Big Data meets Fast Data: an scalable hybrid real-time transactional and analytics solution

Authors: Fred Melo & Dan Baskette
Speaker: William Markito
Failing data copies are replaced transparently

Failed function executions are restarted automatically

Network segmentations are identified and fixed automatically

Data is replicated to other clusters and sites (WAN)

Data is persisted on local disk for ultimate durability

Client and cluster disconnections are handled gracefully
China Railway Corporation

- 5,700 train stations
- 4.5 million tickets per day
- 20 million daily users
- 1.4 billion page views per day
- 40,000 visits per second

Indian Railways

- 7,000 stations
- 72,000 miles of track
- 23 million passengers daily
- 120,000 concurrent users
- 10,000 transactions per minute

World: ~7,349,000,000
~36% of the world population
Up to Petabytes

Master Host

Interconnect

Segment Host

Segment Host

Segment Host

Segment Host

Locator

Up to Terabytes

GREENPLUM DATABASE

APACHE GEODE

Dell Certified Memory
Data Science, Analytics & ML

ANSI SQL

Parallel Configurable Data Load

Transaction data write behind

Push updates

Rest / HTTP

Data Temperature

Hot

Warm

Pivotal
Map based on Longitude (generated) and Latitude (generated). Color shows sum of Transaction Value. Size shows sum of Possible Fraud. Details are shown for Zipcode.
Fraud Qty and Financial Impact

Zipcode: Color shows sum of Possible Fraud. Size shows sum of Transaction Value. The marks are labeled by Zipcode.
Source code

https://github.com/Pivotal-Open-Source-Hub

Join our communities!

https://github.com/apache/incubator-geode
Java, Distributed Systems

https://github.com/greenplum-db/gpdb
C/C++, Distributed Systems
References

https://github.com/Pivotal-Open-Source-Hub

http://greenplum.org

http://geode.incubator.apache.org

Come talk to us at booth and grab a sticker
Thank you

https://github.com/Pivotal-Open-Source-Hub