Think Data; not Database

Arup Nanda

VP, Data Services

Priceline.com





booking.com priceline.com kayak.com agoda.com rentalcars.com opentable.com

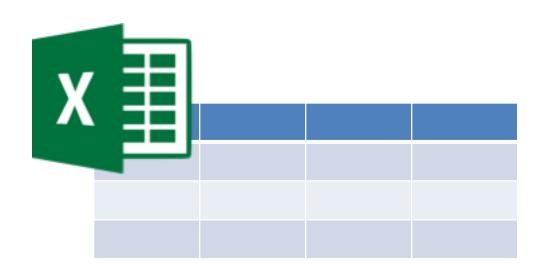


Data

Oracle
MySQL
Postgress
Cassandra



Data



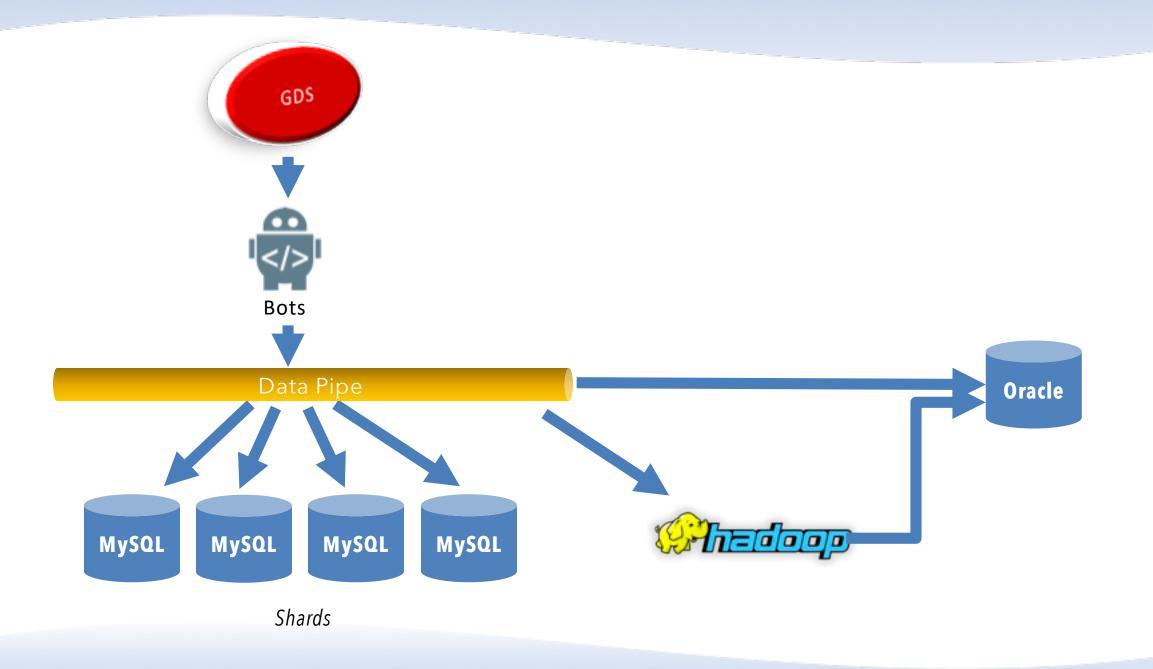


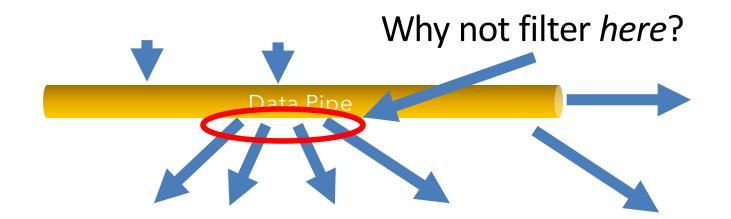
Java Programs

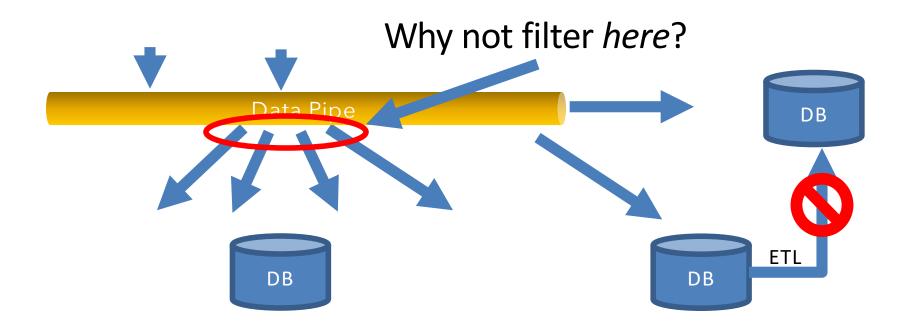
```
// Java Program to Print Prime Numbers from 1 to N
     public class PrintPrimeNumbers1
         public static void main(String[] args)
              int i, number, count;
             System.out.println(" Prime Numbers from 1 to 100 are : ");
              for(number = 1; number <= 100; number++)
                 for (1 = 2; 1 <= number/2; 1++)
                      if(number % i == 0)
                          count++;
                          break;
                 if(count == 0 && number != 1 )
                     System.out.print(number + " ");
                                   ©tutorialgateway.org
