2nd Workshop on Software Challenges to Exascale Computing (SCEC-2018) Workshop,

December 13-14, 2018 New Delhi, India

High Impact Applicationsof Optimization and Statistics (Big-Data) on Multi-Petaflop Systems enabled by MPPLAB(e-Teacher)

MPPLAB *is acronym or*

= Mathematical **P**rogramming in **P**arallel **LAB**oratory

By

Dr.V.C.V. Rao (C-DAC, Pune)

Dr.Narendra K Karmarkar

High Impact Applications of Optimization and Statistics (Big-Data) on Multi-Petaflop Systems enabled by MPPLAB(e-Teacher)

MPPLAB *is acronym or*
= Mathematical **P**rogramming in **P**arallel **LAB**oratory

Project Proposed To National Supercomputing Mission (NSM) as an Application

Members Involved

- **Chief Architect -Dr. Narendra Karmarkar**
  Consulting Scientist, CDAC, Pune

Principal Investigators

CDAC-Pune : Dr.VCV.Rao
CDoT -Delhi : Shri Jayant Bhatnagar, Director
IISc-Bengalore : Prof. Bharadwaj Amrutur,

Overview

Object Oriented Programming
Fortran 2008

Outline

The project involves application of Unified architecture based methodology to very large scale problems of national importance, including in particular telecommunication traffic engineering.

e-Teacher is name of the system to be used for training manpower to under- stand concepts underlying the new system and use it effectively.

High Impact Applications of Optimization and Statistics (Big-Data) on Multi-Petaflop Systems enabled by **MPPLAB(e-Teacher)**

Origin of the Proposal

- Optimisation & Statistics (Big Data) has several applications for large-scale problems in
  - Network optimization,
  - Distributed control,
  - Complex resource allocation and optimization problems,

Due to Large-Scale Systems and Big Data requirement, Top-down Design at HPC based on unified architecture necssary.

Outline

The overall project has three **Stages**

Stage 1 : Unified Object Oriented Interface

Stage 2 : Development of Scalable Libraries

Stage 3 : Implementation for Sample Application Telecom

Stage 3

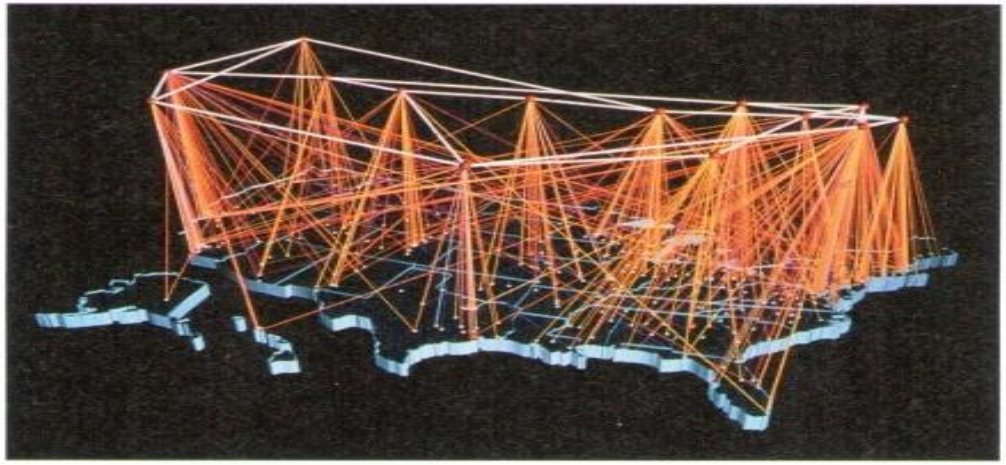
About Telecom Application Math Model

HPC-Big Data Processing:Linear Programming App.

A transportation Problem Delivering Perishables

- Internet traffic among a collection of major sites, with the volume of traffic color coded. routing traffic over immense networks **- Need of linear prog. tech. and high-speed computers.**
  - ●Email ● Data packets, ●Streaming video

Challenge :
Process Graphs
(Directed /Undirected)
Few Million to Billion



Source :
http://www.macmillanhighered.com/Catalog/product/forallpracticalpurposes-tenthedition-comap

MPPLabe-Teacher: An Overview of Application Framework

Examples of Systems represented by Graph

- **Social Networks**
  Internet, facebook, twitter, google plus, whatsapp etc.

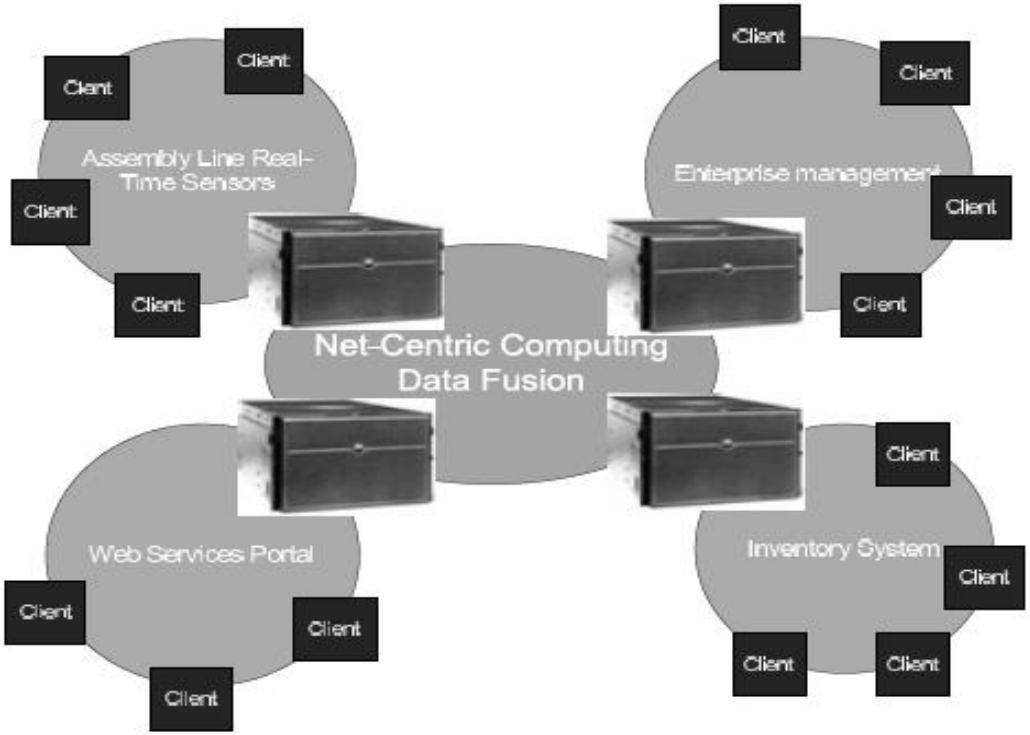
- **Communication Networks**
  Telephone network, Mobile, email etc

