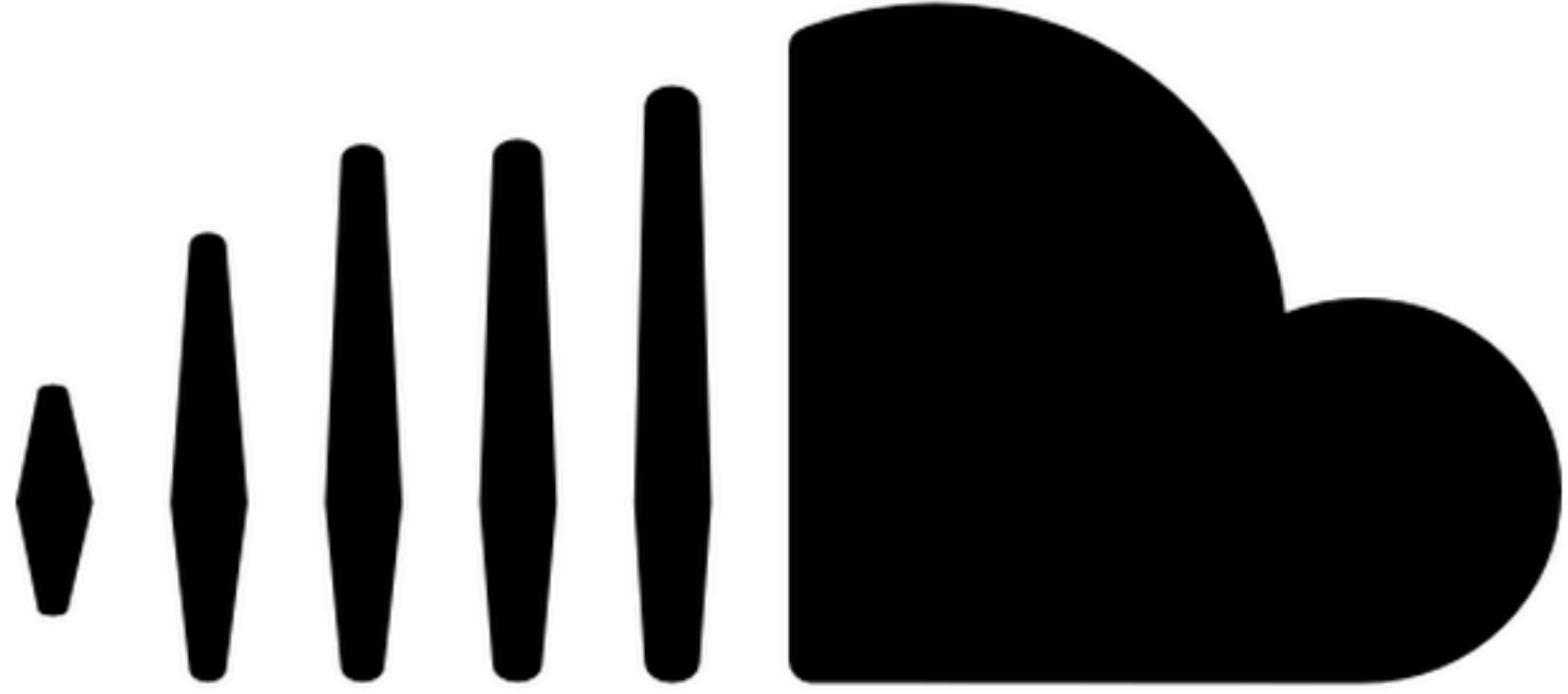


Mechanics of Data Pipelines

Berlin Buzzwords 2017

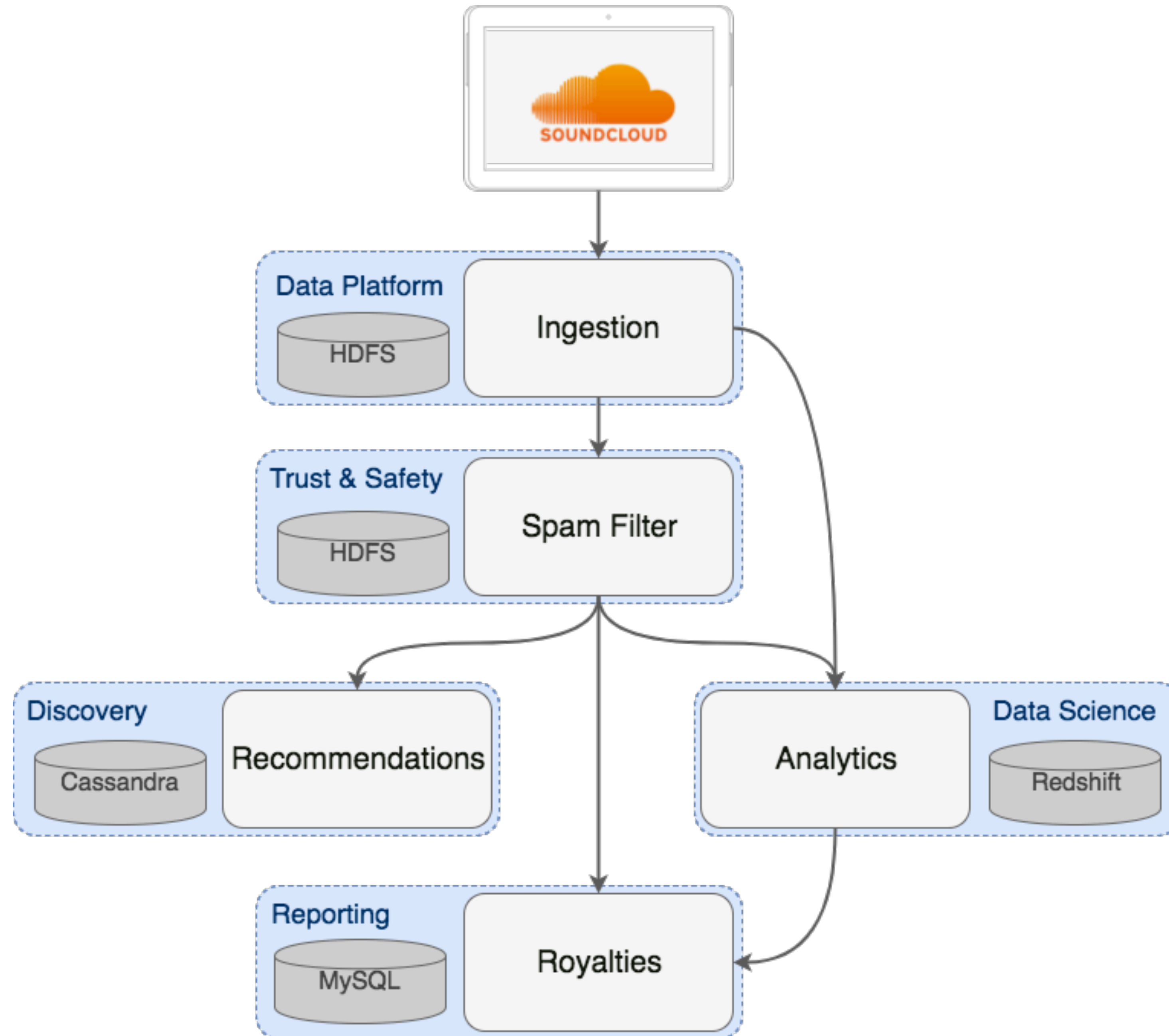
@brapse



Data Pipelines

The software structures which emerge to process and disseminate information.

A connected set of map reduce jobs for loading data into (a) database(s).



Why Data Pipelines?

To integrate diverse perspectives.

Enable and empower collaboration between diverse sets of domain experts.

How do we build them?

Often Badly.

Misunderstood domain.

Misunderstood integration.

Misunderstood coordination.

Thesis:

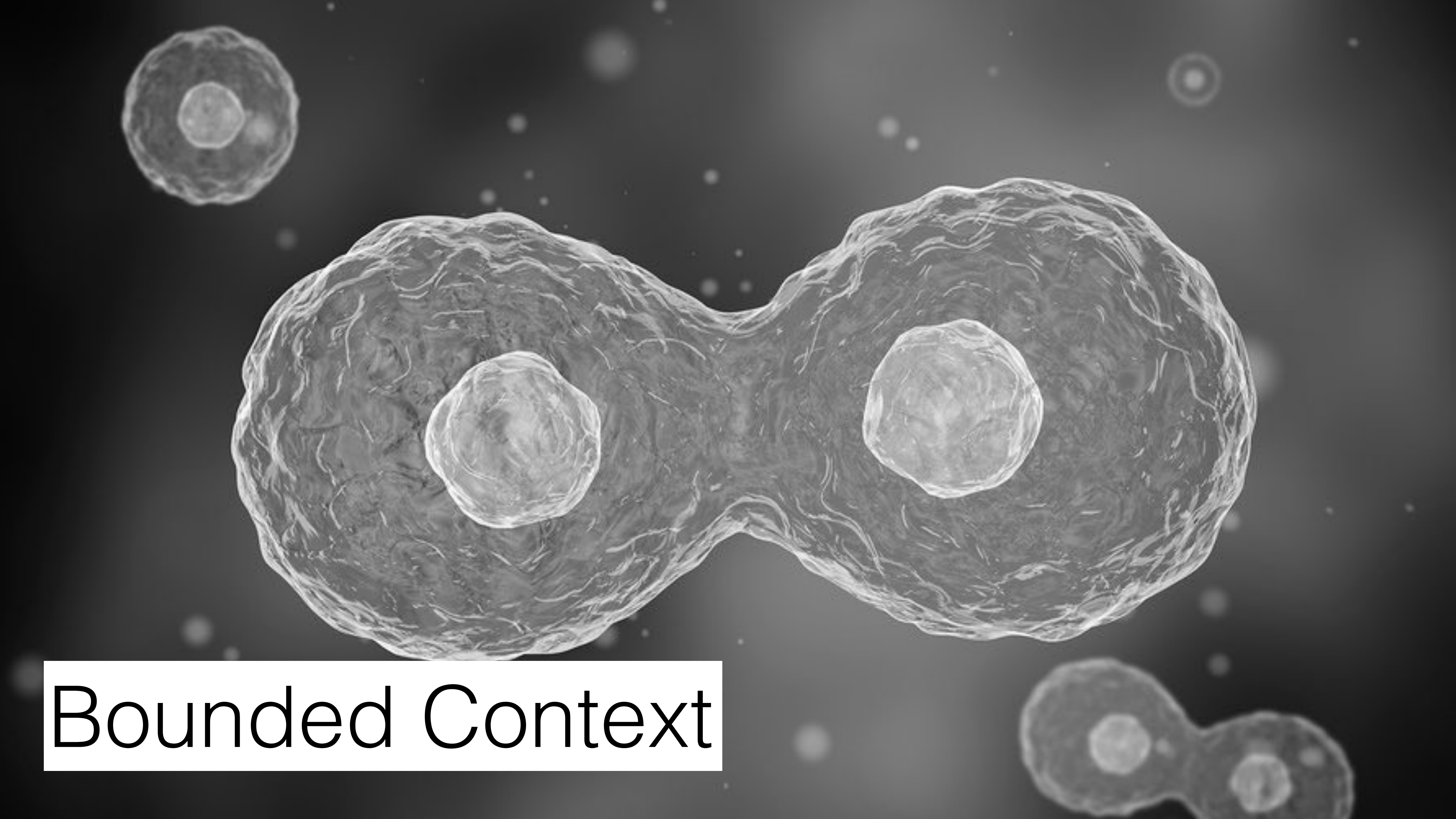
Data Pipelines emerge and grow to reflect collaboration between domains and are impeded by incidental coordination.