<terminated> PrintPrimeNumbers1 [Java Application] C\Program Files\Java\jre1.8.0_161\bin\javaw.e
 Prime Numbers from 1 to 100 are :
2 3 5 7 11 13 17 19 23 29 31 37 41 43 47 53 59 61 67 71 73 79 83 89 97
```

Variables; not data











Querying in DB

SQL Problem:
Its relies on static data



Querying in Kafka

KSQL Realtime Streaming Query

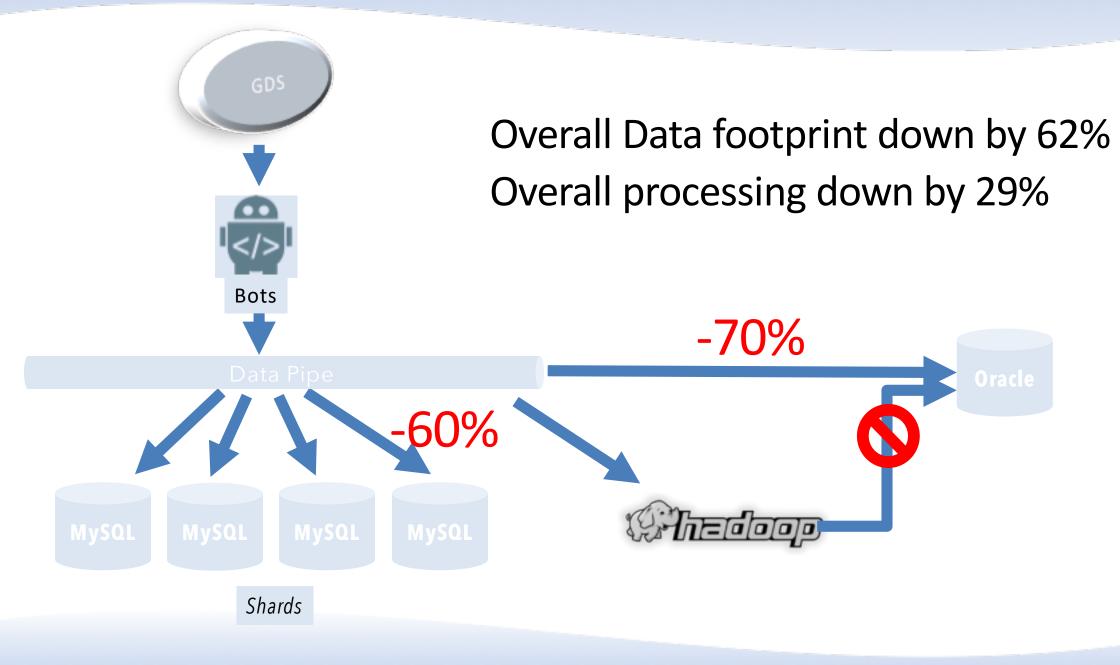


CREATE STREAM searchpath (sigtime BIGINT, cguid VARCHAR, sigkv VARCHAR, vip VARCHAR) WITH (kafka_topic='searchpath', value_format='JSON');

CREATE TABLE all_signals (sigtime BIGINT, cguid VARCHAR, sigkv VARCHAR, vip VARCHAR) WITH (kafka_topic='searchpath', value_format='DELIMITED');

CREATE TABLE vip_signals AS SELECT cguid, count(*)
FROM all_signals WINDOW TUMBLING (SIZE 1 MINUTE)
WHERE vip = 'Y';





Takeaways

- Think "data"; not "database"
- Take advantage of event processing where realtime is not important
- Filter at the pipe
- Don't think in terms of ETL; think data engineering



Thank You!

Blog: arup.blogspot.com

Tweeter: @ArupNanda

Facebook.com/ArupKNanda