



Ingesting clicks data for analytics

Francesco Furiani, C.T.O. @ ClickMeter (PositiveADV srl)



ClickMeter

- Take control of marketing links and maximize conversion rates
- Tool to monitor, compare and optimize all their links in one place

Some stats:

- 100k+ customers
- Getting events for customers from 10 to 3000 req/sec (raw are way higher)
- Parse all of those :)



Getting the data

ClickMeter receives data anytime someone:

- Click on our links
- View our pixels

Our customers uses links/pixels:

- Inside a famous app the day of the big release ✓
- Advertising on an extremely big video portal ✓
- A tiny travel blog ✓
- A physical device for advertising ✓



The challenge

These type of situation are not really predictable

- Unless the customer informs us beforehand (unlikely to happen)

We need to:

- Scale up (customers get angry in case of errors or data not showing)
- Parse data to show to the customers for better insight (they love it)
- Do it as fast as possible
- Do it as cheap as possible



How to do it

Or how we thought to do it...

We obviously need some edge servers to keep answering to those HTTP events

- Beanstalk

We need to write this stuff somewhere

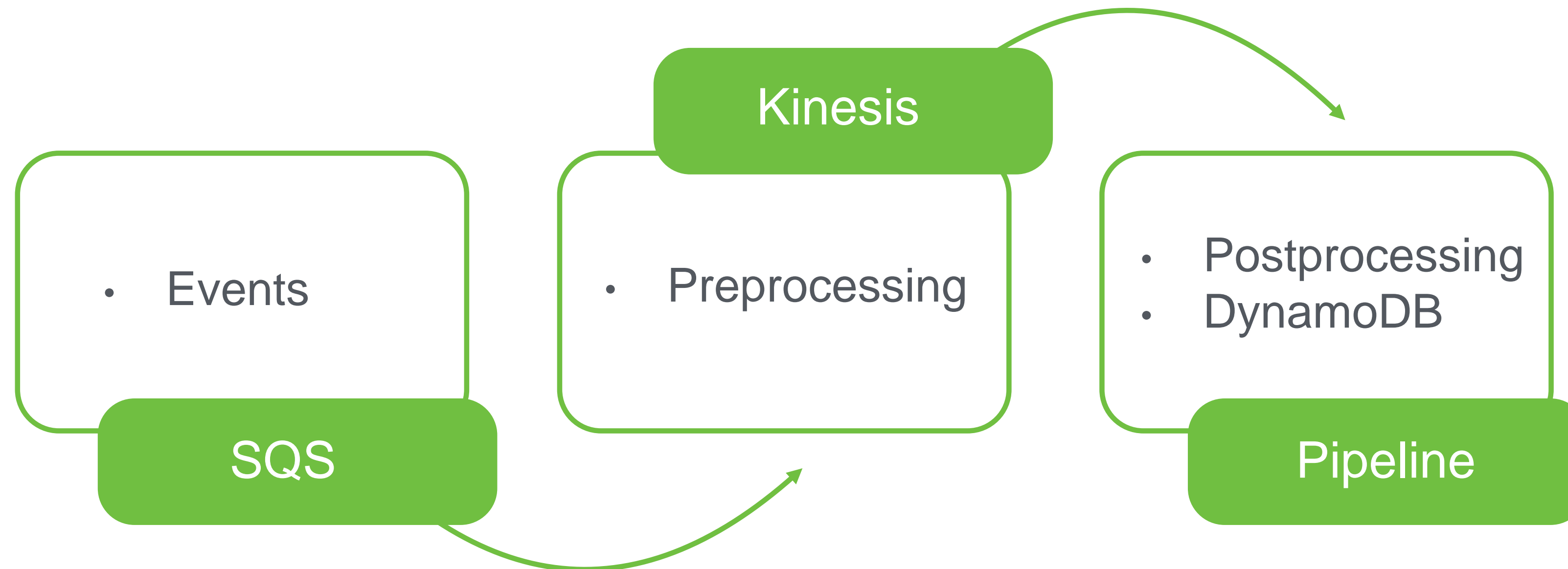
- Kinesis, SQS, DynamoDB

We need to parse/enrich this data (either in real-time, either in batch)