Abstract



Evolution

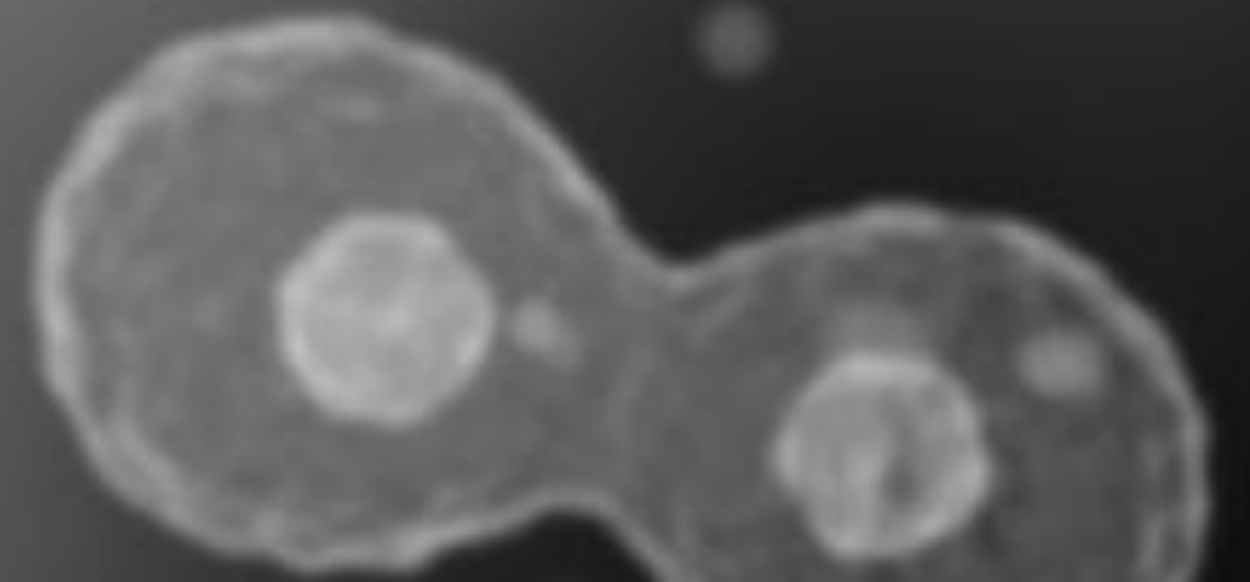
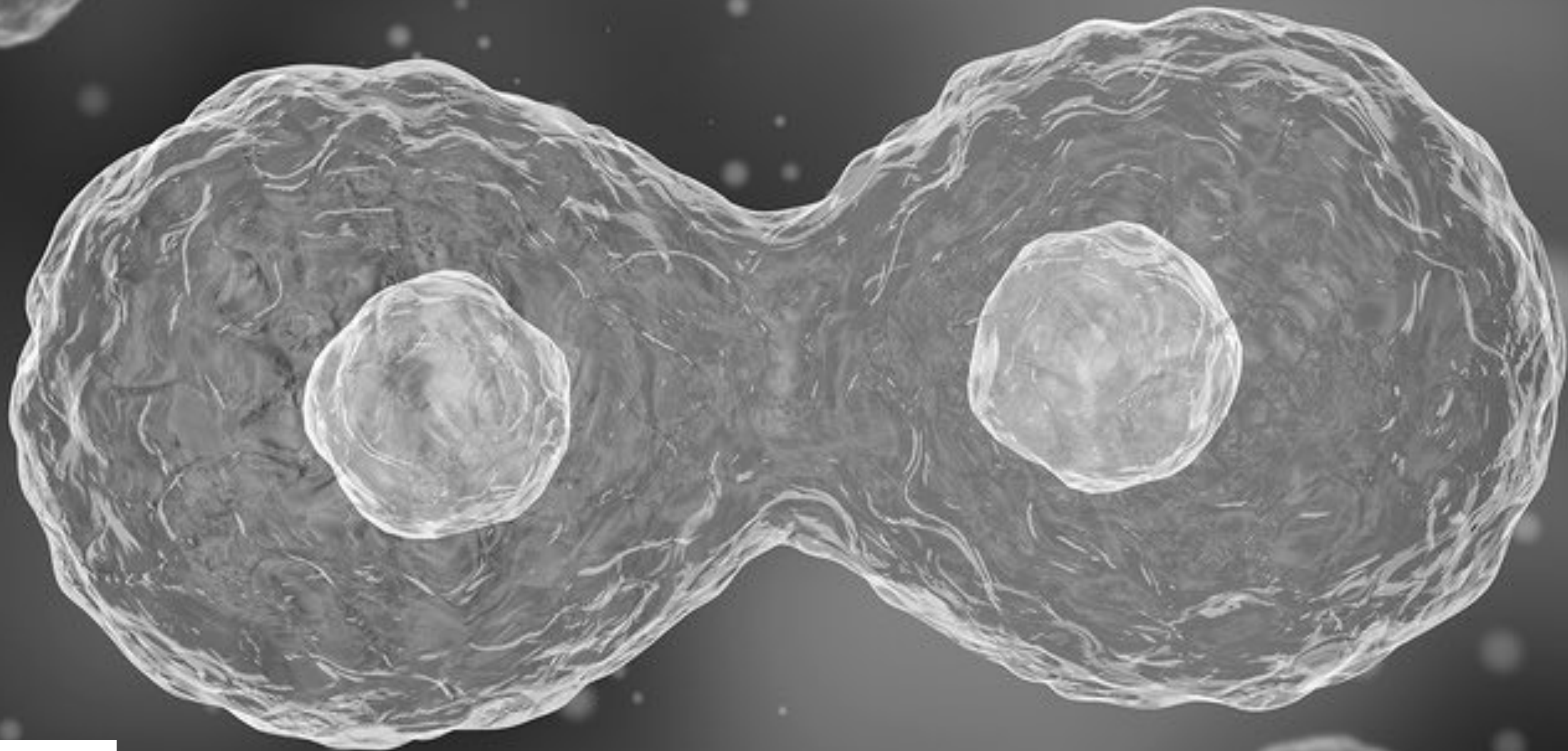
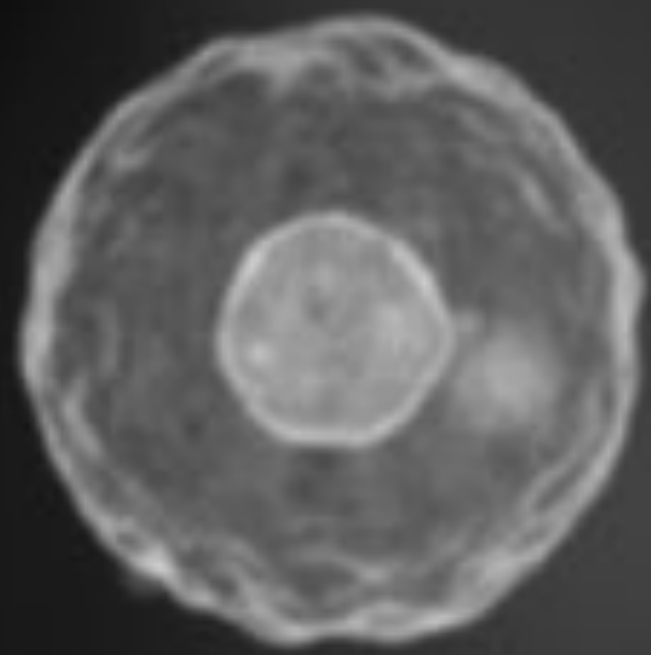


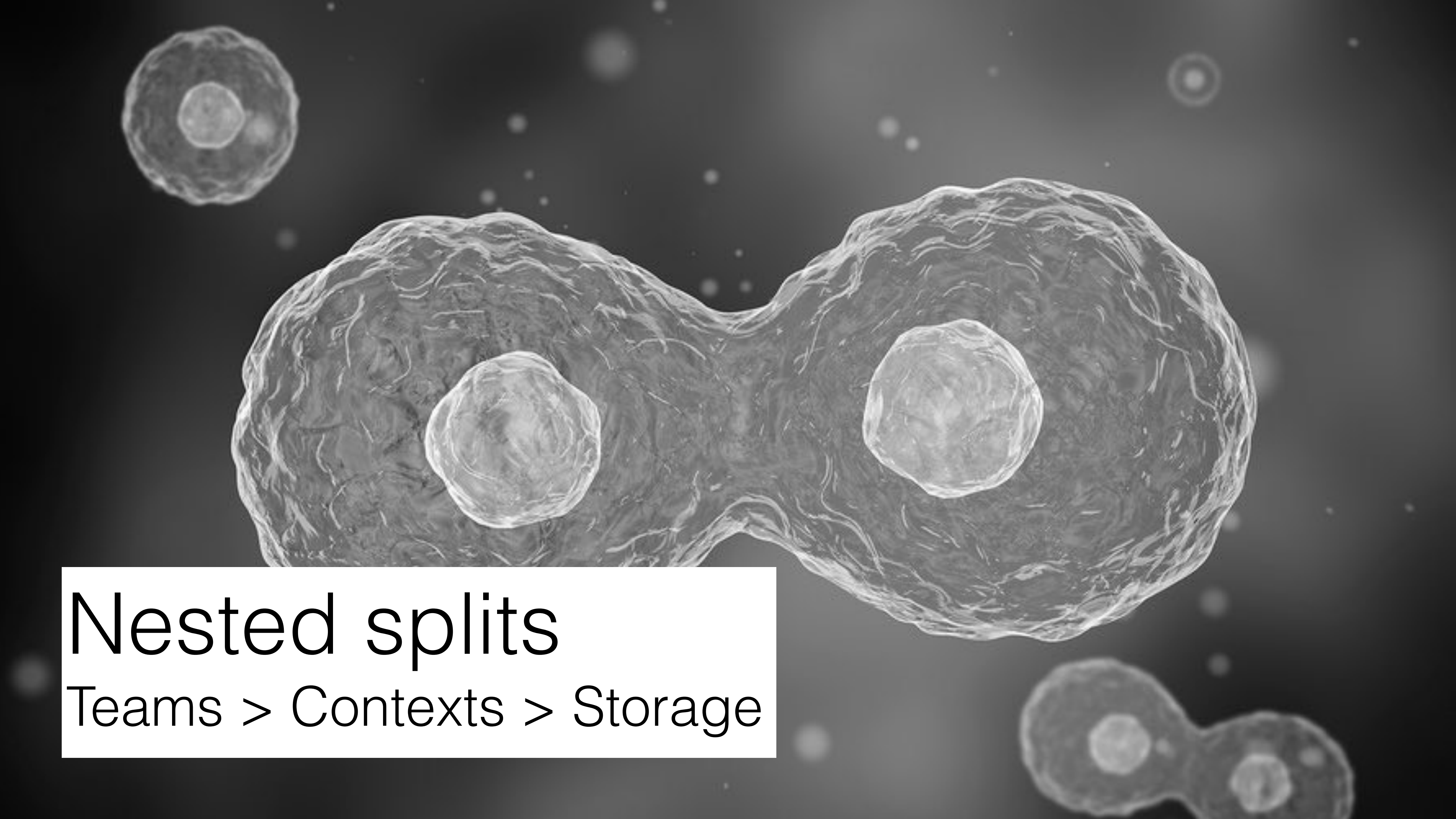
Bounded Context

A grayscale micrograph showing a large, irregularly shaped cell with a highly textured, wrinkled surface. The cell is elongated and has a central constriction, giving it a dumbbell-like appearance. Inside the cell, two large, rounded, and somewhat translucent structures are visible, likely representing storage granules or organelles. The background is dark with some out-of-focus, smaller cells and particles. In the bottom left corner, the word "Storage" is written in a white box.

Storage

Teams



A grayscale microscopic image of cells. The central focus is a large, irregularly shaped cell with a textured, almost fibrous surface. Inside this cell are two smaller, more spherical cells. In the upper left and lower right corners, there are other smaller cells, some appearing to be in the process of dividing or budding. The background is dark with some out-of-focus light spots.