- **Biological Networks**
  Genomes, Protein interactions, disease

- **Electrical Networks**
  Power grid, network on chip etc

Network Centric Computing (NCC) : Environment

- An emerging technology Architecture and an evolutionary stage of client/server computing.



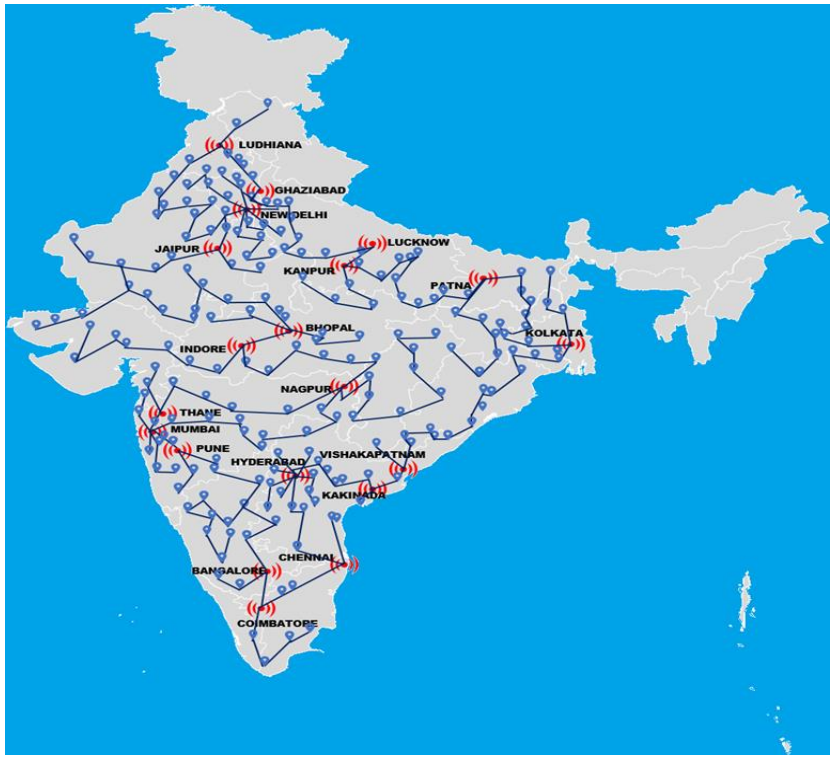
Architecture : supports access to information through multiple electronic channels (personal/network computers, cell phones, kiosks, telephones, etc.).

High Impact Applications of Optimization and Statistics (Big-Data) on Multi-Petaflop Systems enabled by **MPPLAB(e-Teacher)**

Mathematical Modeling – Alg./Lib; Telecomm App. in progress

- **App. Software Design :** Map of India; Map Module; Activity Module; Time Zone; Activity Date Collection; Affinity Types – Data Collection

3. Mathematical Modeling – Alg./Lib; Telecomm App. in progress



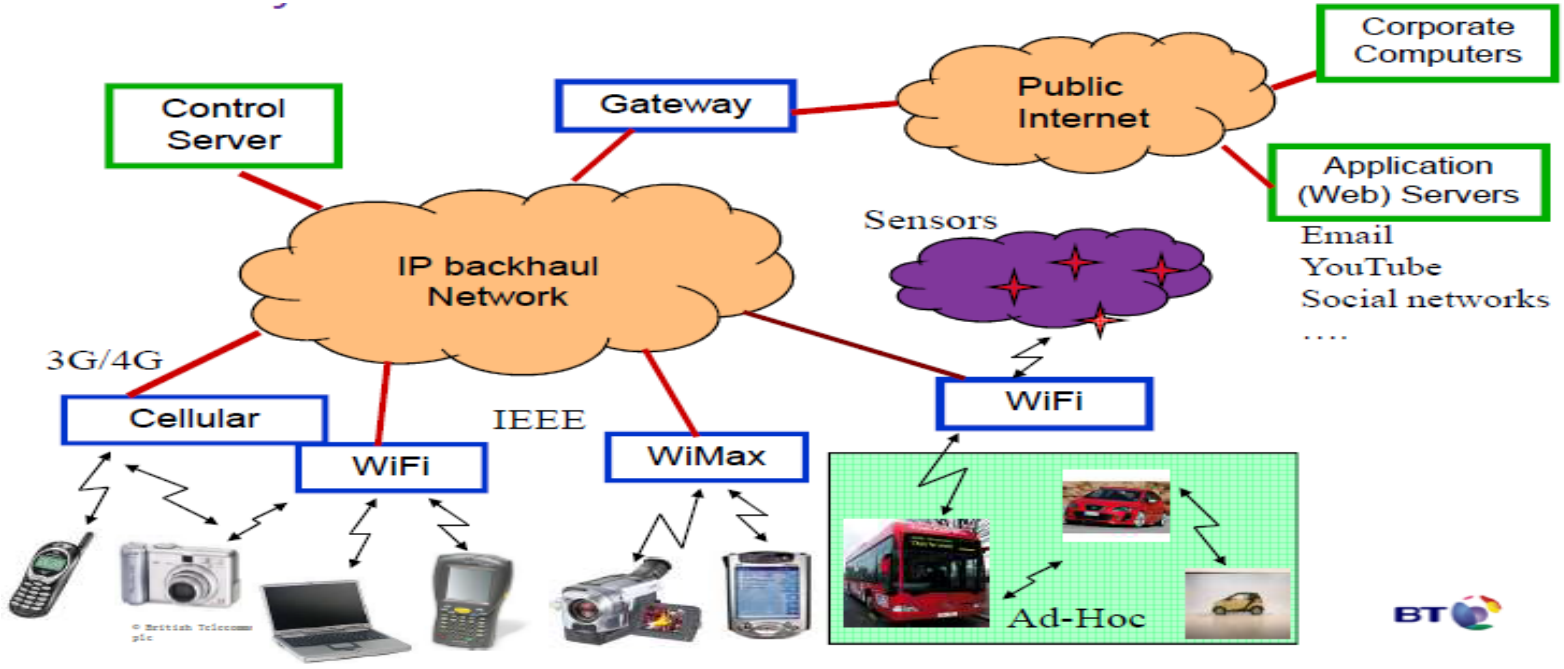
Illustrative Example of Graph Representation of TeleComm Links

