BIG DATA IN HIGH ENERGY PHYSICS

Igor Mandrichenko

Big Data meeting

4/3/2015

What is Big Data ?

- For different industries and areas of science it means different things
  - Clicks, ad exposures, movies preferences, hyper text links, genome sequences, weather patterns, stock trades, etc.
  - Different data structures, different complexity, different requirements
  - Common: no moving data between fast (random access) and slow (sequential access) storage

Big Data as a paradigm shift

- Big Data is the data processing methodology where *all* interesting data are *immediately* available for fast analysis at *any* time

- Wikipedia: Big data is a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale.

Benefits of Big Data

- Fast data analysis
  - No competition over resources
  - No data retrieval latency
  - High parallelism
- Broader set of problems available for solution

What does it mean for HEP ?

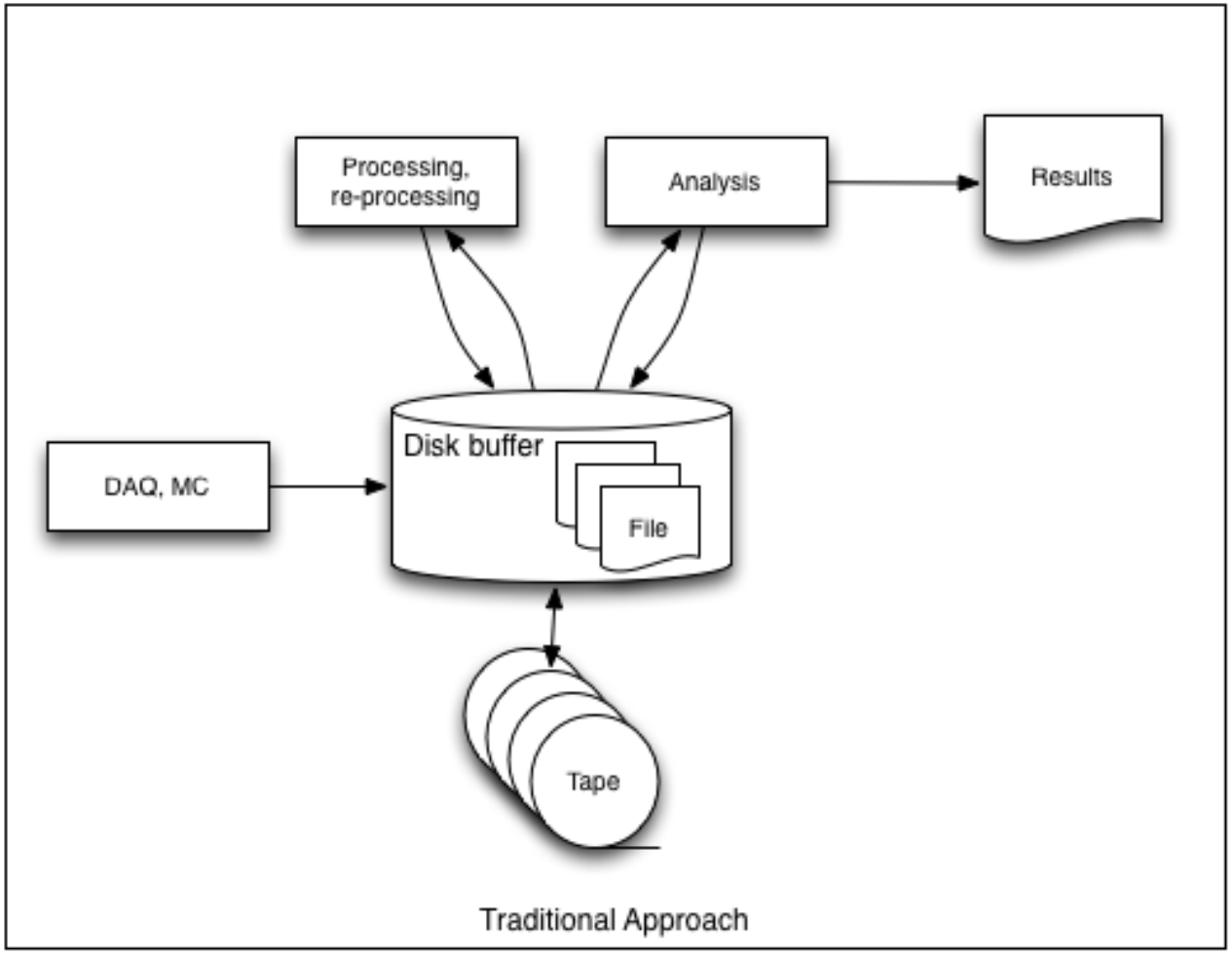
- To have all raw and processed data permanently stored in a scalable random access storage with fast, efficient data lookup (indexing) capabilities

- Benefits:
  - more efficient use of computational resources (CPU) – no need to wait for data staging
  - Fast, agile data (re-)processing, analysis
  - Additional areas of research

Traditional HEP approach

- Collect or produce data (DAQ, MC)
- Write raw data to tape as fast as you can, via disk buffer
- Process or reduce data (reconstruct events)
  - Read data from tape to disk
  - Write reduced data to disk and then back to tape
- Analyze data
    - Read reduced data (and often raw data) from tape to disk
    - Skimming: Filter “interesting” data – read each event, decide whether it is interesting
      - Save “interesting” data for future analysis (write to disk and then to tape)
        - Every group or individual saves their own “interesting” sets, duplicating data
    - Further reduce data into physical results (histogram, mass, cross-section, etc.)

Traditional HEP approach



Traditional HEP approach

- Tape is primary storage of all data from raw to intermediate results of the analysis, “write once, read multiple times”
- Disk is a buffer through which tape storage is scanned. Not considered reliable
- Data need to travel between tape and disk many times
- It is important to be able to filter and save “interesting” data once and reuse it many times so that there is no need to re-scan whole dataset again and again

Big Data technologies

- Storage
  - Scalable, redundant, efficient -> distributed
  - Elastic
- Databases ?
  - SQL vs. noSQL
- Map/reduce
  - Exists since MPI, at least
  - Successfully used by Google for data management and analysis

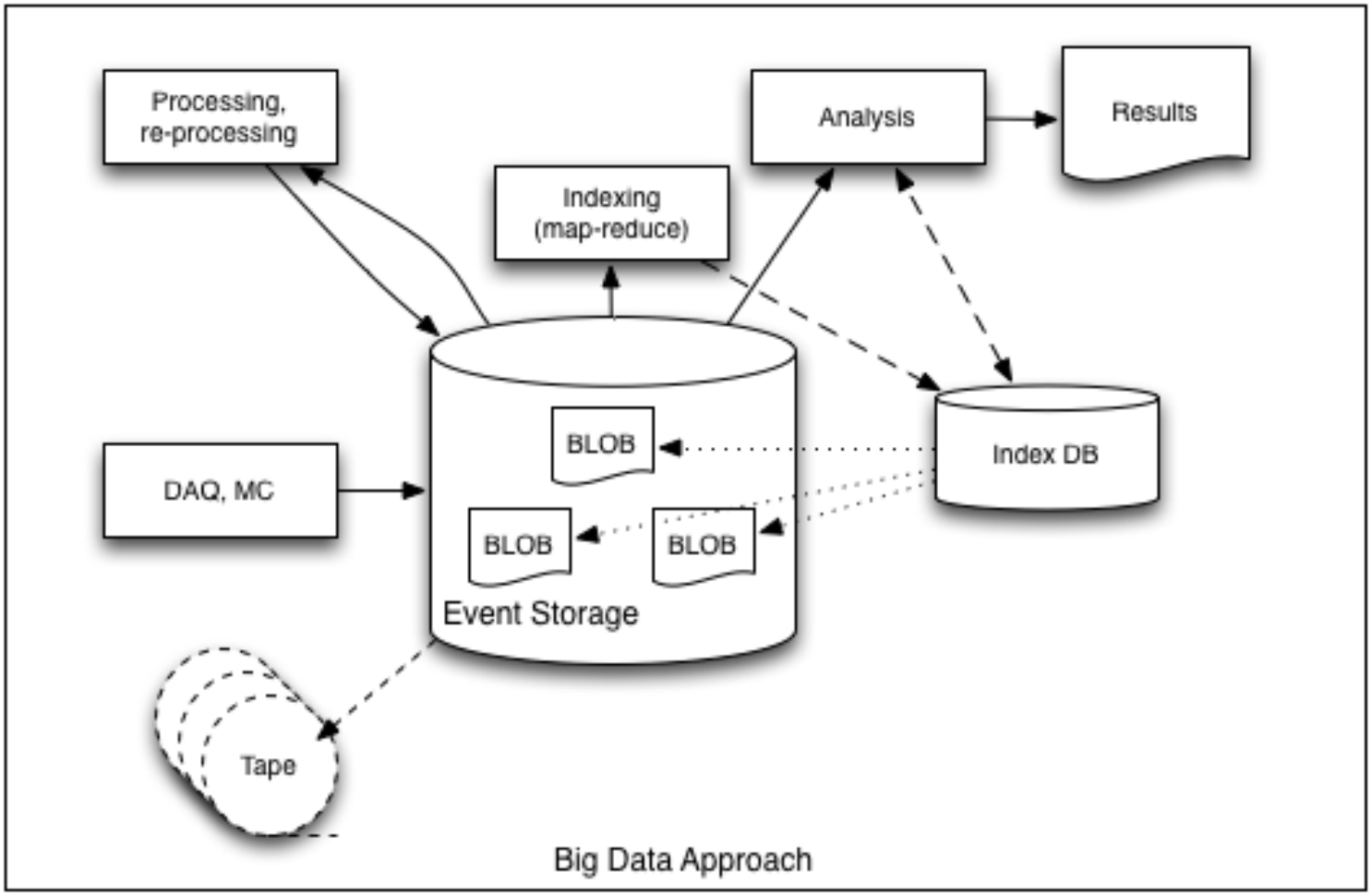
SQL or no SQL ?

- Relational model:
  - Powerful data representation, management and analysis tool
  - Move some simple calculations to the server side
    - selection, sorting, aggregation
  - Confined to single server architecture -> limited scalability
  - ~10-100TB limit
- Non-relational model:
  - Key-value storage plus some extras from RDB concept
    - Secondary indices
    - ACID
- Instead of choosing one or the other, we can use both
  - Store data in a key-value storage
  - Store metadata, structural information in a RDB

Proposed Big Data Approach

- Collect data, store on disk
  - Index data (batch map/reduce task)
- Data processing (event reconstruction)
  - Read new raw data from disk – use index to define what is new
  - Reconstruct
  - Write reconstructed data to disk (associate with raw data)
  - Index data
- Analysis
  - Use index to find potentially interesting events
  - Create and populate your own index
  - Analyze interesting data
    - Update your index
  - Produce physical results

Proposed Big Data Approach



Big Data vs. Traditional

- Disk is primary storage of data, instead of tape
  - 3-5 times replication
  - Tape – last resort backup, “write once, read hopefully never”
  - Data Representation – BLOB, application specific format
- Whole data set is always immediately available from disk
  - Direct, random access as opposed to sequential in case of tape
- Indexing instead of filtering and copying
  - Each piece of data exists in one (x number of replicas) instance, usable by anyone
  - Group or an individual user can create arbitrary indexes

Sizing

- Event Storage - ~1-5 PB
  - ~1000 nodes by ~3TB
  - Up to ~2PB effective size (2-3 times replication)
- Index Database
  - ~10 TB easily
  - Maybe up to ~100 TB

Big Data Approach - pieces

- Event Storage
  - Key (event id) -> BLOB (event data) database or storage, no SQL database
  - Disk based, distributed, replicated, elastic, scalable
  - Some computational capabilities
  - Backup to tape
- Index Database
  - User defined criteria -> event ID
  - Small because it does not contain event data
  - Can be a relational database – multidimensional indexes, complex queries executed by the database
- Indexing process
  - Map/reduce approach seems to fit well
  - Periodically running jobs, which analyze event data and populate indexes
  - Trick: make sure to process only one replica of each data item

What is Big Data

- Big Data is *not* about size
  - It has been known for decades how to collect and record petabytes of data
- Big Data is about the ability to quickly analyze large amounts of data
- Data can be collected and stored quickly, and then always "immediately" available for the processing (reduction) and analysis