Marketo's Data Journey

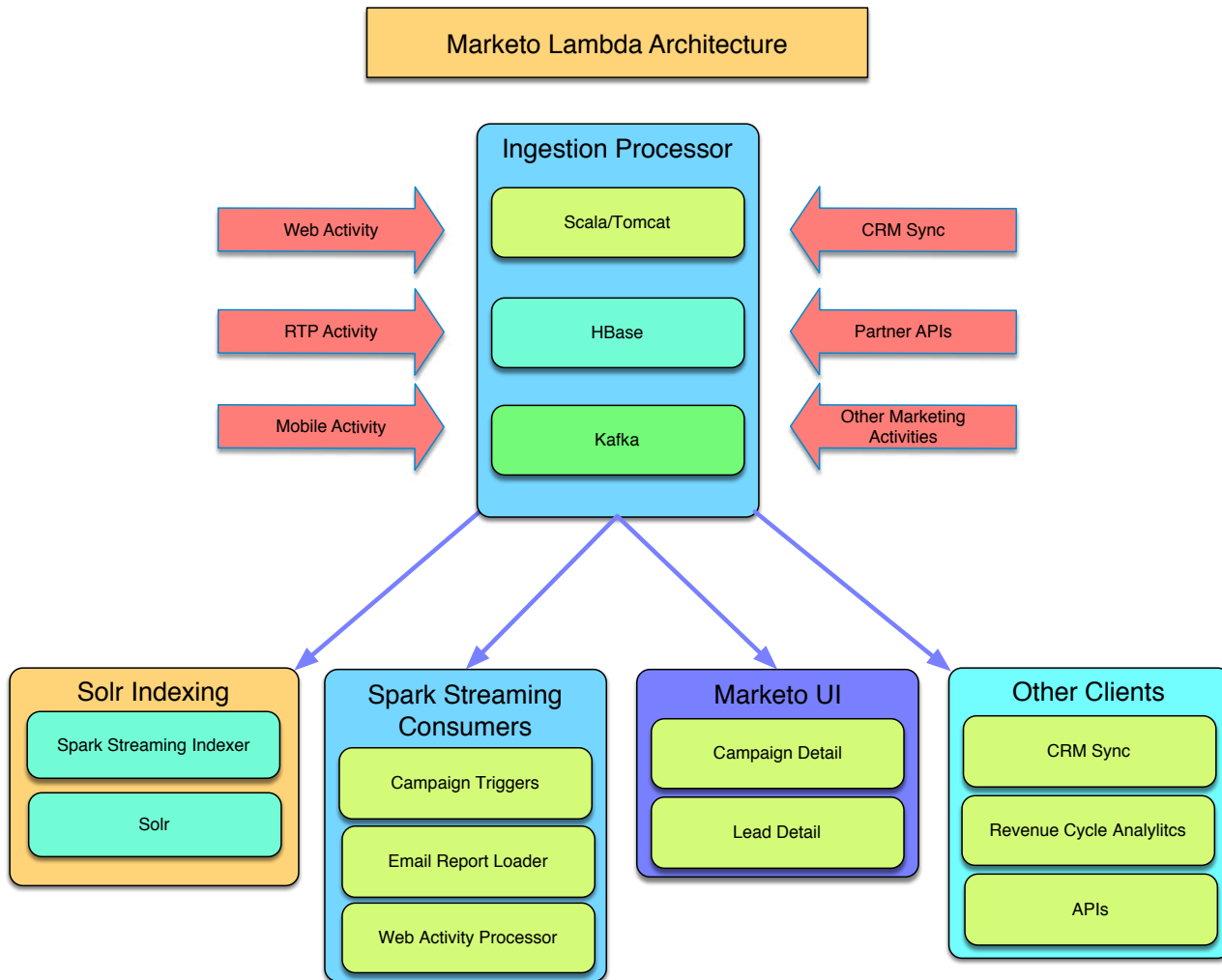The Vision

Business Requirements

- Near real-time activity processing

- 1 billion activities per customer per day

- Improve cost efficiency of operations while scaling up

- Global enterprise grade security and governance

Architecture & Design

Architecture Requirements

- Maximize utilization of hardware

- Multitenancy support with fairness

- Encryption, Authorization & Authentication

- Applications must scale horizontally

Marketo Lambda Architecture

Ingestion Processor
- Scala/Tomcat
- HBase
- Kafka

Web Activity →
RTP Activity →
Mobile Activity →

← CRM Sync
← Partner APIs
← Other Marketing Activities

Solr Indexing
- Spark Streaming Indexer
- Solr

Spark Streaming Consumers
- Campaign Triggers
- Email Report Loader
- Web Activity Processor

Marketo UI
- Campaign Detail
- Lead Detail

Other Clients
- CRM Sync
- Revenue Cycle Analylitcs
- APIs

High Level Architecture

- Enhanced Lambda Architecture

- Inbound activities written to Ingestion Processor
  - High volume (e.g. web) activities
  - CRM sync
  - Mobile Activities
  - etc

- In parallel, events are written to event queue and permanently stored
  - Events stored to facilitate simple historical queries, and keep a permanent system of record
  - Events consumers subscribe to event queue and process injected events in near real time
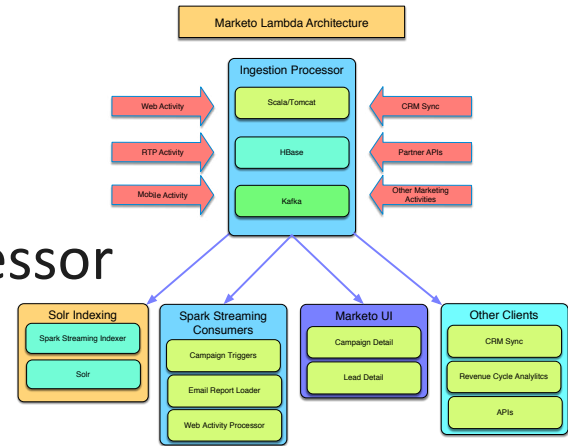
Ingestion

- Ingest interaction data at scale via a REST service

- Persist data into Kafka and HBase in parallel

- Kafka is the queue for downstream consumer to ingest activities

  - We have a data pipeline where some activities (e.g. web activities) are enriched prior to storage and triggering

- HBase is the source of truth for all interactions and activities

New Data Architecture