- **App. Software Design :** Map of India; Map Module; Activity Module; Time Zone; Activity Date Collection; Affinity Types – Data Collection

HPC-BIG Data : Computational modelling telecom

To-day's wireless telecom Network

- ❖ Integration of Sensors, Switches, Devices, Compute Servers
- ❖ Real Time Data Processing – Data Analytics
- ❖ Data Segregation, Data Fusion, Data Access



High Impact Applications of Optimization and Statistics (Big-Data) on Multi-Petaflop Systems enabled by **MPPLAB(e-Teacher)**

Telecommunications – HPC

- Telecom applications require higher class of machine providing hardware based cache-coherence.
- Routers need hardware modification

  Extract Data from network as input to optimization algorithm on supercomputer
- Apply Control to the network based on output of the Algorithms

- Demonstrates the broader applicability of architecture based on Top-down approach being undertaken in implementation part of the project in combination with software from stage one.

MPPLabe-Teacher: An Overview of Application Framework

Examples of Systems represented by Graph

- **Transportation Networks**
  Railways, airways, roadways etc

- **Financial transaction network**
  Bank transactions, Trading Stock market

- **Academic**
  Publications / citation graph

Stage 2

Development of Scalable Libraries

Object Oriented Programming
Fortran 2008

MPPLab(e-teacher) : Comp. Algorithms

- Linear Programming Application to Telecom Traffic Engineering (LPATTE)
- Graph Theory Computations (GTC)
- Massively parallel iterative algorithm (MPIA)
- **Linear Programming for telecom sector**
- Linear Programming in Financial Modelling (LPFM)
- Linear Programming in a Big Data Framework (LPB)

MPPLab(e-teacher) : Comp. Algorithms

- The first stage focuses on generic applications of large -scale linear programming,

- Graph theory, and massively parallel iterative algorithms.

  - These include Linear Programming-Generic Applications (LPGA, Graph Theory Computations (GTC) and Massively parallel iterative algorithm (MPIA)

Stage 1

Computing Systems

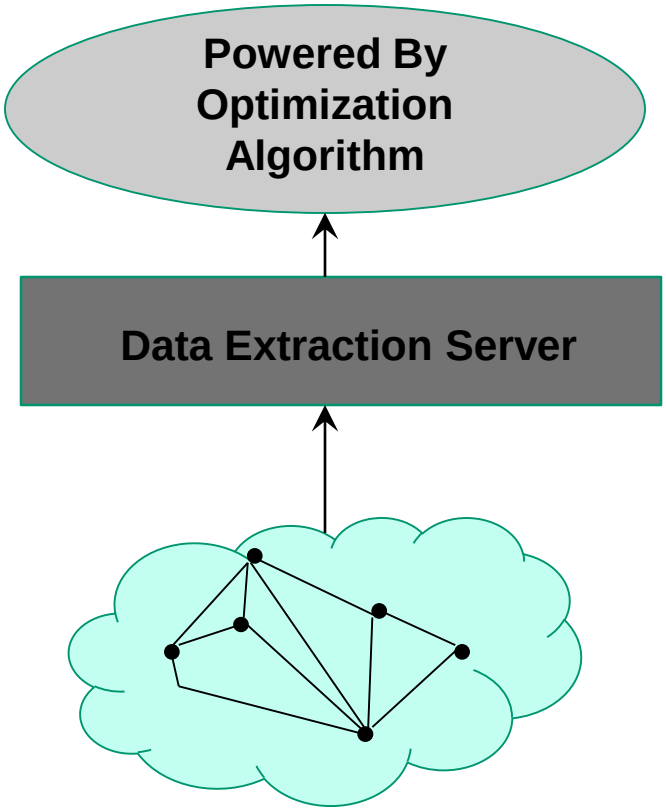
Unified Object Oriented Interface

Approach – Top-Down Approach is considered.