Nested splits

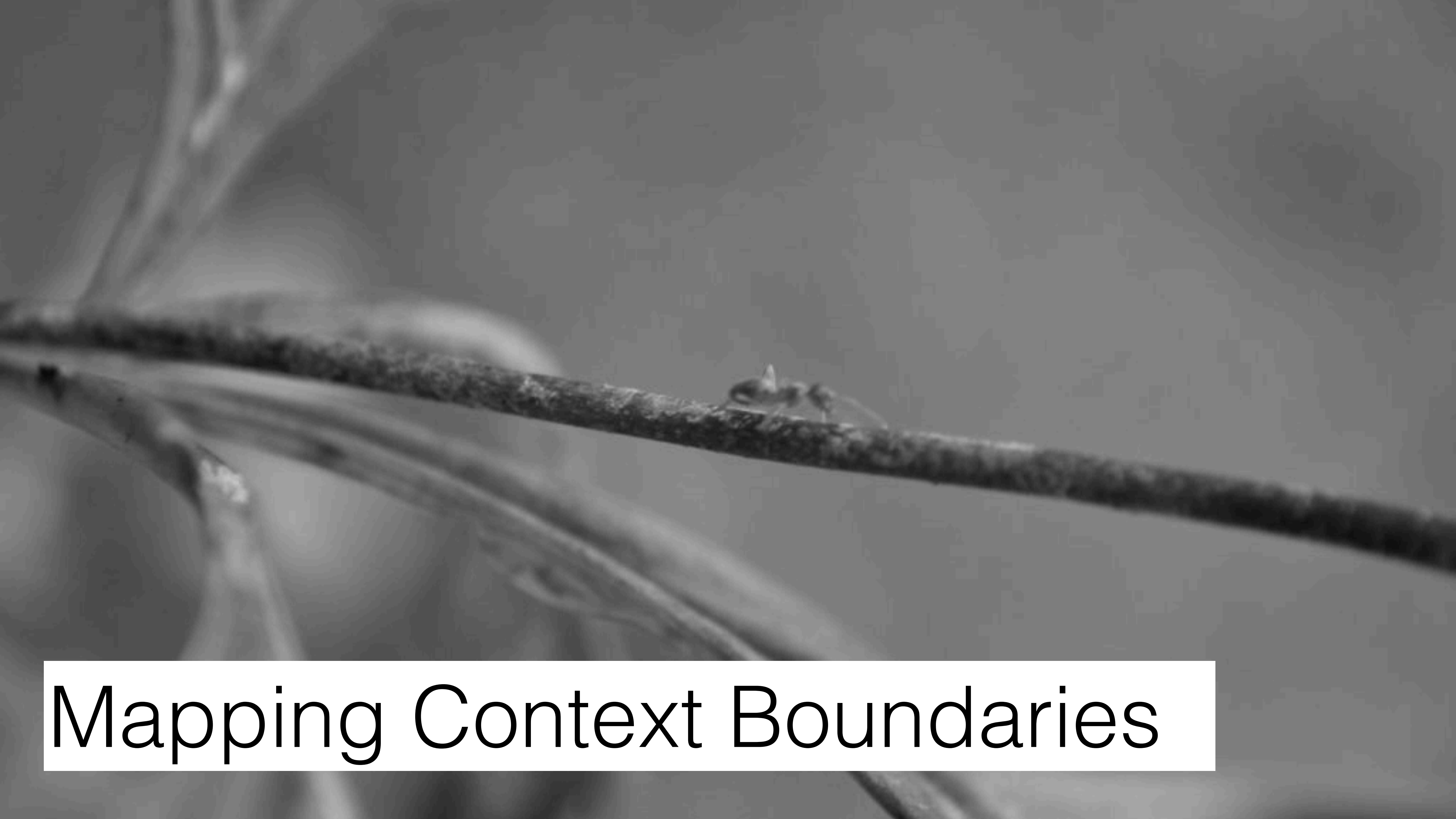
Teams > Contexts > Storage



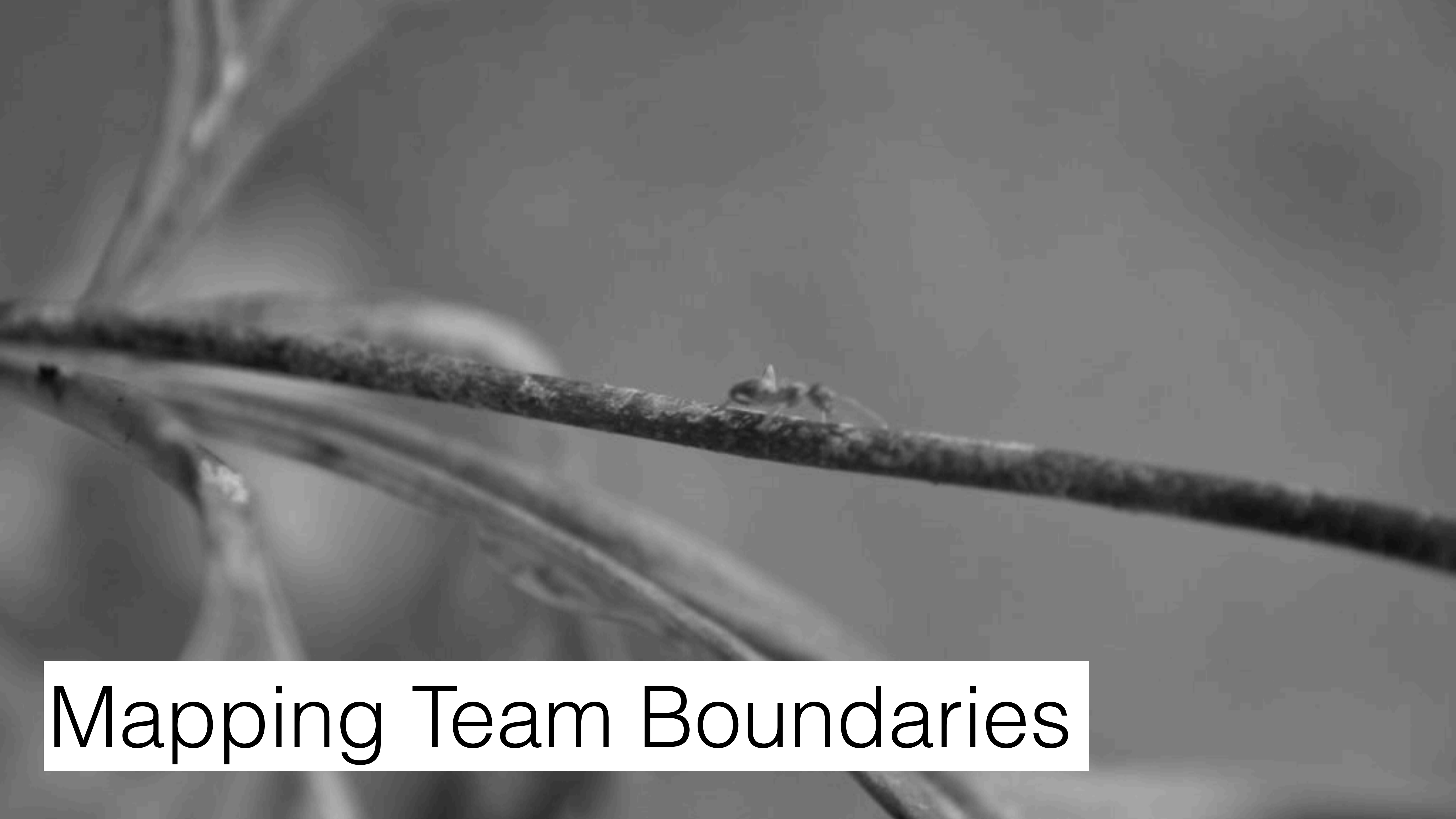
Mapping



Mapping Storage Boundaries



Mapping Context Boundaries



Mapping Team Boundaries

A black and white silhouette of a man in a dynamic, athletic pose. He is shown from the waist up, leaning forward with his right arm extended and hand open, as if reaching for something. In his left hand, he holds a chainsaw, which is positioned horizontally above his right hand. The background is a plain, light color, creating a high-contrast silhouette effect.

