Provenance is a key feature
• Keeps track of everything that happens during execution
• A log that can be queried
• Allows for high-level and domain-specific queries
• What are the maximum values for velocity and pressure on a given CFD simulation?
• A powerful association: Experiment meta-data with strategic experimental results
• Runtime analysis
• Good for time-consuming experiments
• User-steering for convergence analysis
• User interacts with data of interest at runtime
• Interface in the workflow execution to adjust

Provenance enriched visualization may support runtime and systematic analysis on results from large-scale experiments.

### Visualized and User-Steering

- Go fast to the results you need to analyze
  - Navigation over data results (partial) added by provenance (log)
  - Filter results and stage out only the data you need to analyze
  - Interact with a web-based user interface
- Export and analyze partial results to take decisions
  - Efficient data consolidation and data staging features to aggregate and stage out only relevant information
  - Use statistical and visualization tools empowered by provenance data
  - Draw preliminary conclusions

We are currently exploring experiments from CFD [5] with Uncertainty Quantification and Bioinformatics [6,7].

### Related Work

- **Visualization and User-Steering**
  - vírtual
  - Strong support for provenance & provenance visualization on tiled wall displays
  - No HPC support integrated to visualize data directly from clusters or clouds
  - HPC in very large scale with provenance
  - No runtime provenance support
  - Hard to enrich provenance results with provenance
  - ParaView Co-Processing Libraries [12]
  - Not related with SWIMS
  - No solution that integrates HPC execution with visualization enriching analysis with runtime provenance data

### Conclusion

- **Web Application to visualize experiment results enriched with provenance data**
  - Navigate through data, select desired results and visualize it on your workstation or on your tiled wall display environment
  - Keep track of which data produced each result
  - Query provenance data and visualize related results
  - Organize and associate data using tags
  - Provide a personalized view of the results
  - Easier to navigate on large-scale datasets
  - Good for fragmented outputs
  - Personalized actions to visualize data on specific formats

### Large-Scale Scientific Data Visualization using Provenance

- **Features**
  - Visualizes experiment data through the workflow
    - Displays provenance data for information browsing
    - Allows for selections and filtering over provenance data
    - Let scientists select just the results that interest them
    - Nest output files to visualize fragmented data formats
    - Putting tags on result files enables selective data staging for data formats that store information across separate files but need them together during visualization
    - The scientists can specify actions to visualize data with specific tags
    - The action can be programs or scripts defined by the scientists
    - Scientists can choose to run the action on server side or client side

- **Stage-out only the data the scientists selected**
  - Display the results locally or on a display technologies
  - Integrated with tiled wall displays technologies
  - Currently implemented to use TACDisplayCluster
  - Can be extended to other platforms such as CGXL or SAGE
  - Integration with ParaView

### Architecture

- **Web interface connected to Provenance Database and visualization environment**
  - Rich and interactive interface
  - Navigate through its activities and i/o data
  - Filtering options to select the desired results
  - Stage data out of the execution environment
  - Text, images, PDF, video and Paraview visualization
  - Web service in the visualization cluster to display staged output data in the Tiled Wall Display
    - Implemented as a Web Service
    - Interface accessed by the web module in runtime
    - Can also stage data out of the execution environment
    - Pluggable to support different tiled wall displays middlewares.

Action in tags can be one of paraview’s recorded scripts that reads the experiment results and produces a video, for example.

### References