- Kinesis + ReactorKinesis, Pipeline + EMR, S3

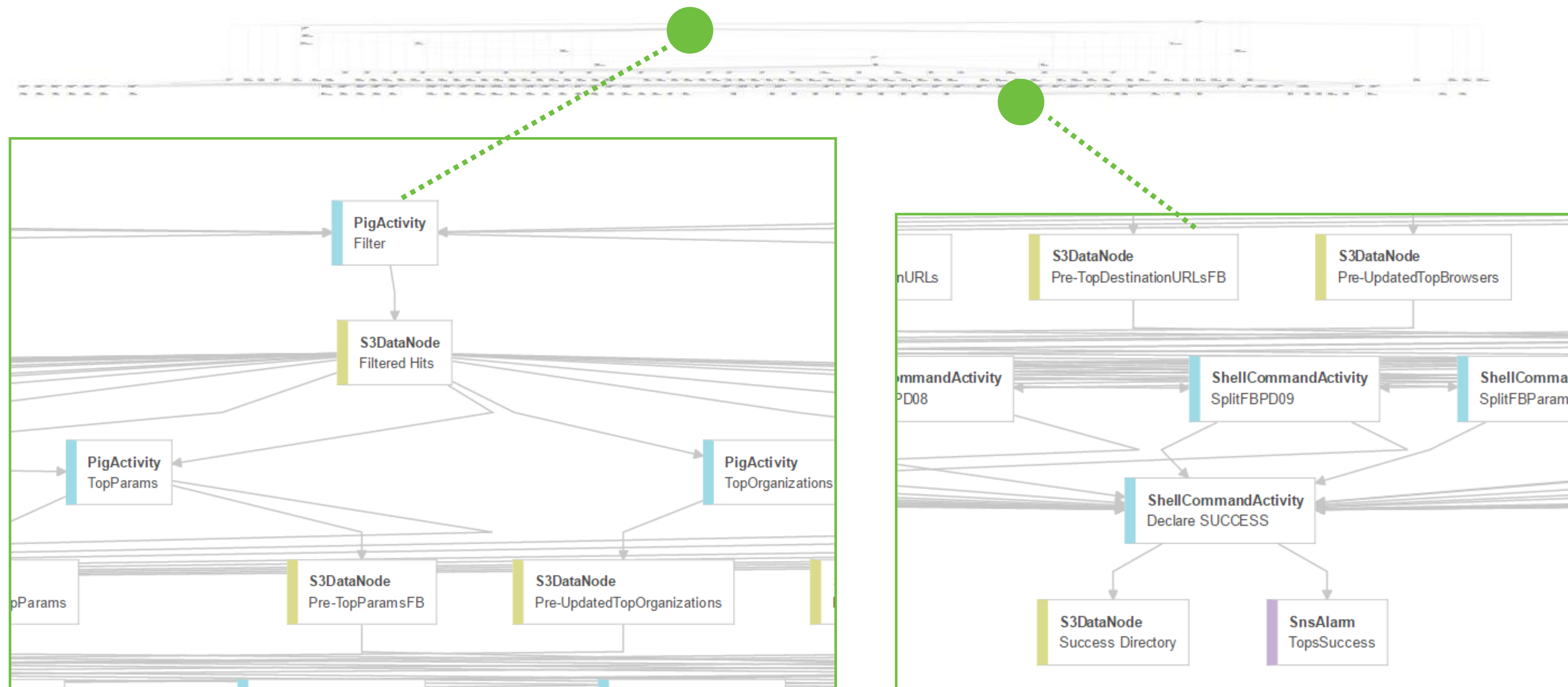


SQS + Kinesis + DynamoDB





Pipeline + EMR





Benefits

- Architecture is pretty much scalable via Cloud Watch Metrics and Scaling Groups
- Customers are happy and they bring more customers
- More data incoming, means a better way for us to research and deliver more insight
- Sleep at night or as the teach says it «High Availability»



Benefits

AWS take care of some operations that would require a dedicated DevOps

- We're faster in delivering new features

Also it gives us the possibility to scale up without increasing (much) the IT budget

- We don't need to buy machines in advance to scale when I receive a peak

Route53 + ELB are very helpful in making our customers have the best experience

- Services and Instances can go down but they get replaced/rerouted easily



Future plans on AWS

EMR + Spark

- Better than plain Hadoop, waiting for PIG on Spark compatibility

DymannoDB Streams + Lambda

- Seems a very nice integration to explore

AWS Machine Learning (prototypes to be ported upon)

- Better data insight

Thank You

Any questions?

 @il_furio