Coordination



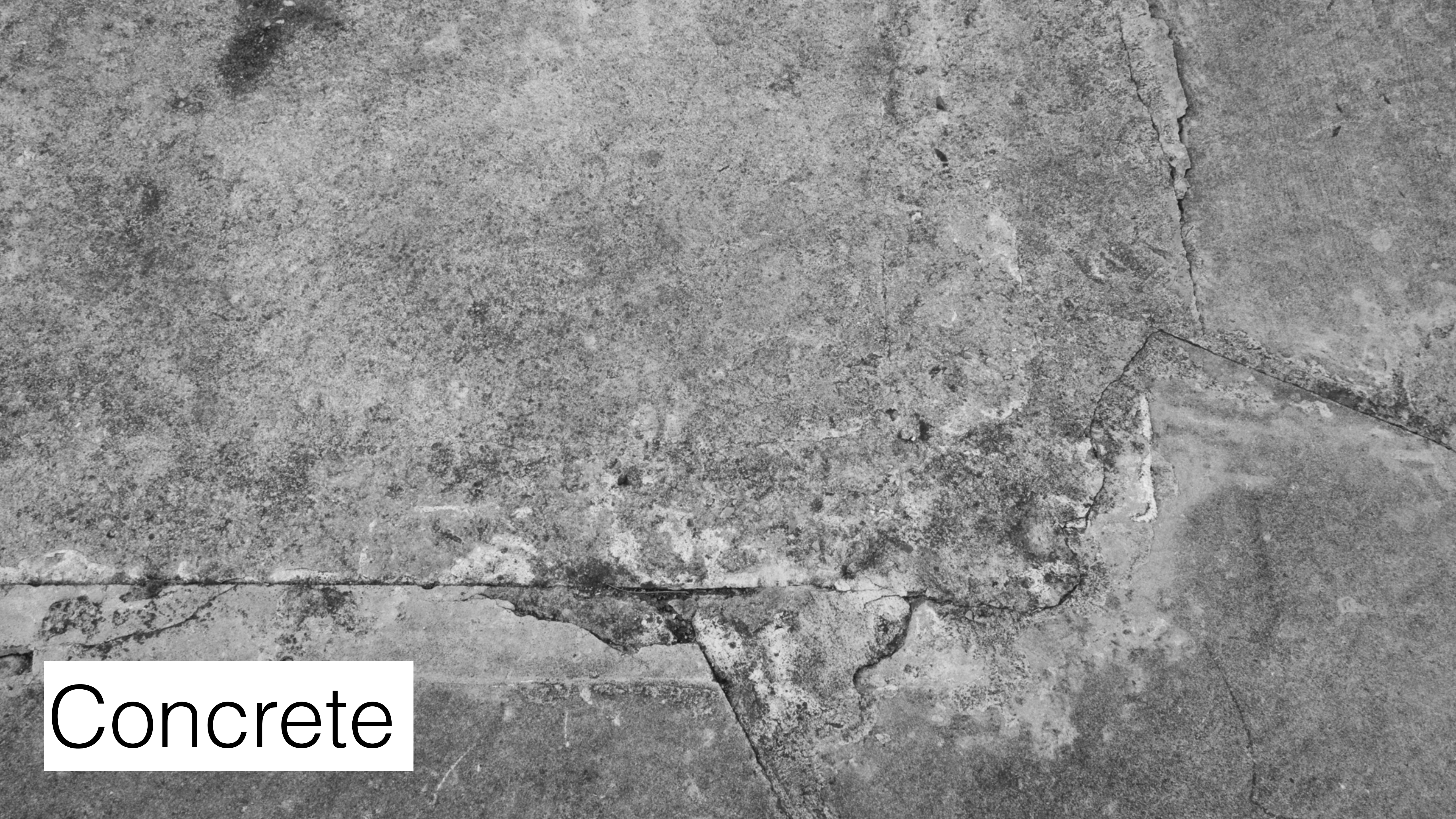
Coordinating Definitions



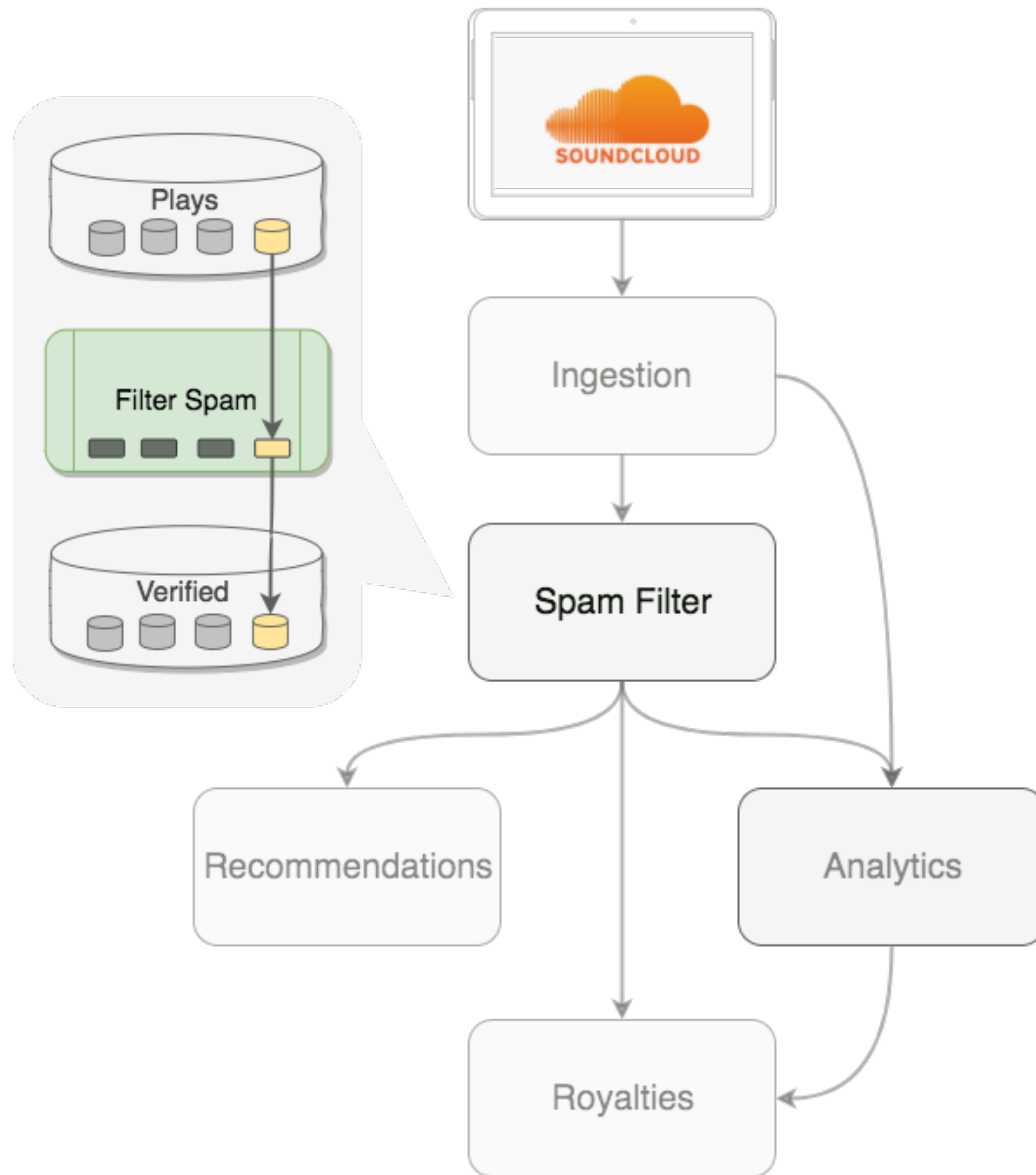
Coordinating Correctness



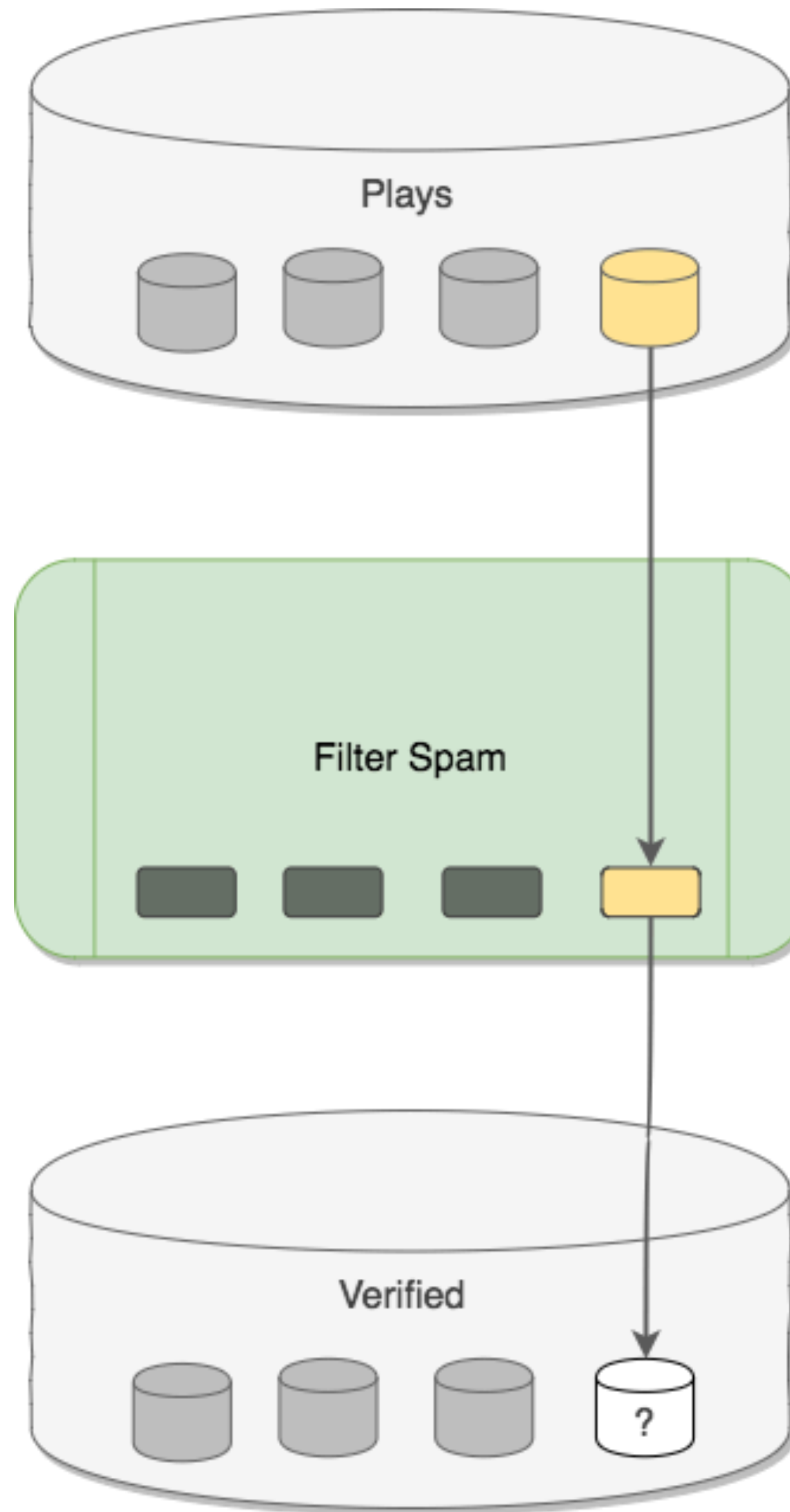
Coordinating Failure



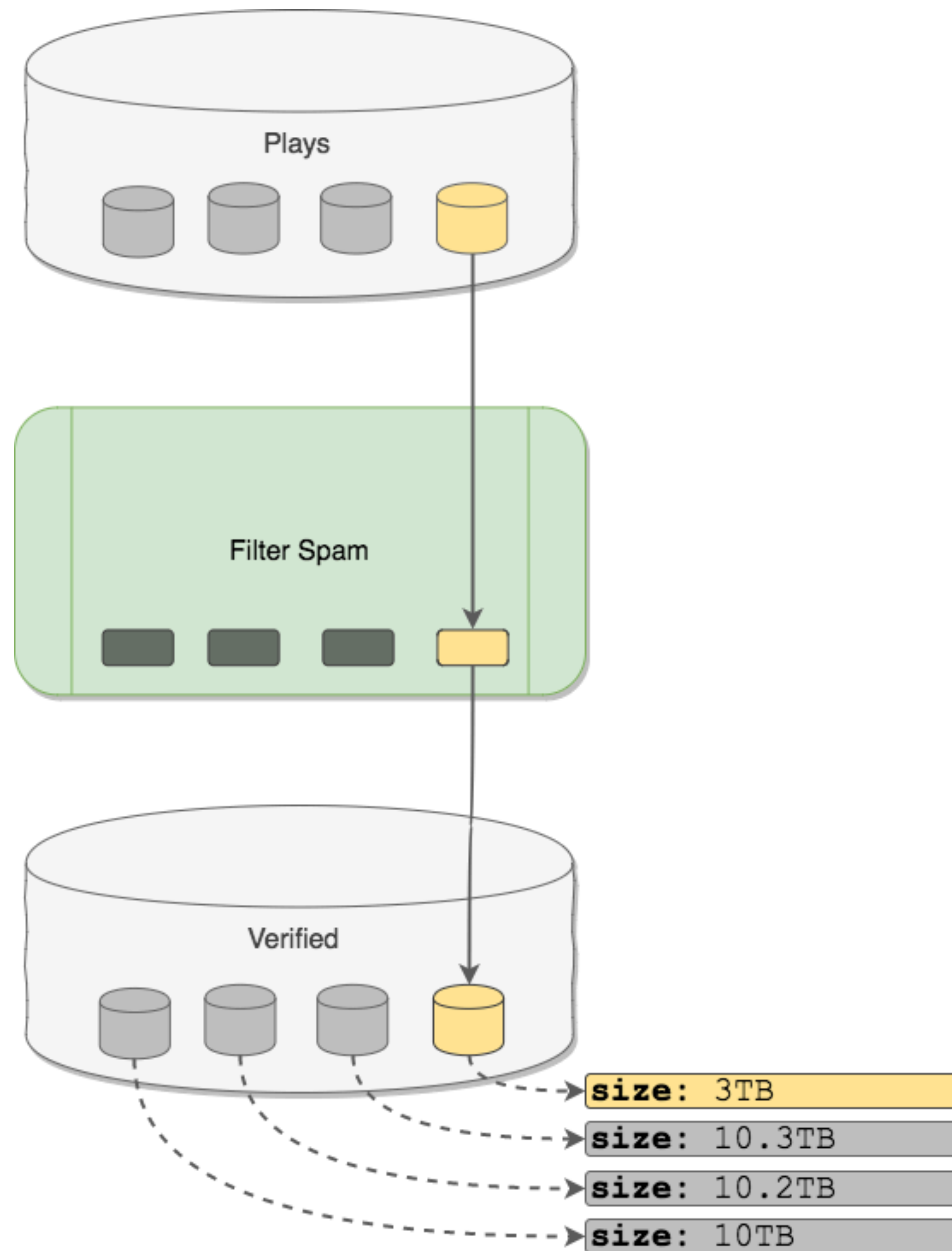
Concrete