❖ **Bottom-Up and Top-Down**

- **In the Bottom-Up** approach,
  - ✓ hardware aspects come first and
  - ✓ constraints imposed by hardware dictate how application software is to be ”ported” or
  - ✓ made to fit the peculiar features of hardware.
- **In the Top-Down**,
  - ✓ **Algorithm-driven approach,**
  - ✓ **implementation aspects of Application** Algorithm come first and hardware system is built and
  - ✓ Configured to fit the requirement for most effective implementation

MPPLAB(e-Teacher) : Unified Architecture - A single, highly integrated & Coherent Implementation framework : Data Extraction



Volume of total Data from network is gigantic

needs reduction based on potential relevance

Preliminary analysis of relevance by

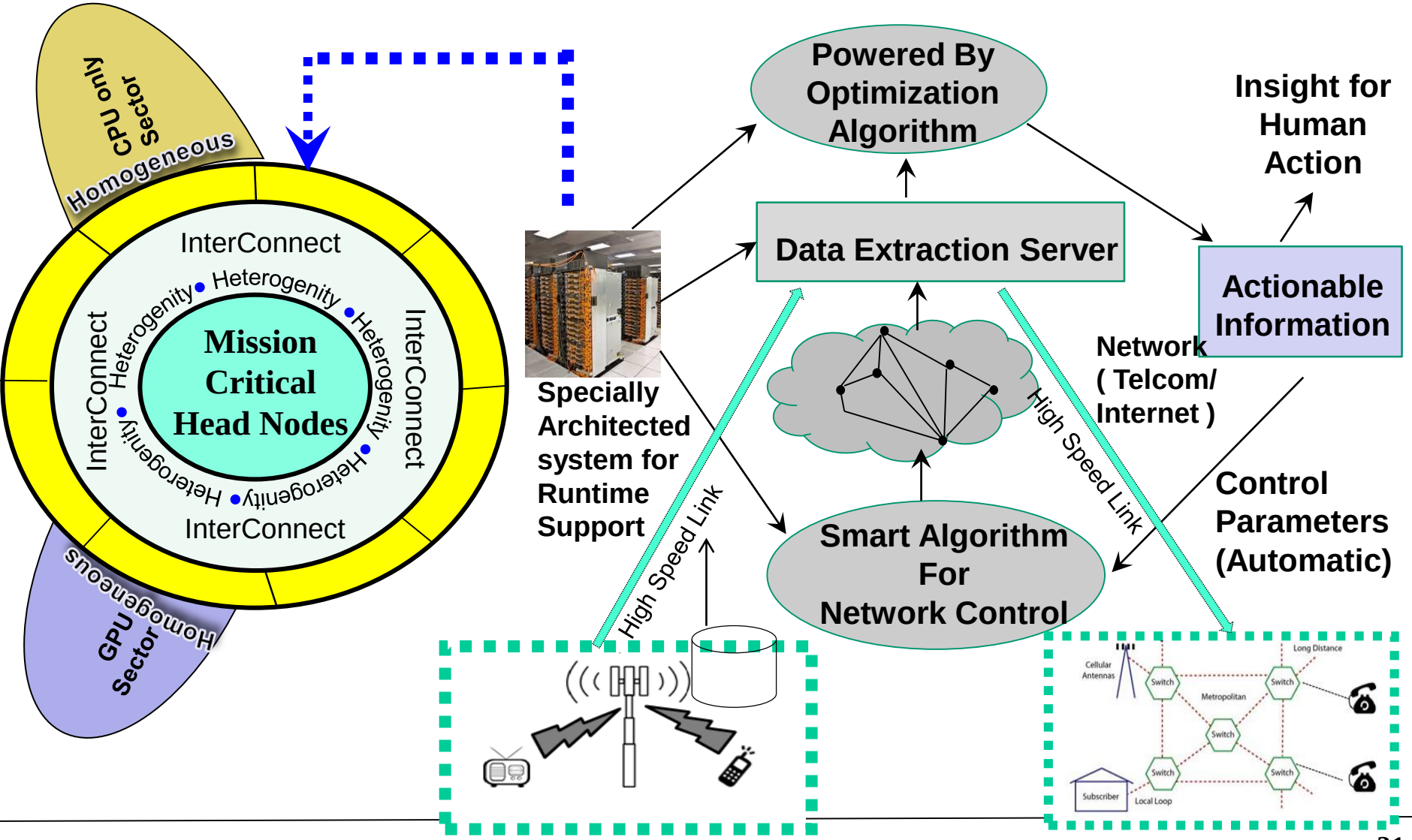
Smart Algorithms

Runtime support on special class of

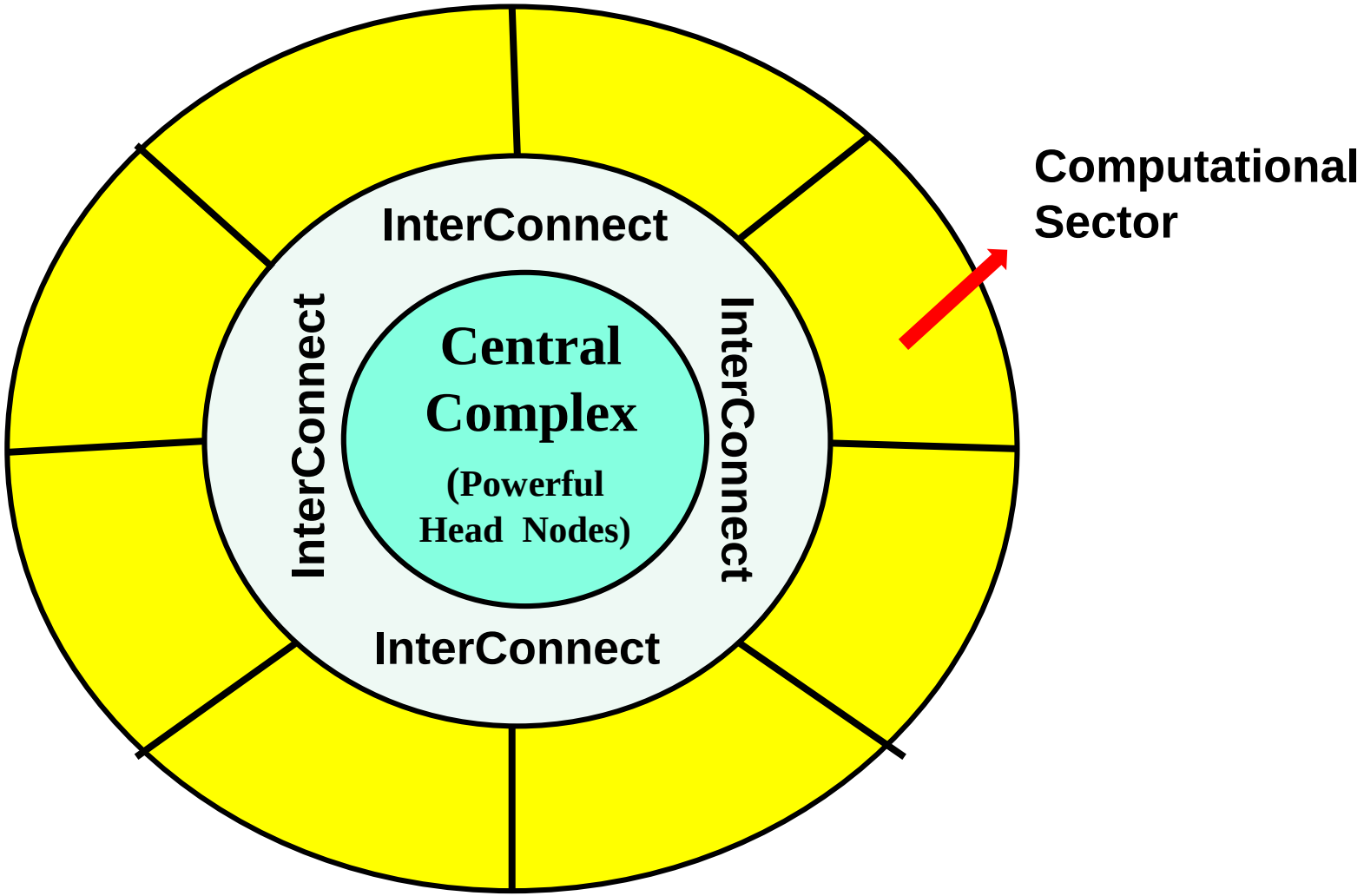
"Data Extraction" servers

High Impact Applications of Optimization and Statistics (Big-Data) on Multi-Petaflop Systems enabled by **MPPLAB(e-Teacher)**

MPPLab (e-Techer) Architecture



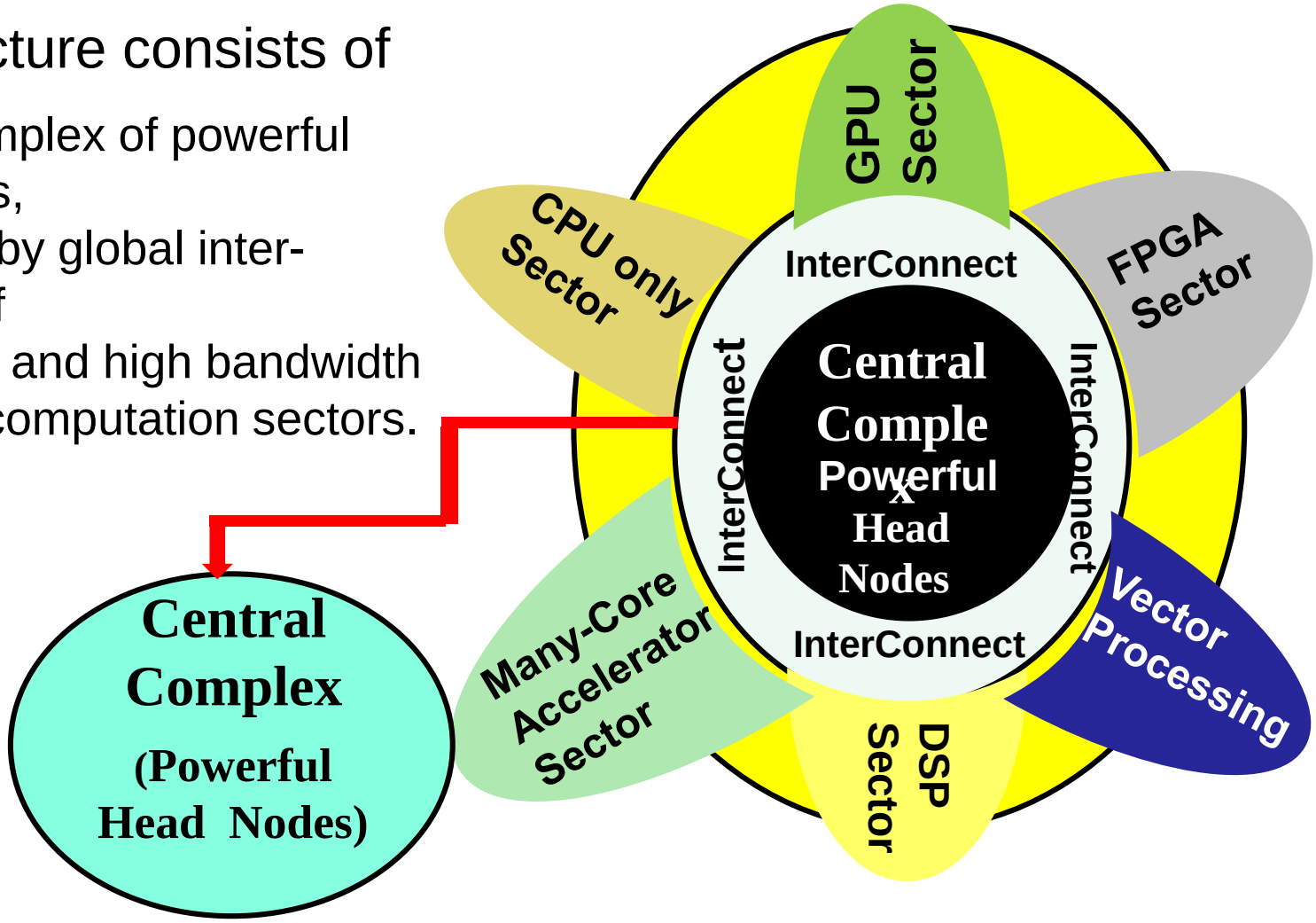
MPPLab(e-Teacher) : Unified Architecture - A single, highly integrated & Coherent Implementation framework



MPPLab(e-Teacher) : Unified Architecture - A single, highly integrated & Coherent Implementation framework

The architecture consists of

- Central complex of powerful head nodes,
- connected by global inter-connects of
- low latency and high bandwidth to several computation sectors.



