Data Management in the Human Brain Project

Thomas Heinis

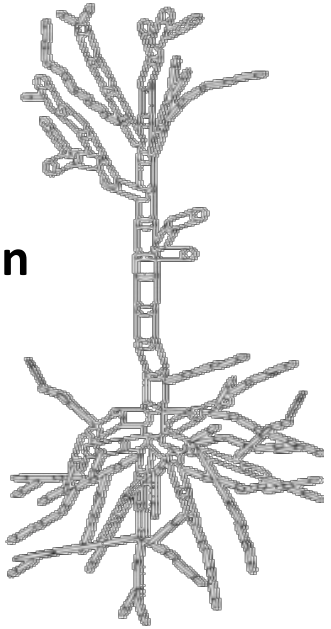
Ecole Polytechnique Federale de Lausanne

DATA-INTENSIVE APPLICATIONS AND SYSTEMS

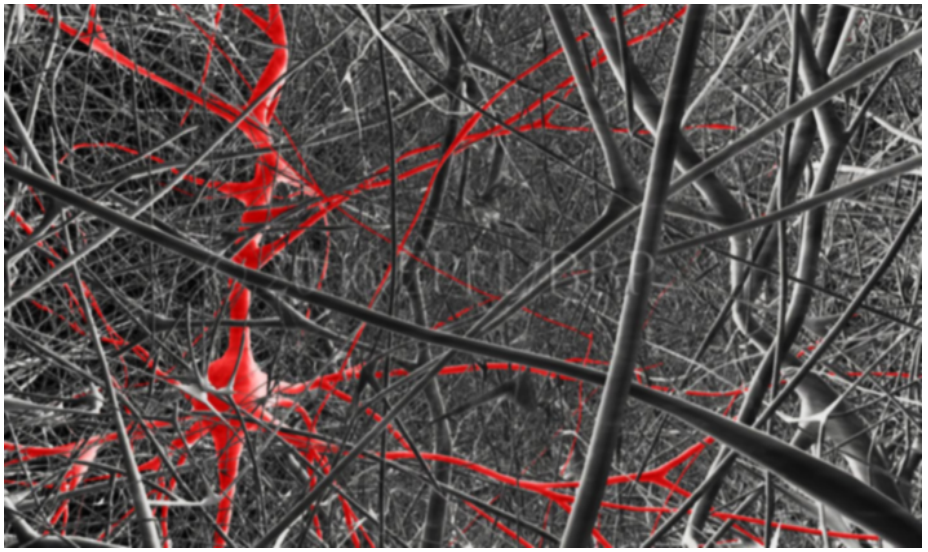
Blue Brain Project

Human Brain Project

- Goal: develop platform to simulate human brain!
- 10 year and 1 billion Euro project to integrate all available knowledge/ data
- Involves coordination of more than 200 interdisciplinary partners
- Lead by Henry Markram
- Based on EPFL's Blue Brain Project

