In Memory
- Small to modest data
- Interactive or batch work
- Might have many thousands of jobs
- Excel, R, SAS, Stata, SPSS

In Server
- Small to medium data
- Interactive or batch work
- Hosted/shared and transactional data
- SQL / NoSQL
- Hosted data pipelines
- iRODS / Globus
- Document databases

Big Data
- Medium to huge data
- Batch work
- Full table scans
- Hadoop, Spark, Flink
Flux - High Performance Computing

- Available today
- ~25,000 Cores providing both batch and interactive use
- Common software packages R, SAS, Stata/MP
- Cannot host long running processes (servers / databases)
- Upto 1,500GB of memory in single node
- 100Gbps network
- 1.5PB Scratch Filesystem
- Soon: HIPAA Aligned
Turbo - High Performance Storage

- Available today (NFS v3/v4 Linux, Mac, Globus)
- 2PB total space
- Sensitive Data OK (HIPAA, FERPA, etc)
- Snapshots, Replication, Globus Sharing
- 40Gbps
- Soon: Windows (CIFS) clients, unreplicated
- Later: Swift (Object), Hadoop Native, CIFS+NFS
Hadoop / Spark - Big Data Discovery

- Available today
- 300TB HDFS (100TB Replicated)
- Cloudera and common tools installed
  - Hive, Pig, MapReduce, RMR, Sqoop
- 1 Minute scan of 1TB Hive Table
- Soon: Significant expansion
Hosted Databases and Services

- Pilot Today (MongoDB), Assume available
- Send us requests for what to host:
  - MySQL, MariaDB, Postgresql
  - Elasticsearch, Solr
  - Vertica, Cassandra, SciDB, Impala
  - iRODS
- Soon: Significant expansion, Solid state drives, Self managed machines
Other Services

- Consulting and support
- Contact for Pilot Credits

- Fire and forget transfer
- Share data off campus
- Cost covered by Campus
Pricing

- Compute and cloud: 1 CPU + 4GB: $11.72/month
- Storage: 1TB $240.30/yr replicated $120.15/yr single copy
- Hadoop Spark: based solely on HDFS space used
Brock Palen - brockp@umich.edu - @brockpalen
http://arc-ts.umich.edu
hpc-support@umich.edu