Different Sectors can have Different Accelerators

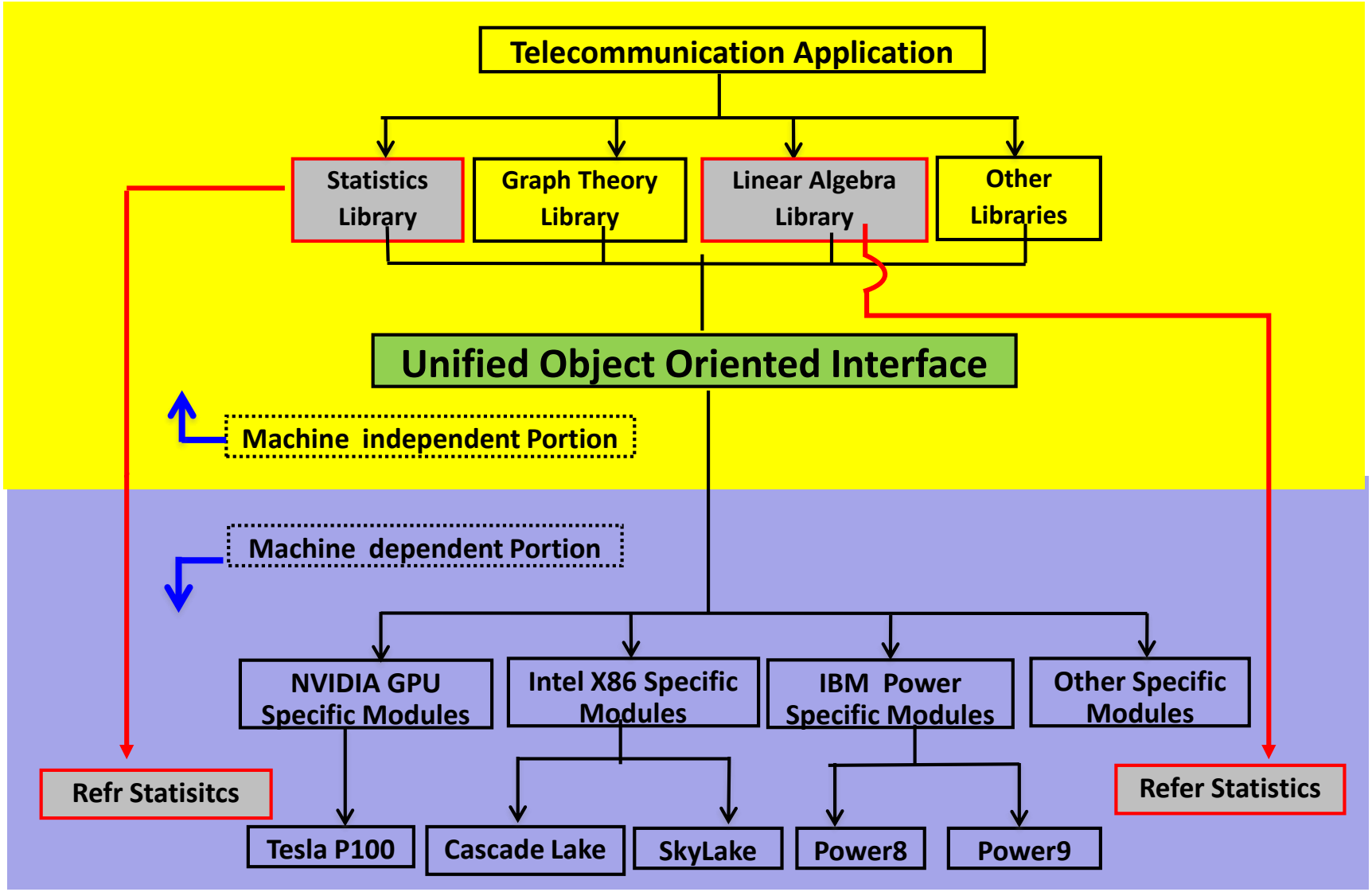
●Vector Processors ●DSP ●GPUs ●FPGAs ●Many-core processors

Stage 2

Development of Scalable Libraries

Object Oriented Programming
Fortran 2008

MPPLab(e-Teacher) : Unified Architecture - A single, highly integrated & Coherent Implementation framework