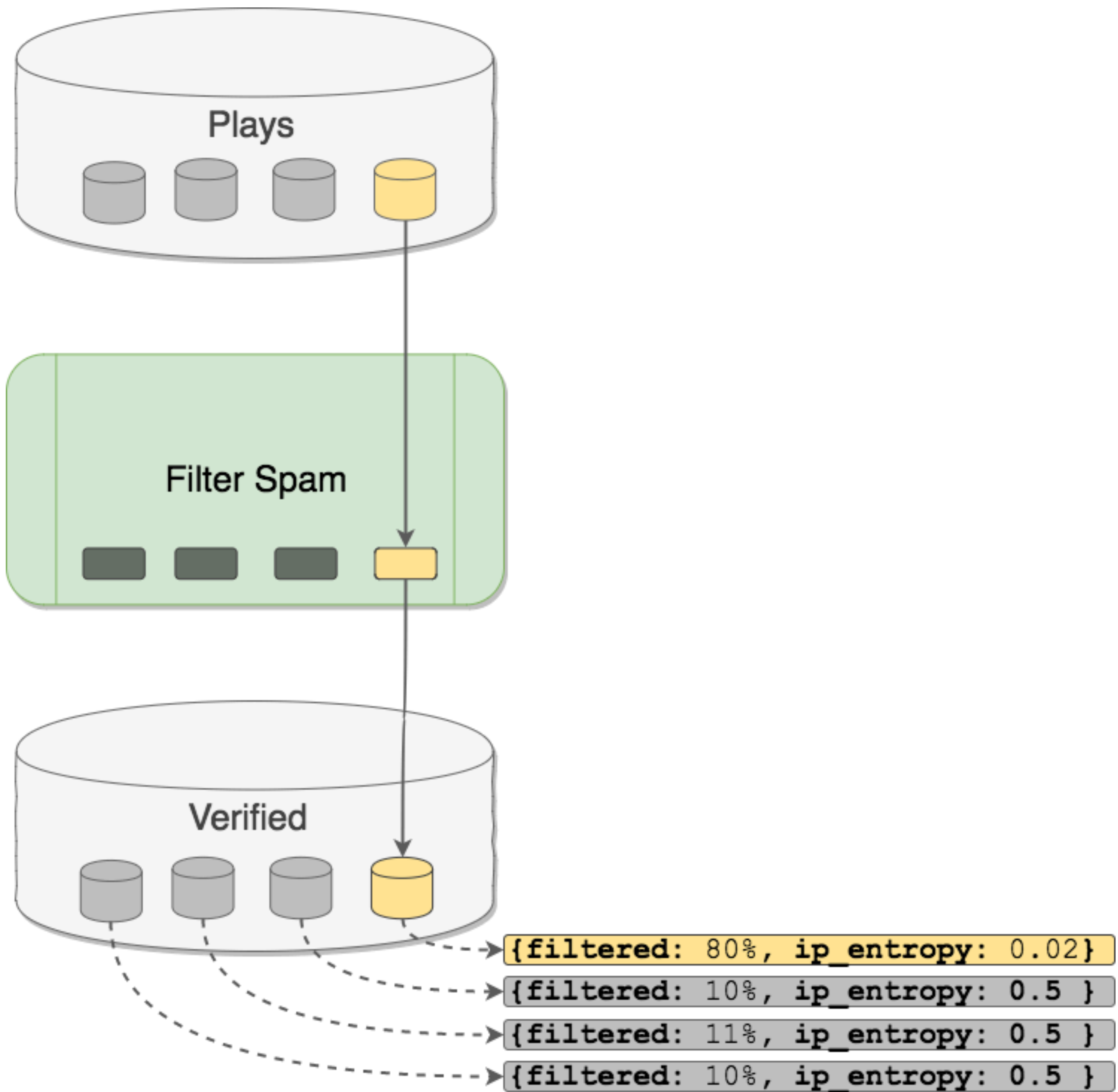
Correctness



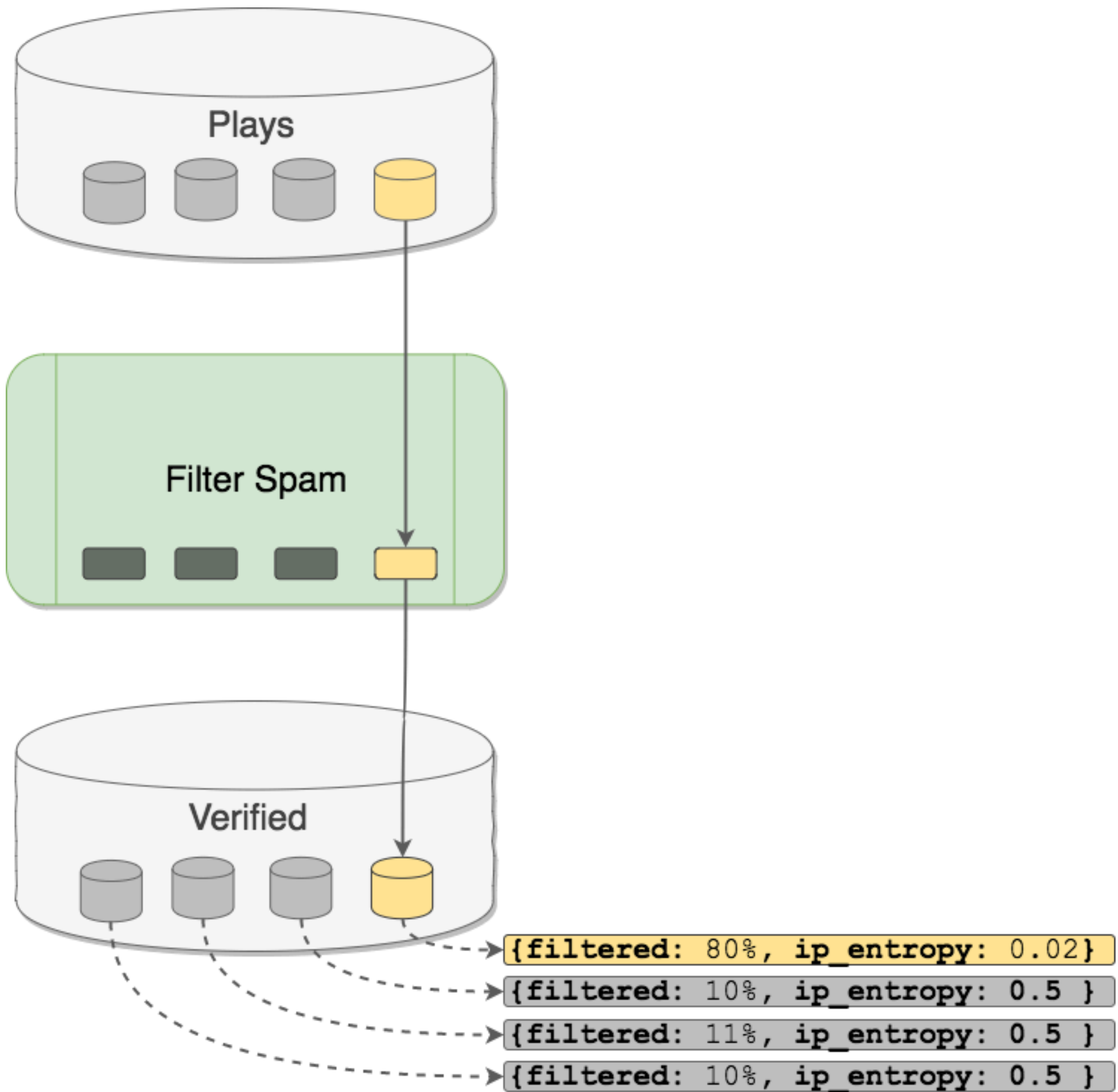
Fingerprints

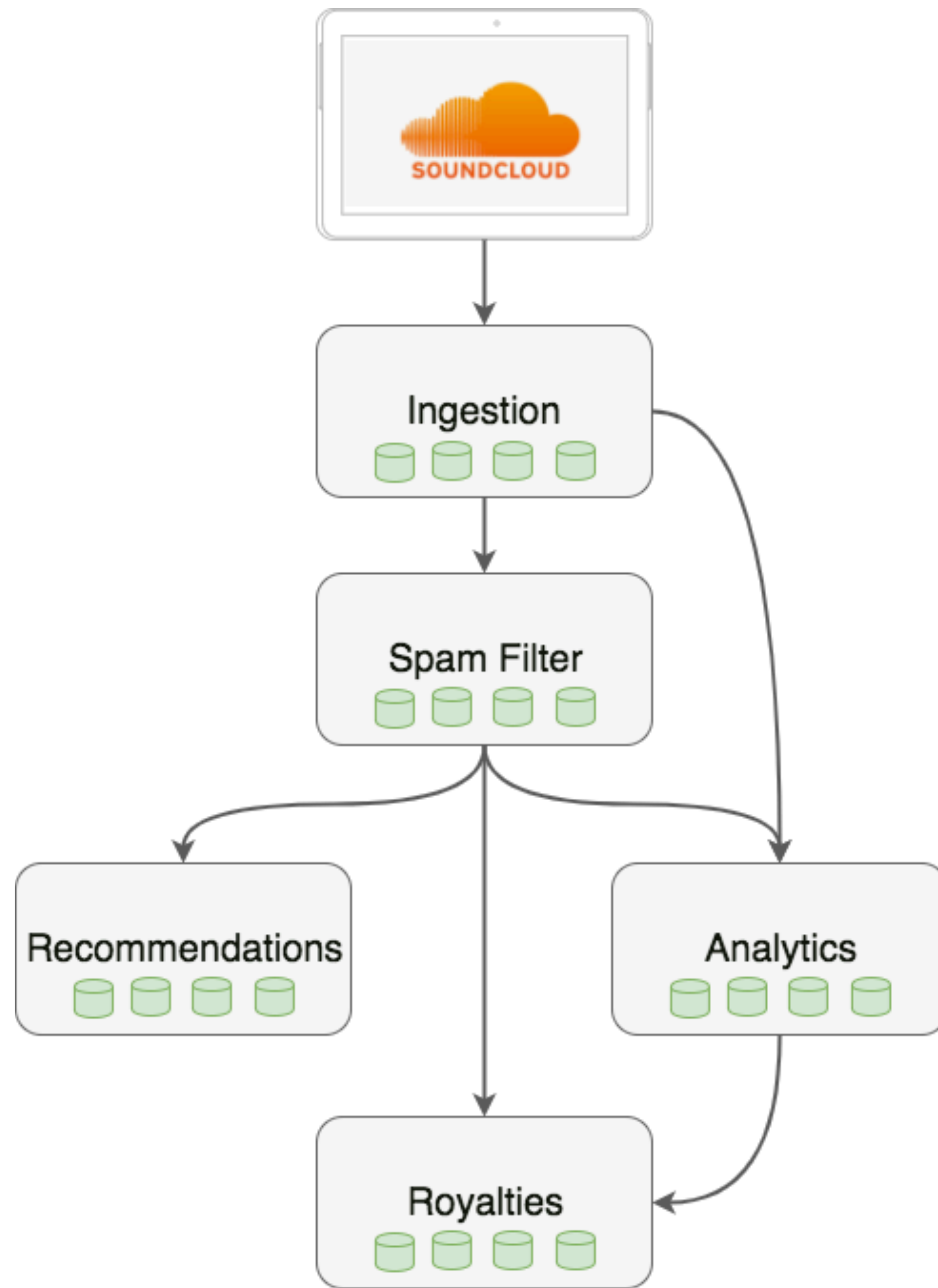


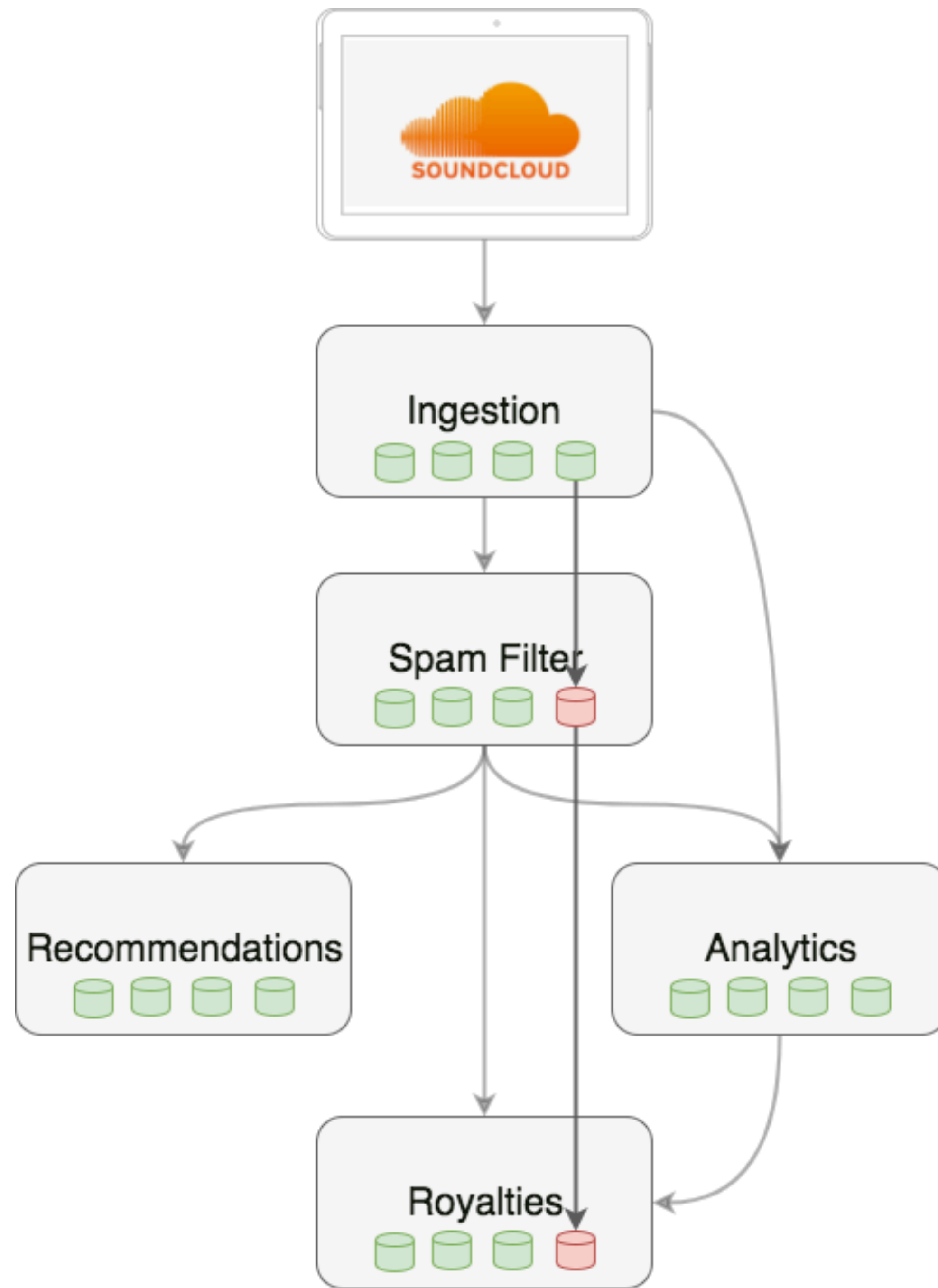
Fingerprints

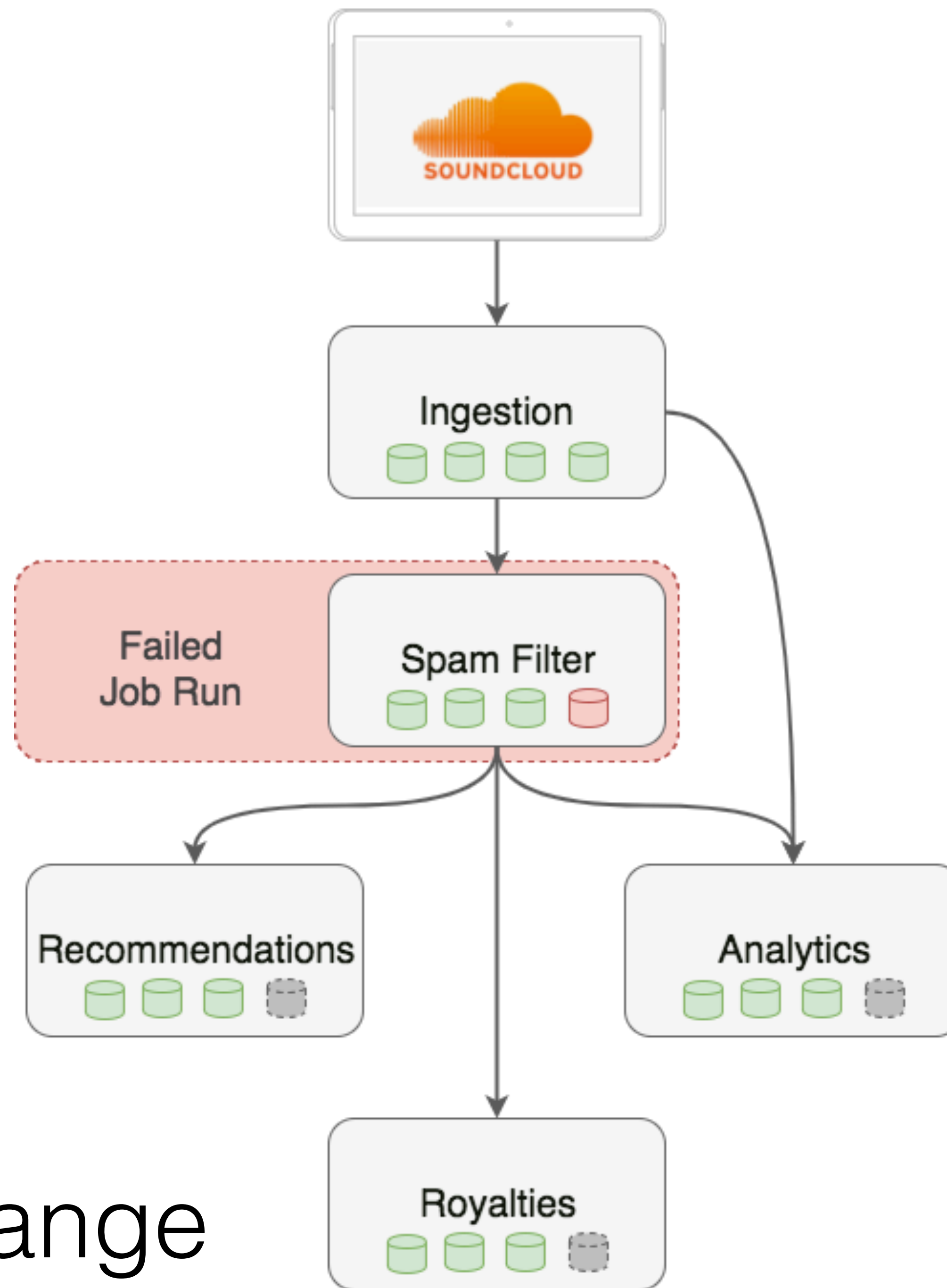


Fingerprints