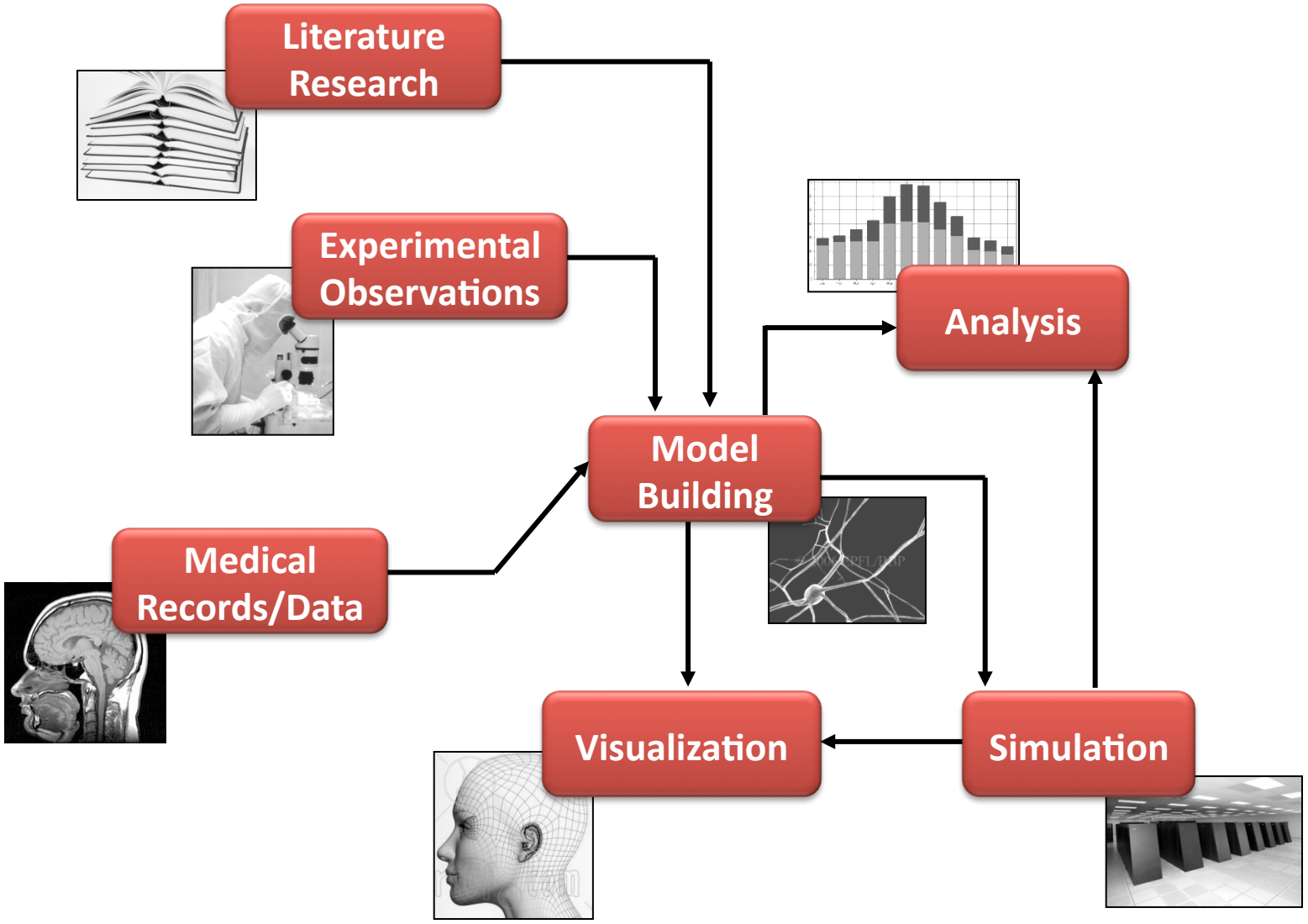


Single Neuron 3D Model



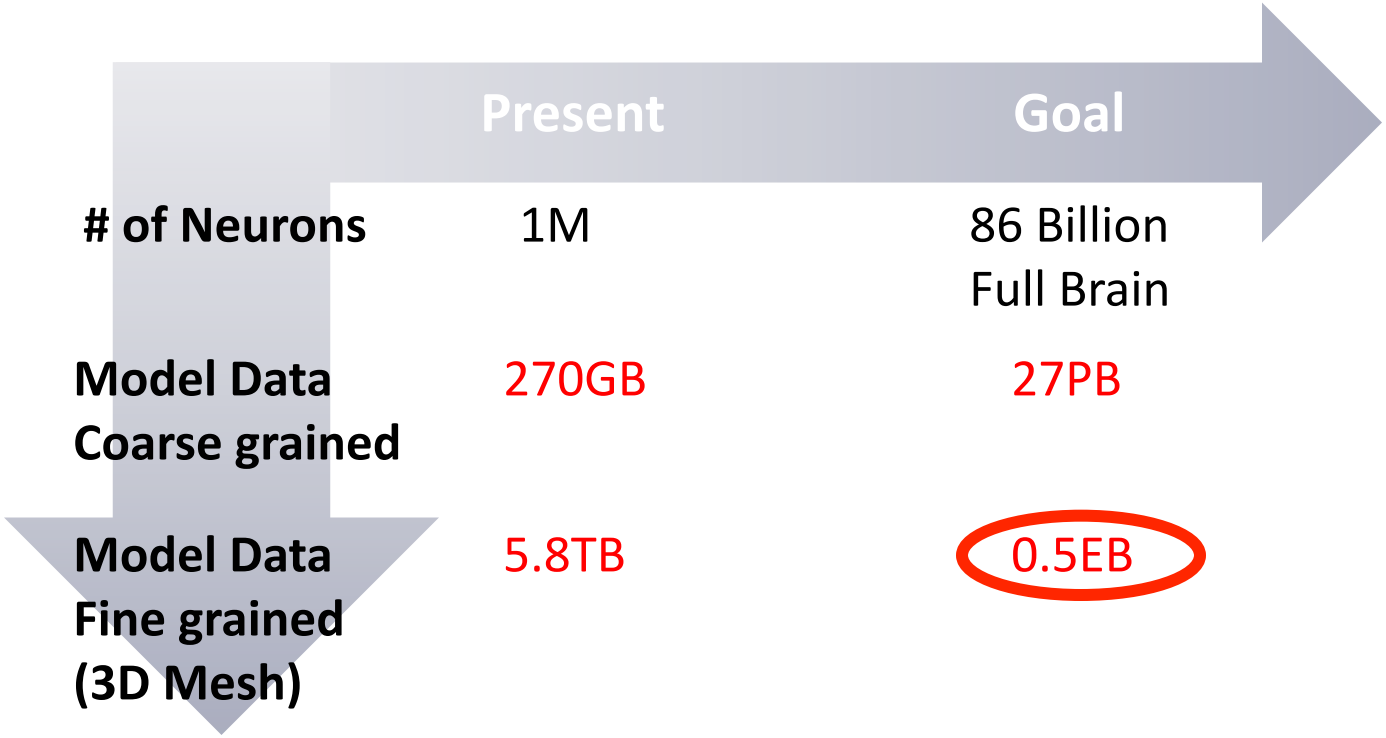
Neocortex Model

Human Brain Project Workflow



Data Deluge

- Medical Data from hundreds of sources
- Model Data alone:

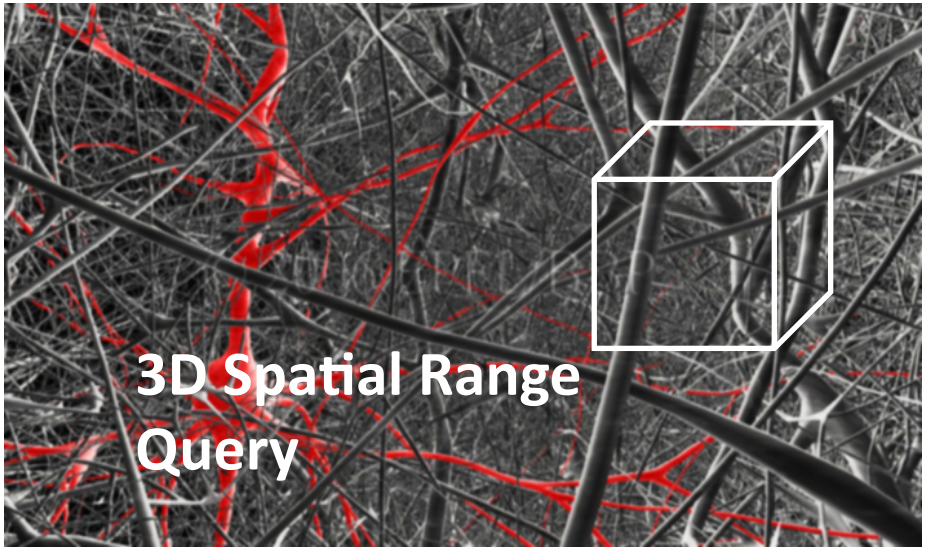


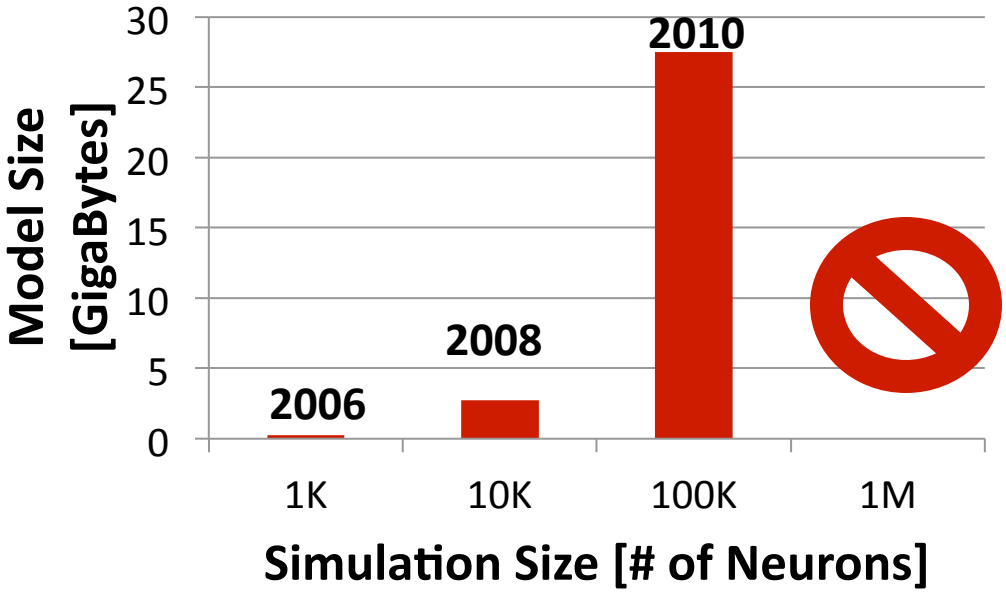
- Simulation Trace: potentially infinite

Spatial Indexing

Use Cases:

- Model Building
- Visualization

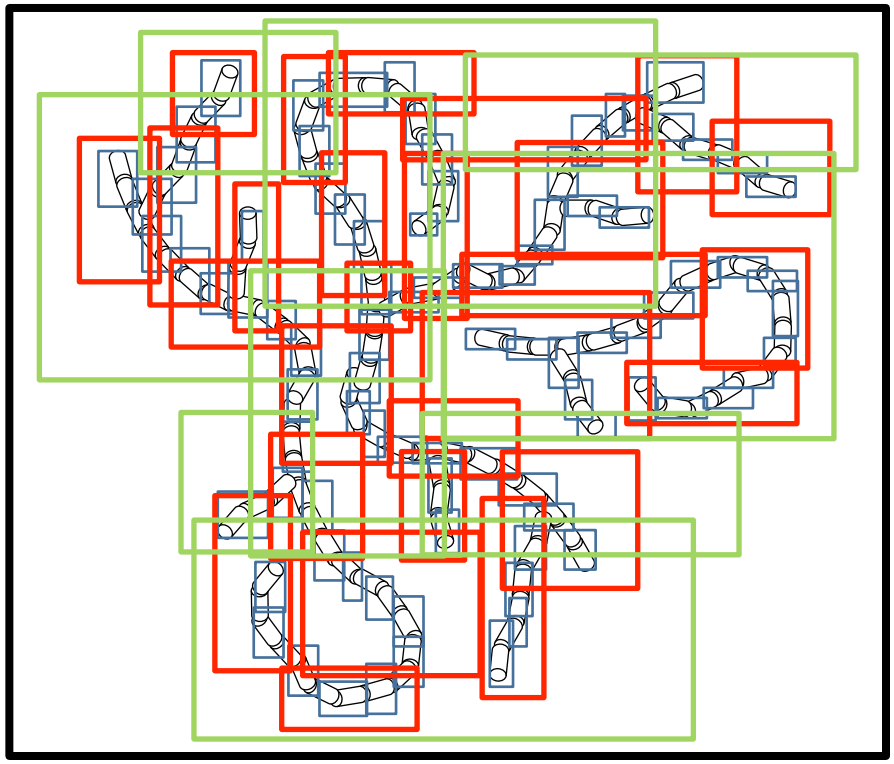




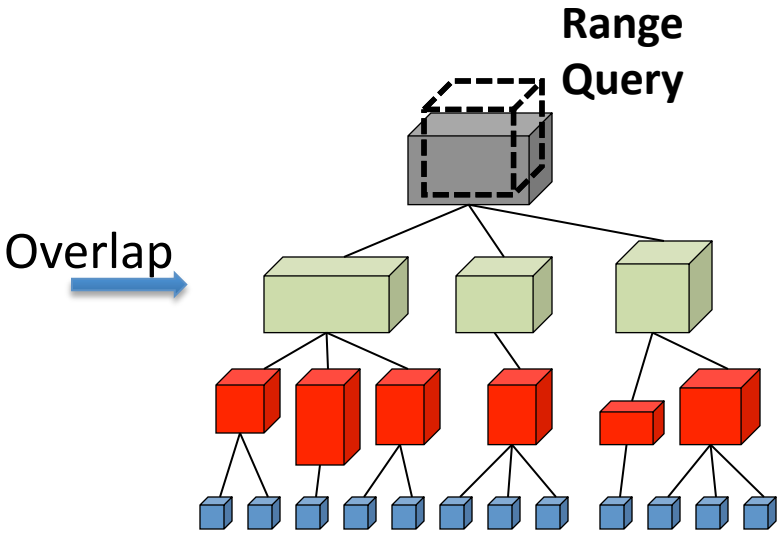
Bottleneck in Spatial Analysis

Overlap Problem

R-Trees: Hierarchy of Minimum Bounding Rectangles (**MBR**)