MPPLabe-Teacher : Developing Parallel Software by Evolution

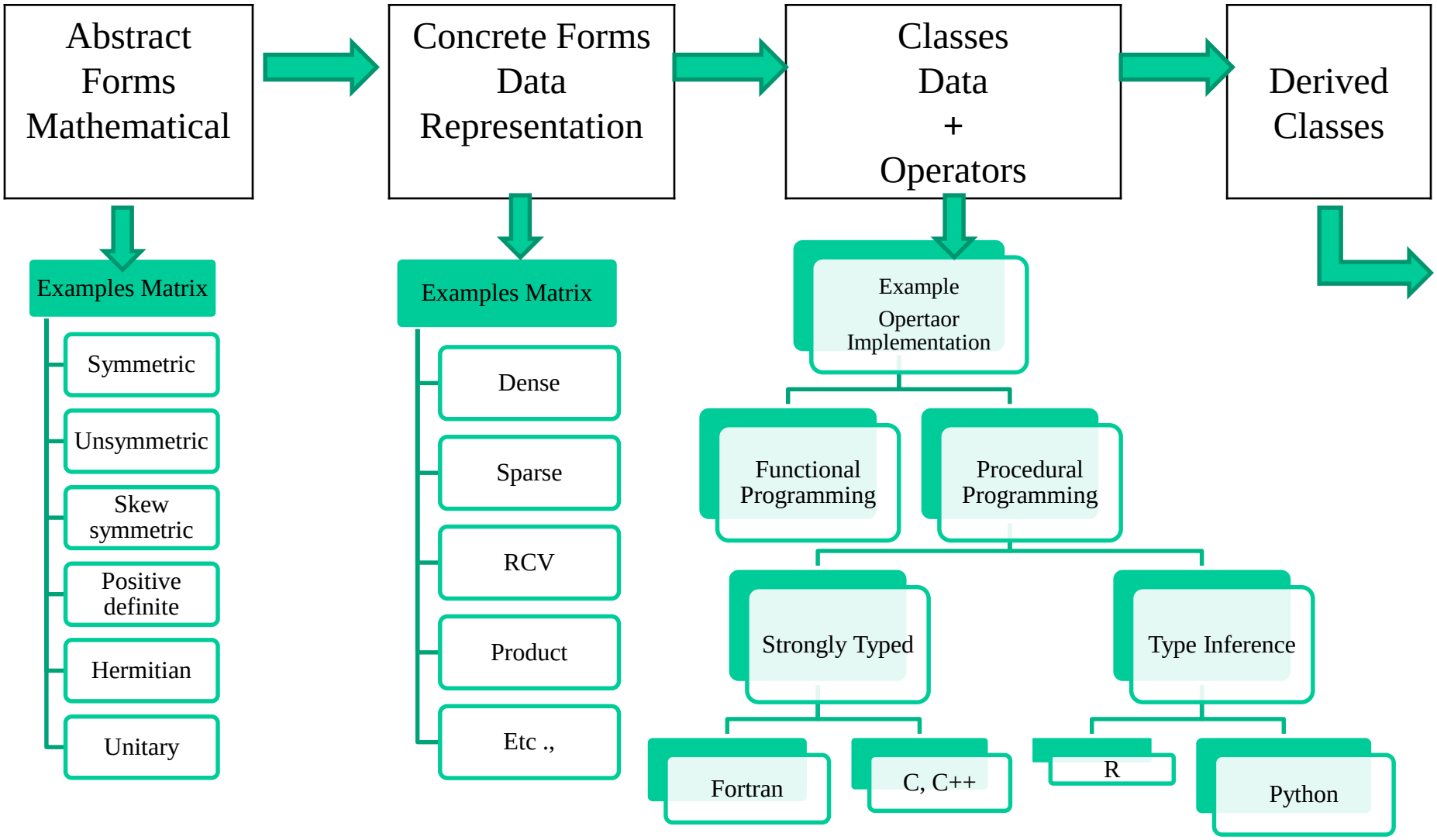
Parallel Application Software Composition : Chain of Step–Wise Refinements

For composing parallel software the system provides support for the following in a language independent manner

- **Static & automatic Parallel callable**
- **Identifying communication patterns & Grouping them into :**
  - **point – to – point**
  - **Broadcast**
  - **Multicast**
- **Make communication Groups**
- **Operations :**
  - **Load Balancing**
  - **bin-packing algorithm**
  - **applied to fine-grain level**
- **Data :**
  - **identify read only data single use/ multiple use**
    - **→ for a routine**
    - **→ global i.e. constant (parameter data)**
  - **data modified only once (hence by only one pro by multiple potential for conflict**
- **Re-oraganize loop for all**
- **Nested loops**
  - **single var → arrays**
  - **operators for cross product of sets**

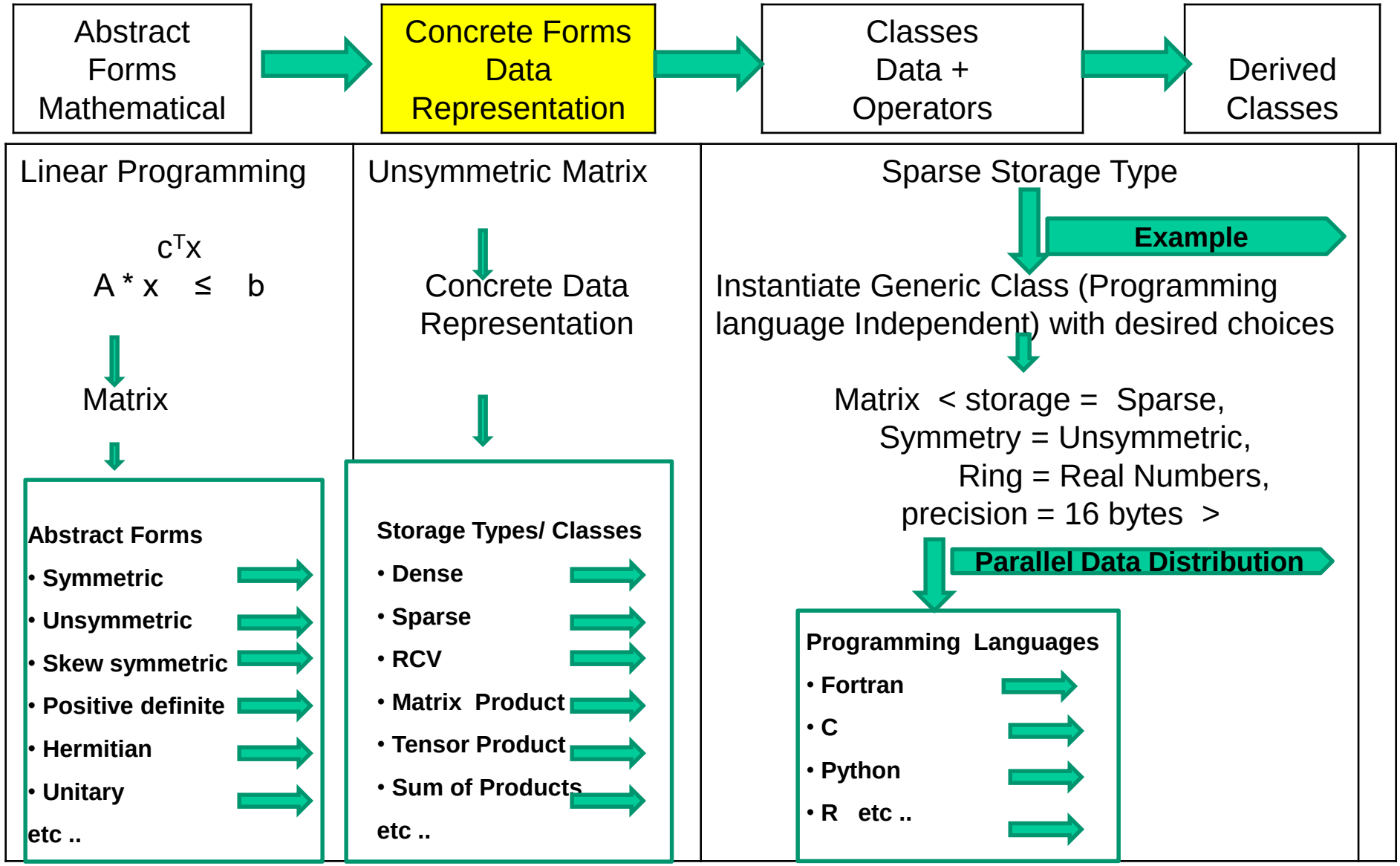
MPPLab e-Teacher:Parallel Application Software Composition