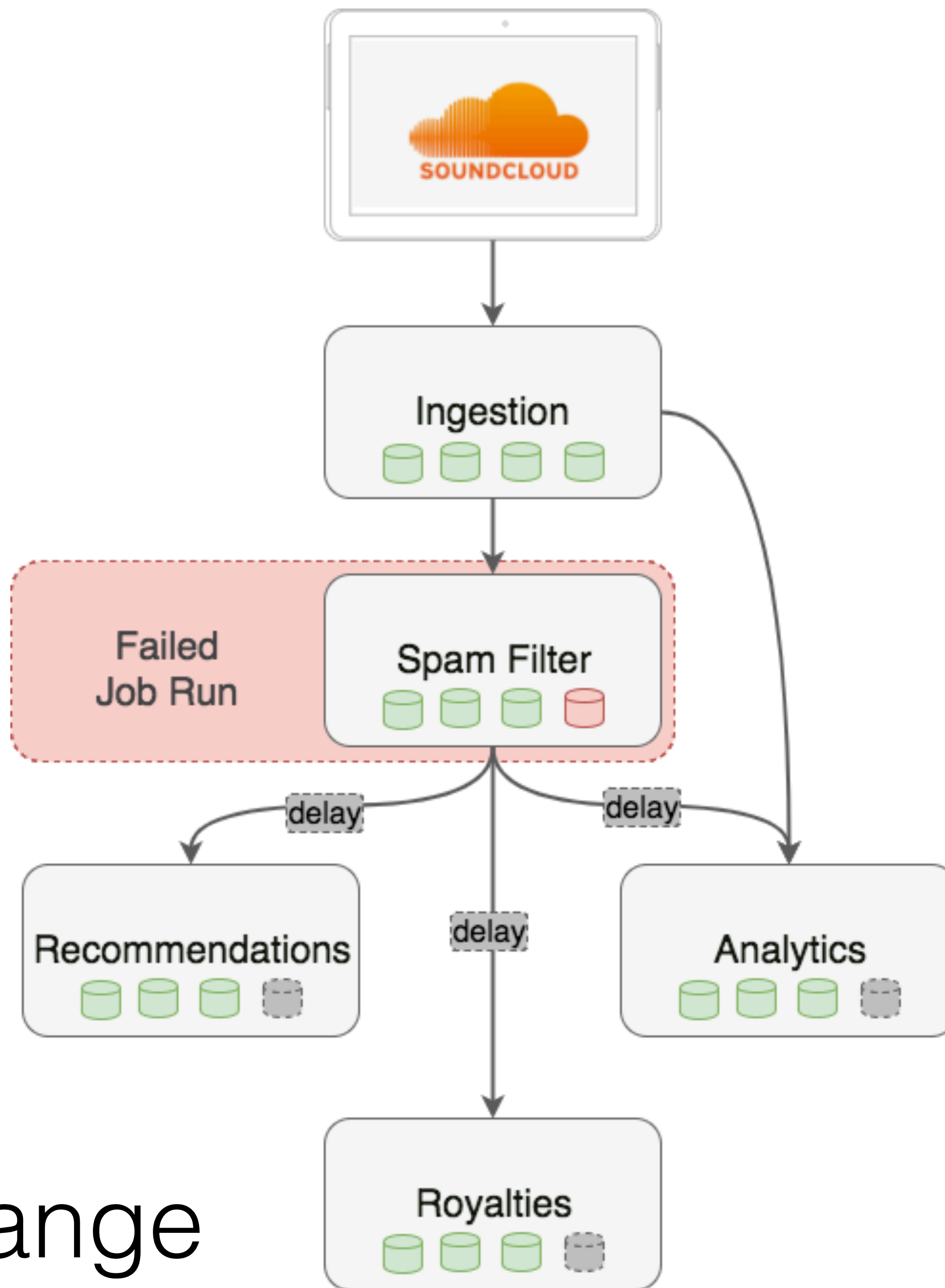




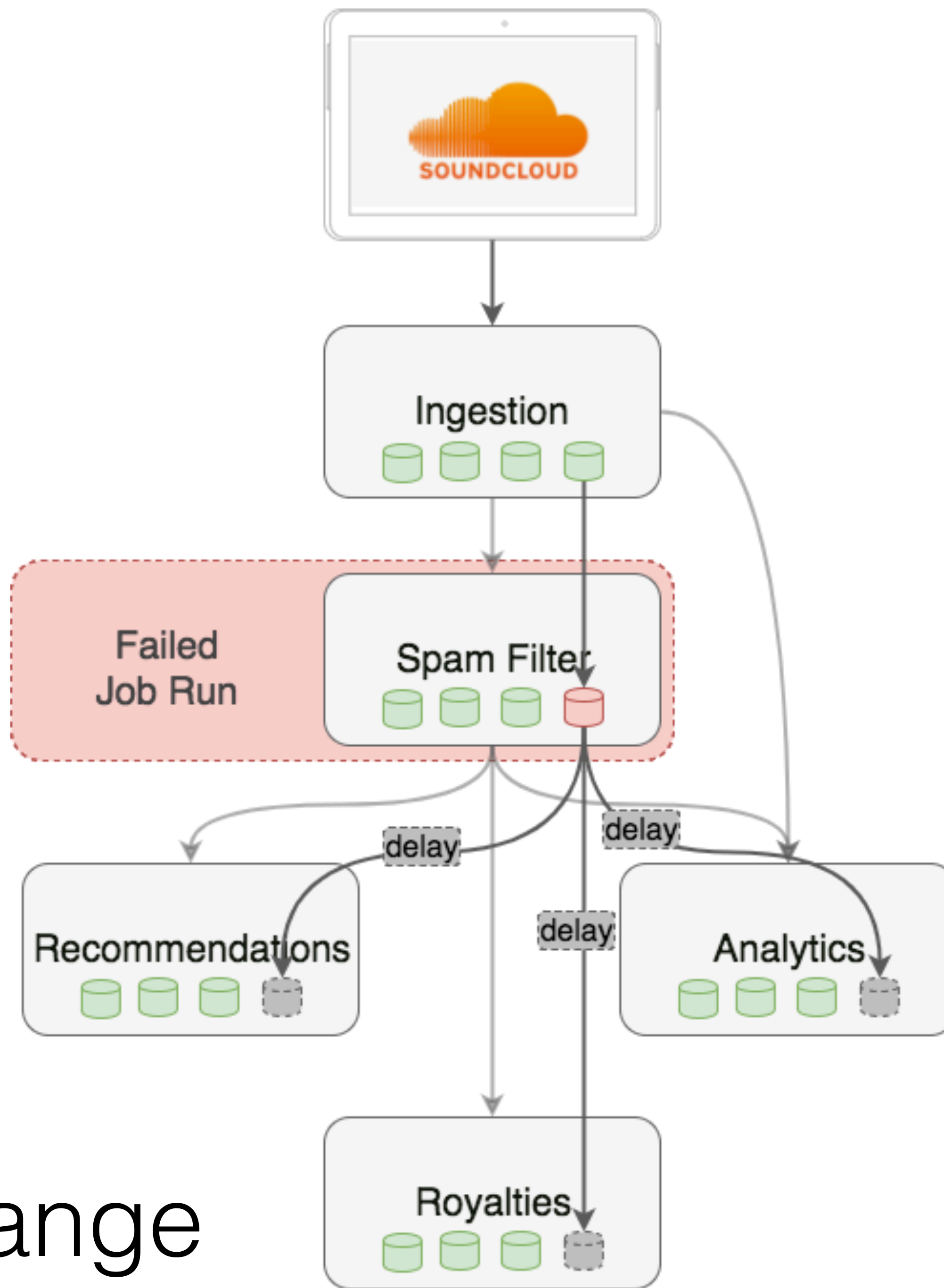




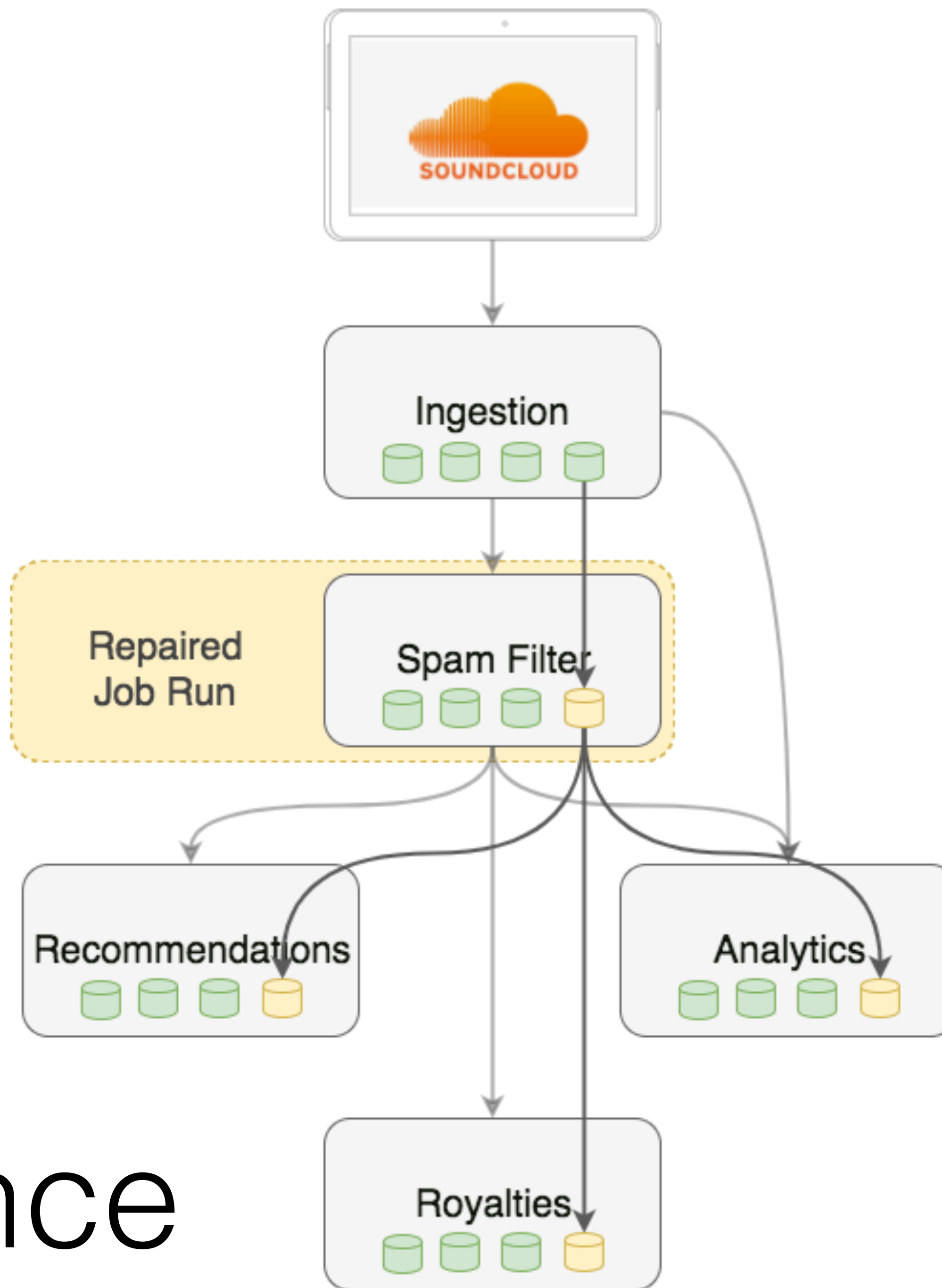
Coordinate Change



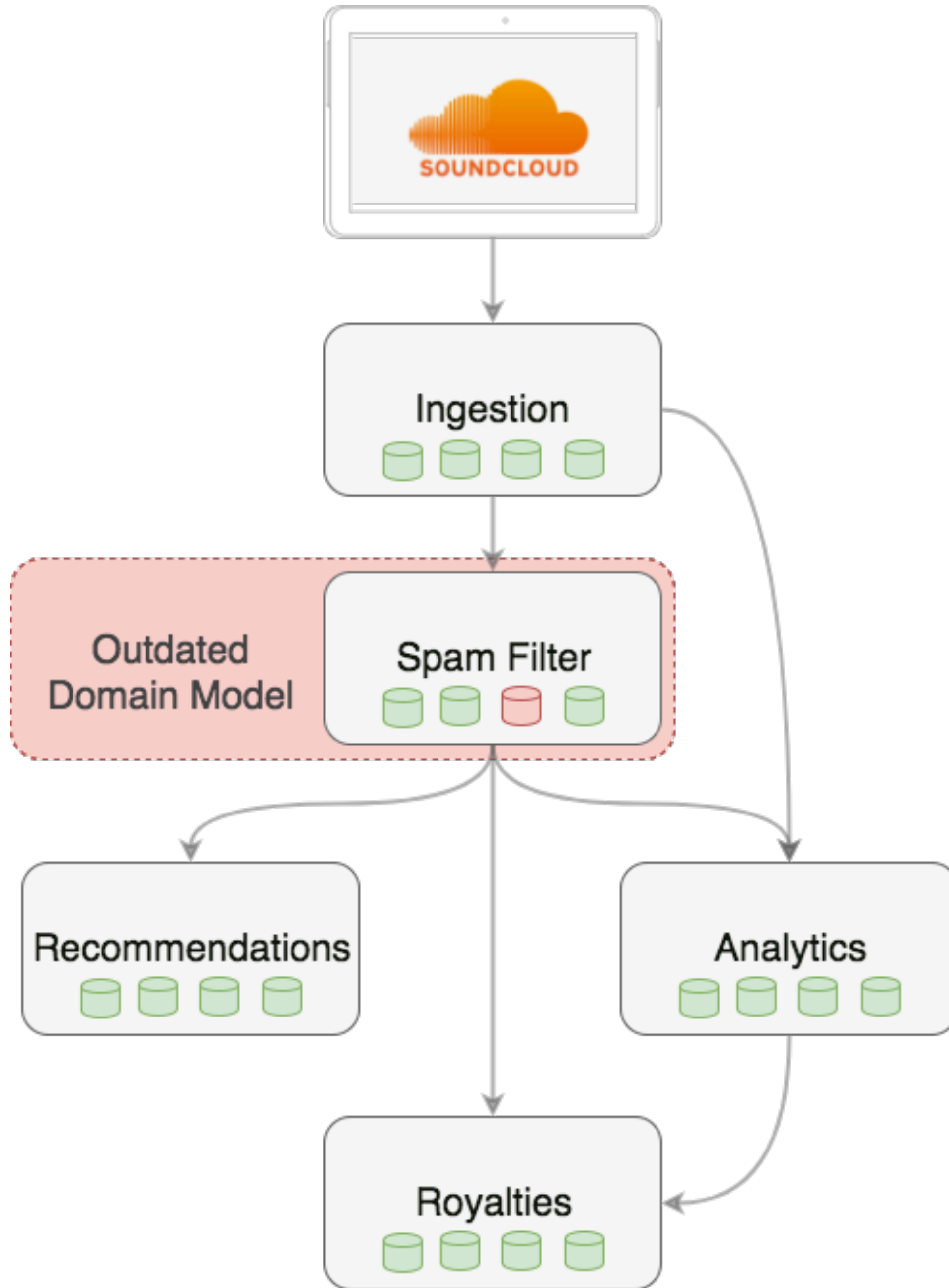
Coordinate Change



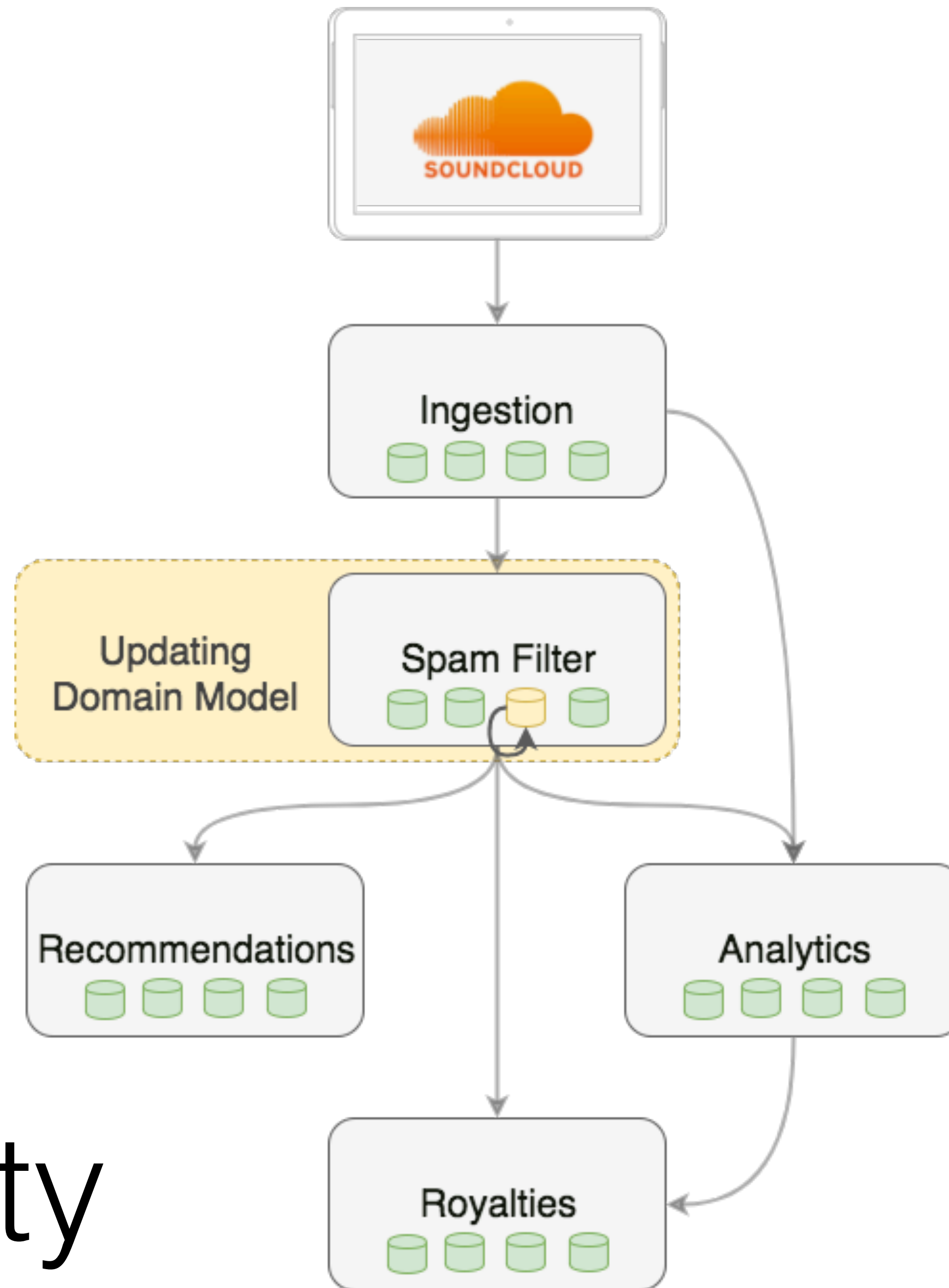
