facebook

Presto (2015)
Past, Present, and Future

SELECT now() - INTERVAL ’10’ MONTH

By The Numbers

- 10 months
- 30 releases (0.68 to 0.98)
- 42 contributors (57 total)
- 1761 commits (4583 total)
- 2406 files changed
- 198,680 insertions(+) 96,833 deletions(-)

Presto@Facebook

- Scan PBs of data every day
- Execute millions of queries each month
- Process trillions rows a day
- 1000s of internal daily active users

New SQL Features

- Structural types (array, map, row)
- UNNEST (like Hive's LATERAL VIEW)
- Views
- Aggregate window functions (rolling avg)
- Tons of new functions (HLL, ML, etc)
- Session properties

Hive

- ORC, DWRF and Parquet
- Real structural types
- DATE type
- Null partition keys
- Improved partition pruning

New Connectors

- Kafka
- JDBC (not sharded)
  - MySQL
  - Postgres
  - Generic JDBC

Internal Changes

- New query queueing system
- Upgrade to Java 8
- New ANTLR4 parser
- New Bytecode compiler framework
- New aggregation and window framework
- Partition aware planner
- IPv6 support (verified)

Optimizations

- New ORC reader
  - Columnar reads, push down, and lazy
- Reuse hash calculation across operators
- Better work load balancing
- Add “Big Query” support
- Use partition metadata for simple queries
- More parallel table writing

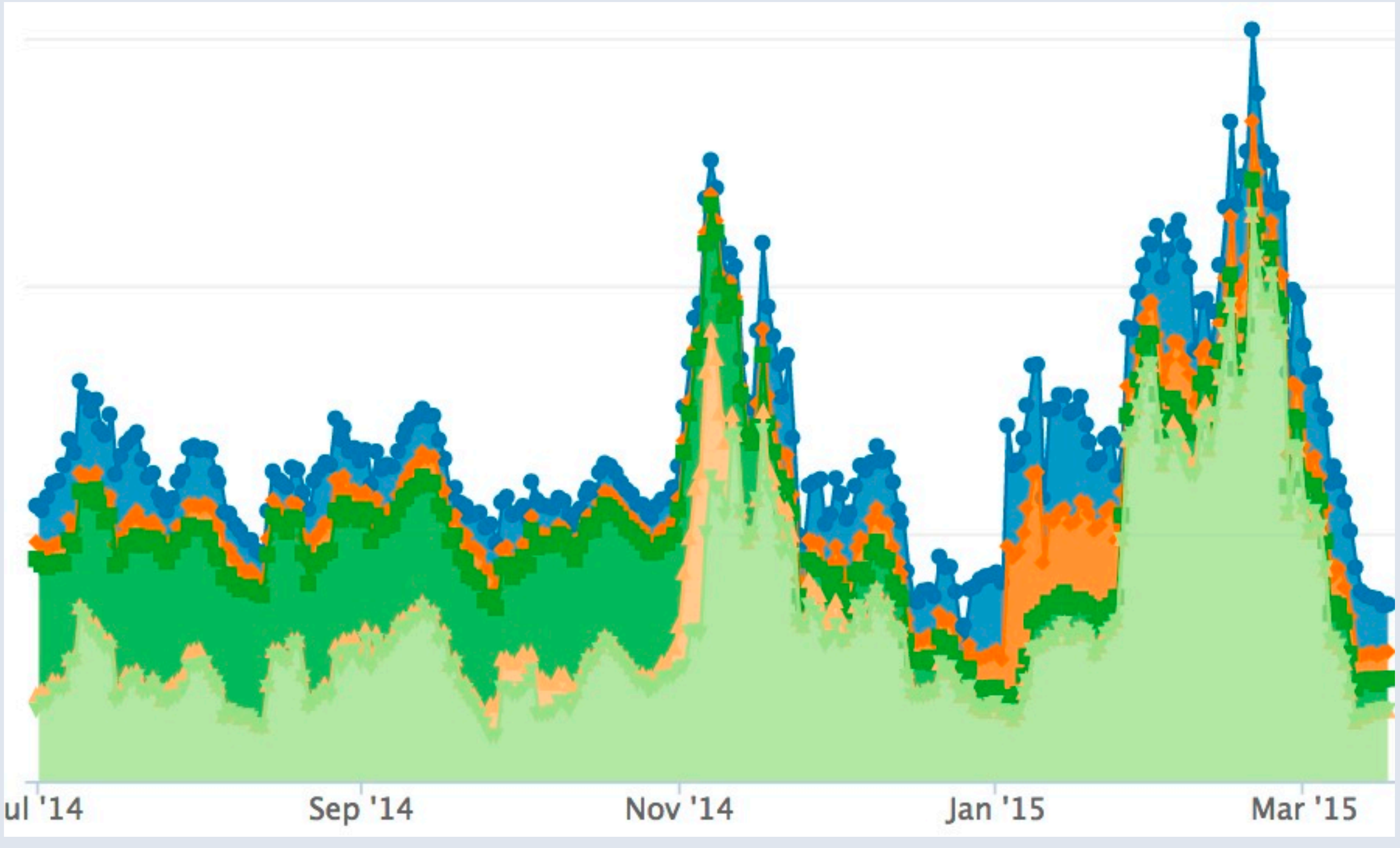
Optimizations

- Wall and CPU efficiency improvements
  - 50% for complex queries (joins, etc)
  - 300% for simple queries (scan, filter, agg)
- ORC Data
  - 2-4x wall and CPU time speedup
  - 4x+ speedup with lazy reads
  - 30x+ speedup with predicate push down

2014 Roadmap Checkin

- [x] Structural types
- [x] Native store
- [ ] Create partition
- [ ] Security
- [x] Distributed joins
- [ ] Native ODBC
- [ ] Huge joins
- [ ] Plugin repository
- [ ] Task recovery
- [ ] Full pushdown
- [ ] Work stealing
- [ ] Optimizer plugins

SELECT now()


ul '14
Sep '14
Nov '14
Jan '15
Mar '15

Reliability

- Presto resource leaks
- External bugs (Jetty, JVM)
- Obscure Presto bugs
- “G1 friendly” memory allocations
  - Avoid “humongous” allocations
  - Eliminated full GCs

Resource Management

- New queueing system
- Full global resource tracking
- Per query limits -- not per task
- Block queries until memory is available

Raptor

- Initial use cases
  - Near real-time loads (every 5-15 minutes)
  - 3 TB/day, 80B rows/day
  - 5 second query over 1 day of data
- Stores data in flash on worker nodes
- Metadata is stored in MySQL

Physical Aware Planner

- Multiple physical table layouts
- SPI changes
  - Clustering and partitioning
  - Simplify physically organized connectors
- Grouping (clustering) aware planning
- New physical aware operators

Security

- Authentication
  - Single sign-on: Kerberos or SSL client cert
- Authorization
  - Simple allow/deny check

SELECT now() + INTERVAL ‘1’ YEAR

APPROXIMATE AT 95.0 CONFIDENCE

SELECT now() + INTERVAL ‘1’ YEAR APPROXIMATE AT ~~95.0~~ CONFIDENCE 33.0

Resource Management

- Automatic query scaling (no “Big Query”)
- Better resource control
- Resource take back
- Bursting
  - Add and remove “extra” resources

New Planner

- Search/exploration
- “Cost-based”
- Better modeling of:
  - Correlated subqueries
  - Nested loops
- Connectors can participate

Security

- Authentication
  - Pluggable backends (LDAP, Kerberos, etc)
- Authorization
  - Pluggable backends (LDAP, JDBC, etc)
  - Integration with plugins
- Grant permissions from SQL

Execution Engine

- Result set caching support
- Adaptive execution
- Dictionary aware execution
- Columnar structural types
- Failure recovery
- Draining (possibly work stealing)

SQL Features

- Types with scale, precision, and length
  - varchar, char, varbinary, binary, decimal
- Full DDL support
  - CREATE, ALTER, DROP
- CUBE and ROLLUP
- Scalar and correlated subqueries (EXISTS)

Raptor

- Full production rollout
- Background storage optimization
- Multiple table layouts
- Indexes
- Real time loading
- UPDATE queries

Other

- Open source more internal connectors
  - Sharded MySQL
  - Generic Thrift

```
SELECT question
FROM audience
WHERE is_awesome(question)
```

facebook

(c) 2007 Facebook, Inc. or its licensors. "Facebook" is a registered trademark of Facebook, Inc.. All rights reserved. 1.0