Bloglines BETA

Paul Querna

BloglinesFS

- Task: Storage for "every blog post ever."
- Inspirations: MogileFS, GFS
- A little bit different....

PodServer

- Data Location Server
- Stateless
  - On startup, indexes all data nodes
- Hash Partitioned
- Also Provides:
  - Local and Cross-D.C. Replication
  - Spigots for Blog Search and other proejcts

ItemDB

- ‘Dumb’ data nodes.
- A single site is a chunk.
  - Ex: All of Slashdot is stored in one file.
  - Each chunk has several local indexes
    - Date, Post URL, etc

Read Example

- App -> PS: Fetch last 5 Articles from BoingBoing
- PS -> App: BoingBoing is on [....]
- App -> ItemDB: Fetch last 5 from BoinBoing

Web Applicaiton

PodServer01

PodServer02

ItemDB

ItemDB

ItemDB

ItemDB

ItemDB

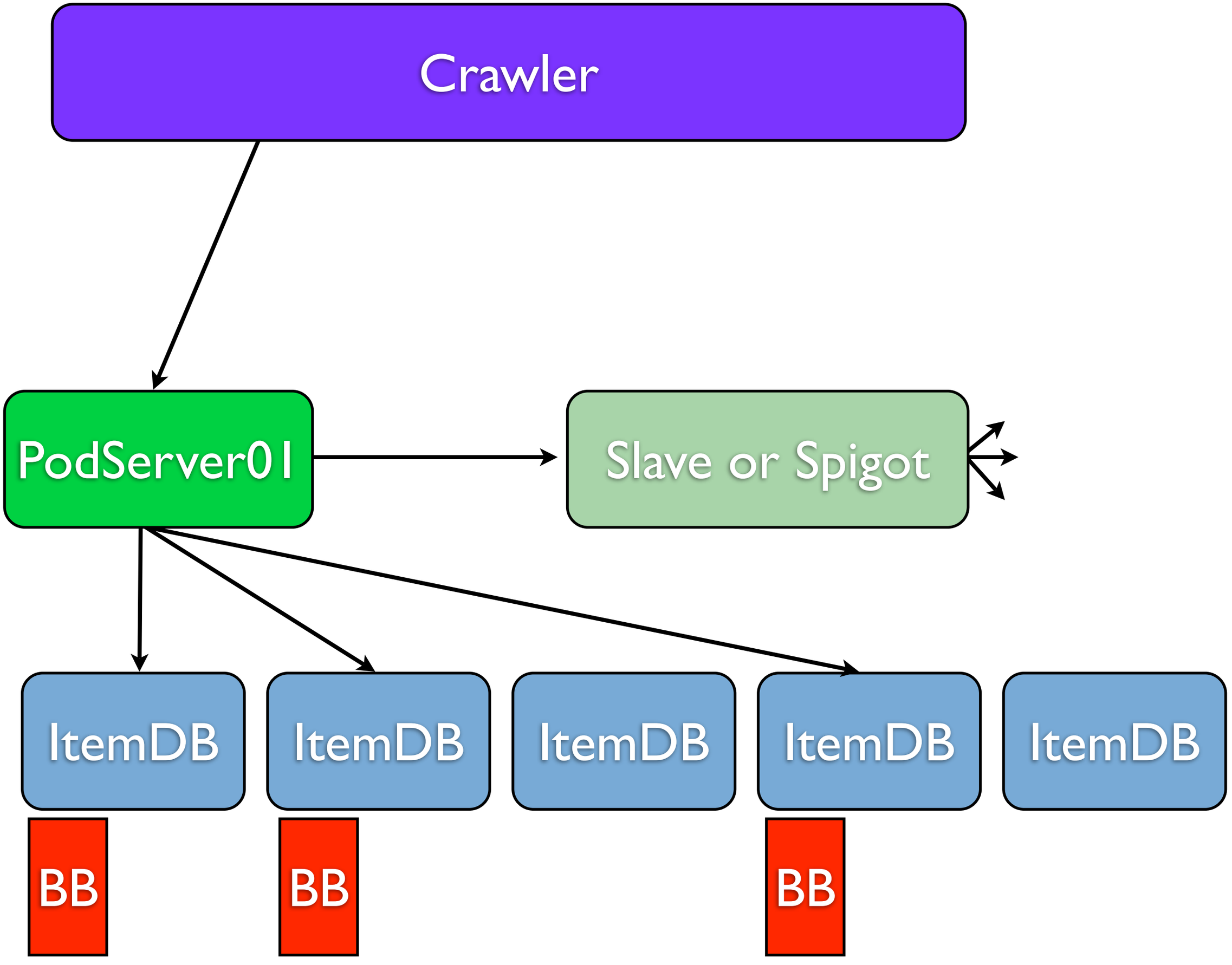
BB

BB

BB

Write Example

- Crawler -> PS: Add new item to BoingBoing
- PS -> [ItemDBs with BB]: Add item (versioned)
- PS -> Slaves/Spigot: Replicate


Crawler
PodServer01
Slave or Spigot
ItemDB
ItemDB
ItemDB
ItemDB
ItemDB
BB
BB
BB

In Production

- 75 Data machines
- 2x 300G disks each.
- One ItemDB per disk, no RAID.
- Thousands Concurrent Writes+Reads per second.
  - Crawl Everything Every 30 minutes...