- Took a traditional MySQL relational database and split it into three specialized data streams

- Kafka for high speed queuing

- Hbase for massive data storage

- Solr for sophisticated query processing

Compute Model

- Phoenix is used as our SQL query layer

- Spark streaming is our pipeline processing engine to trigger campaign execution

- Spark framework allows easy horizontal scale out of compute workload

- Multitenant architecture at scale

Marketo®

Downstream Consumers

- Real-time triggered campaign processing
- Solr indexing for complex query processing
- Email report generation
- Account Based Marketing
- Microsoft CRM integration

Monitoring and Alerting

- Needed to monitor hundreds of new Hadoop and other infrastructure servers

- Our custom Spark Streaming applications required all new metrics and monitors

- Capacity planning requires trend analysis of both the infrastructure and our applications

Marketo

Monitoring and Alerting

- Built a monitoring infrastructure using OpenTSDB and Tcollectors

- Added business SLA metrics to our Sirius console to provide real-time alerts

- Added comprehensive Hadoop monitors into our pre-existing production monitoring system

- Created a common metrics library for use by application components

Marketo®

Business Analytics Reports

- Infrastructure built using Druid, Kafka and Spark batch jobs

- Allows for the ability to report on key performance indicators for email campaigns
  - Per customer
  - Per campaign

- Reports across the marketer's customers to detect campaign effectiveness

Marketo®

Impact and Results

- Rolling out the new platform to existing customers

- Immediate positive performance impact 100x!

- Higher hardware utilization

- Faster delivery and adoption of new applications on the new platform

- Top talent loves modern architecture

We are hiring!
Q & A

Marketo®