Sample Example

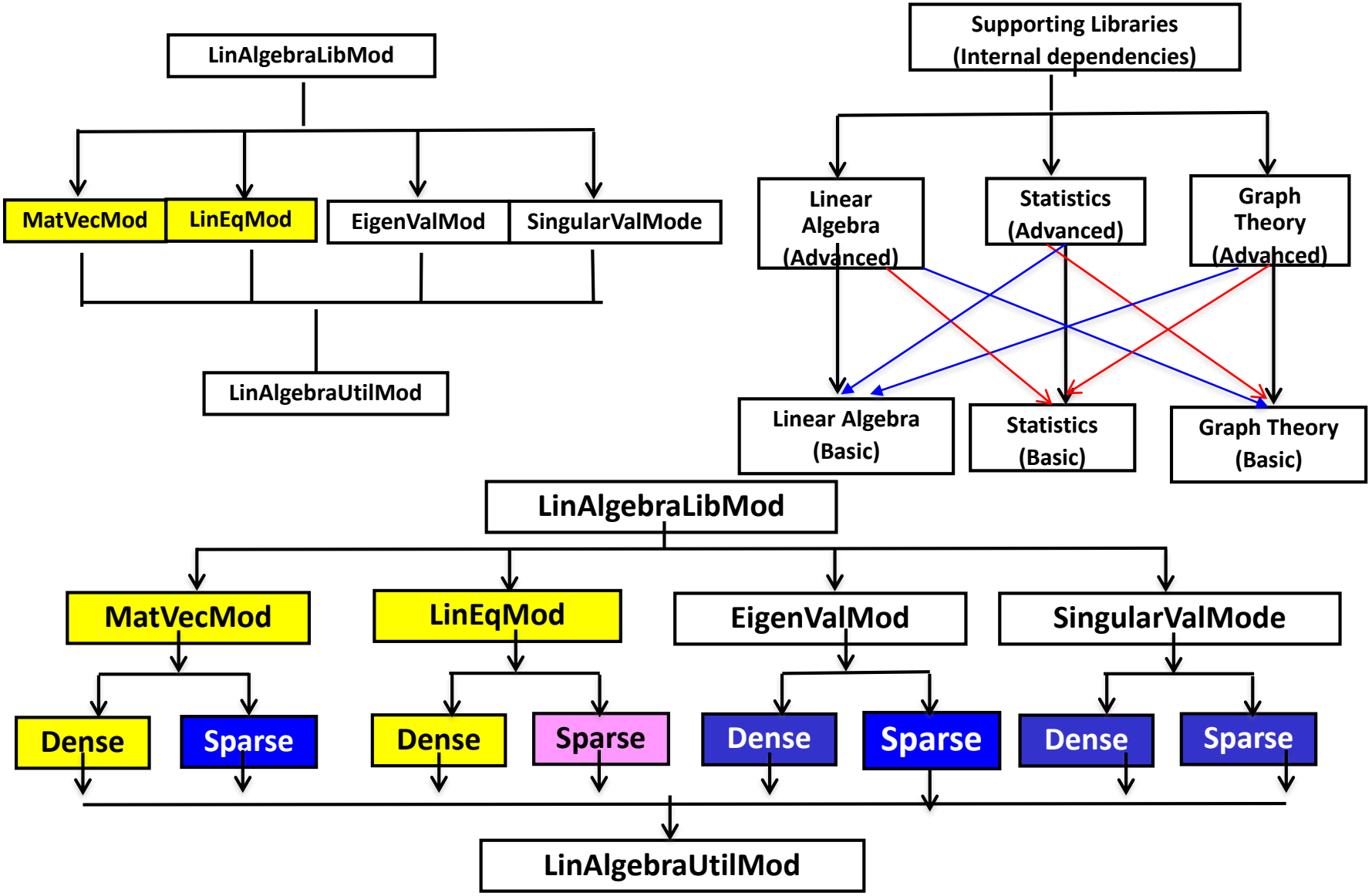


MPPLab e-Teacher:Parallel Application Software Composition

Sample Example



MPPLab(e-Teacher) : Numerical Linear Algebra Library



Conclusions

- An overview of the project titled “ High Impact Applications of Optimization and Statistics (Big-Data) on Multi-Petaflop Systems enabled by MPPLAB(e-Teacher)” is discussed

Thank You

Any Questions ?