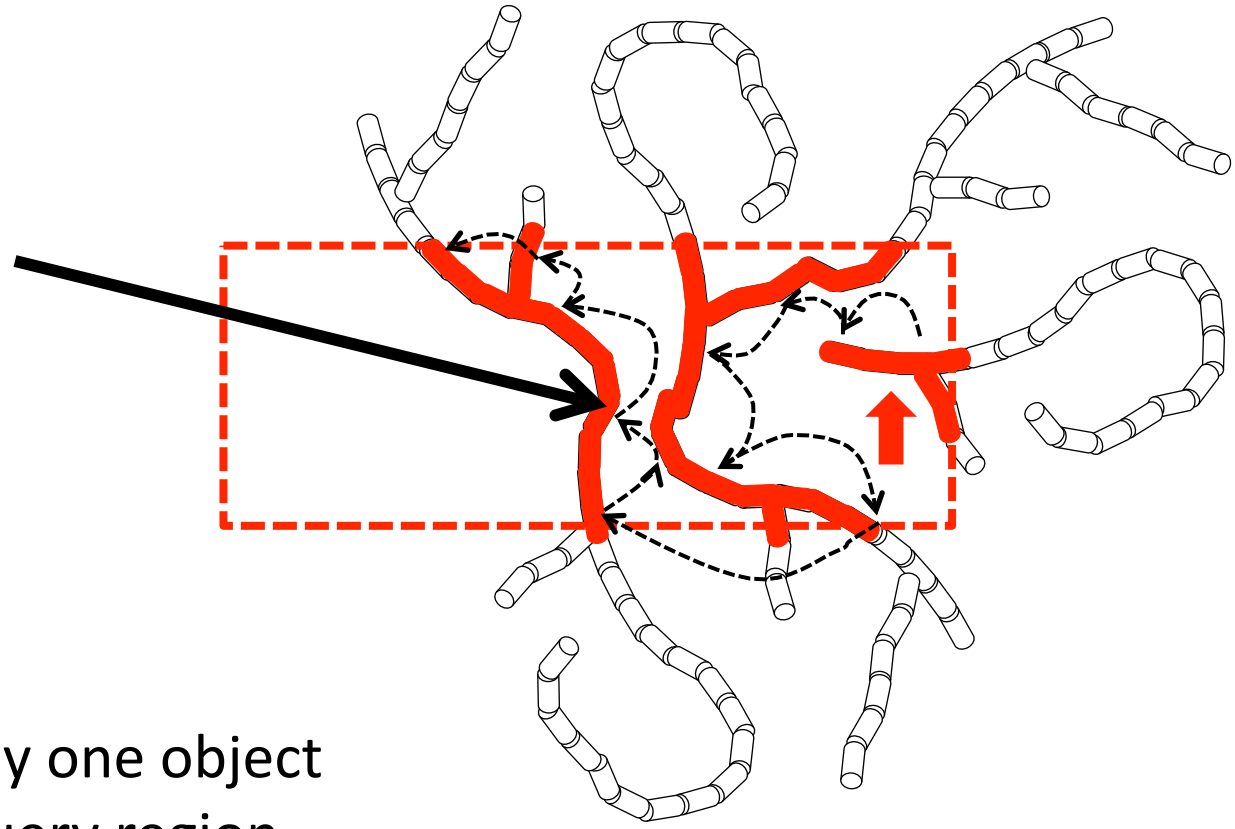
STEP2: Query Execution



Overlap **reduces performance,** increases with Data density

"**FLAT**", A Two Phase Spatial Index*

Add reachability information during index construction

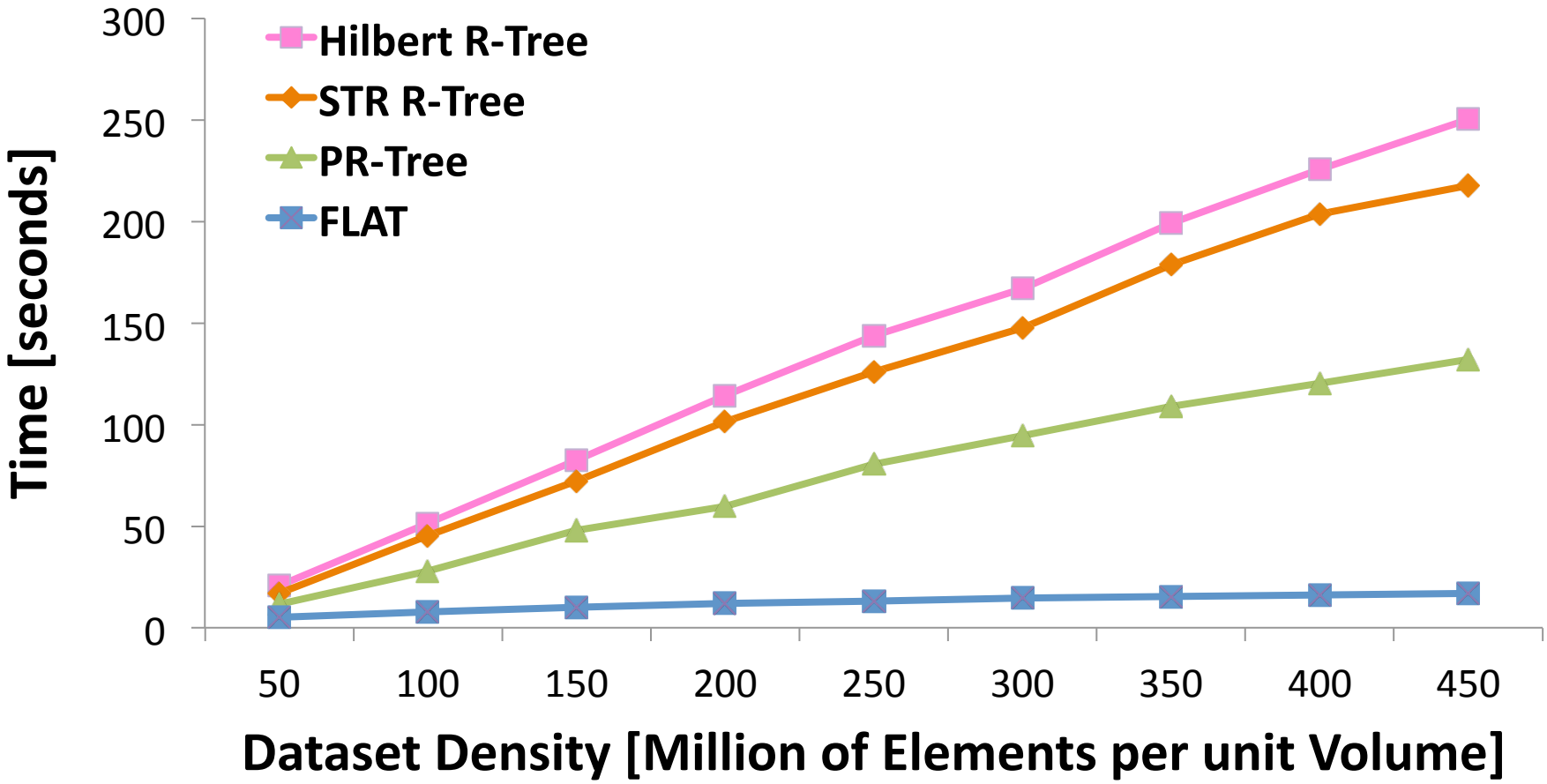


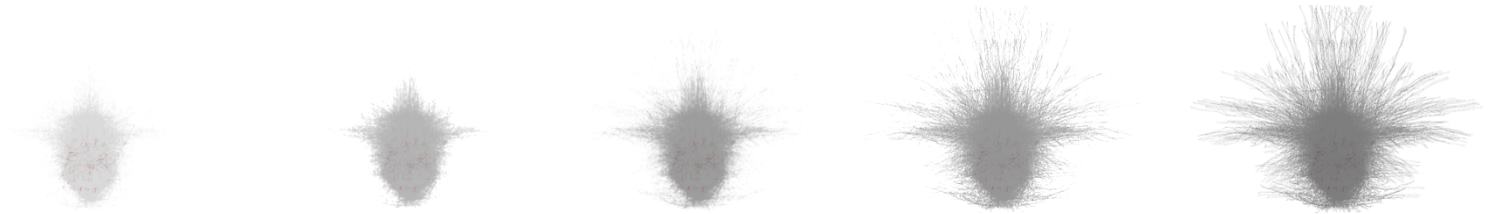
1) SEEDING: Find any one object arbitrarily inside the query region.

2) CRAWLING: Retrieve remaining results by traversing the neighbors.

*joint work with Farhan Tauheed, Laurynas Biveinis and Anastasia Ailamaki

Performance Evaluation





Impact

Blue Brain Project:

