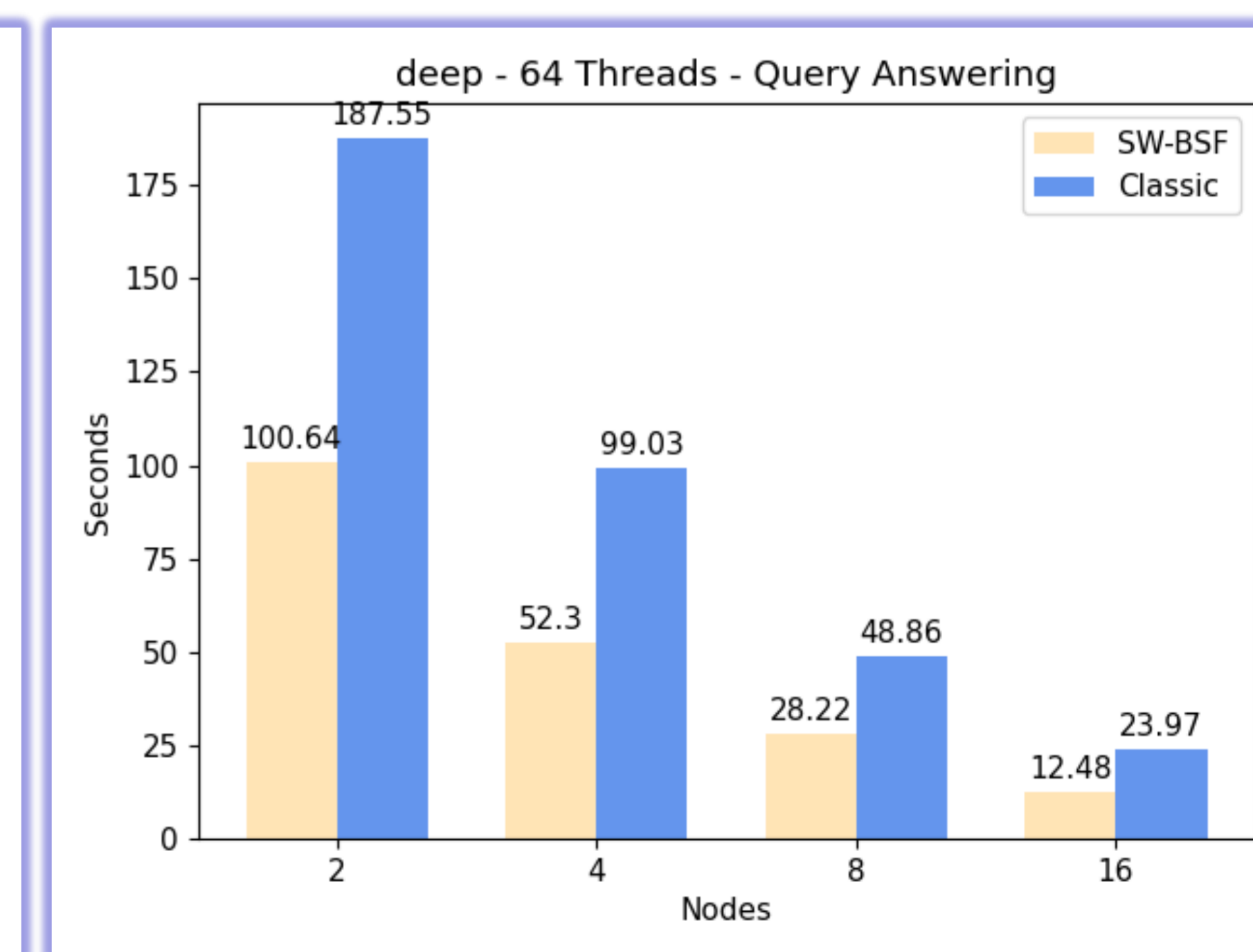
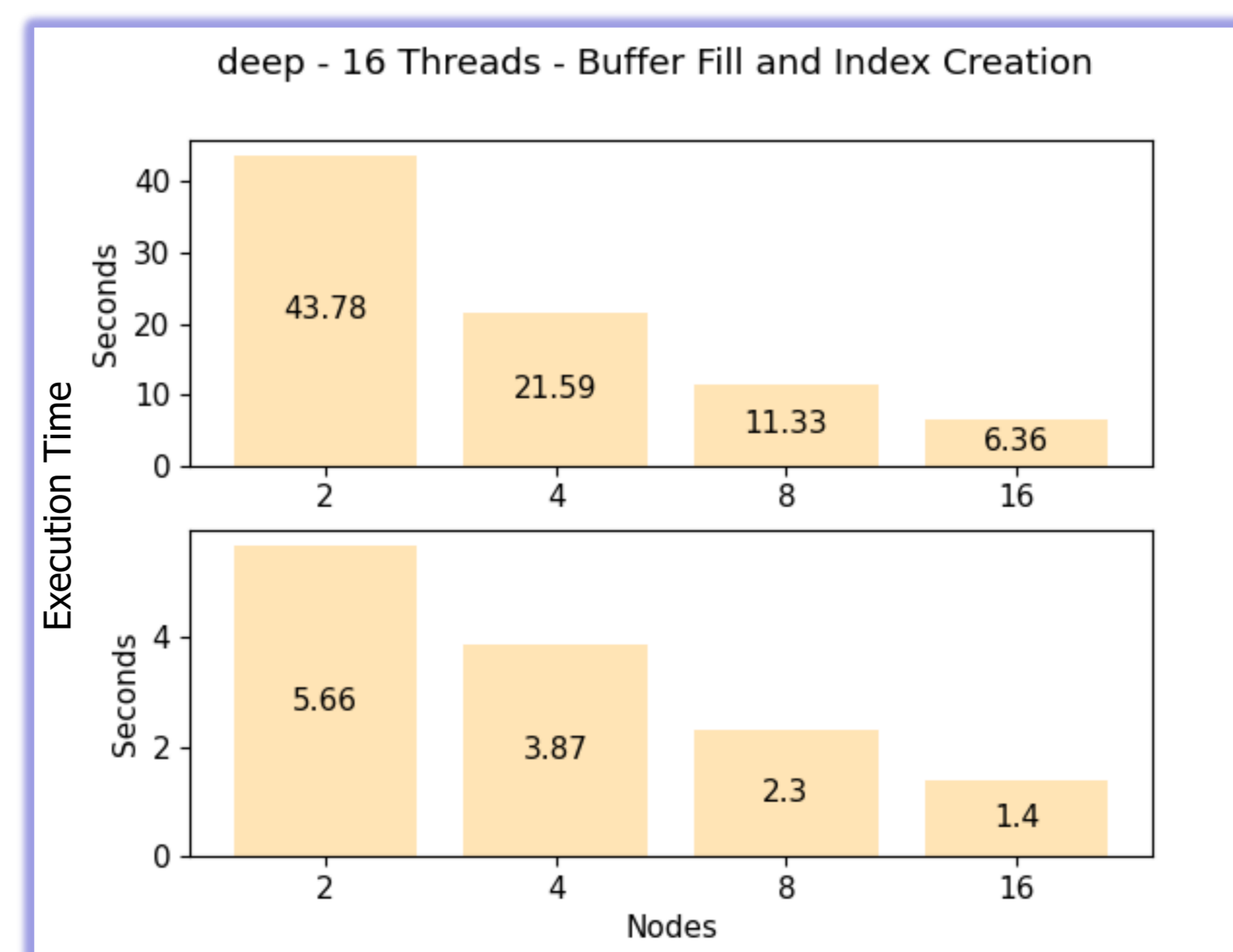
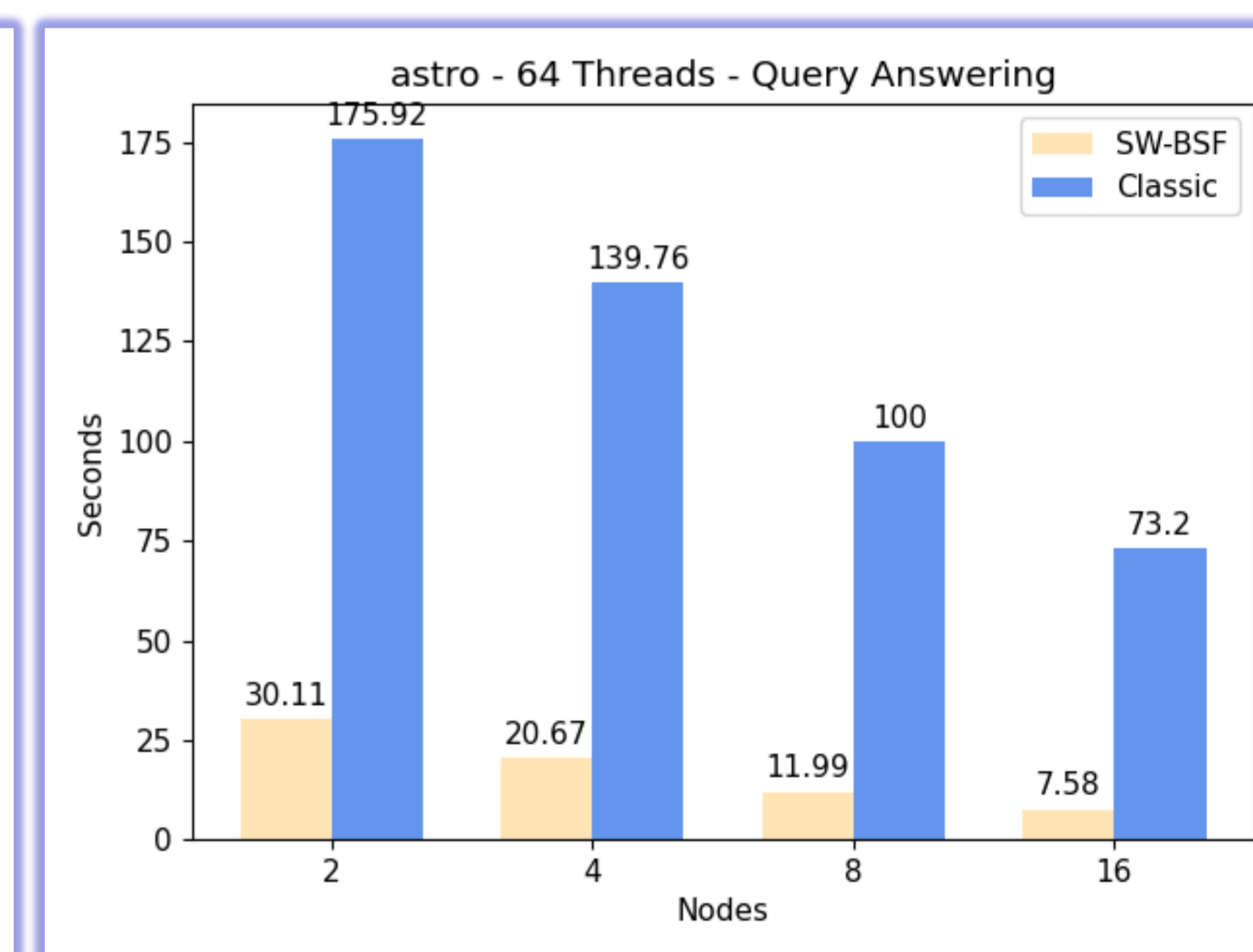
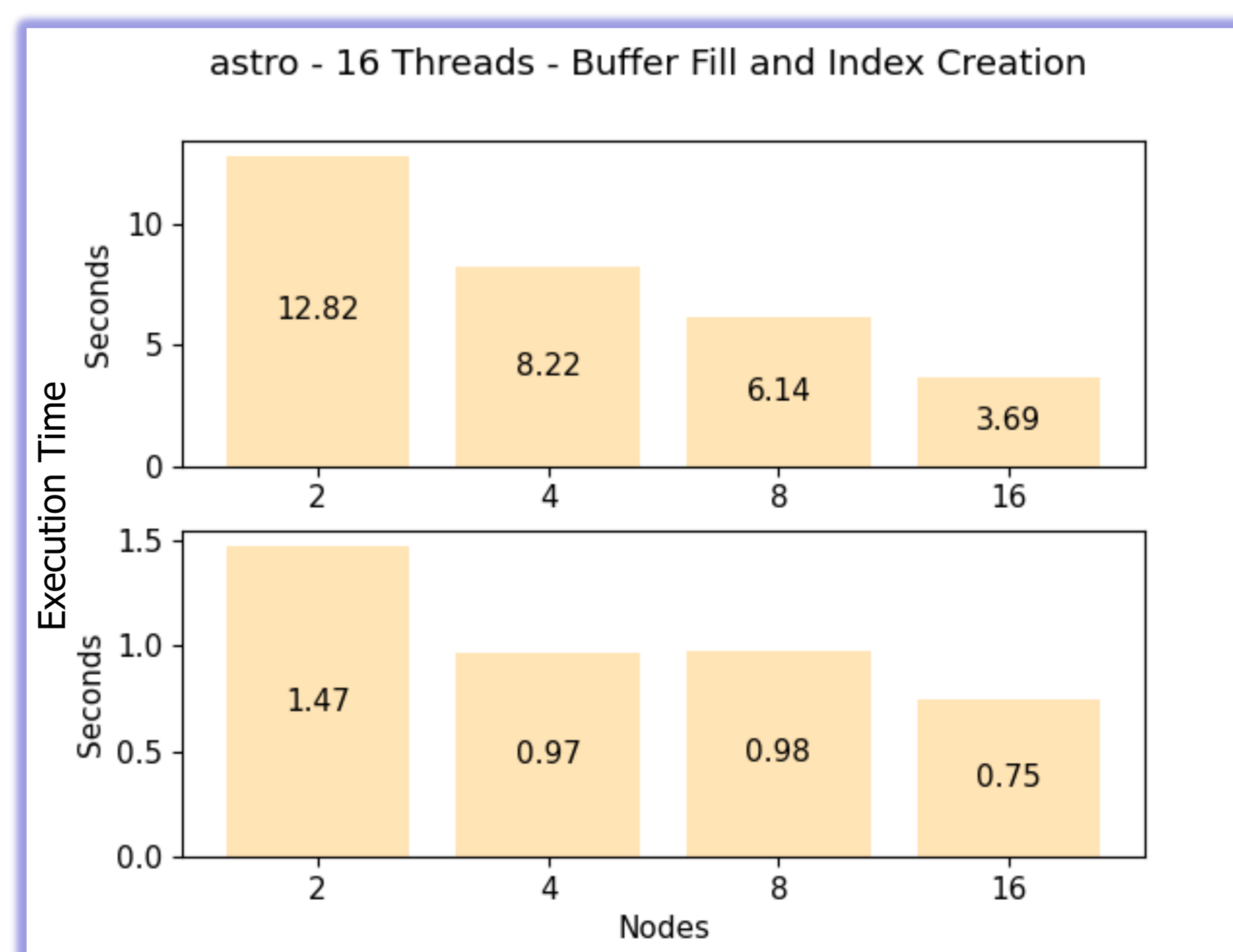
### Results

#### Configurations and Datasets

- System of 16 Nodes
- Astro Dataset: 270 Million Data Series of Astronomical Data
- Deep Dataset: 1 Billion Data Series of image descriptors
- 100 Queries of varying difficulty

#### Experimental Findings

- Index Creation performance scales for big datasets!
- Best-So-Far sharing (SW-BSF) improves significantly the performance of Query Answering
- Nodes may build local indices that do not have relevant data to a query: Less pruning
- Every node in the network receives the best answer possible



**Contact Information**  
 Manos Chatzakis  
 University of Crete, ICS-FORTH  
 Tel.: +30 6978 640371  
 Email: chatzakis@ics.forth.gr