HBaseHackathon Notes
Sunday, August 9, 2009

- Master Rewrite
- What does master do?
    - Job Splitting
        - Distributed
    - Region Assignment
        - Balancing
        - Scan root/meta
            - Online
            - Delete parents if no reference
    - Schema / Alter/online/offline
    - Admin
        - Close region
        - Flush
        - Compact / major compact
    - Watches ZK
        - Itself
        - For RS
- Worst case calculation
    - 200 regionservers
    - 32 logs per regionserver
    - 200 regions written to
    - 2GB or 30 hour full log roll
    - 10MB/sec write speed
    - 1.2M edits per 2G
        - 7k writes/sec across cluster
    - 1.2M edits per 30 hours
        - 100 writes/sec across cluster
- Assignment / balancing
    - RS publish load into ZK
        - /hbase/rsload/startcode({'json':'load'})
        - Configure period it is refreshed
    - Assignment inputs
        - Load
        - Requests / sec, regions online
    - Distribute tables randomly across cluster
        - Never give table back to who unassigned it
        - During split, bottom half to the same server, top half reassigned
    - Assignment Queue
        - Candidate Queue
            - Master watches ZK candidate queue
            - /hbase/rsassign/region(last_owned_by_rs)

                - When new nodes come in, it assigns the out
                - Regionservers put regions into the candidate queue when they unassign/close
            - To Open Queue
                - Regionservers watch their own to open queues
                - /hbase/rsopen/region(extra_info, which hlogs to replay or it's a split, etc)
- Administrative functions
    - Hadoop RPC listeners on Master and Regionservers
    - Clients and Master can talk to RS
- Safe-mode assignment
    - Collect all regions to assign
    - Randomize and assign out in bulk, one msg per RS
    - NO MORE SAFE-MODE
    - Region assignment is always
        - Look at all regions to be assigned
        - Make a single decision for the assignment of all of these regions
- META to ZK
    - New feature in ZK
        - Sorted tree or list node
        - Has a getClosestBefore / getRegionForRow(row)
    - No ROOT
        - META location stored in ZK
        - META table now only contains historian information
        - All other in ZK
- Worker pool region closing
    - Parallel flushes
- No more CHANGE TABLE STATE
- Process server shutdown after RS crash
    - Separate META scan?
    - To figure out regions on an RS
        - Separate map of RS -> region
    - Trade-off between two-writes during assignment
- Table Schema Information
    - Online schema edits?
        - If complex, punt to 0.22
    - Rather than storing with each region, stored once in ZK
- Uptime on UI
- Admin stuff
    - Straight from client ->regionserver
    - No more heartbeat piggyback
- Scaling documentation
    - More conf settings for block cache
    - How to adjust knobs