Part of the toolset used every day

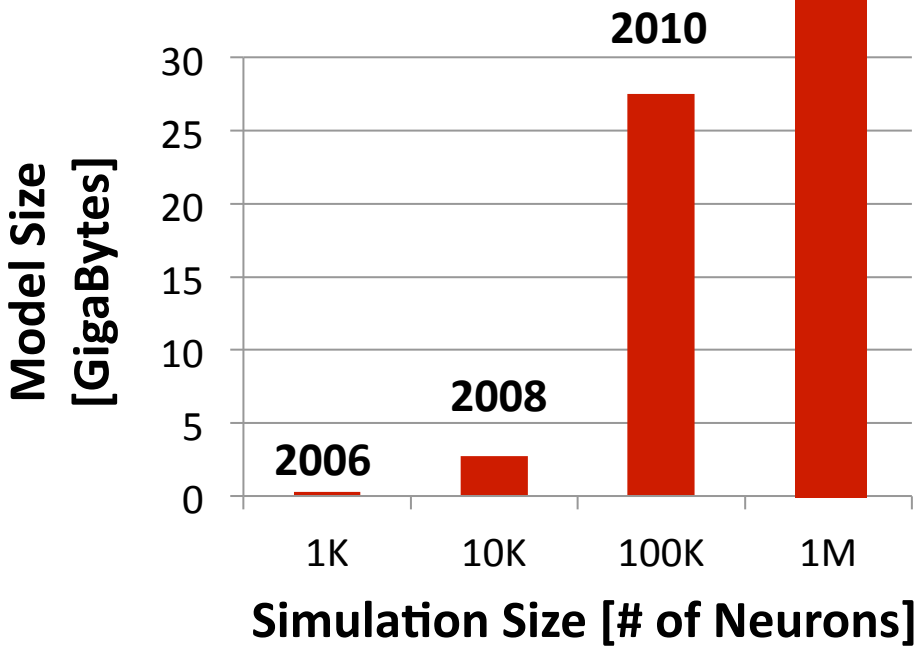
February 2012: first 1 million neuron model indexed

Still 5 orders of magnitude smaller than human brain

General Applicability:

3D surface mesh models

N-Body simulation data set



HBP Data Management Challenges

- Data Privacy/Anonymization
- Querying of Petascale Data
- Cloud Analytics
- Quick & efficient access to raw data
- Distributed Workflow Execution
- Provenance/Reproducibility
- Data Personalization

Thank you

Data Management in the Human Brain Project

Thomas Heinis