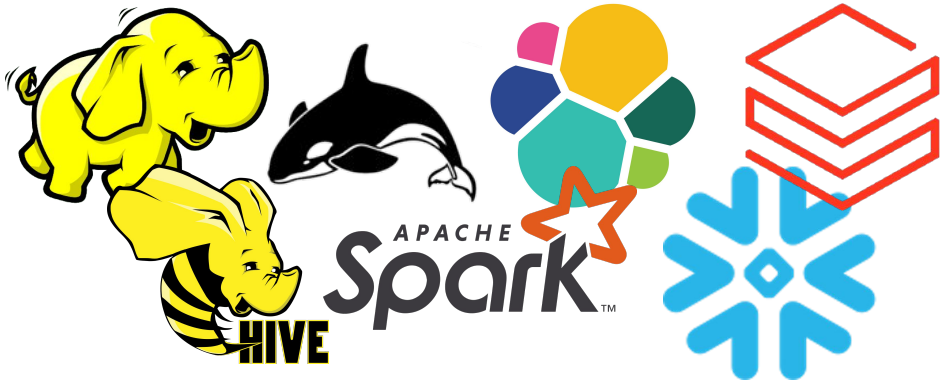
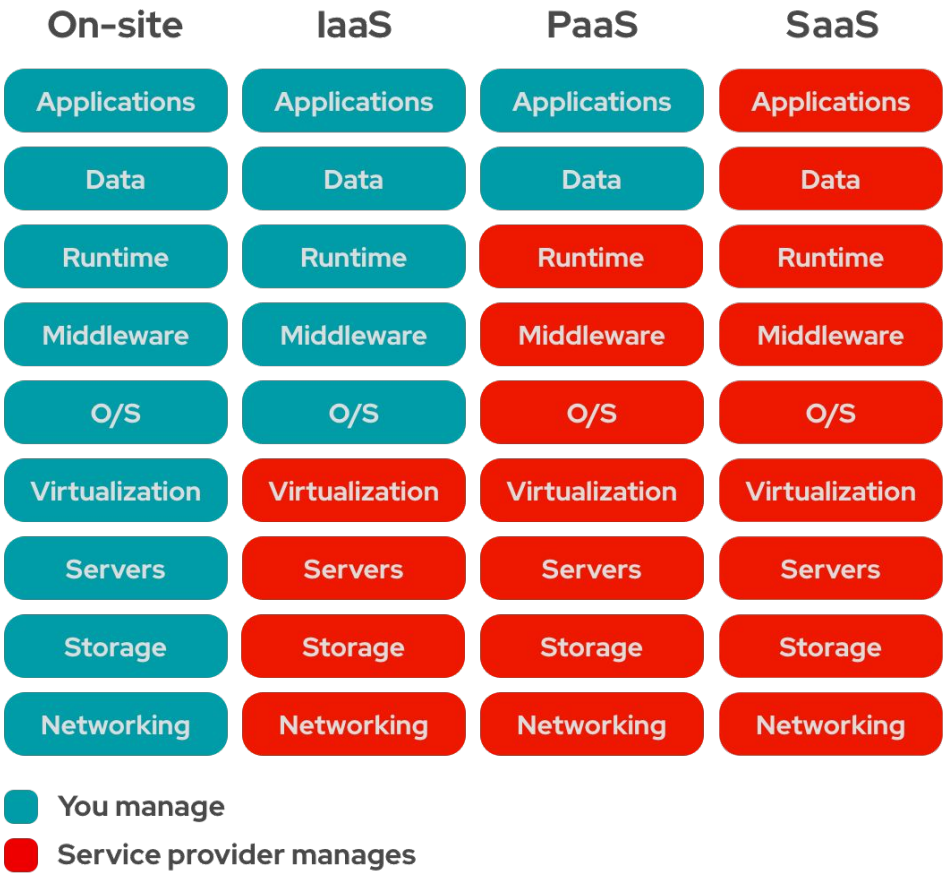
ADALTAS

Big Data Ecosystem



9. Alternative Solutions and Cloud


On-site
IaaS
PaaS
SaaS
Applications
Data
Runtime
Middleware
O/S
Virtualization
Servers
Storage
Networking
You manage
Service provider manages

Bare-metal vs Cloud

- Elastic computing:
  - Cloud storage
  - Elastic compute
  - Pay at usage

Bare-metal vs Cloud

- Time to market
- Total Cost of Ownership
  - It’s f***ing easy
  - It’s f***ing expensive
- CLOUD Act

Solutions and tools - IaaS

- **Bare-metal** server renting:
  - Private and public clouds (IBM, OVH)
  - Enterprise distribution (pay for support)
  - Scalable but slow
- **Big Data Cloud** solutions:
  - Amazon EMR, Azure HDInsight, Google Dataproc
  - Elastic infrastructure, dynamic VM allocation

Solutions and tools - PaaS

- **Data Warehouse** platforms: Snowflake, Teradata
- **Data Analysis** platforms: Databricks (Azure, AWS)
- **Managed Big Data** platforms: Elastic Cloud, Cloudera Data Platform

Solutions and tools - ETL/Dataflow/Streaming

- **ETL:** Talend, Informatica
- **Dataflow:** NiFi, StreamSets, SnapLogic
- **Messaging:** Kafka (Confluent), RabbitMQ
- **Streaming frameworks:** Spark Streaming (Databricks), Flink (Vervica), Kafka Streams (Confluent)

Solutions and tools - ETL/Dataflow/Streaming

- **ETL:** Talend, Informatica
- **Dataflow:** NiFi, StreamSets, SnapLogic
- **Messaging:** Kafka (Confluent), RabbitMQ
- **Streaming frameworks:** Spark Streaming (Databricks), Flink (Vervica), Kafka Streams (Confluent)

Solutions and tools - BI & Monitoring

- **Business Intelligence:** Qlikview, Tableau, Power BI, Micro Strategy
- **Monitoring:** Datadog, Splunk

Solutions and tools - ML Platforms

- Dataiku
- H2O
- MLFlow (Databricks)

What's next in Big Data? Data mesh

How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh