



Boston Hadoop User Group Meetup, July 7, 2015

Kamil Bajda-Pawlikowski
Matt Fuller

Who are we? - Teradata Center for Hadoop!



- History of Teradata Center for Hadoop
 - Formerly Hadapt Founded in July, 2010 by Borgman, Bajda-Pawlikowski, and Abadi
 - Pioneered SQL-on-Hadoop market
 - Based on work done by database research group in Yale Computer Science Department
 - Hybrid of Hadoop scalability and DBMS performance
- Today
 - Acquired by Teradata in July, 2014, renamed Teradata Center for Hadoop
 - 30 developers with deep Hadoop and database expertise
 - Headquarters in Boston, MA
 - Contributors to open source project Presto



Talk Agenda



- What is Presto?
- What is Teradata doing?
- Can I see a Demo?
- How can I contribute?

What is Presto?



- 100% open source distributed ANSI SQL engine for Big Data
 - Modern code base
 - Proven scalability
 - Optimized for low latency, Interactive querying
- Cross platform query capability, not only SQL on Hadoop
- Distributed under the Apache license, now supported by Teradata
- Used by a community of well known, well respected technology companies

facebook



GROUPON™

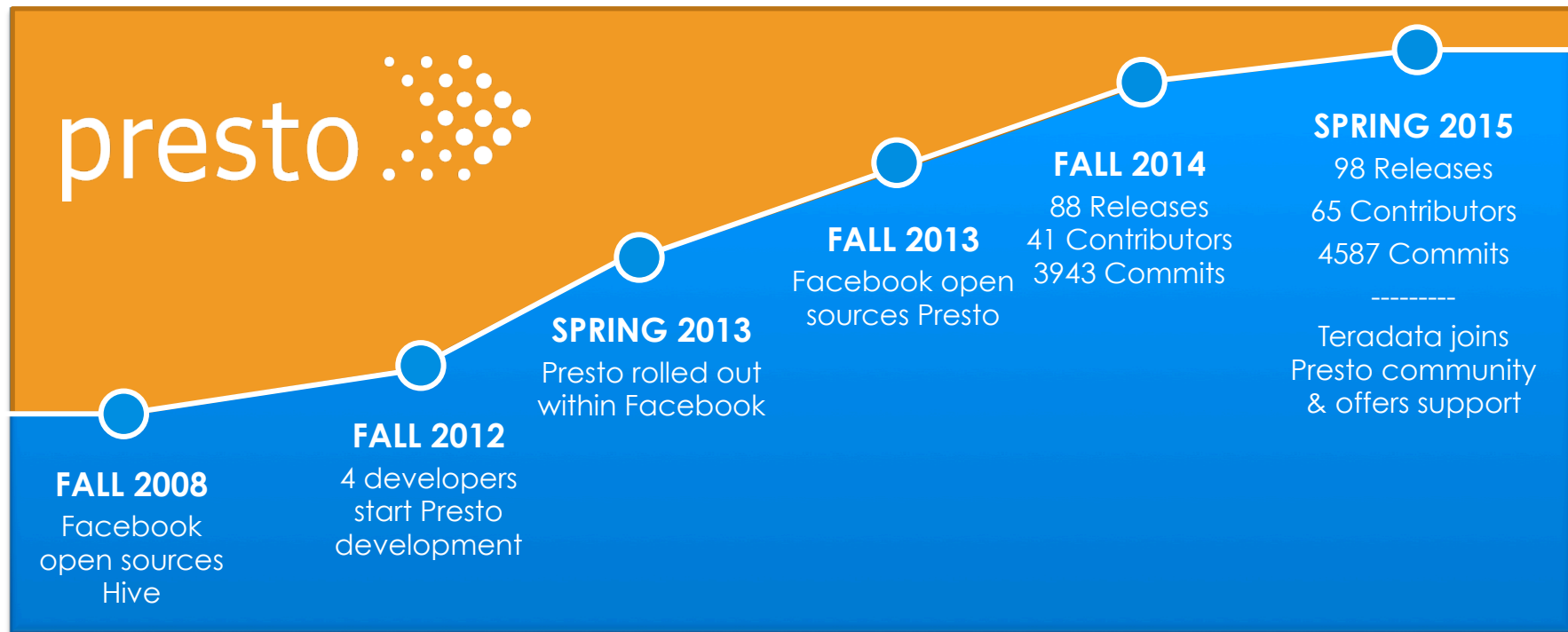
NETFLIX



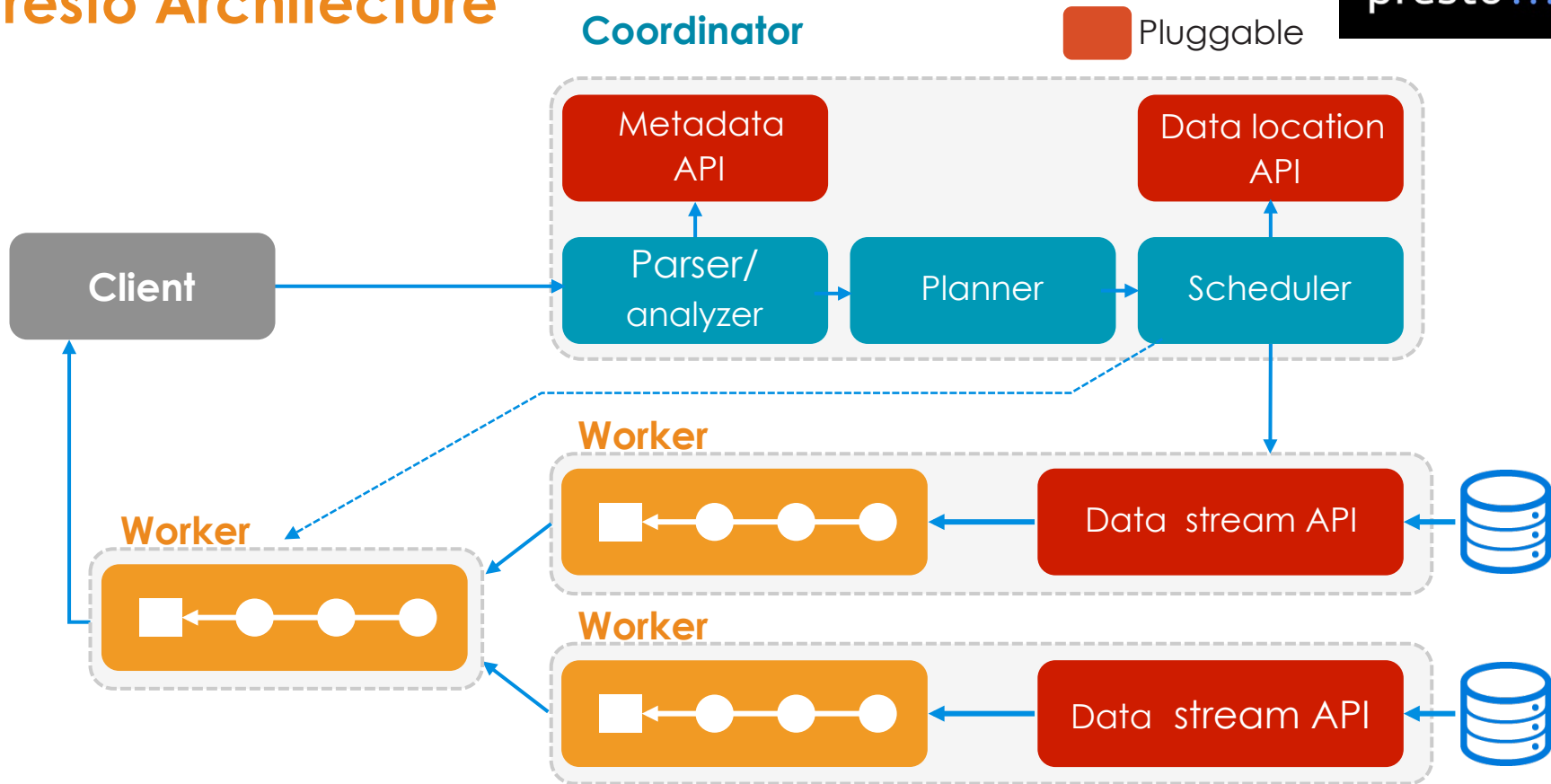
GREE

TERADATA

History of Presto

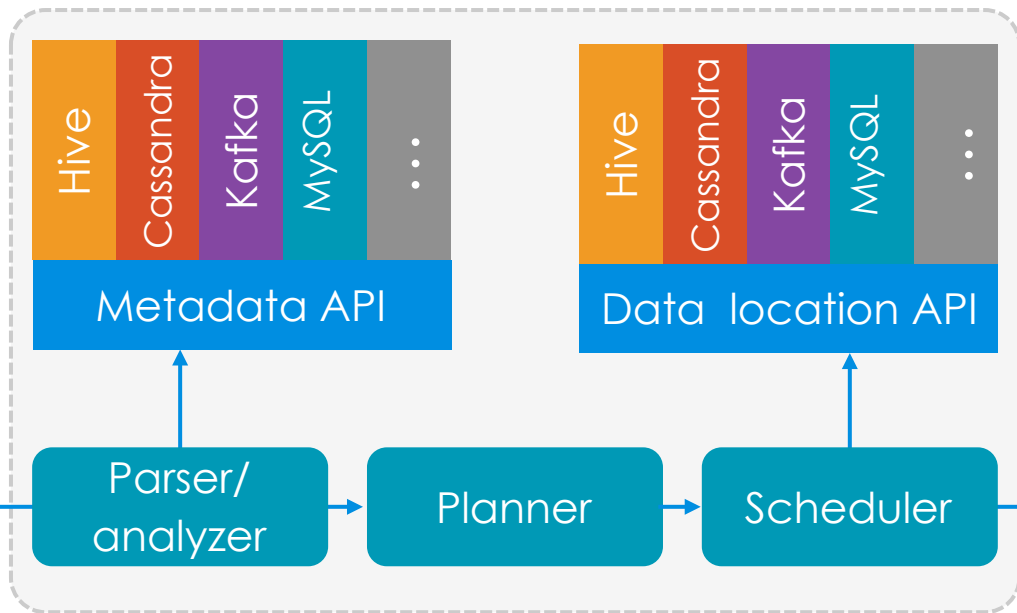


Presto Architecture

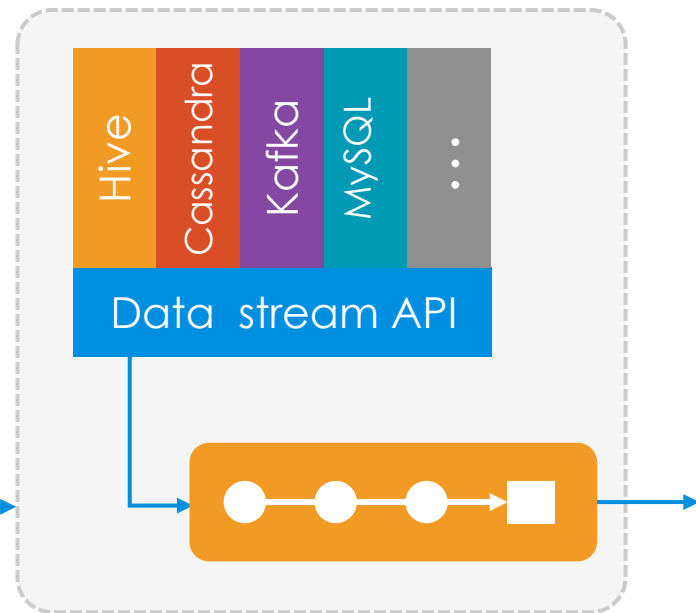


Presto Extensibility – connectors

Coordinator



Worker



Presto = Performance



- Data stays in memory during execution and is pipelined across nodes MPP-style
- Vectorized columnar processing
- Presto is written in highly tuned Java
 - Efficient in-memory data structures
 - Very careful coding of inner loops
 - Bytecode generation
- Optimized ORC reader

Presto in Production



- Facebook
 - Multiple production clusters (100s of nodes total)
 - Including 300PB Hadoop data warehouse
 - 1000s of internal daily active users
 - Millions of queries each month
 - Multiple PBs scanned every day
 - Trillions of rows a day
- Netflix
 - Over 200-node production cluster on EC2
 - Over 15 PB in S3 (Parquet format)
 - Over 300 users and 2.5K queries daily

What is Teradata Doing?



- **100% open source contributions** to Presto to increase adoption in the enterprise
- A **multi-year roadmap** commitment to phased enhancements of the open source code
- The first ever commercial **support offering** for Presto

Teradata Certified Presto
www.teradata.com/presto

Why is Teradata Contributing to Presto?

- Hadoop Distro Agnostic
- Modern Code Base
 - Presto is well-designed open source software with proper database architecture
- Strong Like-Minded Community
- Push down processing across multiple data platforms
- Leverage Teradata expertise to make SQL for Hadoop viable

Demo Time!



Teradata Contributions to Presto



Phase 1

June 8, 2015



Phase 2

Q4 2015



Phase 3

2016

Implement



Installer



Documentation



Monitoring & Support Tools

Integrate



Management Tool Integration



YARN Integration

Proliferate



ODBC / JDBC Drivers



BI Certification



Security



Connectors

Expanding ANSI SQL Coverage

Commercial Support



Teradata's Contributions

- Ease of install and management via Presto-Admin tool
 - www.github.com/prestodb/presto-admin
 - Packaging Presto as an RPM
- Testing Framework for Presto
 - www.github.com/prestodb/tempto
 - Added large number of tests
- Improvements to JDBC driver
 - To be open sourced on www.github.com/prestodb soon!
- Various SQL improvements

Teradata's Contribution Product Roadmap



- YARN Integration
- Ambari Integration
- ODBC & JDBC Drivers that actually work
- Security – Authentication & Authorization
- Continued SQL Improvements
- BI tool certifications – e.g. Tableau
- More Connectors – e.g. Hbase
- Open Source our Docker based Dev Env
- Open our Continuous Integration platform to the community

How can I contribute?

www.github.com/facebook/presto

www.github.com/prestodb

Certified Distro: www.teradata.com/presto

Website: www.prestodb.io

Presto User's Group: www.groups.google.com/group/presto-users

Facebook Page: www.facebook.com/prestodb

Twitter: [#prestodb](https://twitter.com/prestodb)

Presto 101t certified by Teradata



Available for Download

- Presto 101t Server, CLI, JDBC
- Presto-Admin 0.1
- Documentation
- HDP w/ Presto VM Sandbox
- CDH w/ Presto VM Sandbox

www.teradata.com/presto