Coordinate Change



Convergence

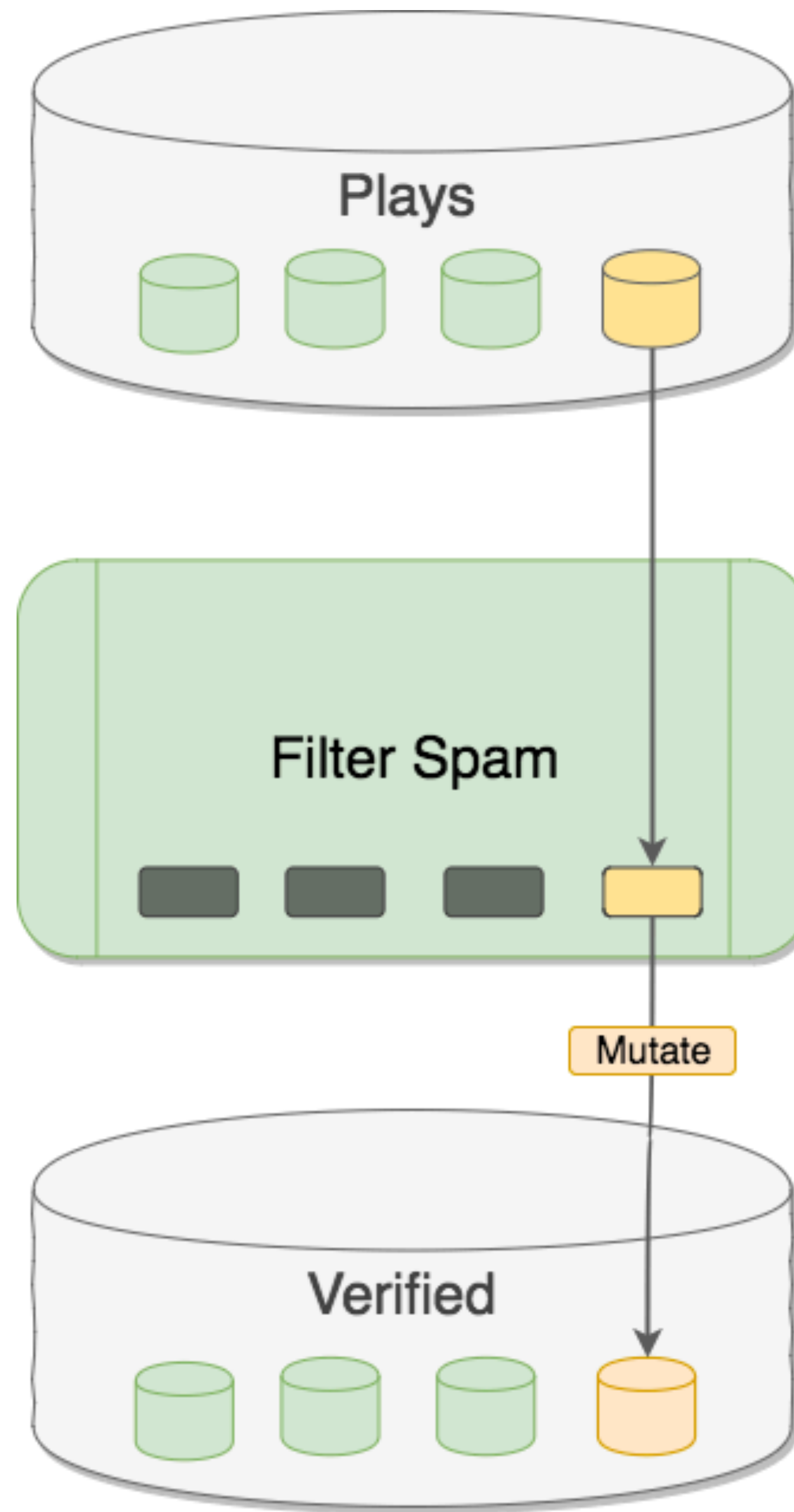


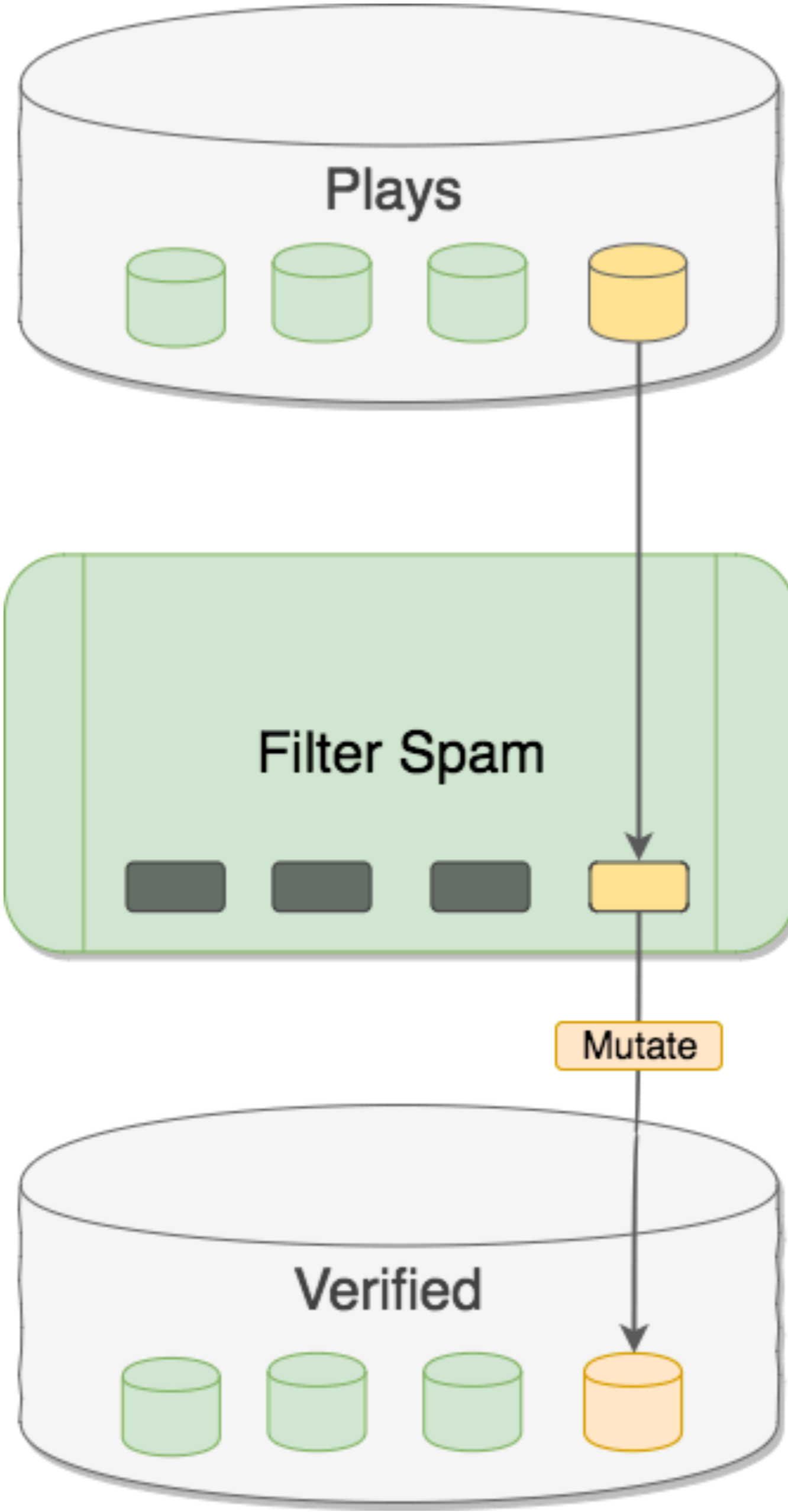
Failure



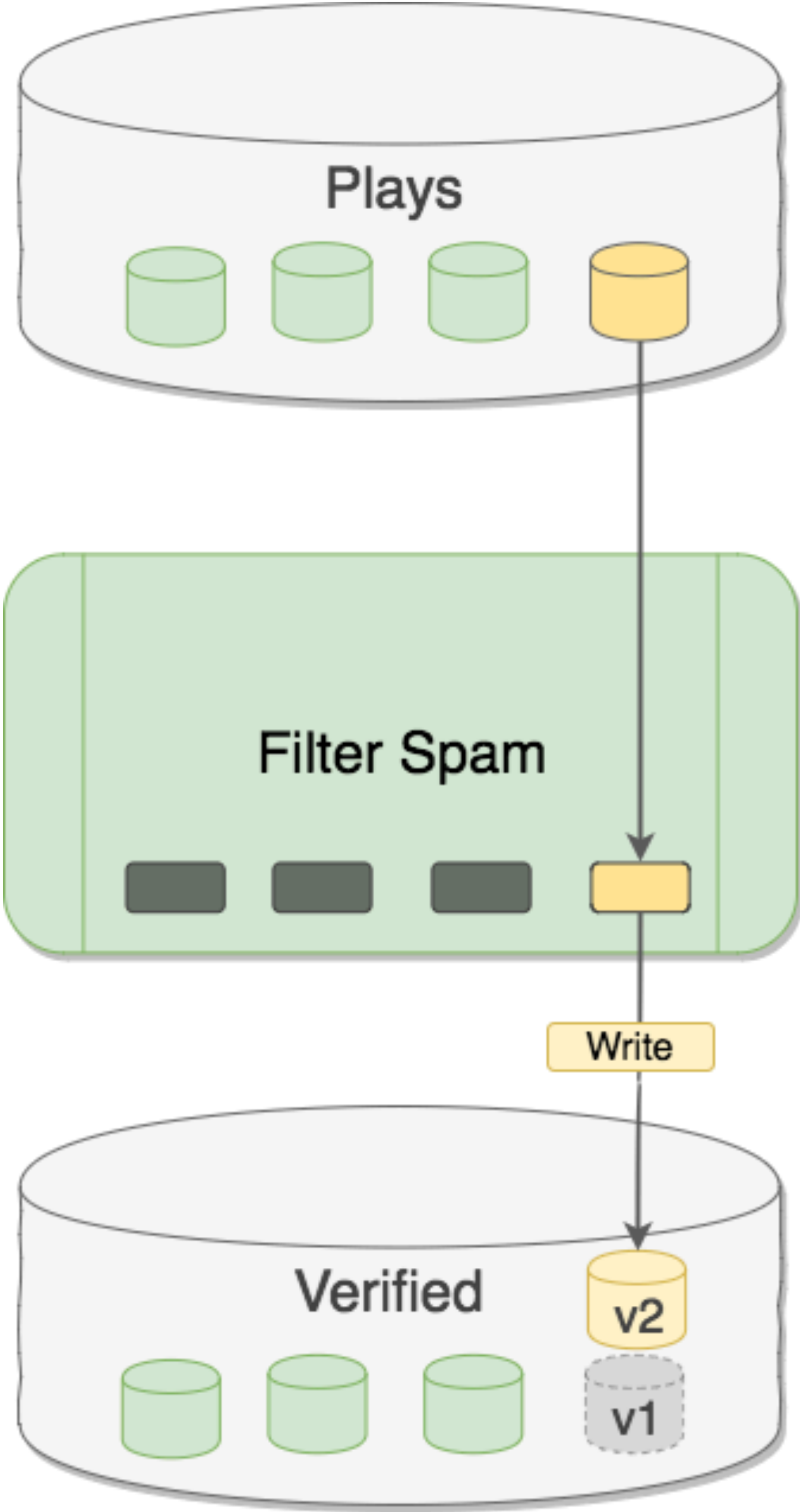
Retroactivity

Mutation

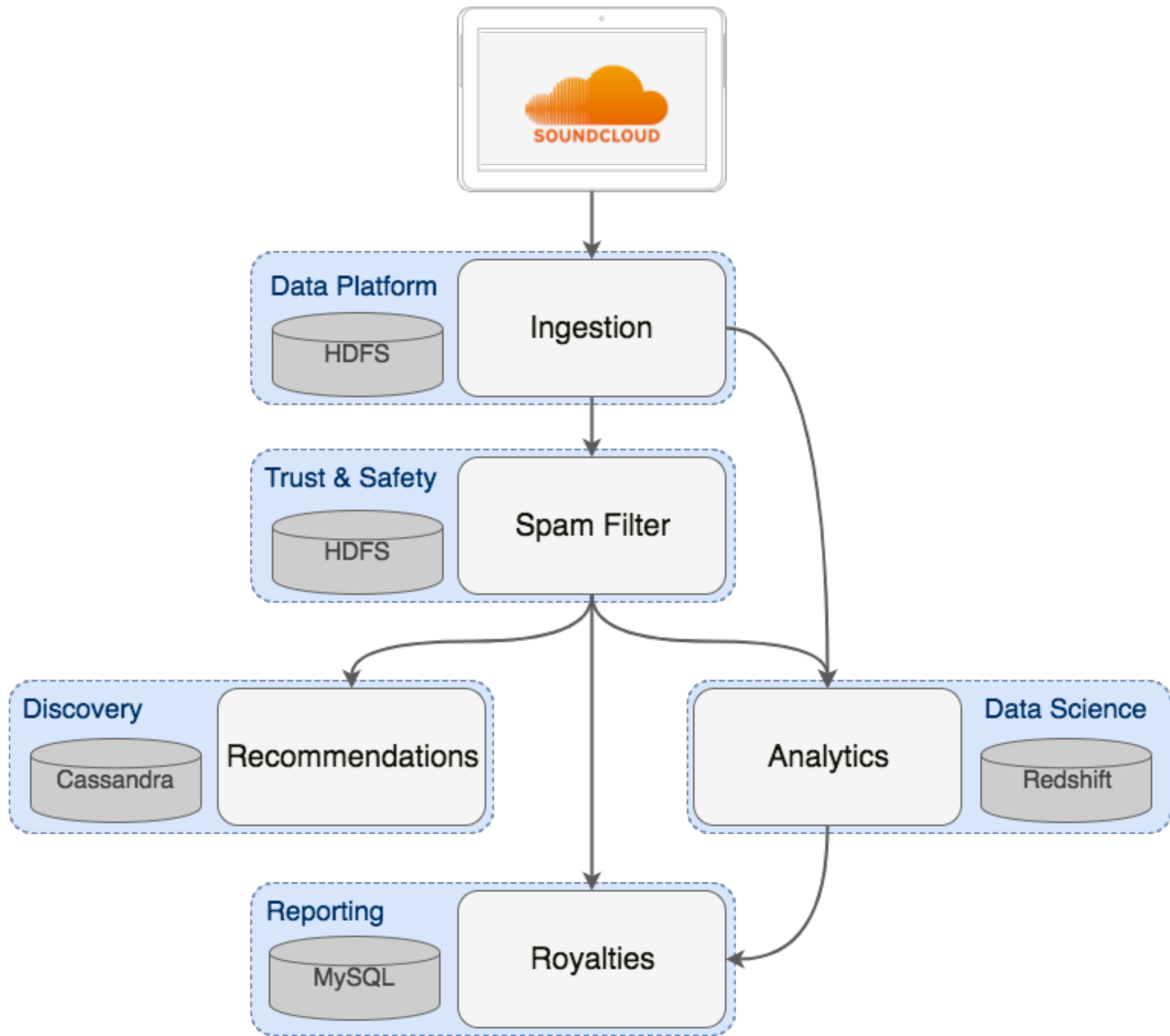




Mutation



Immutability



Conclusions

Data pipelines enable coordination but require shared protocols to determine when and how we read and write data.

Emily Green Omid Aladini

Sebastian Ohm Fronx

Wurmus Matthias Georgi

Thank You David Whiting

Lorand Kasler Gavin Bell Jon

Glover Erik Bartels